

# A Generic Method for the Production of Recombinant Proteins in *Escherichia coli* Using a Dual Hexahistidine-Maltose-Binding Protein Affinity Tag

Joseph E. Tropea, Scott Cherry, Sreedevi Nallamsetty,  
Christophe Bignon, and David S. Waugh

## Summary

A generic protocol that utilizes a dual hexahistidine-maltose-binding protein (His<sub>6</sub>-MBP) affinity tag has been developed for the production of recombinant proteins in *Escherichia coli*. The MBP moiety improves the yield and enhances the solubility of the passenger protein while the His-tag facilitates its purification. The fusion protein (His<sub>6</sub>-MBP-passenger) is purified by immobilized metal affinity chromatography (IMAC) on nickel-nitrilotriacetic acid (Ni-NTA) resin and then cleaved in vitro with His<sub>6</sub>-tobacco etch virus protease to separate the His<sub>6</sub>-MBP from the passenger protein. In the final step, the unwanted byproducts of the digest are absorbed by a second round of IMAC, leaving nothing but the pure passenger protein in the flow-through fraction. Endogenous proteins that bind to the Ni-NTA resin during the first IMAC step also do so during the second round of IMAC. Hence, the application of two successive IMAC steps, rather than just one, is the key to obtaining crystallization-grade protein with a single affinity technique.

**Key Words:** Maltose-binding protein; MBP; immobilized metal affinity chromatography; IMAC; His-tag; affinity tag; affinity chromatography; structural genomics; high throughput; tobacco etch virus protease; TEV protease.

## 1. Introduction

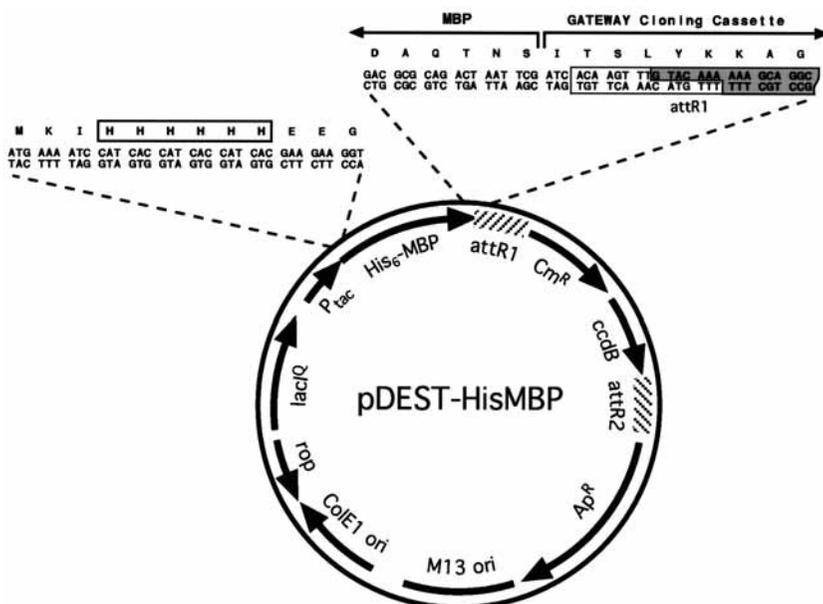
Because of its remarkable ability to enhance the solubility and promote the proper folding of its fusion partners, *Escherichia coli* maltose-binding protein (MBP) has emerged as an attractive vehicle for the production of recombinant proteins (1,2). However, MBP fusion proteins do not always bind efficiently to amylose resin, and even when they do amylose affinity chromatography typically does not produce samples of sufficient purity for structural studies. To address this problem, we identified several locations in which small affinity

tags can be inserted within the framework of an MBP fusion protein without compromising its solubility-enhancing properties (3). In this chapter, we describe how one such configuration, in which a hexahistidine tag (His<sub>6</sub>) is added to the N-terminus of MBP, forms the foundation of an entirely generic strategy for protein production in *E. coli*. The MBP moiety improves the yield and enhances the solubility of the passenger protein, whereas the His-tag facilitates its purification. The soluble fusion protein (His<sub>6</sub>-MBP-passenger) is purified by immobilized metal affinity chromatography (IMAC) on nickel-nitrilotriacetic acid (Ni-NTA) resin and then cleaved *in vitro* with His<sub>6</sub>-tagged tobacco etch virus (TEV) protease (His<sub>6</sub>-TEV protease) to separate the His<sub>6</sub>-MBP from the passenger protein. In the final step, the unwanted byproducts of the digest, as well as any impurities that eluted from the Ni-NTA resin along with the fusion protein in the first IMAC step, are absorbed by a second round of IMAC, leaving nothing but the pure passenger protein in the flow-through fraction. Hence, the application of two successive IMAC steps, rather than just one, is the key to obtaining crystallization-grade protein with a single affinity technique. This simple generic protocol should be readily amenable to automation for high-throughput applications.

## 2. Materials

### 2.1. Recombinational Vector Construction

1. The Gateway<sup>®</sup> destination vector pDEST-HisMBP (*see Fig. 1*).
2. Reagents and thermostable DNA polymerase for PCR amplification (*see Note 1*).
3. Synthetic oligodeoxyribonucleotide primers for PCR amplification (*see Fig. 2*).
4. TE buffer: 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA.
5. TAE-agarose and an apparatus for submarine gel electrophoresis of DNA (*see Note 2*).
6. QIAquick<sup>™</sup> gel extraction kit (Qiagen, Valencia, CA) for the extraction of DNA from agarose gels.
7. Chemically competent DB3.1 cells (Invitrogen, Carlsbad, CA) for propagating pDEST-HisMBP and pDONR201.
8. Competent *gyrA*<sup>+</sup> cells (e.g., DH5 $\alpha$ , MC1061, HB101) (*see Note 3*).
9. Gateway PCR Cloning System (Invitrogen).
10. Luria Bertani (LB) medium and LB agar plates containing ampicillin (100  $\mu$ g/mL). LB medium: add 10 g bacto tryptone, 5 g bacto yeast extract, and 5 g NaCl to 1 L of H<sub>2</sub>O and sterilize by autoclaving. For LB agar, also add 12 g of bactoagar before autoclaving. To prepare plates, allow medium to cool until flask or bottle can be held in hands without burning, then add 1 mL ampicillin stock solution (100 mg/mL in H<sub>2</sub>O, filter-sterilized), mix by gentle swirling, and pour or pipet ca. 30 mL into each sterile Petri dish (100-mm diameter).
11. Reagents for small-scale plasmid DNA isolation (*see Note 4*).
12. An incubator set at 37°C.



**Fig. 1.** Schematic representation of the Gateway<sup>®</sup> destination vector pDEST-HisMBP. This vector can be recombined with an entry vector that contains an open reading frame of interest, via the LR reaction, to generate a His<sub>6</sub>-MBP fusion protein expression vector.

## 2.2. Pilot Expression Experiment

1. Competent BL21PRO cells (Clontech, Palo Alto, CA) containing the TEV protease expression vector pRK603 (4) (see **Notes 5** and **6**).
2. A derivative of pDEST-HisMBP that produces a His<sub>6</sub>-MBP fusion protein with a TEV protease recognition site in the linker between MBP and the passenger protein (see **Subheading 3.1.**).
3. LB agar plates and broth containing both ampicillin (100 µg/mL) and kanamycin (25 µg/mL). See **Subheading 2.1., item 10** for LB broth, LB agar, and ampicillin stock solution recipes. Prepare stock solution of 25 mg/mL kanamycin in H<sub>2</sub>O and filter-sterilize. Store at 4°C for up to 1 mo. Dilute antibiotics 1000-fold into LB medium or molten LB agar.
4. Isopropyl-thio-β-D-galactopyranoside (IPTG), analytical grade (Anatrace Inc., Maumee, OH). Prepare a stock solution of 200 mM in H<sub>2</sub>O and filter-sterilize. Store at -20°C.
5. Anhydrotetracycline (ACROS Organics/Fisher Scientific, Springfield, NJ). Prepare a 1000X stock solution by dissolving in 50% ethanol at 100 µg/mL. Store in a foil-covered tube at -20°C.
6. Shaker/incubator.



7. Sterile baffle-bottom flasks (Bellco Glass, Inc., Vineland, NJ).
8. Cell lysis buffer: 20 mM Tris-HCl (pH 8.0) and 1 mM EDTA.
9. Sonicator (with microtip).
10. 2X Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer (Invitrogen) and 2-mercaptoethanol (Sigma Chemical Co., St. Louis, MO).
11. SDS-PAGE gel, electrophoresis apparatus, and running buffer (*see Note 7*).
12. Gel stain (e.g., Gelcode<sup>®</sup> Blue from Pierce, Rockford, IL, or PhastGel<sup>™</sup> Blue R from Amersham Biosciences, Piscataway, NJ).
13. Spectrophotometer.
14. 1.5-mL Microcentrifuge tubes.

### 2.3. Large-Scale Cell Growth and Protein Purification

#### 2.3.1. Cell Growth

1. LB broth (*see Subheading 2.1., item 10*).
2. Sterile 500-mL and 4-L baffled-bottom flasks (Bellco Glass, Inc.).
3. Sterile ampicillin solution (100 mg/mL; *see Subheading 2.1., item 10*).
4. A stock solution of 30 mg/mL chloramphenicol in ethanol. Store at  $-20^{\circ}\text{C}$  for up to 6 mo.
5. Competent BL21 cells (Novagen, Madison, WI).
6. 20% (w/v) D(+)-glucose in H<sub>2</sub>O. Filter-sterilize and store at  $4^{\circ}\text{C}$ .
7. IPTG (*see Subheading 2.2., item 4*).
8. A derivative of pDEST-HisMBP that produces a His<sub>6</sub>-MBP fusion protein with an intervening TEV protease recognition site.
9. Shaker/incubator.
10. Spectrophotometer.

#### 2.3.2. Protein Purification

1. ÄKTA Explorer Chromatography System (Amersham Biosciences).
2. Ni-NTA Superflow (Qiagen).
3. Column XK 16/10 (Amersham Biosciences).
4. 0.22- $\mu\text{m}$  Polyethersulfone filter (Corning Inc., Corning, NY).
5. 1 L of 25 mM HEPES (pH 8.0) and 200 mM NaCl (*see Note 8*).

---

**Fig. 2.** Construction of a His<sub>6</sub>-MBP fusion vector using PCR and Gateway<sup>®</sup> cloning technology. The open reading frame of interest is amplified from the template DNA by PCR, using primers N1, N2, and C. Primers N1 and C are designed to base pair to the 5' and 3' ends of the coding region, respectively, and contain unpaired 5' extensions as shown. Primer N2 base pairs with the sequence that is complementary to the unpaired extension of primer N1. The final PCR product is recombined with the pDONR201 vector to generate an entry clone, via the BP reaction. This entry clone is subsequently recombined with pDEST-HisMBP using LR Clonase to yield the final His<sub>6</sub>-MBP fusion vector.

6. 1 L of 25 mM HEPES (pH 8.0), 200 mM NaCl, and 25 mM imidazole (see **Note 8**).
7. 1 L of 25 mM HEPES (pH 8.0), 200 mM NaCl, 250 mM imidazole (see **Note 8**).
8. Benzamidine hydrochloride: hydrate (Sigma).
9. Complete EDTA-free protease inhibitor cocktail tablets (Roche Diagnostics GmbH, Mannheim, Germany).
10. SDS-PAGE gel, 2X SDS-PAGE sample buffer, electrophoresis apparatus, and running buffer (see **Subheading 2.2., items 10–12**).
11. His<sub>6</sub>-TEV protease (**5**).

### 3. Methods

#### 3.1. Construction of His<sub>6</sub>-MBP Fusion Vectors by Recombinational Cloning

The Gateway recombinational cloning system is based on the site-specific recombination reactions that mediate the integration and excision of bacteriophage  $\lambda$  into and from the *E. coli* chromosome, respectively. For detailed information about this system, the investigator is encouraged to consult the technical literature supplied by Invitrogen (<http://www.invitrogen.com>).

##### 3.1.1. pDEST-HisMBP

To utilize the Gateway system for the production of His<sub>6</sub>-MBP fusion proteins, one must first construct or obtain a suitable “destination vector.” Currently there are no commercial sources for such vectors. An example of a destination vector that can be used to produce His<sub>6</sub>-MBP fusion proteins (pDEST-HisMBP), which is available from the authors, is shown in **Fig. 1**. pDEST-HisMBP was constructed by inserting an in-frame hexahistidine sequence between codons 3 and 4 of MBP in pKM596 (**6**).

The Gateway cloning cassette in pDEST-HisMBP carries a gene encoding the DNA gyrase poison CcdB, which provides a negative selection against the destination vector, the donor vector, and various recombination intermediates so that only the desired recombinant is obtained when the end products of the recombinational cloning reaction are transformed into *E. coli* and grown in the presence of ampicillin. pDEST-HisMBP and other vectors that carry the *ccdB* gene must be propagated in a host strain with a *gyrA* mutation (e.g., *E. coli* DB3.1) that renders the cells immune to the action of CcdB.

##### 3.1.2. Gateway Cloning Protocol

The easiest way to construct a His<sub>6</sub>-MBP fusion vector by recombinational cloning is to start with a PCR amplicon wherein the open reading frame (ORF) of interest is bracketed by *attB* 1 and *attB* 2 recombination sites on its N- and C-termini, respectively, which can be generated by amplifying the target ORF with PCR primers that include the appropriate *attB* sites as 5' unpaired exten-

sions (see Fig. 2). The 3' ends of the PCR primers are chosen so that they will be able to form 20–25 bp with the template DNA. A recognition site for TEV protease is incorporated between the N-terminus of the ORF and the *attB1* site in this PCR amplicon, enabling the passenger protein to be separated from the N-terminal His<sub>6</sub>-MBP tag. Although this can be accomplished with a single, long N-terminal PCR primer for each gene, we normally perform the PCR amplification with two overlapping N-terminal primers instead, as outlined in Fig. 2. Two gene-specific primers (N1 and C) are required for each ORF. The C-terminal primer (C) includes the *attB2* recombination site as a 5' extension. The 5' extension of the N-terminal primer (N1) includes a recognition site for TEV protease. The PCR product generated by these two primers is subsequently amplified by primers N2 and C to yield the final product. Primer N2 anneals to the TEV protease recognition site and includes the *attB1* recombination site as a 5' extension. This generic PCR primer can be used to add the *attB1* site to any amplicon that already contains the TEV protease recognition site at its N-terminal end. The PCR reaction is performed in a single step by adding all three primers to the reaction at once (see Note 9). To favor the accumulation of the desired product, the *attB*-containing primers are used at typical concentrations for PCR but the concentration of the gene-specific N-terminal primer (N1) is 20-fold lower.

1. The PCR reaction mix is prepared as follows (see Note 10): 1  $\mu$ L template DNA (~10 ng/ $\mu$ L), 10  $\mu$ L thermostable DNA polymerase 10X reaction buffer, 16  $\mu$ L dNTP solution (1.25 mM each), 2.5  $\mu$ L primer N1 (~1  $\mu$ M, or 13 ng/ $\mu$ L for a 40mer), 2.5  $\mu$ L primer N2 (~20  $\mu$ M, or 260 ng/ $\mu$ L for a 40mer), 2.5  $\mu$ L primer C (~20  $\mu$ M, or 260 ng/ $\mu$ L for a 40mer), 1  $\mu$ L thermostable DNA polymerase, and 65.5  $\mu$ L H<sub>2</sub>O (to 100  $\mu$ L total volume).
2. The reaction is placed in the PCR thermal cycler with the following program: initial melt for 5 min at 94°C; 30 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 60 s (see Note 11); hold at 4°C.
3. Purification of the PCR amplicon by agarose gel electrophoresis (see Note 2) is recommended to remove *attB* primer dimers.
4. To create the His<sub>6</sub>-MBP fusion vector, the PCR product is recombined first into pDONR201 to yield an entry clone intermediate (BP reaction), and then into pDEST-HisMBP (LR reaction; see Note 12).
  - a. Add to a microcentrifuge tube on ice: 300 ng of the PCR product in TE or H<sub>2</sub>O, 300 ng of pDONR201 DNA, 4  $\mu$ L of 5X BP reaction buffer, and enough TE to bring the total volume to 16  $\mu$ L. Mix well.
  - b. Thaw BP Clonase enzyme mix on ice for 2 min and then vortex briefly for 2 s twice (see Note 13).
  - c. Add 4  $\mu$ L of BP Clonase enzyme mix to the components in step a and vortex briefly twice.
  - d. Incubate the reaction at room temperature for at least 4 h (see Note 14).

- e. Add to the reaction: 1  $\mu\text{L}$  of 0.75 M NaCl, 3  $\mu\text{L}$  (ca. 450 ng) of the destination vector (pDEST-HisMBP), and 6  $\mu\text{L}$  of LR Clonase enzyme mix (see **Note 13**). Mix by vortexing briefly.
  - f. Incubate the reaction at room temperature for 3–4 h.
  - g. Add 2.5  $\mu\text{L}$  of the proteinase K stop solution and incubate for 10 min at 37°C.
  - h. Transform 2  $\mu\text{L}$  of the reaction into 50  $\mu\text{L}$  of chemically competent DH5 $\alpha$  cells (see **Note 3**).
  - i. Pellet the cells by centrifugation, gently resuspend pellet in 100–200  $\mu\text{L}$  of LB broth and spread on an LB agar plate containing 100  $\mu\text{g}/\text{mL}$  ampicillin, the selective marker for pDEST-HisMBP (see **Fig. 1**). Incubate the plate at 37°C overnight (see **Note 15**).
5. Plasmid DNA is isolated from saturated cultures started from individual ampicillin-resistant colonies and screened by PCR, using the gene-specific primers N1 and C, to confirm that the clones have the expected structure. Alternatively, plasmids can be purified and screened by conventional restriction digests using appropriate enzymes. At this stage, we routinely sequence putative clones to ensure that there are no PCR-induced mutations.

### 3.2. Pilot Expression Experiment

Prior to large-scale cell growth and purification, the fusion protein is overproduced on a small scale to assess its solubility. The amount of fusion protein in the soluble fraction of the crude cell lysate is compared by SDS-PAGE with the total amount of fusion protein in the cells, and the results are analyzed by visual inspection of the stained gel. In a parallel experiment, the fusion protein is cleaved *in vivo* to ascertain whether or not it is an efficient substrate for TEV protease and to evaluate the solubility of the passenger protein after it is cleaved from the His<sub>6</sub>-MBP tag. If the passenger protein remains soluble after intracellular processing, then it is also likely to be soluble after the fusion protein has been purified and processed *in vitro*. Conversely, poor solubility after intracellular processing indicates that troubleshooting will be required before production can be scaled up (see **Subheading 3.2.6**).

#### 3.2.1. Selecting a Host Strain of *E. coli*

The production of TEV protease from the expression vector pRK603 (4) is initiated by adding anhydrotetracycline to the cell culture. This allows it to be regulated independently of the IPTG-inducible His<sub>6</sub>-MBP fusion vector, which is important because sometimes delaying the induction of TEV protease until the fusion protein substrate has had time to accumulate in the cells results in greater solubility of the passenger protein after cleavage (4,6). To achieve regulated expression of TEV protease, the *in vivo* processing experiment must be performed in a strain of *E. coli* that produces the Tet repressor, such as BL21PRO or DH5 $\alpha$ PRO (Clontech). We prefer BL21Pro because of its robust

growth characteristics and the fact that it lacks two proteases (Lon and OmpT) that are present in most *E. coli* K12 strains.

### 3.2.2. Protein Expression

1. Transform competent BL21PRO or DH5 $\alpha$ PRO cells that already contain pRK603 (see **Notes 5** and **6**) with the His<sub>6</sub>-MBP fusion protein expression vector and spread them on an LB agar plate containing 100  $\mu$ g/mL ampicillin and 25  $\mu$ g/mL kanamycin. Incubate the plate overnight at 37°C.
2. Inoculate 2–5 mL of LB medium containing 100  $\mu$ g/mL ampicillin and 25  $\mu$ g/mL kanamycin in a culture tube or shake-flask with a single colony from the plate. Grow to saturation overnight at 37°C with shaking.
3. The next morning, inoculate 50 mL of the same medium in a 250-mL baffled-bottom flask with 0.5 mL of the saturated overnight culture.
4. Grow the cells at 37°C with shaking to mid-log phase (OD<sub>600nm</sub> ~0.5).
5. Add IPTG (1 mM final concentration) and adjust the temperature to 30°C (see **Note 16**).
6. After 2 h, divide the culture into two separate flasks (ca. 20 mL in each). Label one flask “+” and the other “–”.
7. Add anhydrotetracycline to the “+” flask (100 ng/mL final concentration).
8. After 2 more h, measure the OD<sub>600nm</sub> of the cultures (dilute cells 1:10 in LB to obtain an accurate reading). An OD<sub>600nm</sub> of about 3.5 is normal, although lower densities are possible. If the density of either culture is much lower than this, it may be necessary to adjust the volume of the samples that are analyzed by SDS-PAGE.
9. Transfer 10 mL of each culture to a 15-mL conical centrifuge tube and pellet the cells by centrifugation at 4000g at 4°C.
10. Resuspend the cell pellets in 1 mL of lysis buffer and then transfer the suspensions to a 1.5-mL microcentrifuge tube.
11. Store the cell suspensions at –80°C overnight. Alternatively, the cells can be disrupted immediately by sonication (after freezing and thawing) and the procedure continued without interruption, as described in **Subheading 3.2.3**.

### 3.2.3. Sonication and Sample Preparation

1. Thaw the cell suspensions at room temperature, then place them on ice.
2. Lyse the cells by sonication (see **Note 17**).
3. Prepare samples of the total intracellular protein from the “+” and “–” cultures (T+ and T–, respectively) for SDS-PAGE by mixing 50  $\mu$ L of each sonicated cell suspension with 50  $\mu$ L of 2X SDS-PAGE sample buffer containing 10% (v/v) 2-mercaptoethanol.
4. Pellet the insoluble cell debris (and proteins) by centrifuging the sonicated cell suspension from the each culture at maximum speed in a microcentrifuge for 10 min at 4°C.
5. Prepare samples of the soluble intracellular protein from the “+” and “–” cultures (S+ and S–, respectively) for SDS-PAGE by mixing 50  $\mu$ L of each supernatant

from **step 4** with 50  $\mu\text{L}$  of 2X SDS-PAGE sample buffer containing 10% (v/v) 2-mercaptoethanol.

### 3.2.4. SDS-PAGE

We typically use precast Tris-Glycine or NuPAGE gradient gels for SDS-PAGE to assess the yield and solubility of MBP fusion proteins (*see Note 7*). Of course, the investigator is free to choose any appropriate SDS-PAGE formulation, depending on the protein size and laboratory preference.

1. Heat the T $-$ , T $+$ , S $-$ , and S $+$  protein samples at 90°C for about 5 min and then spin them at maximum speed in a microcentrifuge for 5 min.
2. Assemble the gel in the electrophoresis apparatus, fill it with SDS-PAGE running buffer, load the samples (10  $\mu\text{L}$  each), and carry out the electrophoretic separation according to standard lab practices. T and S samples from each culture (“+” and “-”) are loaded in adjacent lanes to allow easy assessment of solubility. Molecular weight standards may also be loaded on the gel, if desired.
3. Stain the proteins in the gel with GelCode Blue reagent, PhastGel Blue R, or a suitable alternative.

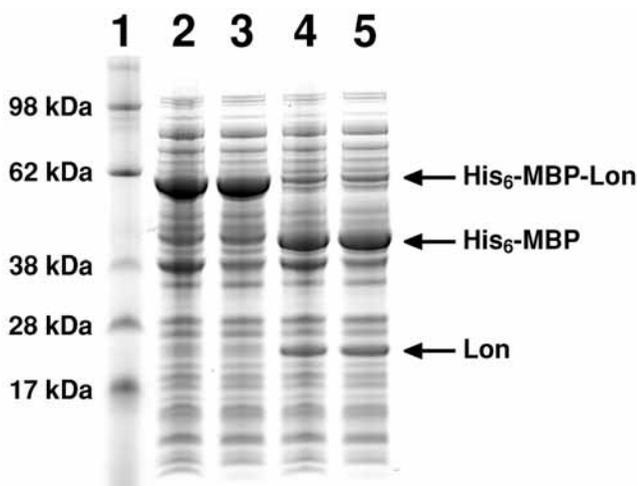
### 3.2.5. Interpreting the Results

The MBP fusion protein should be readily identifiable in the T $-$  sample after the gel is stained because it will normally be the most abundant protein in the cells. Molecular weight standards can also be used to corroborate the identity of the fusion protein band. If the S $-$  sample contains a similar amount of the fusion protein, this indicates that it is highly soluble in *E. coli*. On the other hand, if little or no fusion protein is observed in the S $-$  sample, then this is an indicator of poor solubility. Of course, a range of intermediate states is also possible.

If the fusion protein is an efficient substrate for TEV protease, then little of it will be present in the T $+$  and S $+$  samples. Instead, one should observe a prominent band at ca. 42 kDa that corresponds to the His<sub>6</sub>-MBP moiety and another prominent band migrating with the expected mobility of the passenger protein. If the fusion protein is a poor substrate for the protease, then the “+” samples will look similar to the “-” samples.

If the passenger protein is soluble after it is released from His<sub>6</sub>-MBP, then a similar amount will be present in the T $+$  and S $+$  lanes. On the other hand, some or all of the passenger protein may precipitate at this stage. If a substantial fraction of the passenger protein is insoluble, then troubleshooting may be necessary. Alternatively, an acceptable yield might still be obtained by scaling up cell production.

An example of a pilot expression experiment is shown in **Fig. 3**. In this case the fusion protein (MBP-Lon) was highly soluble and readily cleaved in vivo



**Fig. 3.** Intracellular processing of a His<sub>6</sub>-MBP fusion protein by TEV protease. The catalytic domain of *Escherichia coli* Lon protease was expressed from a derivative of pDEST-HisMBP in BL21PRO cells that also contained the TEV protease expression vector pRK603 as described (see **Subheading 3.2.**). Lane 1: molecular weight standards. Lane 2: T-. Lane 3: S-. Lane 4: T+. Lane 5, S+ (see **Subheading 3.2., steps 4 and 5.**)

by TEV protease. Note also that the Lon protease catalytic domain remained soluble after it was cleaved from the dual His<sub>6</sub>-MBP tag.

### 3.2.6. Troubleshooting

Not every MBP fusion protein will be highly soluble. However, solubility usually can be increased by reducing the temperature of the culture from 30 to 25°C or even lower during the time that the fusion protein is accumulating in the cells (i.e., after the addition of IPTG). In some cases, the improvement can be quite dramatic. It may also be helpful to reduce the IPTG concentration to a level that will result in partial induction of the fusion protein. The appropriate IPTG concentration must be determined empirically, but is generally in the range of 10 to 20  $\mu$ M. Under these conditions, longer induction times (18–24 h) are required to obtain a reasonable yield of fusion protein.

If the fusion protein is a poor substrate for TEV protease *in vivo*, then the same is likely to be true *in vitro*. However, in most cases it is still possible to obtain a sufficient quantity of the pure passenger protein by scaling up production (e.g., from 4 to 6 L of cells or more). In especially problematic cases, the efficiency of the protease digest can be improved by inserting additional amino acid residues between the TEV protease recognition site and the N-terminus of

the passenger protein. We have used both polyglycine and a FLAG-tag epitope in this position with good results (7,8).

Occasionally, a passenger protein may accumulate in a soluble but biologically inactive form after intracellular processing of an MBP fusion protein. Exactly how and why this occurs is unclear, but we suspect that fusion to MBP somehow enables certain proteins to evolve into kinetically trapped, folding intermediates that are no longer susceptible to aggregation. Therefore, although solubility after intracellular processing is a useful indicator of a passenger protein's folding state in most cases, it is not absolutely trustworthy. For this reason, we strongly recommend that a biological assay be employed (if available) at an early stage to confirm that the passenger protein is in its native conformation.

### 3.3. Large-Scale Cell Growth and Protein Purification

#### 3.3.1. Cell Growth

1. Transform competent BL21 cells (Novagen) with the His<sub>6</sub>-MBP-passenger expression vector and select transformants on LB agar plates containing 100 µg/mL ampicillin. Alternatively, if the passenger protein ORF contains codons that are rarely used in *E. coli* (see **Note 6**) then transform BL21 CodonPlus™ RIL (Stratagene) or Rosetta™ (Novagen) cells instead, and select transformants on LB agar plates containing both ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL).
2. Inoculate 100 mL of LB broth in a 500-mL baffle-bottom shake flask (Bellco Glass or the equivalent) containing the appropriate antibiotic(s) with a single drug-resistant colony from the transformation in **step 1**. Shake overnight at 37°C.
3. Add 10 mL of the saturated overnight culture to 1 L of fresh LB broth plus antibiotic(s) in a 4-L baffle-bottom shake flask. To ensure that there will be an adequate yield of pure protein at the end of the process, we ordinarily grow at least 4 L of cells at a time. Sterile glucose can be added to 0.2% to increase biomass production. Shake the flask(s) at 37°C (ca. 250 rpm) until the cells reach mid-log phase (OD<sub>600nm</sub> ~0.5).
4. Shift the temperature to 30°C (see **Note 18**) and then add IPTG to a final concentration of 1 mM. Continue shaking for 4–6 h.
5. Recover the cells by centrifugation. Freeze the cell pellet(s) at –80°C.

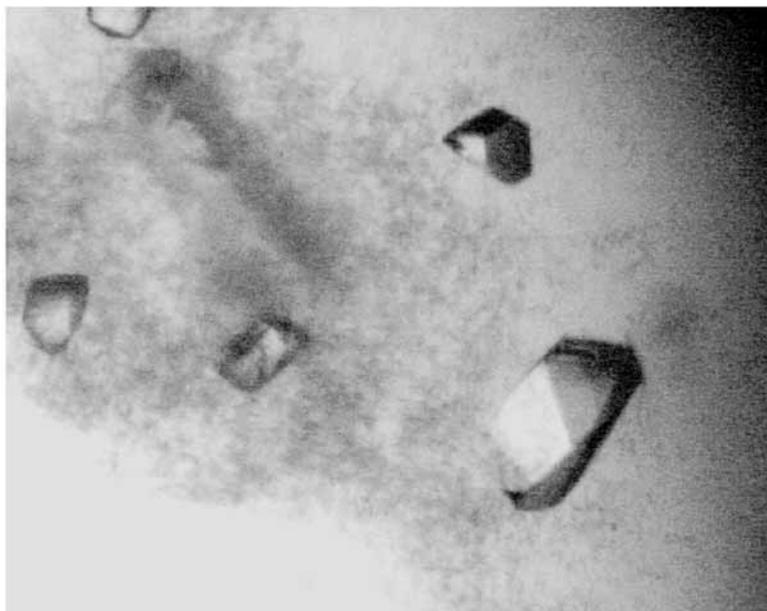
#### 3.3.2. Protein Purification

1. All procedures are performed at 4–8°C. Thaw the cell pellet(s) on ice and suspend in ice-cold 25 mM HEPES (pH 8.0), 200 mM sodium chloride, 25 mM imidazole buffer containing protease inhibitors, using at least 10 mL/g of wet weight of cell paste (see **Note 8**). Lyse the cells and clarify the crude cell extract by high-speed centrifugation and filtration (see **Note 19**).
2. Apply the supernatant to a column of Ni-NTA resin (Qiagen) equilibrated in 25 mM HEPES (pH 8.0), 200 mM sodium chloride, and 25 mM imidazole buffer without protease inhibitors (see **Notes 8** and **20**). Wash the column with this buffer until a



**Fig. 4.** Generic purification of *Escherichia coli* Lon protease catalytic domain (see **Subheading 3.3.**). Lane 1: molecular weight standards. Lane 2: Soluble cell extract. Lane 3: Peak fractions from the first round of immobilized metal affinity chromatography (IMAC), eluted with an imidazole gradient. Lane 4: tobacco etch virus protease digest. Lane 5: flow-through fractions after second round of IMAC. Lane 6: material absorbed to Ni-NTA column during second round of IMAC.

- stable baseline is reached and then elute the bound fusion protein (His<sub>6</sub>-MBP-passenger) with a linear gradient over 10 column volumes into 25 mM HEPES (pH 8.0), 200 mM sodium chloride, and 250 mM imidazole buffer (see **Note 8**). The fusion protein usually elutes between 100 and 150 mM imidazole. The purity of the fusion protein at this stage is typically in the range of 70 to 80% (see **Fig. 4**, lane 3).
3. To reduce the imidazole concentration to ca. 25 mM, the peak fractions containing the fusion protein are pooled and diluted with 25 mM HEPES (pH 8.0) and 200 mM sodium chloride buffer that does not contain protease inhibitors or imidazole (see **Notes 8** and **21**). The concentration of the fusion protein is estimated by measuring the A<sub>280</sub>.
  4. His<sub>6</sub>-TEV protease is added to the pooled, diluted fractions from **step 3** and incubated overnight at 4°C (see **Note 22**). The progress of the cleavage reaction can be monitored by SDS-PAGE (see **Fig. 4**, lane 4).
  5. The His<sub>6</sub>-TEV protease-treated pool from **step 4** is applied to a second Ni-NTA column under the same conditions used during application to the first column (see **Note 23**). Pure passenger protein passes through the Ni-NTA column while residual undigested fusion protein, His<sub>6</sub>-MBP, His<sub>6</sub>-TEV protease, and any contaminants are retained (see **Fig. 4**, lanes 5 and 6). The purity of the passenger protein at this stage



**Fig. 5.** Crystals of *Escherichia coli* Lon protease catalytic domain (ca.  $0.2 \times 0.1 \times 0.1$  mm) obtained after the protein was purified as described (see **Subheading 3.3.** and **Note 24**). Space group:  $P3_2$ . Diffraction limit: 1.75 Å (beam line X-9B at NSLS, Brookhaven, NY).

is greater than 90% and is of sufficient quality for crystallization trials (see **Note 24** and **Fig. 5**).

#### 4. Notes

1. We recommend a proofreading polymerase such as *Pfu* Turbo (Stratagene), Platinum *Pfx* (Invitrogen), or Deep Vent (New England Biolabs) to minimize the occurrence of mutations during PCR.
2. We typically purify fragments by horizontal electrophoresis in 1% agarose gels run in TAE buffer (40 mM Tris-acetate, pH 8.0, and 1 mM EDTA). It is advisable to use agarose of the highest possible purity (e.g., Seakem-GTG from FMC BioPolymer). Equipment for horizontal electrophoresis can be purchased from a wide variety of scientific supply companies. DNA fragments are extracted from slices of the ethidium bromide-stained gel using a QIAquick gel extraction kit (Qiagen) in accordance with the instructions supplied with the product.
3. Any *gyrA*<sup>+</sup> strain of *E. coli* can be used. We prefer competent DH5 $\alpha$  cells (Invitrogen) because they are easy to use and very efficient.
4. We prefer the Wizard miniprep kit (Promega) or the QIAprepSpin miniprep kit (Qiagen), but similar kits can be obtained from a wide variety of vendors.

5. Although any method for the preparation of competent cells can be used (e.g.,  $\text{CaCl}_2$ ) (9), we prefer electroporation because of the high-transformation efficiency that can be achieved. Detailed protocols for the preparation of electrocompetent cells and electrotransformation procedures can be obtained from the electroporator manufacturers (e.g., Bio-Rad, BTX, Eppendorf). Briefly, the cells are grown in 1 L of LB medium (with antibiotics, if appropriate) to mid-log phase ( $\text{OD}_{600} \sim 0.5$ ) and then chilled on ice. The cells are pelleted at  $4^\circ\text{C}$ , resuspended in 1 L of ice-cold  $\text{H}_2\text{O}$  and then pelleted again. After several such washes with  $\text{H}_2\text{O}$ , the cells are resuspended in 3–4 mL of 10% glycerol, divided into 50- $\mu\text{L}$  aliquots, and then immediately frozen in a dry ice/ethanol bath. The electrocompetent cells are stored at  $-80^\circ\text{C}$ . Immediately prior to electrotransformation, the cells are thawed on ice and mixed with 10–100 ng of DNA (e.g., a plasmid vector or a Gateway reaction). The mixture is placed into an ice-cold electroporation cuvet and electroporated according to the manufacturers' recommendations (e.g., a 1.8-kV pulse in a cuvet with a 1-mm gap). One milliliter of SOC medium (9) is immediately added to the cells and they are allowed to grow at  $37^\circ\text{C}$  with shaking (ca. 250 rpm) for 1 h; 5–200  $\mu\text{L}$  of the cells are then spread on an LB agar plate containing the appropriate antibiotic(s).
6. If the ORF encoding the passenger protein contains codons that are rarely used in *E. coli* (<http://www.doe-mbi.ucla.edu/Services/RACC/>), this can adversely affect the yield of an MBP fusion protein. In such cases, it is advisable to introduce an additional plasmid into the host cells that carries the cognate tRNA genes for rare codons. The pRIL plasmid (Stratagene) is a derivative of the p15A replicon that carries the *E. coli* *argU*, *ileY*, and *leuW* genes, which encode the cognate tRNAs for AGG/AGA, AUA, and CUA codons, respectively. pRIL is selected for by resistance to chloramphenicol. In addition to the tRNA genes for AGG/AGA, AUA, and CUA codons, the pRARE accessory plasmid in the Rosetta host strain (Novagen) also includes tRNAs for the rarely used CCC and GGA codons. Like pRIL, the pRARE plasmid is a chloramphenicol-resistant derivative of the p15A replicon. Both of these tRNA accessory plasmids are compatible with derivatives of pDEST-HisMBP. On the other hand, they are incompatible with the vector pRK603 that we use for intracellular processing experiments (see **Subheading 3.2.**). Nevertheless, because pRK603 and the tRNA accessory plasmids have different antibiotic-resistance markers, it is possible to force cells to maintain both plasmids by simultaneously selecting for kanamycin and chloramphenicol resistance. Alternatively, the kanamycin-resistant TEV protease expression vector pKM586, a pRK603 derivative with the replication machinery of a pSC101 replicon, can be maintained as stable in conjunction with p15A-type tRNA plasmids.
7. We find it convenient to use precast gels for SDS-PAGE gels (e.g., 1.0-mm thick, 10 well, 10–20% Tris-glycine gradient), running buffer, and electrophoresis supplies from Invitrogen.
8. Buffers compatible with IMAC (HEPES, phosphate, and Tris buffers (25–50 mM, pH 7.0–8.0) must be used. We include sodium chloride (100–500 mM) and a low concentration of imidazole (25 mM) in our cell lysis and IMAC equilibra-

tion buffers to decrease nonspecific adsorption to the resin. In addition, protease inhibitors (1 mM PMSF or AEBSF, or complete EDTA-free protease inhibitor cocktail tablets and 1 mM benzamidine hydrochloride) are added to the cell lysis buffer unless they are contraindicated (e.g., for the purification of an *active* recombinant protease, the inhibitors may have to be excluded). We always avoid EDTA and other divalent metal chelators in our cell lysis and IMAC buffers. The 25 mM HEPES (pH 8.0), 200 mM sodium chloride, 25 mM imidazole buffer used for cell lysis, and IMAC is prepared by mixing 5.96 g of HEPES (BioChemika grade, Fluka Chemical Corp.), 11.69 g of sodium chloride, and 1.70 g of imidazole (BioChemika grade, Fluka Chemical Corp.) with distilled water to ca. 950 mL. The solution is adjusted to pH 8.0 with 10 N sodium hydroxide and the volume increased to 1 L with H<sub>2</sub>O. The buffer is filtered through a 0.22- $\mu$ m polyethersulfone membrane and stored at 4°C. Protease inhibitors (for cell lysis buffer only) are added immediately before use (7.83 mg benzamidine hydrochloride and one tablet of complete EDTA-free protease inhibitor cocktail per 50 mL of buffer). The 25 mM HEPES (pH 8.0) and 200 mM sodium chloride buffer is prepared in the same manner, except that the imidazole is omitted. The 25 mM HEPES (pH 8.0), 200 mM sodium chloride, and 250 mM imidazole buffer used for IMAC is prepared by mixing 5.96 g of HEPES, 11.69 g of sodium chloride, and 17 g of imidazole with distilled water to ca. 950 mL. The solution is adjusted to pH 8.0 with hydrochloric acid and then the volume is increased to 1 L with H<sub>2</sub>O. The buffer is filtered through a 0.22- $\mu$ m polyethersulfone membrane and stored at 4°C.

9. Alternatively, the PCR reaction can be performed in two separate steps, using primers N1 and C in the first step and primers N2 and C in the second step. The PCR amplicon from the first step is used as the template for the second PCR. All primers are used at the typical concentrations for PCR in the two-step protocol.
10. The PCR reaction can be modified in numerous ways to optimize results, depending on the nature of the template and primers. See **ref. 9** (Vol. 2, Chapter 8) for more information.
11. PCR cycle conditions can also be varied. For example, the extension time should be increased for especially long genes. A typical rule-of-thumb is to extend for 60 s/kb of DNA.
12. This “one-tube” Gateway protocol bypasses the isolation of an “entry clone” intermediate. However, the entry clone may be useful if the investigator intends to experiment with additional Gateway destination vectors, in which case the BP and LR reactions can be performed sequentially in separate steps; detailed instructions are included with the Gateway PCR kit. Alternatively, entry clones can easily be regenerated from expression clones via the BP reaction, as described in the instruction manual.
13. Clonase enzyme mixes should be thawed quickly on ice and then returned to the –80°C freezer as soon as possible. It is advisable to prepare multiple aliquots of the enzyme mixes the first time that they are thawed in order to avoid repeated freeze-thaw cycles.

14. At this point, we remove a 5- $\mu$ L aliquot from the reaction and add it to 0.5  $\mu$ L of proteinase K stop solution. After 10 min at 37°C, we transform 2  $\mu$ L into 50  $\mu$ L of competent DH5 $\alpha$  cells (see **Note 3**) and spread 100–200  $\mu$ L on an LB agar plate containing 25  $\mu$ g/mL kanamycin, the selective marker for pDONR201. From the number of colonies obtained, it is possible to gage the success of the BP reaction. Additionally, entry clones can be recovered from these colonies in the event that no transformants are obtained after the subsequent LR reaction.
15. If very few or no ampicillin-resistant transformants are obtained after the LR reaction, the efficiency of the process can be improved by incubating the BP reaction overnight.
16. 30°C is the optimum temperature for TEV protease activity. At 37°C, the protease does not fold properly in *E. coli* and little processing will occur. Reducing the temperature also improves the solubility of some MBP fusion proteins.
17. We routinely break cells in a 1.5-mL microcentrifuge tube on ice with two or three 30-s pulses using a VCX600 sonicator (Sonics and Materials, Inc.) with a microtip at 38% power. The cells are cooled on ice between pulses.
18. We have found that decreasing the induction temperature to 30°C increases the quality and solubility of the fusion protein without significantly decreasing the yield, especially in the presence of glucose.
19. For large-scale protein purification we routinely break cells using a Gaulin APV fluidizer at 10,000–11,000 psi for two to three rounds. Other homogenization techniques such as French press, sonication, or manual shearing should yield comparable results. Centrifugation of the disrupted cell suspension for at least 30 min at 30,000g is recommended. Filtration through a 0.2- $\mu$ m polyethersulfone or cellulose acetate membrane is helpful to remove residual particulates and fines prior to chromatography.
20. We use Ni-NTA Superflow (Qiagen) and an Amersham Biosciences ÄKTA Explorer chromatography system. Ten- to twenty-five-milliliter columns are employed, depending on the amount of fusion protein produced by the cells. A properly poured 10 mL Ni-NTA Superflow column (in an Amersham Biosciences column XK16/20) can be run at 2–4 mL/min (backpressure less than 0.4 MPa) and will bind up to 100 mg of fusion protein. If a chromatography system is not available, the IMAC can be performed using a peristaltic pump or manually by gravity. If the latter is used, Ni-NTA agarose should be substituted for Superflow and the elution performed with step increases of imidazole in 25-mM increments. Binding and elution profiles can be monitored spectrophotometrically at 280 nm and by SDS-PAGE. Care must be taken to properly zero the spectrophotometer because imidazole has significant absorption in the UV range.
21. It is convenient to reduce the imidazole concentration to 25 mM at this step in preparation for the second round of IMAC. If the volume is too large, it can be reduced using any commercially available concentrating units (e.g., an Amicon Stir-Cell Concentrator with a YM membrane). Reduction of volume is not critical, however, because IMAC and the subsequent cleavage reaction (see **Note 22**) are insensitive to this variable.

22. His<sub>6</sub>-TEV protease is active in all buffers compatible with IMAC. We use 1 mg of protease per ca. 150 mg of fusion protein (estimated by measuring the A<sub>280</sub>). Using more His<sub>6</sub>-TEV protease than required will have no deleterious effect. Volumes as large as 500 mL can be used. Incubations can be performed between 4 and 30°C with equivalent results, although an overnight incubation at 4°C is most convenient. His<sub>6</sub>-TEV protease can be purchased from Invitrogen. However, for large-scale applications it is far more cost effective to overproduce and purify the enzyme in-house as described (5).
23. A Ni-NTA column of equal or greater volume relative to the first column is required. After dilution of the imidazole to 25 mM, all other buffer components should be the same. In instances where saturation of the Ni-NTA resin during the first round of IMAC occurs, then a column of greater volume is recommended to avoid contamination of the passenger protein emerging in the column effluent during the second round of IMAC.
24. Although the passenger protein is typically free of contaminants after the second IMAC step, it may still exist as a mixture of oligomeric forms including some high molecular weight aggregates. For this reason, we recommend that size exclusion (gel filtration) chromatography be employed as a polishing step. This is also a good way to exchange the protein into an appropriate buffer for crystallization trials. Additionally, we recommend that the molecular weight of the final product be verified by mass spectrometry if possible.

## Acknowledgments

The authors wish to express their gratitude to Dr. Istvan Botos for providing us with a photograph of Lon protease crystals (*see Fig. 5*). The authors also wish to thank Rachel Kapust, Karen Routzahn, and Jeff Fox for their valuable contributions to the development of these methods and Invitrogen for granting us early access to Gateway Cloning technology.

## References

1. Kapust, R. B. and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674.
2. Fox, J. D., Routzahn, K. M., Bucher, M. H., and Waugh, D. S. (2003) Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Letts.* **537**, 53–57.
3. Routzahn, K. M. and Waugh, D. S. (2002) Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. *J. Struct. Funct. Genomics* **2**, 83–92.
4. Kapust, R. B. and Waugh, D. S. (2000) Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* **19**, 312–318.
5. Kapust, R. B., Tozser, J., Fox, J. D., et al. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic efficiency. *Protein Eng.* **14**, 993–1000.

6. Fox, J. D. and Waugh, D. S. (2003) Maltose-binding protein as a solubility enhancer. *Methods Mol. Biol.* **205**, 99–117.
7. Evdokimov, A. G., Tropea, J. E., Routzahn, K. M., Copeland, T. D., and Waugh, D. S. (2001) Structure of the N-terminal domain of *Yersinia pestis* YopH at 2.0 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 793–799.
8. Evdokimov, A. G., Tropea, J. E., Routzahn, K. M., and Waugh, D. S. (2002) Three-dimensional structure of the type III secretion chaperone SycE from *Yersinia pestis*. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 398–406.
9. Sambrook, J. and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.



## Cloning, Production, and Purification of Proteins for a Medium-Scale Structural Genomics Project

Sophie Quevillon-Cheruel, Bruno Collinet, Lionel Trésaugues, Philippe Minard, Gilles Henckes, Robert Aufrère, Karine Blondeau, Cong-Zhao Zhou, Dominique Liger, Nabila Bettache, Anne Poupon, Ilham Aboulfath, Nicolas Leulliot, Joël Janin, and Herman van Tilbeurgh

### Summary

The South-Paris Yeast Structural Genomics Pilot Project (<http://www.genomics.eu.org>) aims at systematically expressing, purifying, and determining the three-dimensional structures of *Saccharomyces cerevisiae* proteins. We have already cloned 240 yeast open reading frames in the *Escherichia coli* pET system. Eighty-two percent of the targets can be expressed in *E. coli*, and 61% yield soluble protein. We have currently purified 58 proteins. Twelve X-ray structures have been solved, six are in progress, and six other proteins gave crystals. In this chapter, we present the general experimental flowchart applied for this project. One of the main difficulties encountered in this pilot project was the low solubility of a great number of target proteins. We have developed parallel strategies to recover these proteins from inclusion bodies, including refolding, coexpression with chaperones, and an in vitro expression system. A limited proteolysis protocol, developed to localize flexible regions in proteins that could hinder crystallization, is also described.

**Key Words:** Yeast proteins; protein expression; structural genomics; inclusion bodies; co-expression.

### 1. Introduction

Structural genomics aims at the systematic structure determination of proteins, driven either by structural and/or functional objectives (1). The principal goals of the South-Paris Yeast Pilot Project are to express, purify, and systematically determine the structure of soluble single-domain proteins of the yeast *Saccharomyces cerevisiae* (<http://www.genomics.eu.org> [2]). At the present stage 240 yeast open reading frames (ORFs) have been cloned using a standard

protocol in a unique expression system, with constructs containing a hexahistidine (His<sub>6</sub>)-tag at the 3' end of the target genes. In a single-pass experiment, 82% of these could be expressed in *Escherichia coli*, and 61% were soluble. We have currently purified 58 proteins. Twelve X-ray structures have been solved, six are in progress, and six additional protein crystals are being optimized. The resolution of two structures by nuclear magnetic resonance (NMR) is in progress.

One of the main tasks of the project is to set up efficient strategies for (1) the cloning of yeast ORFs, (2) the overexpression in *E. coli* of the corresponding recombinant proteins, and (3) their purification for structural studies. We have adopted a systematic approach that allows us to compare the efficiency of cloning and purification strategies on a large ensemble of proteins, all are prepared using the same protocol. Although cloning and expression were, in general, met with success, the low solubility of a large number of target proteins caused a considerable drop in the overall efficiency of the process that goes from a gene clone to a protein structure. We have, therefore, developed parallel strategies to recover proteins from inclusion bodies, including in vitro refolding or coexpression with chaperones in addition to in vitro expression techniques. A second predicted bottleneck is the low crystallization success rate for otherwise well-behaved and soluble proteins. We also describe here a simple limited proteolysis protocol, which localizes flexible parts of proteins and can be used to design shorter constructs that are more likely to crystallize.

## 2. Materials

1. Purified genomic DNA from yeast S288C, used as starting material for PCR and ORF cloning. This is the strain used for the *S. cerevisiae* genome sequencing project (3).
2. Genomic DNA purification buffer: 100  $\mu$ L of 50 mM Tris-HCl pH 8.0, 20 mM EDTA and 0.6% sodium dodecyl sulfate (SDS).
3. Oligonucleotide primers (MWG Biotech, Roissy CDG, France).
4. DyNAzyme (Finzymes [Ozyme], St. Quentin en Yvelines, France).
5. DNA modification enzymes (Taq DNA polymerase, restriction enzymes, T4 DNA ligase) (New England Biolabs [Ozyme]).
6. Agarose gel equipment.
7. Expression vectors pET (Stratagene [Ozyme]).
8. Vector pCRT7/CT-TOPO (Invitrogen, Cergy Pontoise, France).
9. Plasmid pGKJE3 for overexpression of chaperones (4).
10. *E. coli* strains: XL1-Blue, BL21(DE3)pLysS, Rosetta(DE3)pLysS (Stratagene), C41(DE3), and C43(DE3) (Avidis, St. Beauzire, France).
11. Transformation buffer: 50 mM CaCl<sub>2</sub>, at 4°C.
12. 2xYT (BIO101) and M63-derived minimum media (VWR-Prolabo, Fontenay-sous-Bois, France).
13. LeMaster amino acid mix (5) (final concentration in milligrams per liter): L-Ala 250, L-Arg 290, L-Asp 200, L-Gly 270, L-Cys 17, L-Pro 50, L-Ser 1080, L-Tyr 84, L-His 30, L-Gln 170, and L-Glu 330 (Sigma, St. Quentin Fallavier, France).
14. L-Selenomethionine (Se-Met) (Acros Organics, Noisy le Grand, France).

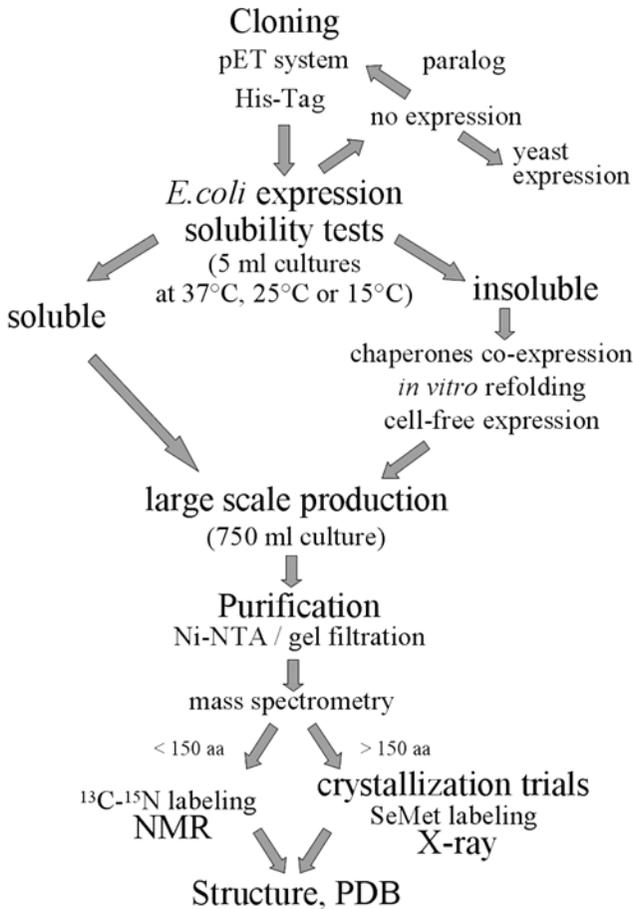
15.  $^{15}\text{N}$  ammonium chloride and  $^{13}\text{C}$  glycerol (Martek Biosciences Corporation, Columbia, MD).
16. Kanamycin, chloramphenicol, tetracyclin (Sigma).
17. IPTG (isopropyl- $\beta$ -D-thio-galactopyranoside) (Sigma).
18. Bacterial incubators (Multitron, Infors, Massy, France) and bioreactors (1.5-L capacity; Applikon System, Les Mureaux, France).
19. Rodhorsyl (VWR-Prolabo).
20. Sonicator.
21. Liquid nitrogen.
22. Ni-nitroloacetic (NTA) resin (Qiagen, Courtaboeuf, France).
23. Lysis buffer: 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, and 5 mM  $\beta$ -mercaptoethanol ( $\beta$ -SH), at 4°C.
24. Wash buffer: lysis buffer supplemented with 20 mM imidazole, at 4°C.
25. Elution buffer: lysis buffer supplemented with 200 mM imidazole, at 4°C.
26. Inclusion bodies resolubilizing buffer: 6 M guanidinium hydrochloride (GndCl), 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 5 mM  $\beta$ -SH, at 4°C.
27. In vitro refolding buffers: 200 mM NaCl or 20% glycerol or 0.6 M arginine or “cocktail buffer” (a mix of 50 mM each of  $\text{CuSO}_4$ ,  $\text{ZnCl}_2$ ,  $\text{MgCl}_2$ ,  $\text{MnCl}_2$ , ADP, NADH, biotin, and thiamine), each at pH 6.5, 7.0, and 8.5.
28. SDS-polyacrylamide gel (PAGE) equipment.
29. Superdex 75 and 200 (16/60) (Amersham Biosciences, Orsay, France), chromatography equipment.
30. Vivaspin 6 and 20 concentrators (Sartorius, Palaiseau, France).
31. Proteases (Sigma).

### 3. Methods

The methods described next follow the chronological steps of our experimental flowchart, shown in [Fig. 1](#) and comprise: (1) the selection of 250 ORFs from the S288C *S. cerevisiae* genome, (2) their cloning in an *E. coli* expression vector, (3) the strategy for rapid testing of their overexpression and solubility, including the comparison of the efficiency of expression strains, (4) the multiple option strategy developed for recovery of inclusion bodies, (5) large-scale production of recombinant proteins, including (6) optimization of synthetic culture medium and labeling of proteins for X-ray or NMR, (7) purification and characterization of proteins, and, finally, (8) the development of a simple and rapid limited proteolysis protocol to localize nonstructured regions within purified proteins.

#### 3.1. Target Selection

Not all proteins are equally well suited for a high-throughput structure determination approach. In order to test technologies and to establish protocols in our structural genomics project, a subset of yeast proteins was selected ([2](#)). Membrane proteins (detected by TMpred [[6](#)]) and multiple-domain proteins were excluded, as well as proteins containing low-complexity regions and coiled coil domains. ORFs were classified into three categories



**Fig. 1.** Simplified experimental flowchart adopted within our Yeast Structural Genomics Project (<http://genomics.eu.org>).

by homology search using sequence comparisons (DARWIN [7], FASTA [8], and BLAST [9]): (1) those that are homologous to a protein of known structure, (2) those that are homologous to proteins whose structure is unknown, and (3) those that do not have a clearly identified homolog. This bioinformatics filter, combined with motif scans such as PRODOM [10], allows the presence of multiple-domain proteins to be detected. The third filter used a motif search algorithm (ProfileScan, PFAM [11]). These tools are actively used by our group to identify domains within large proteins for structural studies. The fourth filter is the search for homologies using multiple and/or iterative alignments (HMMER, PFAM, PSI-BLAST [12]), which are more sensitive than pairwise sequence alignments. The last filter used fold recog-

nition techniques (3DPSSM [13] and FROST [14]), which can be more powerful than standard sequence comparison methods alone, when sequence homology falls below detection level.

The list was further trimmed for the presence of transmembrane segments or the presence of a low-complexity segment (using Hydrophobic Cluster Analysis [15]), “sticky” proteins (with coiled coil regions for instance), or proteins that were already targeted by other structural genomics project.

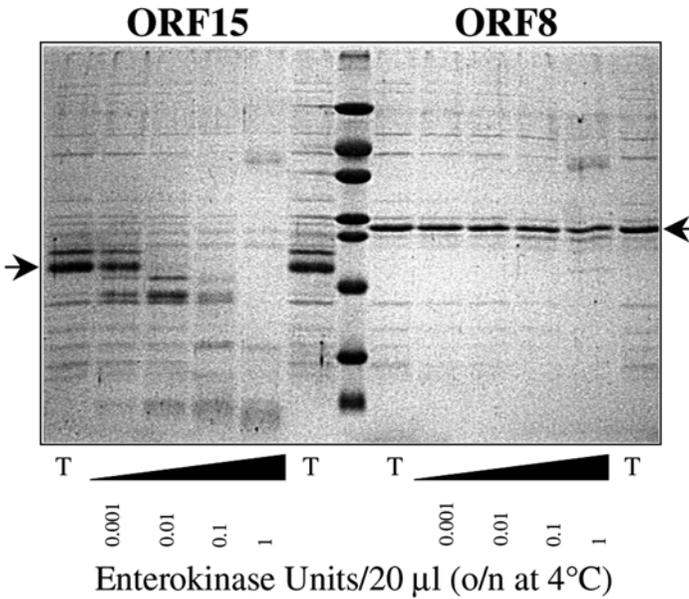
### 3.2. Cloning

A general cloning strategy for a large-scale structural genomics project, based on a first step of PCR amplification from yeast genomic DNA, has to satisfy different criteria including (1) the choice of the prokaryotic expression system (T7 promoter), (2) the cloning strategy (classical restriction/ligation in pET vectors [16] or ligase free cloning method, e.g., “Topo-TA cloning” [17]), (3) the presence or not of an additional copy of the *lacI* gene on the vector, limiting the production of proteins before IPTG induction, and (4) the nature and position of a tag (at the N- or C-terminus of the protein). One also has to decide whether the tag will be cleaved or not after affinity purification (Fig. 2). Some thoughts and considerations around these strategic options are gathered in Notes 1 and 2, and some characteristics of the four vectors used in this study are listed in Table 1.

1. The genomic DNA is purified from a 0.5-mL overnight yeast culture (the cells are resuspended in genomic DNA purification buffer, incubated at 100°C during 10 min, and centrifuged for 10 min at 14,000g. The supernatant is diluted 10-fold in water) and is used as a template for PCR reactions (35 cycles, “hot-start” protocol).
2. The selected ORFs were inserted between a 5'-oligonucleotide containing a *NdeI* site in place of the AUG codon and a 3'-oligonucleotide. This sequence is immediately followed by six histidine codons, a stop codon, and, finally, a *NotI* sequence. The PCR reaction mixture, 50 µL in total, is composed of 0.5 µL of genomic DNA, 0.5 U of DyNAzyme EXT (Finnzymes), 30 pmol of each primer, and 0.01 mM dNTP in the suitable enzyme buffer.
3. The PCR products are purified with the PCR Purification Kit from Qiagen. The digestion with restriction enzymes is performed overnight at 37°C. The inserts are ligated after a second step of purification in a derivative pET-9 or pET-29 vectors (Table 1 and ref. 18). When a *NdeI* site already existed in the selected ORF, the cloning is made in a pET-28 vector between *NcoI* and *NotI* sites.
4. The standard DNA manipulations are made in XL1-Blue strain (Stratagene).
5. The plasmids are purified with the Plasmid Purification Kit from Qiagen.
6. The DNA sequence of the constructs is checked.

### 3.3. Protein Production

The overexpression of the *S. cerevisiae* proteins in *E. coli* should be tested first on a small-scale culture (5 mL), in order to select (1) the best expression



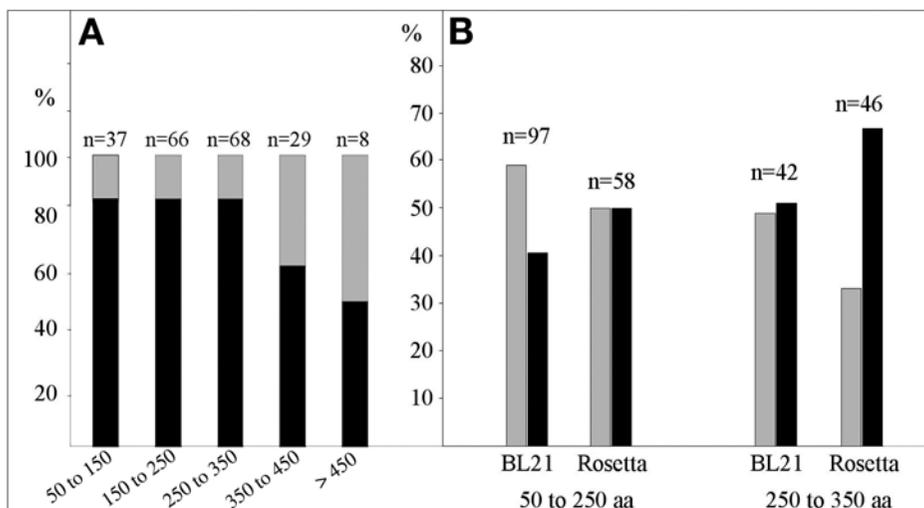
**Fig. 2.** Sodium dodecyl sulfate-polyacrylamide gel analysis of His-tag cleavage by enterokinase digestion on partially purified proteins containing the cleavage site DDDDK between the His-tag and the protein target. Two  $\mu\text{g}$  protein was incubated overnight at  $4^\circ\text{C}$  in  $20\text{-}\mu\text{L}$  reaction mixtures with increasing amounts of protease. The figure illustrates the case of a protein completely degraded by proteolytic treatment (ORF15), and another that gave poor digestion yields (ORF8). T, lane without protease. The arrows point to nondigested proteins.

**Table 1**  
**Characteristics of the Plasmids Used in This Study**

	Company	Promoter	Cloning sites	<i>lacI</i> gene	Resistance	Commercial tag
Derived pET-9 <sup>a</sup>	Novagen	T7 prom	<i>NdeI/NotI</i>	no	Kan	not used
pET-29	Novagen	T7 prom	<i>NdeI/NotI</i>	yes	Kan	not used
pET-28	Novagen	T7 prom	<i>NcoI/NotI</i>	yes	Kan	not used
pCR <sup>R</sup> T7/ CT-TOPO	Invitrogen	T7 prom	TA-cloning	no	Amp	not used

<sup>a</sup>Modified polylinker *NdeI* – *SfiI* – *NotI*.

strain and (2) the optimal temperature of induction for soluble expression. We have routinely used the following two strains: BL21(DE3)pLysS and Rosetta(DE3)pLysS. Rosetta coexpresses tRNAs corresponding to codons



**Fig. 3.** Small-scale expression tests. **(A)** Percentage of expressed (black) vs not expressed (gray) proteins for targets of different sizes (number of amino acids). **(B)** Comparison of the expression efficiency for yeast proteins in two different *Escherichia coli* strains: BL21(DE3)pLysS and Rosetta(DE3)pLysS. Gray, low and medium expression levels; black, high expression levels. The targets are divided into two groups: constructs made of 50–250 amino acids, and those composed of 250–350 amino acids.

rarely used in *E. coli* but frequently used in eukaryotes. Some trials using the C41(DE3) and C43(DE3) strains, originally developed for the expression of membrane or toxic proteins (19), were carried out for proteins not expressed by the aforementioned system, but were not met with success. Nevertheless 80% of protein targets smaller than 350 amino acids were successfully expressed. This percentage drops to 50% for larger proteins (Fig. 3A). An interesting observation resides in the comparison of expression rates in the BL21 and Rosetta strains as a function of the length of the protein targets. Even if some bias exists (the tests were not always performed on the same set of proteins) it is clear that the presence of rare tRNAs in the Rosetta strain favored a higher expression of larger proteins compared to BL21 (see Fig. 3B and Note 3).

### 3.3.1. Transformation of *E. coli* Expression Strains and Protein Induction

1. Competent cells are prepared with standard transformation protocols: heat-shock (cold  $\text{CaCl}_2$ ) or electroporation protocol. It is important to note that the use of freshly transformed cells is mandatory for obtaining an efficient and reproducible expression. Previously transformed expression strains that were kept as glycerol stocks are not reliable as starting expression material. It seems that the expression strain cells transformed with pET vectors are not stable during extended storage at  $-80^\circ\text{C}$ .

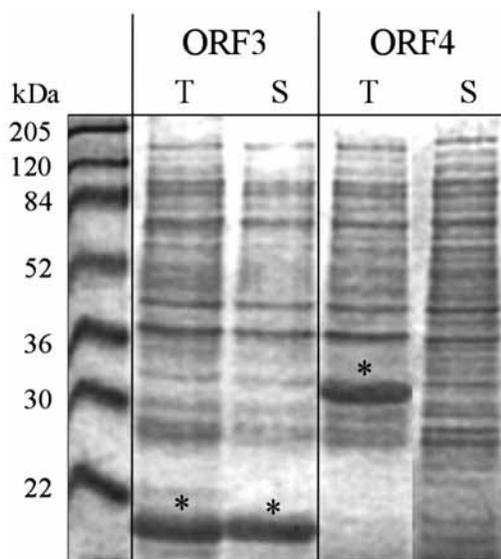
2. From a culture of untransformed expression strain, a stock of competent cells is prepared, aliquoted in 500  $\mu\text{L}$ , and stored at  $-80^{\circ}\text{C}$ . For expression, an aliquot of competent cells is thawed and 50  $\mu\text{L}$  are transformed with about 100 ng of each pure plasmid. The transformed cells are not plated but directly grown overnight at  $37^{\circ}\text{C}$  in 5 mL liquid 2YT medium supplemented with Kan or Amp. This is subsequently used as a preculture.
3. The following day 10 mL of medium are inoculated with 250  $\mu\text{L}$  of preculture and cells grown until  $A_{600\text{nm}}$  is reached at 1. Protein expression is induced with 0.3 mM IPTG. The culture is then divided into two 5-mL aliquots, and each are incubated at 37 or  $25^{\circ}\text{C}$ . Protein expression is allowed to take place for 4 h (or alternatively overnight when a lower expression temperature is chosen).
4. The cultures are centrifuged at 5000g during 10 min at  $4^{\circ}\text{C}$ . The cells are resuspended in 1 mL of lysis buffer and stored at  $-20^{\circ}\text{C}$  overnight. This freezing step will help the subsequent lysis step.

### 3.3.2. Screen for Protein Expression Level and Solubility

1. The suspended cells are thawed at room temperature and sonicated for 15–30 s at  $4^{\circ}\text{C}$ . The solution then becomes less viscous presumably because of the breakage of *E. coli* genomic DNA.
2. An aliquot of the “total extract” is analyzed by SDS-PAGE according to standard protocol.
3. The rest of the lysed cells are centrifuged at 13,000g,  $4^{\circ}\text{C}$ , during 30 min. An aliquot of the clear supernatant (“soluble extract”) is analyzed by SDS-PAGE.
4. 14% Acrylamide gels are loaded with 5 or 10  $\mu\text{L}$  of samples and are stained with Brilliant Blue (Sigma). An example of the expression of two ORFs is shown [Fig. 4](#), ORF3 is expressed in a soluble form, and ORF4 is expressed as inclusion bodies.

### 3.4. Strategies for Recovery of Inclusion Bodies

Because 37% of the 204 expressed proteins form inclusion bodies in *E. coli*, we developed a procedure for the recovery of these proteins. The strategy has been described in detail elsewhere ([18](#)), and only a brief overview will be given here. A set of 20 representative proteins expressed as inclusion bodies was studied in parallel. The strategy is made of three different options, adapted from refolding protocols for structural genomics (high-throughput) purposes. (1) In vitro refolding by dilution: after purification of the inclusion bodies in denaturing conditions (6 M GdnCl), the proteins are refolded by dilution using a screen of refolding buffers following a procedure adapted to a 96-well plate format; folding is followed by measuring light scattering at 390 nm (a high absorbance is an indication of the formation of protein aggregates); (2) coexpression of the target protein with bacterial chaperones (DnaK-DnaJ-GrpE-GroEL-GroES), using the plasmid developed by Nishihara et al. ([4](#)); and (3) cell-free expression, which is a different way to express the proteins. We chose to use the technology developed in Dr. Yokoyama’s laboratory ([20,21](#)),



**Fig. 4.** Sodium dodecyl sulfate-polyacrylamide gel of crude extracts of *Escherichia coli* showing an example of a soluble protein (ORF3) and an example of expression in inclusion bodies (ORF4). T, total cell extracts obtained after freezing–thawing and sonication; S, soluble extracts after centrifugation for 30 min at 13,000g of total extracts.

in a batch scale (50- $\mu$ L reaction volume) or dialysis scale (1–2 mL reaction volume), producing, respectively, micrograms or milligrams of protein. According to the ORF under study, the three approaches were useful for recovering inclusion bodies, and complemented each other. Some proteins were rescued by all three protocols, whereas others were refolded by only one or two of them. The chaperones' coexpression approach is easily adaptable to a pre-existent expression protocol and, therefore, is particularly useful for high-throughput structural genomics. To complete this short overview, see **Note 4** discussing some refolding strategies developed in other structural genomics projects, especially those that use the expression of fusion proteins (22). Other important strategies consist in directed evolution (23) or in switching to eukaryotic expression systems (24).

### 3.5. Large-Scale Production

Basically, 750 mL of 2xYT medium are inoculated in flasks at 37°C with 10 mL of freshly transformed overnight precultures. The target proteins are expressed after IPTG induction for 4 h, the optimal temperature determined during expression and solubility screen. This procedure typically yields between 5 and 50 mg of recombinant proteins. When overexpression is too low or when a large scale of

protein is needed for crystallization screens or biochemical studies, bioreactor facilities are used. The pH in the reactor is maintained at 7.0 by adding either NaOH or H<sub>2</sub>SO<sub>4</sub>. The dissolved oxygen content is maintained greater than 30% air saturation by increasing the agitation speed from 800 to 1500g. Aeration is kept to 1 v.v.m. (1 vol of air per 1 vol of culture per minute). Foaming is controlled by addition of one-tenth diluted Rodhorsyl (VWR-Prolabo). In flasks or bioreactors, the growth process consists in two steps, the biomass production achieved at 37°C, and the protein production performed at optimal temperature (*see Subheading 3.3*). The induction period (2 h, 4 h, or overnight) is dependent on the incubation temperature (37, 25, or 15°C, respectively). Synthetic media should be optimized for scaling up as described in *ref. 25*.

### 3.6. Labeling

#### 3.6.1. Se-Met Labeling for Crystallography Studies

Multiwavelength anomalous diffraction (MAD) phasing using selenomethionine-substituted protein crystals is the method of choice for the determination of X-ray structures (*5*). Two alternative strategies are frequently used for the incorporation of Se-Met into proteins. The use of *E. coli* B834 strain (Novagen), which is auxotrophic for methionine (*26*), was not satisfactory in our hands because we obtained low growth rates, owing to the toxicity of Se-Met and poor protein yields. The method we finally adapted in our project relies on the metabolic inhibition of the methionine pathway to obtain Se-Met incorporation using a standard expression strain (*27,28*).

1. 500 mL of M63mGly5 culture (this medium derived from M63 medium described in *refs. 29* and *30* and supplemented with the LeMaster amino acid solution known to activate the general cell metabolism (*see Subheading 2* and *ref. 5*) of *E. coli* expression strain transformed with the pET construct is grown at 37°C.
2. At OD<sub>600</sub>=1 to 1.5, the suspension is supplemented by a cocktail of amino acids (L-Lys, L-Phe, and L-Thr at 125 mg/L each; L-Ile, L-Leu, L-Val at 62.5 mg/L each), to repress the methionine biosynthesis pathway. L-Se-Met is added at 62.5 mg/L.
3. The production of the recombinant protein is induced 30 min later by addition of 0.3 mM IPTG, for 2 h at 37°C, 4–6 h at 25°C, or overnight at 15°C.
4. Se-Met incorporation into the protein is assayed by mass spectrometry (MS) after purification of the protein.

#### 3.6.2. <sup>13</sup>C and <sup>15</sup>N Labeling for NMR Studies

The resolution of a structure by NMR requires uniform labeling of the protein nitrogens (<sup>15</sup>N) and carbons (<sup>13</sup>C). Culture media volumes are kept to a minimum (250–500 mL), mainly because of the cost of the labeled carbon source (<sup>13</sup>C-glycerol). In order to choose when to induce, the consumption of labeled glycerol is followed by HPLC during bacterial growth.

1. The first overnight inoculum is cultivated in 10 mL of 2xYT medium at 37°C and 200 g.
2. An overnight preseed culture in the appropriate medium (50 mL of M63m<sup>15</sup>NGly5 or 20 mL of M63m<sup>15</sup>N/<sup>13</sup>C-Gly5) is inoculated at an initial OD<sub>600</sub> of 0.1.
3. The totality of the preseed culture is added into the final culture composed of the same medium. The cells are grown at 37°C until the exponential growth phase.
4. At an OD<sub>600</sub> of about two, the temperature is eventually reduced prior to the addition of 0.3 mM IPTG. The induction is maintained 2 h to overnight depending on the temperature.
5. The labeling is assayed by MS after purification of the protein.

### 3.7. Protein Purification and Biophysical Controls

The high-throughput nature of a structural genomics project demands a general and simple purification protocol (*see Note 5; 31–34*). We, therefore, chose to add a His<sub>6</sub>-tag to the recombinant proteins. The first purification step is a Ni<sup>++</sup> affinity column and the tag is generally not removed for the subsequent crystallization experiments (*see Note 6 and Fig. 2*). From a few test experiments to cleave the His-tag by proteolytic digestion, we concluded that it would be very difficult to integrate this step into a systematic and rapid protocol. We speculated that the crystallization of the majority of proteins will not be affected by the presence of the short tag. The affinity step is systematically followed by a gel filtration chromatography step to remove contaminant proteins and aggregates and to estimate the monodispersity and oligomeric state of the proteins. In most cases, this protocol yielded sufficient quantities of purified protein.

1. Cells obtained from a 750-mL culture are stored at –20°C at least overnight in 40 mL of lysis buffer, and broken by three cycles of freezing/thawing and sonication at 4°C. The suspension is centrifuged at 13,000g for 30 min at 4°C.
2. The supernatant is loaded on 2 mL of Ni-NTA equilibrated in the lysis buffer. The flow-through is kept on ice for SDS-PAGE control. The resin is washed with 20 mL of the buffer supplemented with 20 mM imidazole. The protein is eluted in three steps with 8 mL of the buffer containing respectively 100, 200, and 400 mM imidazole. An aliquot of each fraction is loaded on a SDS-PAGE to localize the protein.
3. The protein-containing fraction(s) are concentrated by centrifugation with a Vivaspin concentrator (Vivascience). The protein is immediately applied (*see Note 6*) to a Superdex 75 or 200 and is eluted at 1 mL/min for each protein in a suitable buffer in terms of pH (an electrofocusing analysis may be necessary) and NaCl concentration (*see Note 6*).
4. A SDS-PAGE of the fractions containing the protein is performed in order to control the purity and to correctly choose the fractions to be pooled.
5. The pure protein is concentrated for crystallization trials (usually around 10 mg/mL). Crystallization trays are set up as quickly as possible after protein sample preparation; best results for crystallization are obtained with very fresh protein samples.

6. An aliquot of pure protein is systematically assayed by MS, in order to control the integrity of the sample and/or the correct incorporation of various labels (Se-Met or  $^{13}\text{C}/^{15}\text{N}$ ). In some ambiguous cases, we carry out one-dimensional or two-dimensional NMR spectra to control the correct folding of the proteins. Other biophysical data (circular dichroism, microcalorimetry, isothermal calorimetry, small angle X-ray scattering, or fluorescence) sometimes complement our standard analysis protocol.

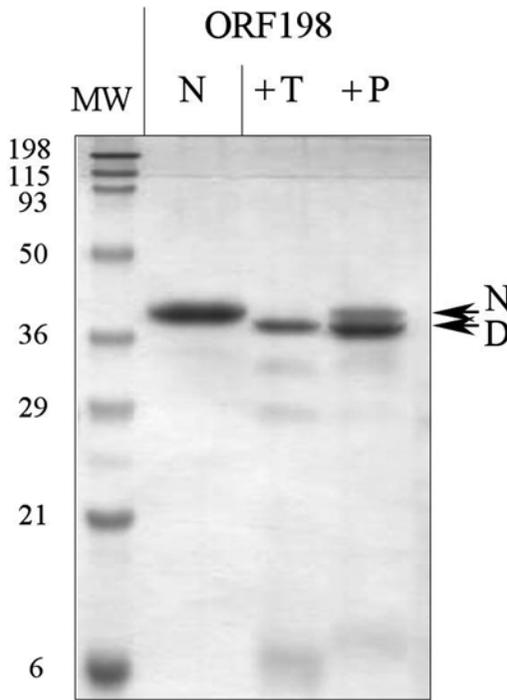
### 3.8. Limited Proteolysis

Many well-structured proteins contain regions of high conformational mobility, often situated at the N- or C-terminus of the protein. It is well established that these regions often hinder crystallization. This fact that we are not removing the intrinsically mobile terminal His-tag might actually make things worse (as previously mentioned, we found that omitting the proteolysis step considerably speeds up the purification process and also results in higher purification yields). Although we successfully crystallized 37% of the 59 purified proteins, we wanted to test biochemical protocols to increase the yield and quality of protein crystals. We developed a simple and small-scale limited proteolysis protocol to generate large subfragments of the proteins, which are resistant to further proteolysis and may, therefore, correspond to the structured and globular protein cores (**Fig. 5**). The partial cleavage is first measured via SDS-PAGE and if proteolysis has taken place, a binding test onto the Ni-NTA column helps to localize the proteolytic cleavage site. Afterward, precise localization of the digested site is carried out by MS. The fragment is subcloned by PCR and the new construct expressed in *E. coli* for large-scale expression and purification. At this time, four proteins that failed to crystallize have been subcloned. The polypeptides remained well overexpressed and are purified in the same buffers as the natives one. The crystallization trials are in progress.

1. In a 50- $\mu\text{L}$  mixture, 10  $\mu\text{g}$  of pure protein are incubated with 1/10 and 1/200 (w/w) of protease (trypsin, papain, pepsin, and so on), for 30 min at 37°C.
2. 10  $\mu\text{L}$  is immediately analyzed on a SDS-PAGE gel.
3. 5  $\mu\text{L}$  is frozen for analysis by MS.
4. The rest is bound to 20  $\mu\text{L}$  of Ni-NTA, the resin is washed, and the polypeptide is eluted with 400 mM imidazole. All the fractions are analyzed by SDS-PAGE (**Fig. 5**).

### 4. Notes

1. The standard expression system we used is based on the pET system. In addition to the high level T7 promoter common to the pET series, the important features of our standard construct are: (1) a Kan R marker more adaptable to high cell density fermentation than Amp R marker, (2) the expressed protein is strictly limited to the ORF sequence fused to the His-tag, without any linkers that might inhibit crystallization. We decided to keep the tag in place because proteolytic procedures cannot be systematically applied to a large number of targets, as illustrated for two



**Fig. 5.** Sodium dodecyl sulfate-polyacrylamide gel showing the limited proteolysis digestion of purified ORF198. Lane N corresponds to the native protein before incubation. Lane +T, digestion by trypsin; lane +P, digestion by papain. Arrows point to the native protein (N) and to the globular core of the protein identified by limited proteolysis (D).

ORFs in **Fig. 2**. The His-tag was introduced just after the last amino acid of the protein (and not at the N-terminus), in order to retrieve only full-length transcripts during the Ni-NTA purification step. A test experiment on 30 proteins for which we compared the expression and solubility yields between constructs with the His-tag at the N- and C-terminus, respectively, did not allow us to discern any marked differences. For ORF PCR and subsequent cloning, we used 3' primers (50mers) made of the last six codons, six histidine codons, a stop codon, the *NotI* restriction site, and four extra bases (the 5' primers was shorter with just *NdeI* or *NcoI* site, ATG and some coding codons). This strategy requires ordering of long primers (about 50 nt), more prone to sequence errors during chemical synthesis.

2. Even if the restriction/ligation cloning in pET vectors was very efficient (for each construction, four clones were tested and three or four contained the insert), a facilitated and less time-consuming cloning strategy, based on the "Topo-TA cloning" technology (pCRT7/CT-TOPO vector), was tested at the beginning of the project on 59/81 selected ORFs. Surprisingly, we have observed a difference of efficiency for expression of proteins for the ORFs cloned in this TOPO vector when compared

with those cloned in a pET vector: 82% of proteins were expressed in the “pET system,” whereas only 52% were expressed in the “Topo system.” To confirm these observations we subcloned 36 ORFs in the pET vector, which did not express in “Topo vectors,” or only expressed at a very low level. Half of the Topo-unexpressed proteins and all of the 10 low-expressed proteins became highly expressed in the pET vector, whereas 13 proteins remained unexpressed.

3. The systematic approaches applied in structural genomics projects around the world on the production of large numbers of proteins, allow for the first time to compare the efficiency of protocols classically used in laboratories for the production of recombinant proteins. In our project, we focused on the comparison of expression efficiency of commercial *E. coli* strains (see **Subheading 3.3.** and **Fig. 3.**). At this stage we can also provide some concluding remarks concerning the “pET expression system” in general. (1) We initiated the project with a systematic comparison of expression level between systems using plasmids containing the *lacI* gene or not (pET-9 vs pET-29). Because the strict criteria of the target selection led us to a list of *a priori* cytoplasmic proteins (see **Subheading 3.1.**), the presence of a supplementary copy of the *lacI* gene on the vectors was not crucial, and gave in general a slightly lower expression level of the proteins. (2) We systematically verified expression leakage of the recombinant proteins during the growth phase of the cultures before addition of IPTG, a frequent problem (observed in about 50% of the cases) constituting a drawback for the production of Se-Met-labeled proteins. This convinced us to adapt the labeling protocol by growing the cultures as soon as possible in minimum medium complemented with Se-Met in place of Met, at the very beginning of the exponential phase (see **Subheading 3.6.**).
4. The large number of proteins expressed as inclusion bodies in *E. coli* is one of the most important bottlenecks we were confronted with. The most simple experimental parameter to influence solubility of expressed proteins is to lower the induction temperature. On a set of 140 well-expressed proteins we observed (without discriminating between the strains) that 48% of the proteins were soluble when produced at 37°C, and interestingly 22% became soluble when the expression temperature was lowered to 25°C or below. For the remaining 30%, we developed the three-layered strategy as described in **Subheading 3.4.**, and finally decided to routinely coexpress the five chaperones. For instance, the presence of chaperones increased the solubility, between 10 and 90%, for 17/29 insoluble proteins. Alternatively, other structural genomics projects use fusion proteins with soluble domains (green fluorescence protein, maltose-binding protein, glutathione-S-transferase, and other) allowing targets to be kept in the soluble phase (22) (see Chapter 1). For this type of strategy, which requires making several constructs for each target, the “Gateway” cloning technology (Invitrogen) combined with automation of the procedure is recommended. The most important problem is the necessity to release the fused domain before crystallization trials. Even if the crystallization of several proteins fused to maltose-binding protein via a rigid and short spacer has recently been described, this will generally not be the case (35).
5. The results presented herein were part of the Yeast Structural Genomics Pilot Project that took place during years 2001 and 2003. Since then, several other studies have

started (see <http://genomics.eu.org/spip/-Projects->) and the protocols described in this chapter are still in routine use for these new structural genomics projects.

- At this stage of the project, 60% of the 121 proteins expressed under soluble form have been tested for purification in order to obtain milligram quantities of pure protein for crystallization trials. The first affinity purification step is exactly the same for all proteins (see **Subheading 3.7.**). The only difficulty consists in the determination of the optimal pH and salt concentration of the gel filtration buffer, consistent with a mono-disperse protein, in each individual case. Eighty-two percent of 72 proteins were purified to homogeneity and at sufficient quantities for setting up automated crystallization screens (at least 200  $\mu$ L at 2–50 mg/mL). The one-dimensional or two-dimensional NMR spectra obtained for some proteins allowed detection of the existence of very soluble, highly concentrated but “unfolded” proteins (36). We developed a biophysical-based study including small angle X-ray scattering, microcalorimetry, or circular dichroism to better understand these phenomena and to verify if any ligands, cofactors, or nucleic/protein partners are necessary for the protein to adopt a well-defined structure.

## Acknowledgments

This work is supported by grants from the Ministère de la Recherche et de la Technologie (Programme Génopoles).

## References

- Brenner, S. E. (2001) A tour of structural genomics. *Nat. Rev. Genet.* **2**, 801–809.
- Quevillon-Cheruel, S., Collinet, B., Zhou, C. Z., et al. (2003) A structural genomics initiative on yeast proteins. *J. Synchrotron Radiat.* **10**, 4–8.
- Goffeau, A., Barrell, B.G., Bussey, H., et al. (1996) Life with 6000 genes. *Science* **274**, 563–567.
- Nishihara, K., Kitagawa, M., Yanagi, H., and Yura, T. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**, 1694–1699.
- Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.
- Hofmann, K. and Stoffel, W. (1993) Tmbase: A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**, 166.
- Gonnet, G. H., Cohen, M. A., and Benner, S. A. (1992) Exhaustive matching of the entire protein sequence database. *Science* **256**, 1443–1445.
- Pearson, W. R. and Lipman, D. J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Rost B. and Liu J. (2003) The PredictProtein server. *Nucleic Acids Res.* **31**, 3300–3304.

11. Bateman A., Coin L., Durbin R., et al. (2004) The Pfam protein families database. *Nucleic Acids Res.* **32**, 138–141.
12. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
13. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. E. (2000) Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520.
14. Marin, A., Pothier, J., Zimmerman, K., and Gibrat, J. F. (2001) Protein structure prediction: bioinformatic approach. In: *Protein Threading Statistics: An Attempt to Assess the Significance of a Fold*, (Tsigelny, I., ed.), International University Line, La Jolla, CA.
15. Callebaut, I., Labesse, G., Durand, P., et al. (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.* **53**, 621–645.
16. Studier, F. W. and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130.
17. Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* **269**, 32,678–32,684.
18. Trésaugues, L., Collinet, B., Minard, P., et al. (2004) Refolding strategies from inclusion bodies in a structural genomics project. *J. Struct. Funct. Genomics* **5**, 195–204.
19. Miroux B. and Walker J. (1996) Over-production of proteins in *Escherichia coli*: mutants hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298.
20. Kigawa, T., Muto, Y., and Yokoyama, S. (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* **6**, 129–134.
21. Kigawa, T., Yabuki, T., Yoshida, Y., et al. (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **442**, 15–19.
22. Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* **11**, 313–321.
23. Waldo, G. S. (2003) Improving protein folding efficiency by directed evolution using the GFP folding reporter. *Methods Mol. Biol.* **230**, 343–359.
24. Boettner, M., Prinz, B., Holz, C., Stahl, U., and Lang, C. (2002) High-throughput screening for expression of heterologous proteins in the yeast *Pichia pastoris*. *J. Biotechnol.* **99**, 51–62.
25. Bettache, N. A., Quevillon-Cheruel, S., Bondet, V., van Tilbeurgh, H., and Blondeau, K. Determination of optimal cultivation conditions for large scale production of yeast carboxyl methyltransferase (Ppm1) in *Escherichia coli*., submitted.
26. Studts, J. M. and Fox, B. G. (1999) Application of fed-batch fermentation to the preparation of isotopically labeled or selenomethionyl-labeled proteins. *Protein Expr. Purif.* **16**, 109–119.

27. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
28. Doublet, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Meth. Enzymol.* **276**, 523–530.
29. Sambrook, J., Frits, E. F., and Maniatis, T. (eds.) (1989) Preparation and transformation of competent *E.coli*. In: *Molecular Cloning, 2nd ed.*, CSH Laboratory Press, Cold Spring Harbor, NY, pp. I.74–I.84.
30. David, G., Blondeau, K., Renouard, M., and Lewit-Bentley, A. (2002) Crystallization and preliminary analysis of *Escherichia coli* YodA. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1243–1245.
31. Lesley, S. A., Kuhn, P., Godzik, A., et al. (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. USA* **99**, 11,664–11,669.
32. Vincentelli, R., Bignon, C., Gruez, A., et al. (2003) Medium-scale structural genomics: strategies for protein expression and crystallization. *Acc. Chem. Res.* **36**, 165–172.
33. Heinemann, U., Bussow, K., Mueller, U., and Umbach, P. (2003) Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc. Chem. Res.* **36**, 157–163.
34. Matte, A., Sivaraman, J., Ekiel, I., Gehring, K., Jia, Z., and Cygler, M. (2003) Contribution of structural genomics to understanding the biology of *Escherichia coli*. *J. Bacteriol.* **185**, 3994–4002.
35. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **12**, 1313–1322.
36. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–756.



## Baculoviral Expression of an Integral Membrane Protein for Structural Studies

Dean R. Madden and Markus Safferling

### Summary

The baculovirus system has proven successful for the expression of integral membrane proteins for structural studies. A recombinant baculovirus, in which the gene of interest is placed under the control of the late-stage polyhedrin promoter, serves as the starting point for viral expansion and protein expression studies. Using large-scale insect cell culture techniques together with a filter-binding assay for protein function, the conditions of expression, purification, and solubilization can be optimized. As applied to the glutamate receptor ion channel subunit GluR2, this approach yields milligram quantities of pure, active protein, which have been used for single-particle electron microscopic analysis of the receptor structure. Detergent exchange protocols are also discussed, as a prerequisite for two-dimensional crystallization trials.

**Key Words:** Integral membrane protein expression; baculovirus system; ionotropic glutamate receptor; ligand-gated ion channel; insect-cell culture; immunoaffinity purification; filter-binding assay; detergent solubilization; electron microscopy; two-dimensional protein crystallization.

### 1. Introduction

Integral membrane proteins comprise 15–25% of the predicted protein sequences in eukaryotic genomes. As gatekeepers of the cell, they also constitute roughly half of the targets of pharmaceutical research (1,2). However, despite their great physiological importance, fewer than 1% of the high-resolution protein structures determined to date are of integral membrane proteins, all of which are either naturally abundant or of bacterial origin. One of the major stumbling blocks for structural studies of low-abundance eukaryotic membrane proteins has been the expression and purification of sufficient quantities of functionally active material for two-dimensional (2D) or three-dimensional

(3D) crystallization and subsequent electron or X-ray crystallographic analysis. Bacterial expression systems often fail to produce correctly folded eukaryotic membrane proteins, most likely owing to a lack of suitable mechanisms for facilitating protein folding and performing posttranslational modifications. Insect cells perform a full range of eukaryotic protein modifications, and are more easily scaled up than most mammalian systems. They also offer a lipid environment, particularly in terms of sterol composition, that may be more suitable for the function of some integral membrane proteins than that of yeast (3). With the development of transposition-mediated production of recombinant baculoviruses, insect cell expression vectors can now be manipulated in a relatively straightforward fashion. As a result, they have found wide use in the expression of integral membrane proteins (reviewed in refs. 5–20). Recent advances include the coexpression of chaperonins, the development of constitutive insect cell expression systems, and the humanization of insect cell glycosylation patterns (21–24). In this chapter, we describe the basic steps involved in the baculoviral expression and immunoaffinity purification of single milligram quantities of an eukaryotic membrane protein—the glutamate receptor ion channel GluR2—in functional form suitable for electron microscopic analysis and 2D crystallization screening.

## 2. Materials

1. Recombinant baculovirus-encoding gene of interest (here *v506-2*).
2. Standard equipment for tissue culture work, including flow hood.
3. Sf9 and High Five insect cells (Invitrogen, Carlsbad, CA).
4. 75-cm<sup>2</sup> Tissue culture flasks (T-75) (Corning, Corning, NY).
5. Grace's insect cell medium (Invitrogen).
6. Fetal calf serum (heat-inactivated) (Sigma, St. Louis, MO).
7. 250 µg/mL Amphotericin B (Sigma).
8. 10 mg/mL Gentamycin (Sigma).
9. Pluronic F-68 (Invitrogen).
10. TNM-FH medium: 500 mL Grace's insect cell medium supplemented with 50 mL heat-inactivated fetal calf serum, 5 mL 250 µg/mL amphotericin B, 500 µL 10 mg/mL gentamycin, and 5 mL of 10% pluronic F-68.
11. 27°C Incubators (static and shaking).
12. 1.8-L Wide-mouth Fernbach flasks (Schott, Mainz, Germany).
13. Gas-permeable stoppers (e.g., N-42 Biosilico, Schott).
14. Hemocytometer.
15. 0.4% Trypan blue (Sigma).
16. BacPAK titer determination kit (BD Biosciences, San Jose, CA).
17. Equipment for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).
18. M1 α-FLAG antibody (Sigma).
19. HEPES-buffered saline: 20 mM HEPES, pH 7.4, and 150 mM NaCl.

20. Lysis buffer: 20 mM HEPES, pH 7.4, and 5 mM EDTA.
21. Wash buffer: 20 mM HEPES, pH 7.4, 200 mM NaCl, and 0.5 mM EDTA.
22. Phenylmethylsulfonyl fluoride (PMSF): dissolve to 0.1 M in ethanol. Store at  $-20^{\circ}\text{C}$ . **Caution:** toxic; add immediately before use because not stable in water.
23. Complete protease inhibitor tablets (Roche, Indianapolis, IN).
24. Polytron disintegrator (Kinematica, Newark, NJ).
25. Dilution buffer: 20 mM HEPES, pH 7.4, 200 mM NaCl, and 10% (w/v) glycerol.
26. Detergent stock: 20% (w/v) detergent (*see Subheadings 3.2.3 and 3.5*), 20 mM HEPES, pH 7.4, 200 mM NaCl, and 10% (w/v) glycerol. Make fresh, as some detergents are unstable in solution.
27. Detergents: Triton X-100 (TX100), Triton X-114, *n*-octyl- $\beta$ -D-maltoside, *n*-decyl- $\beta$ -D-maltoside (DM), *n*-dodecyl- $\beta$ -D-maltoside (DDM), *n*-octyl- $\beta$ -D-glucoside (OG), thio-OG, *n*-nonyl- $\beta$ -D-glucoside, CHAPS, sodium cholate, Tween-20, octanoyl-*N*-methylglucamide, decanoyl-*N*-methylglucamide, *N*-dodecyl-*N,N*-dimethyl-3-ammonio-1-propanesulfonate, lysophosphatidyl choline, *n*-octanoyl- $\beta$ -D-glucosylamine, Zwittergent-3.10, and SDS.
28. Bicinchoninic acid (BCA) protein determination kit (Pierce, Rockford, IL).
29.  $\alpha$ -Amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA).
30. [ $^3\text{H}$ ]-AMPA (Perkin-Elmer, Wellesley, MA).
31. AMPA binding buffer: 30 mM Tris-HCl, pH 7.2, 2.5 mM  $\text{CaCl}_2$ , 100 mM KSCN, and 0.1% (w/v) TX100. Make fresh. **Caution:** KSCN will react with acids to produce toxic gas.
32. Whatman GF/B filters (25-mm diameter).
33. Polyethylene imine.
34. L-Glutamate.
35. Vacuum filtration manifold 1225 (Millipore, Billerica, MA).
36. Emulsifier-safe (Perkin-Elmer).
37. M1 anti-FLAG immunoaffinity agarose gel (Sigma).
38. Low-pressure chromatography equipment.
39. M1 wash buffer: 3 mM  $\text{CaCl}_2$ , 0.1% (w/v) TX100, 20 mM HEPES, pH 7.4, 300 mM NaCl, and 10% (w/v) glycerol.
40. M1 elution buffer: 3 mM EDTA, 0.1% (w/v) TX100, 20 mM HEPES, pH 7.4, 150 mM NaCl, 10% (w/v) glycerol.
41. M1 glycine wash buffer: 0.1 M glycine, pH 3.5.
42. M1 storage buffer: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.02%  $\text{NaN}_3$ . **Caution:**  $\text{NaN}_3$  is toxic.
43. Bradford reagent (Bio-Rad, Hercules, CA).
44. Carbon-coated electron microscope grids (EMS, Fort Washington, PA).
45. Glow-discharge apparatus.
46. Anticapillary tweezers.
47. EM wash buffer: 25 mM Tris-HCl, pH 7.4, 20 mM NaCl, and 2 mM EDTA.
48. Negative staining solution: 2% (w/v) uranyl acetate, 0.1% (w/v) glycerol.
49. Electron microscope.
50. High-resolution digital scanner (e.g., SCAI scanner/Zeiss).
51. EM image processing software [e.g., SPIDER (25), IMAGIC (26), or EMAN (27)].

52. Exchange buffer: 20 mM HEPES, pH 7.4, 500 mM NaCl, 10% glycerol, and detergent (*see item 27*).
53. Exchange elution buffer: 20 mM HEPES, pH 7.4, 400 mM imidazole, 500 mM NaCl, 10% glycerol, and detergent (*see item 27*).
54. Fast Flow chelating sepharose (Amersham, Piscataway, NJ).
55. Amido black 10B (Bio-Rad).

### 3. Methods

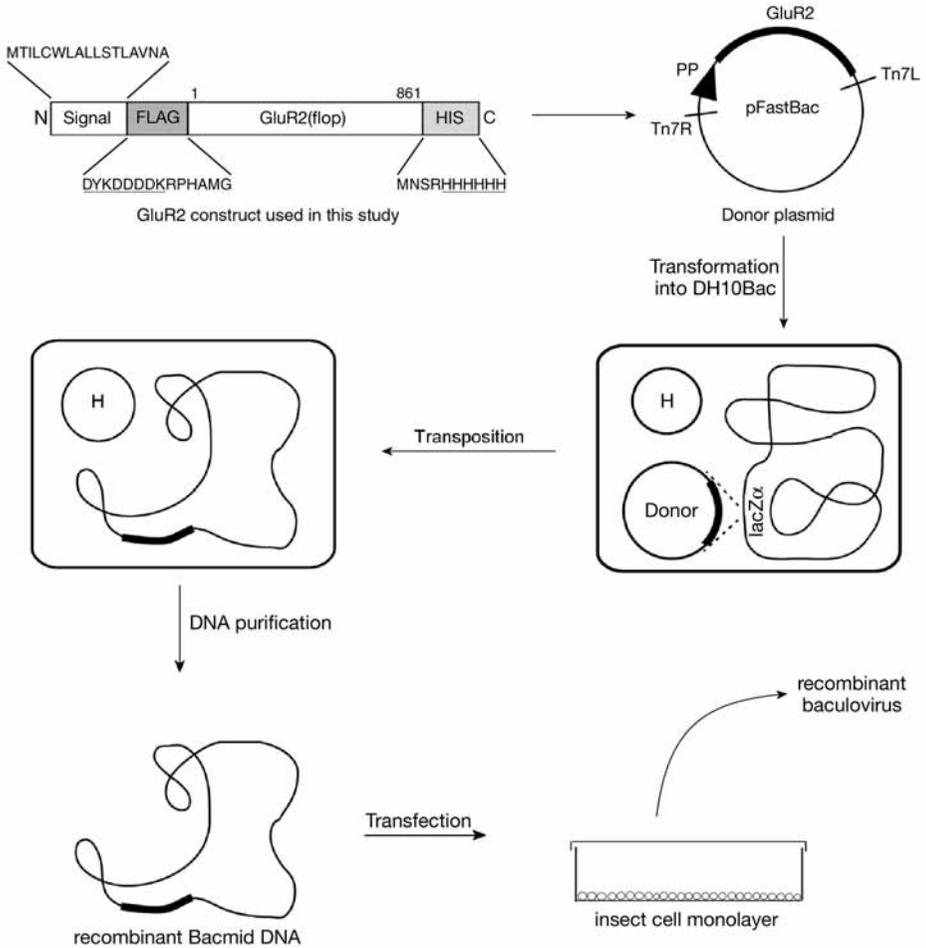
Large-scale baculoviral expression of membrane proteins for structural studies requires the following steps: (1) production and harvest of recombinant baculovirus, (2) optimization of the conditions of solubilization and the parameters of infection for maximal protein expression and recovery, and (3) milligram-scale immunoaffinity purification. The resulting protein is suitable for electron microscopic analysis and—following detergent exchange—for 2D or 3D crystallization trials.

#### 3.1. Producing Recombinant Baculovirus for Large-Scale Expression

In order to infect large volumes of insect cell cultures, corresponding volumes of high-titer viral stock are required. In this section, the characteristics of the recombinant baculovirus used in this study are described, together with techniques for viral expansion in suspension culture and quantification of the viral titer.

##### 3.1.1. Generation of Recombinant Baculovirus

The first step in insect cell expression of a target protein is the generation of an appropriate recombinant baculovirus. In this case, recombinant baculovirus v506-2 was generously provided by Dr. K. Keinänen (University of Helsinki, Helsinki, Finland) (*18*). The expression construct was assembled using standard DNA subcloning techniques in a derivative of the pFastBac1 transfer vector (*28*) (**Fig. 1**) (*see Note 1*). This vector contains the strong AcNPV polyhedrin promoter, which is activated during the late stages of infection, followed by a multiple cloning site and an SV40 polyadenylation site. The expression cassette is flanked on both sides by the left and right arms of the Tn7 transposon, permitting sequence-specific transposition into the baculovirus genome using the Bac-to-Bac system, according to manufacturer's instructions (*see Note 2*). Within the expression cassette, the ecdysteroid UDP-glucosyltransferase signal sequence was fused to a FLAG epitope, followed by the mature coding sequence of the “flop” splice variant of GluR2 (“GluR2[flop],” residues 1-861, omitting Ile862) (*29,30*) and a C-terminal hexahistidine tag (**Fig. 1**) (*see Note 3*).



**Fig. 1.** Generating recombinant baculovirus for GluR2 expression—a schematic overview. The expression cassette was assembled in a derivative of pFastBac1. Its components are described in the text (**Subheading 3.1.1.**). The transfer vector was transformed into DH10Bac cells, which harbor a bacmid—a plasmid encoding the AcNPV baculoviral genome modified to include the *LacZα* gene with internal sequence-specific transposition sites. DH10Bac cells also include a helper plasmid that encodes the required transposase activity. Recombination is detected by blue/white selection on X-gal- or Bluo-gal-containing medium. Bacmid DNA is then purified from overnight culture and introduced into Sf9 cells by lipofection, leading to the production of fully infective recombinant baculoviruses (P1 viral stock). (Adapted from **ref. 47** © 2003 Invitrogen Corporation. All rights reserved. Used with permission.)

### 3.1.2. Baculovirus Expansion

Working from the small starting volume (0.5–2.0 mL) of recombinant baculovirus obtained from the Bac-to-Bac system (P1 stock), a two-stage viral expansion is used to obtain the quantities required for large-scale expression (31). First, an initial expansion is performed using monolayers. The resulting P2 baculovirus stock is then used to infect 500-mL suspension cultures. In each case, a low multiplicity of infection ( $\leq 0.1$ ) is used (see **Note 4**). Following infection, the Sf9 cells are cultured for 8–12 d to permit two cycles of expansion.

Standard insect cell culture techniques (32,33) are used to maintain monolayer cultures of Sf9 cells in T-75 flasks in TNM-FH medium. Typically, cultures are passaged at dilutions of 1:5 or 1:6 every 3–4 d.

Initial expansion (~45 mL):

1. One confluent T-75 flask of Sf9 cells is passaged into three fresh T-75 flasks (1:6 dilution).
2. Cells are allowed to form a semi-confluent monolayer (~2 d).
3. 100  $\mu$ L of baculoviral stock solution is added to each flask. Mix by gently rocking the flask sideways and front-to-back.
4. Flasks are incubated 8–12 d at 27°C, until the monolayer sloughs off.
5. Using sterile technique throughout, the medium from each flask is transferred to a 15-mL falcon tube and centrifuged 15 min at  $\geq 3000g$  to pellet cell debris. The supernatant is recovered as the P2 viral stock.
6. Viral stocks should be stored in the dark at 4°C.
7. Expression of correctly processed protein can be confirmed in the cell pellet by Western blotting using the M1 anti-FLAG antibody, which is specific only for N-terminal FLAG epitopes.

Assuming that protein expression is detected, a large-scale expansion (1–8 L) is then initiated:

1. Sf9 cells from a confluent T-75 flask are placed in a 125-mL Erlenmeyer flask with a gas-permeable stopper (see **Note 5**) in 30 mL of TNM-FH medium.
2. The suspension culture is shaken at 110 rpm at 27°C until it reaches a density of  $2.4\text{--}3.0 \times 10^6$  cells/mL (as determined using a hemocytometer, usually after 3–4 d).
3. The culture is diluted with fresh TNM-FH medium to a density of  $0.4 \times 10^6$  cells/mL.
4. This process is repeated until the desired volume of culture is obtained. For the virus expansion, 500-mL suspension cultures are prepared at a density of  $0.4 \times 10^6$  cells/mL in 1.8-L wide-mouthed Fernbach flasks with gas-permeable stoppers. Fernbach flasks are shaken at 90–110 rpm (25-mm orbital radius) or 60–70 rpm (50-mm orbital radius) at 27°C.
5. When the cells have reached a density of  $1.3\text{--}1.7 \times 10^6$  cells/mL, 1.5 mL of the P2 viral stock is added to the flask.

6. Cultures are incubated 10–12 d. After this point, cell viability is monitored daily. An aliquot of 0.9 mL of cells is removed using sterile technique and incubated with 0.1 mL of 0.4% Trypan blue for 5 min at room temperature. Blue (dead) and clear (viable) cells are counted separately in a hemocytometer.
7. Once cell viability is less than 10%, the cultures are filled into autoclaved centrifuge bottles, maintaining sterile technique. Cells and cell fragments are pelleted at 7500g (e.g., 6000 rpm in a JLA-9.1000 rotor). The virus-containing supernatant is carefully decanted into sterile bottles.
8. Viral stocks should be stored in the dark at 4°C (*see Note 6*).
9. The viral titer is determined using the BacPAK system (BD Biosciences).

### 3.2. Optimization of Protein Expression

To obtain soluble protein, cells are infected, harvested by centrifugation and lysed, and the resulting cell membranes are pelleted. Subsequently, the protein is solubilized by addition of detergent. The choice of solubilizing detergent and the multiplicity and length of infection must be adjusted to maximize the yield of soluble protein.

#### 3.2.1. Large-Scale Baculoviral Infection

1. Inoculate eight 500-mL suspension cultures of High Five cells in wide-mouth 1.8-L Fernbach flasks at a density of  $0.5 \times 10^6$  cells/mL (*see Note 7*).
2. When the cell density reaches  $1.8\text{--}2.2 \times 10^6$  cells/mL (ca. 48 h later), add high-titer virus stock solution to give a multiplicity of infection (MOI) of four (~50–60 mL/flask, assuming a titer of  $8 \times 10^7$  pfu/mL).
3. Harvest cells approx 90 h postinfection (p.i.) by centrifugation at 1300g (e.g., 2500 rpm in JLA-9.1000 rotor) for 25 min (*see Note 8*).

#### 3.2.2. Preparation of Membrane Pellets

1. Prepare the following buffers the day before: 100 mL of HEPES-buffered saline, 250 mL of lysis buffer, and 150 mL of wash buffer.
2. Cell pellets are resuspended in a total of 100 mL of HEPES-buffered saline, placed in four 29 × 104-mm screw-cap centrifuge bottles, and centrifuged for 20 min at 1200g at 4°C.
3. Add 0.1 mM PMSF and one Complete protease-inhibitor tablet to the lysis buffer.
4. The cell pellets should be resuspended in 7–10 mL of lysis buffer each (*see Note 9*). The centrifuge tube should be placed on ice.
5. Cell lysis is performed at 14,000 rpm using a Polytron disintegrator. Place each centrifuge tube on ice in a glass beaker and pulse for 10 s. Cycle through all tubes a total of three times. This allows each tube to cool between pulses.
6. Following lysis, fill each tube with lysis buffer and centrifuge at 37,000g for 25 min at 4°C (e.g., 17,500 rpm in JA-25.50 rotor).
7. **Steps 4–6** are performed a total of three times. The third time, wash buffer is used instead of lysis buffer. Add 0.1 mM PMSF to the wash buffer immediately prior to use.

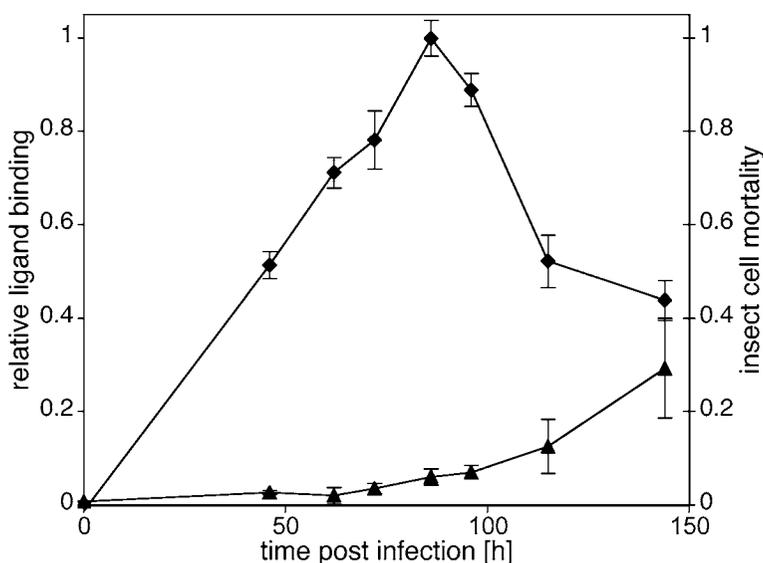
### 3.2.3. Determination of Solubilization Conditions

1. Prepare 500 mL of dilution buffer and 10 mL of each detergent stock (*see Note 10*).
2. Immediately prior to use, add PMSF to the dilution buffer to a final concentration of 0.1 mM.
3. The pellets from the third Polytron treatment are resuspended in dilution buffer (7–10 mL) and homogenized using a 10-s pulse with the Polytron at 10,000 rpm.
4. Pool the resuspended membranes. Determine the protein concentration of 50- and 100-fold diluted protein samples using the BCA assay according to manufacturer's instructions (*see Note 11*).
5. Equal aliquots of the membrane suspension should be diluted to a final concentration of 4–6 mg/mL at a final detergent concentration of 1.0–1.5% (w/v) each. First, add the required detergent to the required volume of dilution buffer, and then add the mixture to the membrane suspension. Reserve one aliquot as a reference for total protein activity. Detergents assayed were: TX100, Triton X-114, DDM, OG, CHAPS, sodium cholate, Tween-20, octanoyl-*N*-methylglucamide, *N*-dodecyl-*N,N*-dimethyl-3-ammonio-1-propanesulfonate, SDS, and lysophosphatidyl choline (**34,35**).
6. Mix the suspension gently at 4°C for 1 h or overnight.
7. Pellet unsolubilized material by centrifugation at 120,000g for 45 min.

### 3.2.4. Assay for Extent of Solubilization

In order to determine the fraction of protein solubilized in active form, it is useful to have a functional assay. In the case of the AMPA receptors, filter binding using the high-affinity agonist AMPA provides a reliable estimate of ligand-binding activity. For each sample, the following protocol is performed to estimate the total amount of protein activity solubilized (*see Note 12*). In the case of the GluR2, TX100 provided the highest yield (93%), and was used for all subsequent solubilizations (**34**).

1. Freshly prepare 1 L of AMPA-binding buffer.
2. Soak Whatman GF/B 25-mm diameter filters in 0.3% polyethylene imine. Prepare four filters for each AMPA concentration to be measured.
3. Prepare 450- $\mu$ L samples of [ $^3$ H]-AMPA in AMPA-binding buffer, such that the final concentration after addition of 50  $\mu$ L of protein sample will be 1, 3, 10, 30, 100, and 300 nM [ $^3$ H]-AMPA (*see Note 13*). Prepare four tubes for each concentration. To one tube at each concentration, add L-glutamate to a final concentration of 1 mM from a 100 mM stock as a control for nonspecific binding.
4. Add 50  $\mu$ L of the appropriate protein sample (supernatant from solubilization or unsolubilized membrane suspension) to each tube. Incubate on ice for 1 h.
5. Place the soaked GF/B filters in a filtration barrel (Millipore) and apply vacuum to the drum.
6. Quickly dispense 5 mL of AMPA-binding buffer into the sample and immediately pour through a filter. Working rapidly, dispense 5 mL of AMPA-binding buffer into the tube and pour through the filter to wash out nonspecifically bound radioligand. Repeat the washing step once.



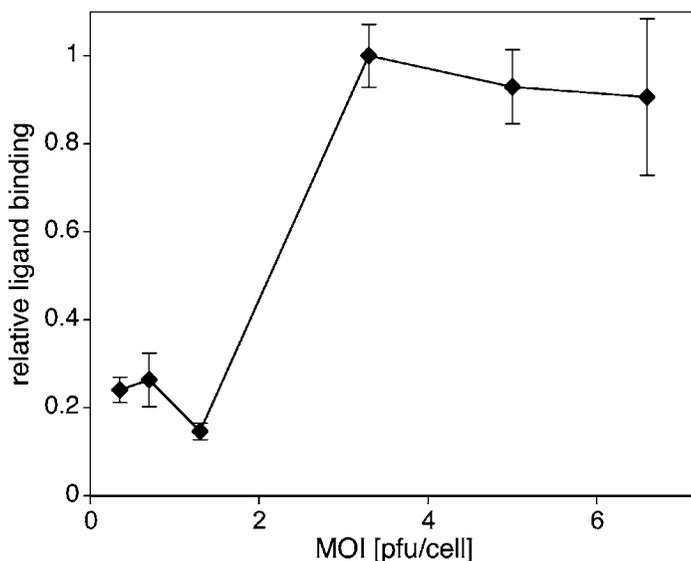
**Fig. 2.** Time course. Level of detergent-soluble [ $^3\text{H}$ ]-AMPA-binding activity (diamonds, measured at 15 nM, left-hand axis) and cell mortality (triangles, right-hand axis), as a function of time of harvest following baculoviral infection.

- Solubilize the filters overnight in 5 mL Emulsifier-Safe in scintillation vials.
- Count the retained radioactivity using a scintillation counter. For each radioligand concentration, subtract the background radioactivity observed in the presence of 1 mM glutamate to determine the specific binding.
- Plot the binding isotherm to determine affinity and binding capacity. Compare the binding capacity of the solubilisates to that of the unsolubilized membrane reference to determine the fractional solubilization for each detergent.

### 3.2.5. Determining Optimal Time of Infection

The polyhedrin promoter drives extremely strong protein expression during the very late stages of infection. As a result, protein accumulation typically begins later than 24 h p.i. As the infection progresses beyond 96 h p.i., significant cell lysis is observed, which can lead to proteolytic degradation of the protein of interest. As a result, it is important to determine the correct time of harvest empirically. For High Five cells expressing GluR2, maximal expression is observed ca. 88 h p.i., with a rapid fall-off in ligand-binding capacity both before and after this time-point (**Fig. 2**) (see **Note 12**).

- Prepare and infect eight 500-mL flasks of High Five cells at MOI = 5, as described in **Subheading 3.2.1**. (see **Note 14**).



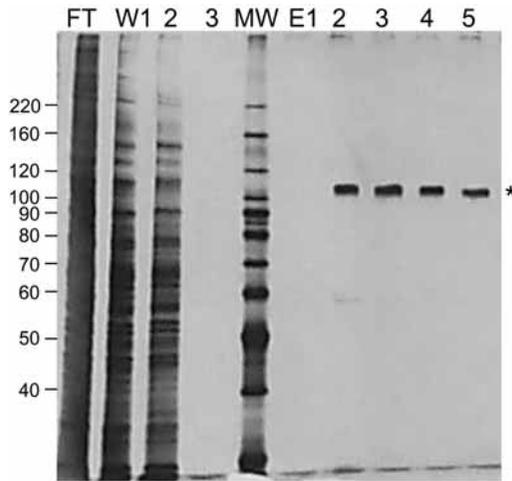
**Fig. 3.** Optimization of multiplicity of infection. Level of detergent-soluble [ $^3\text{H}$ ]-AMPA-binding activity (measured at 15 nM) as a function of multiplicity of infection.

2. Harvest one bottle every 12 h beginning at 48 h p.i., and every 24 h beginning at 96 h p.i. Determine cell density and viability at time of harvest.
3. Prepare membrane pellets and solubilize proteins as described in **Subheadings 3.2.2.** and **3.2.3.**, using the detergent of choice (*see Note 15*).
4. Determine ligand-binding activity by filter binding assay as described in **Subheading 3.2.4.** (*see Note 16*).
5. Harvest subsequent infections at the time and cell viability levels corresponding to the highest ligand-binding activity.

### 3.2.6. Determining Optimal MOI

It is necessary to adjust the MOI to ensure simultaneous and complete infection of the insect cells in the suspension culture. Above a certain threshold, at which all cells are infected, increasing the MOI does not increase protein yield and, therefore, represents a waste of viral stock. For GluR2, MOI  $\geq 3$  was required to achieve maximal expression (**Fig. 3**).

1. Prepare and infect eight 500-mL flasks of High Five cells as described in **Subheading 3.2.1.** Use MOI ranging from approx 0.5 to 8 (*see Note 14*).
2. At the time-point of maximum expression, harvest the cells, prepare membrane pellets, and solubilize proteins (*see Note 15*).



**Fig. 4.**  $\alpha$ -FLAG immunoaffinity purification. Flow-through (FT), wash (W1-W3) and elution (E1-E5) fractions are visualized by silver staining following sodium dodecyl sulfate-polyacrylamide gel electrophoresis on a 10% gel. Molecular mass standards (MW) are indicated at left. The position of the GluR2 protein is indicated by an asterisk on the right.

3. Determine ligand-binding activity by filter binding assay as described in **Subheading 3.2.4.** (*see Note 16*).
4. Select an MOI value at the beginning of the plateau in the ligand-binding activity (here MOI = ~4).

### 3.3. Affinity Purification

Once the conditions of infection and solubilization have been established (*see Subheadings 3.2.4.–3.2.6.*), the protein is purified using tags incorporated in the construct. In the case of GluR2, a single-step immunoaffinity purification is sufficient to yield highly pure samples (**Fig. 4**).

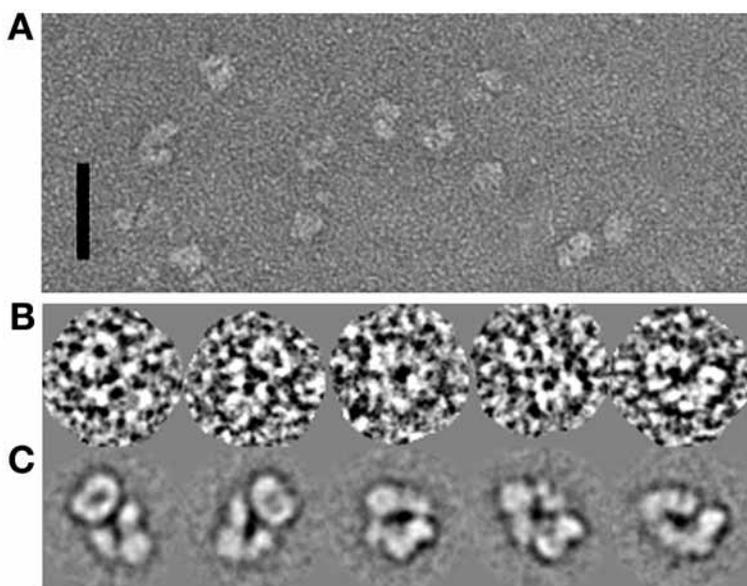
1. Prepare membrane pellets as described in **Subheadings 3.2.1.–3.2.3.**, except that the entire pellet is solubilized using 1.5% (w/v) TX100 for 1 h at 4°C. This requires 50 mL of 20% (w/v) TX100 stock solution.
2. Prepare the following buffers: 150 mL of M1 wash buffer, 75 mL of M1 elution buffer, 25 mL of M1 glycine wash buffer, and 100 mL of M1 storage buffer.
3. All chromatography steps should be performed at 4°C.
4. Pour 1.5 mL of M1 affinity agarose gel in a column with ID = 1 cm, and attach to a peristaltic pump or low-pressure chromatography system. Ensure that the system can accurately deliver flow rates as low as 0.25 mL/min. The column should be freshly packed for each preparation.
5. Wash the column with 10 column volumes (CV) of M1 wash buffer at 2 mL/min.

6. Centrifuge the solubilisate for 45 min at 120,000g. Collect the supernatant.
7. Add  $\text{CaCl}_2$  to a final concentration of 3 mM (0.003 vol of 1 M stock).
8. Load the protein overnight (~0.4 mL/min).
9. Wash the column with approx 50 CV of M1 wash buffer at 1 mL/min.
10. At the end of the wash, collect approx 200  $\mu\text{L}$  of buffer from the column. Using 100  $\mu\text{L}$  of sample, determine protein concentration by Bradford assay. The protein concentration should be less than 20  $\mu\text{g}/\text{mL}$ . If not, continue washing and check the protein concentration every 10 CV.
11. Elute the protein from the column using approx 30 CV of M1 elution buffer at 0.6 mL/min. Collect three CV fractions.
12. Determine the protein concentration in each fraction by Bradford assay (using 25- $\mu\text{L}$  samples). Determine protein purity by SDS-PAGE using silver stain. Pool those fractions containing high concentrations of pure protein, typically fractions two to five.
13. Regenerate the column material by washing with 6 CV of M1 glycine wash buffer at 2 mL/min. Do not leave the M1 affinity gel in M1 glycine buffer longer than necessary.
14. Equilibrate the column material by washing with 20 CV of M1 storage buffer at 2 mL/min.
15. Using a pipet, resuspend and recover the column material in 20 CV of M1 storage buffer. Store at 4°C.
16. Wash the glass column housing and frits using 3 CV 2 N HCl; 5 CV  $\text{H}_2\text{O}$ ; 3 CV 2 N NaOH; 5 CV  $\text{H}_2\text{O}$ .

### 3.4. Electron Microscopic Analysis

The protein obtained by purification can be used directly for single-particle analysis by electron microscopy (EM). This provides a visual control of the structural homogeneity of the sample. It also reveals the molecular dimensions, and possibly the symmetry of the particle. The GluR2 appear as elongated, hollow particles, viewed in a variety of perspectives corresponding to different orientations on the EM grid (**Fig. 5**) (**18,36**).

1. Glow discharge carbon-coated EM grid under vacuum.
2. Working at 4°C, apply 3  $\mu\text{L}$  of the protein solution to the carbon film and allow it to adsorb for 30 s. Blot the drop from the edge of the grid, wash with two droplets of EM wash buffer, and apply 25  $\mu\text{L}$  of negative staining solution, while continuing to blot from the edge of the grid. Allow to dry (*see Note 17*).
3. Insert the grid into the electron microscope and inspect under low-dose conditions. Photograph areas that are uniformly stained at a nominal magnification of  $\times 50,000$ . Protein particles should be visualized as bright, stain-excluding areas within a uniform background of negative stain. Single-particle images can be digitized using a high-resolution scanner, and classified and analyzed using the IMAGIC (**26**), SPIDER (**25**), or EMAN (**27**) software packages (**Fig. 5**) (**18**).



**Fig. 5.** Electron microscopy of GluR2. (A) Typical views of individual GluR2 molecules embedded in negative stain. (B) Single images and (C) class averages showing characteristic perspectives following alignment and classification. (Reproduced with permission from [ref. 18](#). © 2001 American Chemical Society.)

### 3.5. Detergent Exchange for 2D Crystallization

Like 3D crystallization, 2D crystallization is a highly empirical process and can be performed using a number of techniques ([37–46](#)). Several of these approaches require either specific detergents or else detergents with a high critical micelle concentration (CMC) that support removal by dialysis, allowing reconstitution into a lipid bilayer. Although TX100 is well suited as a solubilizing agent for GluR2, it has a very low CMC (0.2–0.9 mM, depending on buffer conditions) and is chemically heterogeneous, which can impede crystallization. In order to identify suitable detergents (i.e., ones that preserve ligand-binding activity and protein homogeneity), GluR2 is bound to an affinity column and is washed and then eluted in buffer containing the new detergent. The eluted protein is then assayed for yield and functional activity. Once suitable alternatives are identified, they can be substituted for TX100 in the wash and elution buffers during immunoaffinity purification (*see Subheading 3.3.*). In this case, both 1.8 mg/mL DM and 1 mg/mL DDM proved to be excellent alternatives to TX100 ([Table 1](#)) ([35](#)). These two maltoside detergents are suitable for detergent removal and/or surface crystallization protocols, respectively.

**Table 1**  
**Recovery and Ligand-Binding Activity of GluR2 Following Detergent Exchange**

Detergent	CMC (mM)	Exchange conc. (mM)	Yield (%)	Activity (%)
TX100	~0.3 (0.2–0.9)	1.6	86.0	96 ± 2.6
OG	20–25	35	63.0	25 ± 0.9
Thio-OG	9	13	56.0	7.6 ± 0.6
NG	6.5	13	57.0	41 ± 1.0
OM	23.4	30	78.0	26.3 ± 0.25
DM	1.6	3.7	99.5	90.5 ± 2.9
DDM	0.1–0.6	2	99.0	85 ± 4.2
MEGA-10	6–7	11.5	55.0	6.8 ± 1.1
NOGA	80	90	12.4	0.1 ± 0.1
Zwittergent-3.10	25–40	50	65.0	0.5 ± 0.4

DM, *n*-decyl- $\beta$ -D-maltoside; DMM, *n*-dodecyl- $\beta$ -D-maltoside; MEGA-10, decanoyl-*N*-methylglucamide; NG, *n*-nonyl- $\beta$ -D-glucoside; NOGA, *n*-octanoyl- $\beta$ -D-glucosylamine; OG, *n*-octyl- $\beta$ -D-glucoside; OM, *n*-octyl- $\beta$ -D-maltoside; TX100, Triton X-100.

1. Prepare 40 mL of exchange buffer and 10 mL of exchange elution buffer for each 25  $\mu$ g of GluR2 to be exchanged. Values are reported in **Table 1** for the following detergents: TX100, OG, Thio-OG, *n*-nonyl- $\beta$ -D-glucoside, *n*-octyl- $\beta$ -D-maltoside, DM, DDM, decanoyl-*N*-methylglucamide, *n*-octanoyl- $\beta$ -D-glucosylamine, and Zwittergent-3.10.
2. Pack a low-pressure column with 1 mL Fast Flow chelating sepharose for each 25  $\mu$ g of GluR2 to be exchanged.
3. Charge and equilibrate the column with 5 CV each of H<sub>2</sub>O, 100 mM Zn(OAc)<sub>2</sub>, H<sub>2</sub>O, and exchange buffer (*see Note 18*).
4. All subsequent chromatography steps should be performed at 4°C.
5. Load the sample onto the column at 0.5 mL/min.
6. Wash the column with exchange buffer, and monitor the removal of TX100 by the absorbance at 275 nm, using the formula %Triton (w/v) = 0.044  $\times$  OD<sub>275</sub> (1-cm path length). Within 10–15 CV, the Triton concentration should be significantly below CMC. Elute with exchange elution buffer, collecting 1-CV fractions. Both steps should be carried out at a flow rate of 0.5 mL/min.
7. Determine the amount of protein recovered by Amido Black assay.
8. Determine the ligand-binding activity of the sample using filter binding with 15 nM [<sup>3</sup>H]-AMPA (*see Subheading 3.2.4.*).
9. The fractional yield and ligand-binding activity recovered is shown in **Table 1** as a function of the detergent used.

#### 4. NOTES

1. Bac-to-Bac transfer vectors are also available within the Gateway system (Invitrogen), permitting subcloning of expression cassettes without use of restriction enzymes.

2. In following the Bac-to-Bac protocols, we typically observed the following points: during the blue/white selection step, it is important to wait long enough to distinguish clearly between the blue and white colonies. In our experience, using X-gal plates, this can require more than 48 h of incubation. Once the colonies are of sufficient size, the plates can be left at room temperature for further color development. In order to obtain larger starting volumes of baculoviral material, we typically transfect  $2 \times 10^6$  cells in a 60-mm dish (rather than  $0.5 \times 10^6$  in a 35-mm dish), and double the volumes of the standard transfection protocol. The P1 viral stock (supernatant) is harvested after 96 h, because the titer increases markedly between 48 and 96 h p.i. As an alternative to the Bac-to-Bac system, the BaculoDirect system (Invitrogen) can also be used, in which the transposition step is performed *in vitro*.
3. For immunoaffinity purification, the M1  $\alpha$ -FLAG antibody was used. This antibody binds only in the presence of  $\text{Ca}^{++}$ , permitting efficient elution upon addition of EDTA. However, the M1 antibody also binds only to FLAG epitopes located at the absolute N-terminus of the protein. An N-terminal Met residue will thus interfere with binding. As a result, this system is particularly well suited to constructs that include a signal sequence that is cleaved upon secretion. Alternatively, the M2  $\alpha$ -FLAG antibody can be used, but requires elution with FLAG peptide or extreme pH.
4. Use of a low MOI ensures that very few cells will be multiply infected. During successive viral passages, this prevents the expansion of “defective, interfering particles,” which cannot replicate independently but can replicate upon coinfection with intact baculovirus. For this reason, it is also best to initiate large-scale viral expansions from a low-passage number stock.
5. Do not use spring-clip aluminum caps, which do not seal the neck tightly enough to prevent contamination of the culture.
6. Large-scale viral stocks frequently develop a sediment during long-term (>6 mo) storage. The sediment can be avoided during pipetting and does not appear to impede infectivity. We typically also freeze 1-mL aliquots of all viral stocks as a backup in case of contamination of the stock solutions stored at 4°C.
7. Suspension culture is initiated 8–9 d prior to infection, beginning with a 60-mL culture seeded at approx  $0.5 \times 10^6$  cells/mL from confluent T-75 monolayer cultures. This is expanded at 48–72 h intervals to 250 mL, then to  $2 \times 500$  mL, then to  $8 \times 500$  mL.
8. It is also possible to use continuous-flow centrifugation to harvest infected cells.
9. Use an artist’s paintbrush with synthetic bristles to resuspend the pellet thoroughly, *i.e.*, without clumps. The pellet is initially resuspended in a small volume to promote efficient cell lysis by the Polytron disintegrator. Monitor the progress of cell lysis by light microscopy. Typically, the membranes must be pelleted, resuspended, and lysed three times to achieve 95% lysis.
10. Mix liquid detergents (*e.g.*, TX100) carefully before preparing the detergent stock.
11. Following the BCA protein determination, the resuspended membrane pellets can be frozen at  $-80^\circ\text{C}$  for long-term storage. The pellet is thawed in a water bath at RT, and divided into four aliquots. Five milliliters of dilution buffer are added to each aliquot, which is then homogenized by treatment with the Polytron (10 s;

- 10,000 rpm). It is also possible to prepare membranes from 8 L of infected cells, and to freeze one-half of the membranes for later use.
12. If a functional assay is not available, Western blotting with an appropriate antibody can be used to estimate the amount of protein expressed and/or the fraction of protein solubilized.
  13. To reduce the amount of radioligand required, it is possible to dilute the radioligand 5- to 15-fold using unlabeled ligand for the highest concentration (here, e.g., 100 and 300 nM) measurements. If the affinity of the solubilized protein has already been established, it is also possible to compare ligand-binding capacity at a single radioligand concentration (e.g., see **Subheading 3.2.5.**).
  14. It is possible to work with smaller cell culture volumes per data point, although this requires more careful technique during membrane preparation and solubilization. The data presented in **Fig. 2** were obtained from 50 mL of culture per time-point.
  15. Filter binding measurements can also be performed on washed membranes without solubilization, permitting optimization of time and multiplicity of infection prior to solubilization conditions. However, the membrane fragments can cause blockage of the filters, leading to uneven filtration and more variable results.
  16. Once a  $K_d$  value has been established, ligand-binding capacity can be compared among samples using a single radioligand concentration. In **Figs. 2** and **3**, comparisons were performed using 15 nM [ $^3\text{H}$ ]AMPA.
  17. As an alternative, grids can be plunged into liquid nitrogen before they dry completely. This can improve specimen preservation, but does require the use of cryo-EM techniques in all subsequent observation steps.
  18.  $\text{Ni}^{++}$  and  $\text{Co}^{++}$  can also be used as counterions for immobilized metal affinity chromatography. Different counterions frequently exhibit different affinities for the polyhistidine tag.

## Acknowledgments

The authors thank Doris Bader for outstanding technical assistance throughout the project, Dr. Willem Tichelaar and Helga Clasen for electron microscopy, and their colleagues in the Ion Channel Structure Group for advice. The authors would also like to thank Drs. Kari Keinänen and Arja Pasternack (University of Helsinki, Helsinki, Finland) for providing the recombinant baculovirus v506-2 for GluR2 expression, and for their on-going collaboration and advice. This work was supported by the Max Planck Society and by HHMI award no. 76200-560801 to Dartmouth Medical School under the Biomedical Research Support Program for Medical Schools.

## References

1. Montelione, G. T. and Anderson, S. (1999) Structural genomics: keystone for a Human Proteome Project. *Nat. Struct. Biol.* **6**, 11–12.
2. Müller, A., MacCallum, R. M., and Sternberg, M. J. E. (2002) Structural characterization of the human proteome. *Genome Res.* **12**, 1625–1641.
3. Tate, C. G. (2001) Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett.* **504**, 94–98.

4. Grisshammer, R. and Tate, C. G. (1995) Overexpression of integral membrane proteins for structural studies. *Q. Rev. Biophys.* **28**, 315–422.
5. Green, T., Stauffer, K. A., and Lummis, S. C. (1995) Expression of recombinant homo-oligomeric 5-hydroxytryptamine<sub>3</sub> receptors provides new insights into their maturation and structure. *J. Biol. Chem.* **270**, 6056–6061.
6. Kuusinen, A., Arvola, M., Oker-Blom, C., and Keinänen, K. (1995) Purification of recombinant GluR-D glutamate receptor produced in Sf21 insect cells. *Eur. J. Biochem.* **233**, 720–726.
7. Dale, W. E., Textor, J. A., Mercer, R. W., and Simchowicz, L. (1996) Expression and characterization of the human erythrocyte anion exchanger in a baculovirus/Sf-9 cell system. *Protein Expr. Purif.* **7**, 1–11.
8. Weill, C., Autelitano, F., Guenet, C., Heitz, F., Goeldner, M., and Ilien, B. (1997) Pharmacological and structural integrity of muscarinic M2 acetylcholine receptors produced in Sf9 insect cells. *Eur. J. Pharmacol.* **333**, 269–278.
9. Autry, J. M. and Jones, L. R. (1998) High-level coexpression of the canine cardiac calcium pump and phospholamban in Sf21 insect cells. *Ann. N. Y. Acad. Sci.* **853**, 92–102.
10. Chinni, C., Bottomley, S. P., Duffy, E. J., Hemmings, B. A., and Stone, S. R. (1998) Expression and purification of the human thrombin receptor. *Protein Expr. Purif.* **13**, 9–15.
11. Fukuzono, S., Takeshita, T., Sakamoto, T., Hisada, A., Shimizu, N., and Mikoshiba, K. (1998) Overproduction and immuno-affinity purification of myelin proteolipid protein (PLP), an inositol hexakisphosphate-binding protein, in a baculovirus expression system. *Biochem. Biophys. Res. Commun.* **249**, 66–72.
12. van Huizen, R., Czajkowsky, D. M., Shi, D., Shao, Z., and Li, M. (1999) Images of oligomeric Kv beta 2, a modulatory subunit of potassium channels. *FEBS Lett.* **457**, 107–111.
13. Huang, Y. W., Lu, M. L., Qi, H., and Lin, S. X. (2000) Membrane-bound human 3beta-hydroxysteroid dehydrogenase: overexpression with His-tag using a baculovirus system and single-step purification. *Protein Expr. Purif.* **18**, 169–174.
14. Rao, U. S., Mehdi, A., and Steimle, R. E. (2000) Expression of amiloride-sensitive sodium channel: a strategy for the coexpression of multimeric membrane protein in Sf9 insect cells. *Anal. Biochem.* **286**, 206–213.
15. Tierney, M. L. and Unwin, N. (2000) Electron microscopic evidence for the assembly of soluble pentameric extracellular domains of the nicotinic acetylcholine receptor. *J. Mol. Biol.* **303**, 185–196.
16. Antaramian, A., Butanda-Ochoa, A., Vazquez-Martinez, O., Diaz-Munoz, M., and Vaca, L. (2001) Functional expression of recombinant type 1 ryanodine receptor in insect cells. *Cell Calcium* **30**, 9–17.
17. Werten, P. J., Hasler, L., Koenderink, J. B., et al. (2001) Large-scale purification of functional recombinant human aquaporin-2. *FEBS Lett.* **504**, 200–205.
18. Safferling, M., Tichelaar, W., Kümmerle, G., et al. (2001) First images of a glutamate receptor ion channel: oligomeric state and molecular dimensions of GluRB homomers. *Biochemistry* **40**, 13,948–13,953.
19. Massotte, D. (2003) G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim. Biophys. Acta* **1610**, 77–89.

20. Warne, T., Chirnside, J., and Schertler, G. F. (2003) Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *Biochim. Biophys. Acta* **1610**, 133–140.
21. Joyce, K. A., Atkinson, A. E., Bermudez, I., Beadle, D. J., and King, L. A. (1993) Synthesis of functional GABAA receptors in stable insect cell lines. *FEBS Lett.* **335**, 61–64.
22. Farrell, P. J., Lu, M., Prevost, J., Brown, C., Behie, L., and Iatrou, K. (1998) High-level expression of secreted glycoproteins in transformed lepidopteran insect cells using a novel expression vector. *Biotechnol. Bioeng.* **60**, 656–663.
23. Higgins, M. K., Demir, M., and Tate, C. G. (2003) Calnexin co-expression and the use of weaker promoters increase the expression of correctly assembled Shaker potassium channel in insect cells. *Biochim. Biophys. Acta* **1610**, 124–132.
24. Jarvis, D. L. (2003) Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production. *Virology* **310**, 1–7.
25. Frank, J., Radermacher, M., Penczek, P., et al. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struc. Biol.* **116**, 190–199.
26. van Heel, M., Harauz, G., and Orlova, E. V. (1996) A new generation of the IMAGIC image processing system. *J. Struc. Biol.* **116**, 17–24.
27. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struc. Biol.* **128**, 82–97.
28. Schiöth, H. B., Kuusinen, A., Muceniece, R., Szardenings, M., Keinänen, K., and Wikberg, J. E. (1996) Expression of functional melanocortin 1 receptors in insect cells. *Biochem. Biophys. Res. Commun.* **221**, 807–814.
29. Keinänen, K., Wisden, W., Sommer, B., et al. (1990) A family of AMPA-selective glutamate receptors. *Science* **249**, 556–560.
30. Sommer, B., Keinänen, K., Verdoorn, T. A., et al. (1990) Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* **249**, 1580–1585.
31. Madden, D. R., Abele, R., Andersson, A., and Keinänen, K. (2000) Large-scale expression and thermodynamic characterization of a glutamate receptor agonist-binding domain. *Eur. J. Biochem.* **267**, 4281–4289.
32. Ausubel, F. M., Brent, R., Kingston, R. E., et al. (eds.) (1999) *Short Protocols in Molecular Biology*, John Wiley, New York, NY.
33. O'Reilly, D. R., Miller, L. K., and Luckow, V. A. (1994) *Baculovirus Expression Vectors: A Laboratory Manual*, Oxford University Press, New York, NY.
34. Keinänen, K., Kohr, G., Seeburg, P. H., Laukkanen, M. L., and Oker-Blom, C. (1994) High-level expression of functional glutamate receptor channels in insect cells. *Biotechnology (N.Y.)* **12**, 802–806.
35. Safferling, M. (2000) *Expression und Charakterisierung von Neurotransmitterrezeptoren am Beispiel des nikotinischen  $\alpha 7$ -Acetylcholinrezeptors und des ionotropen Glutamaterezeptors GluRB*, Ph.D., University of Kaiserslautern, Kaiserslautern, Germany, May 2000.
36. Tichelaar, W., Safferling, M., Keinänen, K., Stark, H., and Madden, D. R. (2004) The three-dimensional structure of a nonotropic glutamate receptor reveals a dimer-of-dimers assembly. *J. Mol. Biol.* **344**, 435–442.

37. Scarborough, G. A. (1994) Large single crystals of the *Neurospora crassa* plasma membrane H<sup>+</sup>-ATPase: an approach to the crystallization of integral membrane proteins. *Acta Crystallogr.* **D50**, 643–649.
38. Auer, M., Scarborough, G. A., and Kühlbrandt, W. (1999) Surface crystallisation of the plasma membrane H<sup>+</sup>-ATPase on a carbon support film for electron crystallography. *J. Mol. Biol.* **287**, 961–968.
39. Kühlbrandt, W. (1992) Two-dimensional crystallization of membrane proteins. *Q. Rev. Biophys.* **25**, 1–49.
40. Hasler, L., Heymann, J. B., and Engel, A. (1998) 2D crystallization of membrane proteins: rationales and examples. *J. Struct. Biol.* **121**, 162–171.
41. Mosser, G. (2001) Two-dimensional crystallogenesis of transmembrane proteins. *Micron* **32**, 517–540.
42. Rigaud, J. -L., Chami, M., Lambert, O., Levy, D., and Ranck, J. -L. (2000) Use of detergents in two-dimensional crystallization of membrane proteins. *Biochim. Biophys. Acta* **1508**, 112–128.
43. Ringler, P., Heymann, B., and Engel, A. (2000) Two-dimensional crystallization of membrane proteins. In: *Membrane Transport: A Practical Approach*, (Baldwin, S. A., ed.), Oxford University Press, Oxford, UK.
44. Stahlberg, H., Fotiadis, D., Scheuring, S., et al. (2001) Two-dimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins. *FEBS Lett.* **504**, 166–172.
45. Lévy, D., Mosser, G., Lambert, O., Moeck, G. S., Bald, D., and Rigaud, J. -L. (1999) Two-dimensional crystallization on lipid layer: a successful approach for membrane proteins. *J. Struct. Biol.* **127**, 44–52.
46. Lebeau, L., Lach, F., Vénien-Bryan, C., et al. (2001) Two-dimensional crystallization of a membrane protein on a detergent-resistant lipid monolayer. *J. Mol. Biol.* **308**, 639–647.
47. Invitrogen Corporation. (2002) Bac-to-Bac<sup>®</sup> Baculovirus Expression System Manual. Version C, p. 5.



## Protein Engineering

Sonia Longhi, François Ferron, and Marie-Pierre Egloff

### Summary

This chapter focuses on protein engineering strategies that aim to increase the chances of obtaining crystals suitable for X-ray diffraction. The chapter is divided into three main parts: one dealing with protein engineering through a bioinformatics approach, the second focusing on DNA modifications via random mutagenesis, and the third describing a nonexhaustive number of *in vitro* modifications based on site-directed mutagenesis.

**Key Words:** Crystallization; mutagenesis; limited proteolysis; error-prone PCR; DNA shuffling; bioinformatics.

### 1. Introduction

The genomics era has provided the scientific community with a realistic estimate of the actual length and complexity of proteins encoded by genomes. In particular, Gerstein (1) has shown that the average length of sequences encoded by eight microbial genomes is 340 amino acids, which is twice the average length (170 residues) of polypeptide chains found in the Protein Data Bank ([www.pdb.org](http://www.pdb.org)). This reflects the bias toward small, soluble, stable proteins used in crystallographic and nuclear magnetic resonance (NMR) studies (2), as well as for proteins that fold naturally when expressed in bacteria or that can be refolded *in vitro*. The recent approach of protein structure determination via structural genomics does not abolish this skewing of our knowledge of protein properties: structural genomics projects have actually demonstrated the feasibility of rapid structure determination but considerable bottlenecks remain because approximately only 10% of the initial targeted proteins actually lead to diffraction-quality crystals (3,4).

Current strategies for overcoming these bottlenecks can be divided into two categories. One is to use native protein sequences and test various expression or

refolding strategies to find conditions that yield satisfactory amounts of soluble proteins. Approximately 30% of soluble, monodisperse proteins do not crystallize or give crystals of insufficient quality for crystallographic analysis and it is well established that the protein itself is the most important parameter in the crystallization process. Consequently, an alternative method is to use protein sequences that have been modified in order to optimize the protein's suitability for structure determination while conserving its native conformation. This approach makes possible an extensive search because for each target protein, a large set of newly engineered proteins can be screened. The main disadvantage of this method lies in the possible loss of functionality of the resulting protein. However, extensive studies have proven that the integrity of internal structure and functional properties can often tolerate the introduction of one or several mutations (5,6) and success for crystallization has been reported to be enhanced by the removal of loops, N- or C-terminal flexible regions, heterogeneous surface features, or introduction of new lattice interactions (7–9). Likewise, truncation of protein domains in multifunction proteins has been successfully used to delineate protein domains, which in most cases can retain enzymatic activity and lead to crystals (10).

This chapter focuses on the latter strategy, i.e., how to modify the protein of interest in order to obtain crystals suitable for X-ray diffraction. The chapter is divided into three main parts, one dealing with protein engineering through a bioinformatics approach, the second focusing on DNA modifications via random mutagenesis, and the third describing a nonexhaustive number of in vitro modifications based on site-directed mutagenesis.

## 2. Materials

### 2.1. In Silico Protein Engineering

1. Computer connected to the Internet.
2. Amino acid sequence of the target protein.

### 2.2. Random Mutagenesis

#### 2.2.1. Nested Deletion Libraries

1. Expression plasmid.
2. Restriction enzymes and appropriate buffers (New England BioLabs, Ipswich, MA).
3. Exonuclease III and appropriate 10X buffer (New England BioLabs).
4. Mung Bean nuclease and appropriate 10X buffer (New England BioLabs).
5. T4 DNA ligase and appropriate 10X buffer (New England BioLabs).
6. *Escherichia coli*–competent cells.
7. 100 and 70% ethanol.
8. Phenol, equilibrated in TE buffer (10 mM Tris-HCl and 1 mM EDTA), pH 8.0.
9. Chloroform.

10. TE buffer, pH 8.0.
11. 1% Sodium dodecyl sulfate (SDS) and 0.2 M Na<sub>2</sub> EDTA.
12. Agarose gel electrophoresis and DNA sequencing equipment.

### 2.2.2. Error-Prone PCR Random Mutagenesis

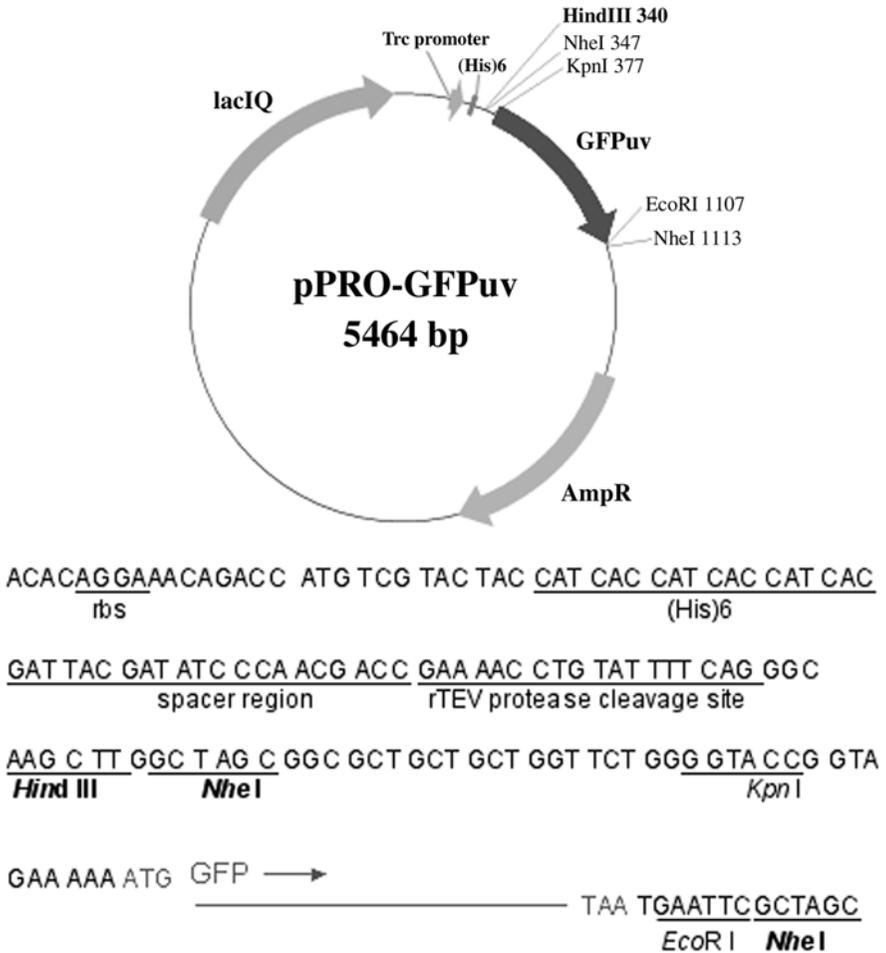
1. cDNA template.
2. Oligonucleotide primers.
3. Mutazyme<sup>®</sup> DNA polymerase and appropriate buffer (Stratagene, San Diego, CA).
4. dNTPs mixtures.
5. Restriction enzymes, T4 DNA ligase and appropriate 10X buffers (New England BioLabs) or, alternatively, appropriate reagents for Gateway<sup>™</sup> Cloning Technology (Invitrogen, Carlsbad, CA).
6. Expression vector bearing either a multiple cloning site for conventional restriction-mediated cloning, or site-specific recombination sites for Gateway Cloning Technology (Invitrogen).
7. *E. coli*-competent cells.
8. PCR thermocycler.
9. Agarose gel electrophoresis equipment.

### 2.2.3. Libraries of Random PCR cDNA Fragments

1. cDNA template.
2. Oligonucleotide primers.
3. *Taq* DNA polymerase and appropriate buffer (Takara Premix *Taq*, Takara BioInc., Shiga, Japan).
4. dNTPs mixtures.
5. Restriction enzymes and appropriate buffers (New England BioLabs).
6. T4 DNA ligase and appropriate 10X buffer (New England BioLabs).
7. pPRO-GFPuv vector (**II**) (see **Fig. 1**).
8. *E. coli*-competent cells.
9. PCR thermocycler.
10. Ultraviolet (UV) transilluminator.
11. Agarose gel electrophoresis, SDS-polyacrylamide gel electrophoresis (PAGE), and DNA sequencing equipment.

### 2.2.4. DNA Shuffling

1. Genes to be shuffled.
2. DNase I and appropriate buffer (Sigma-Aldrich, Dorset, UK).
3. Oligonucleotide primers.
4. *Taq* and *Pfu* DNA polymerases and appropriate 10X buffers (Takara and Stratagene, respectively).
5. dNTPs mixtures.
6. Restriction enzymes, T4 DNA ligase and appropriate 10X buffers (New England BioLabs) or, alternatively, appropriate reagents for Gateway Cloning Technology (Invitrogen).



**Fig. 1.** The pPRO-GFPuv vector.

7. Expression vector bearing either a multiple-cloning site for conventional restriction-mediated cloning, or site-specific recombination sites for Gateway Cloning Technology (Invitrogen).
8. *E. coli*-competent cells.
9. PCR thermocycler.
10. Agarose gel electrophoresis equipment.

### 2.3. Protein Modification

#### 2.3.1. Removal of Flexible Protein Regions

1. cDNA template.

2. Oligonucleotide primers containing the desired mutation.
3. QuikChange site-directed mutagenesis kit (Stratagene).

### 2.3.2. Autoproteolysis

1. 20  $\mu$ L of the protein sample concentrated at 5 mg/mL.
2. SDS-PAGE equipment, appropriate buffers, and molecular weight markers.
3. Glycerol.
4. Dithiothreitol.

### 2.3.3. Limited Proteolysis as a Tool for Protein Engineering

1.  $\alpha$ -Chymotrypsin, elastase, enteproteinase Glu-C V8, subtilisin, and trypsin (lyophilized powder from Sigma).
2. HCl.
3. 30  $\mu$ L of pure protein sample concentrated at 0.5 mg/mL.
4. SDS-PAGE equipment and appropriate buffers and molecular weight markers.
5. Buffer A: 10 mM Tris-HCl, pH 7.5, and 100 mM NaCl.
6. cDNA template.
7. Oligonucleotide primers containing the desired mutation.
8. QuikChange site-directed mutagenesis kit (Stratagene).

### 2.3.4. Modifying the Carbohydrate Content of Glycoproteins

1. cDNA template.
2. Oligonucleotide primers containing the desired mutation.
3. QuikChange site-directed mutagenesis kit (Stratagene).
4. *N*-glycosidase F (Roche, Basel, Switzerland).
5. Purified target protein concentrated at 3 mg/mL.
6. Transferrin from human serum (Roche).
7. Buffer A: 20 mM Tris-HCl and 150 mM NaCl, pH 8.0.
8. SDS-PAGE equipment and appropriate buffers and molecular weight markers.
9. Endoglycosidase H (Roche).

### 2.3.5. Crystal Contacts Engineering Via Surface Residues Mutagenesis

1. cDNA template.
2. Oligonucleotide primers containing the desired mutation.
3. QuikChange site-directed mutagenesis kit (Stratagene).

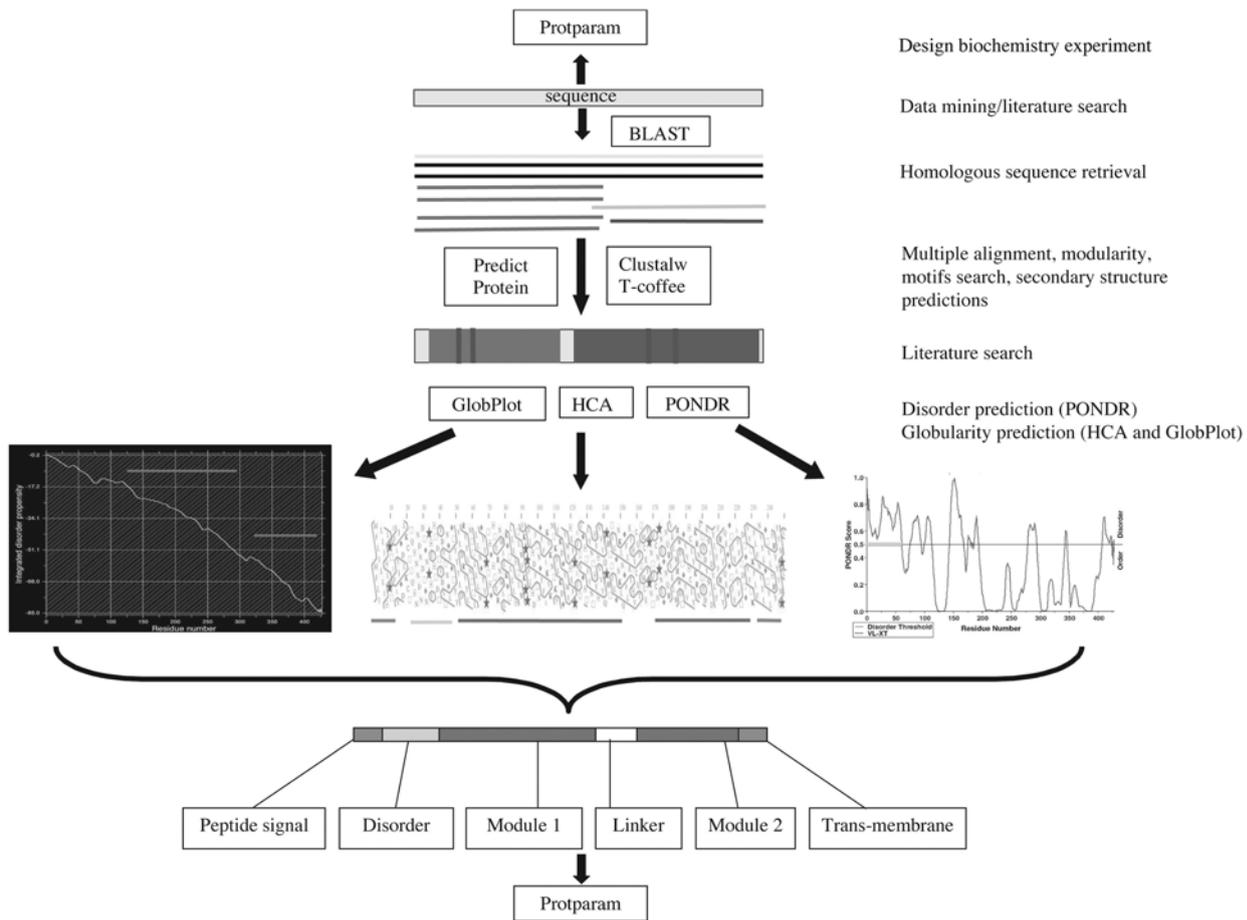
### 2.3.6. Fusion Proteins

1. cDNA template.
2. IMPACTTM-CN System (New England BioLabs).

## 3. Methods

### 3.1. In Silico Protein Engineering

*In silico* approach is one of the tools currently used to engineer proteins in order to make them more likely to crystallize than native target proteins. For



**Fig. 2.** Flowchart for standard *in silico* protein engineering.

instance, dissection of a whole protein into domains is a useful approach both in terms of functional and structural analysis of large proteins comprising two or more functional domains (10). Bioinformatics can be successfully used to define such domains; it can also provide information regarding which residue is worth mutating or which region of the sequence should be removed in order to improve the “crystallizability” of the target protein.

We describe here a simple, step-by-step approach to analyze the target protein sequence using bioinformatics. This approach is illustrated in Fig. 2. Any prior information on the protein or one of its homologs has to be taken into account: protein–protein interactions, oligomeric state, membrane binding, solubility, previous identification of conserved motifs or presence of peptide signal, and so on. This information can be collected from MEDLINE (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) or with web browsers like Google (<http://directory.google.com/Top/Science/Biology/>). Useful programs dedicated to protein sequence analysis are available on the Expert Protein Analysis System server (<http://www.expasy.org/>). For instance, the “ProtParam” program computes some physicochemical properties of the target protein from its amino acid sequence, which may help in designing some biochemical experiments.

### 3.1.1. Gathering of an Informative Cluster of Protein Sequences

#### 3.1.1.1. RETRIEVING HOMOLOGOUS SEQUENCES

Making a set of homologous sequences more or less distant from the target protein sequence is the first step of the bioinformatics method described here. As a rule of thumb, the most similar sequences will highlight the conserved sequence motifs, whereas the less similar sequences will provide information on the modularity of the target protein.

This step is performed using the BLAST program (12) and the BLASTP option at the following URL: <http://www.ncbi.nlm.nih.gov/BLAST/>.

1. Paste your sequence in raw format (i.e., without the first line that starts with a “>” sign).
2. Choose a protein sequence database. The following databases are available:
  - a. Nr: all non-redundant GenBank CDS translations, RefSeq proteins, PDB, SwissProt, PIR, and PRF.
  - b. refseq: the NCBI Reference Sequence is a non-redundant sequence database of genomes, transcripts, and proteins.
  - c. swissprot: the last major release of the SWISS-PROT protein sequence database (no updates).
  - d. pat: protein sequences derived from the Patent division of GenBank.
  - e. pdb: sequences derived from the three-dimensional structure Protein Data Bank (13).
  - f. Month: all new or revised GenBank CDS translation, PDB, SwissProt, PIR, and PRF released in the last 30 d.

3. Press the “BLAST” button.
4. An intermediate page appears. At this step, you have two possibilities:
  - a. Click on the interactive image. The link leads to a first alignment, usually referenced by a publication.
  - b. Continue the BLAST query. Click then on the “Format” button.
5. According to your own criteria, select the most interesting sequences by ticking the box then click on the “Get selected sequences” button.
6. A new page appears. Next to the “Display” button, select FASTA and then click on “Display.” Next to the “Send to” button, select “Text” and then click on “Send to.”
7. Save the page in text format. This file will be your input file to the alignment program.

### 3.1.1.2. SEQUENCE ALIGNMENTS

An alignment of the sequences that have been previously retrieved is highly informative. It highlights conserved motifs, points out the differences between highly and poorly conserved regions, or suggests where loops or linkers are located. We present here two automatic sequence alignment programs.

**3.1.1.2.1. ClustalW (14)** The user interface is available at: <http://bioweb.pasteur.fr/seqanal/interfaces/clustalw-simple.html>.

1. Paste your sequence in FASTA format including the first line that starts with a “>” sign.
2. Fill in your e-mail address.
3. Press the “Run ClustalW” button.
4. Click the link: infile.aln and save in text format.

**3.1.1.2.1. T-Coffee (15)** The user interface is available at: <http://www.ch.embnet.org/software/TCoffee.html>.

1. Paste your sequence in FASTA format including the first line that starts with a “>” sign.
2. Press the “Run T-COFFEE” button.
3. Click the link: clustalw (aln) and save in text format or directly visualize in html or pdf.

For a nice output in postscript format, use the ESPRIPT program (<http://esript.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi>).

### 3.1.1.3. SECONDARY STRUCTURE PREDICTIONS

This step should help the user in defining the modularity of the target protein (linkers usually lack secondary structure). If the BLAST search has revealed a structural homolog, the secondary structure predictions may confirm if your sequence alignment is correct, and may also help in designing modifications (mutations, insertions, or deletions) of the target protein. One of the most com-

prehensive websites for protein structure analysis is provided by the predict protein server (16) located at: <http://www.predictprotein.org/newwebsite/submit.html>.

1. Enter your e-mail address in the top text field.
2. Select the output format by checking a box. (Deselect the default mode and html mode, the result will be sent to you by e-mail, in text format.)
3. Provide a name for your run (optional but useful).
4. Paste your sequence in raw format without the first line that starts with a ">" sign.
5. Click submit/run prediction button.
6. Save the e-mailed result as a text file.

Use the ESPRIPT program (17) (<http://escript.ibcp.fr/ESPrict/cgi-bin/ESPrict.cgi>) to visualize the output:

1. Provide sequence alignment file clicking the "Browse" button.
2. Provide the predict protein file clicking the "Browse" button in the section secondary structure.
3. Press the "submit" button.

### 3.1.2. Domain Definition

#### 3.1.2.1. GLOBULARITY

In this section, two different methods providing indications about protein globularity are described: the GlobPlot and the hydrophobic cluster analysis (HCA) methods. The latter is a powerful method, although it may not be straightforward for nonexperts. A basic, but still useful, analysis of the results is described next.

*3.1.2.1.1. The GlobPlot Method* The GlobPlot method (<http://globplot.embl.de/>) provides a plot where the globularity is reflected by each change in the slope of the plot (18). The graphic output outlines "disordered" and "potential globular domains" according to the author's definition.

1. Paste your sequence in raw format without the first line that starts with a ">" sign.
2. Click on "GlobPlot now" button.
3. Follow the link to download the result file. The result file is a postscript file, therefore, a postscript previewer such as Ghostscript (<http://www.cs.wisc.edu/~ghost/>) is needed.

*3.1.2.1.2. The Hydrophobic Cluster Analysis Method* The HCA method (<http://psb11.snv.jussieu.fr/hca/hca-seq.html>) reflects the environment of each amino acid in the structure and highlights hydrophobic clusters (19).

1. Paste your sequence in raw format without the first line that starts with a ">" sign.
2. Click the "send" button.
3. Follow the link to download the postscript result file (use Ghostscript previewer).

A density of hydrophobic clusters mixed with few charged amino acids reflects the globular parts of the protein. Large hydrophobic clusters may indicate a transmembrane region. A lack of cluster indicates the presence of a linker or a disordered (natively unfolded) region (*see Subheading 3.1.2.2.*).

Results from both programs should be correlated, and should pinpoint regions of the protein that are worth deleting or mutating in order to get rid of the nonglobular, hydrophobic, or disordered regions, which are likely to prevent the crystallization process.

### 3.1.2.2. DISORDER

A number of proteins include natively unfolded/disordered regions, whose presence is obviously incompatible with crystallization. We mention here two servers that predict disorder from the protein amino acid sequence and the physicochemical parameters they extract from this sequence.

#### 3.1.2.2.1. POND<sup>R</sup>

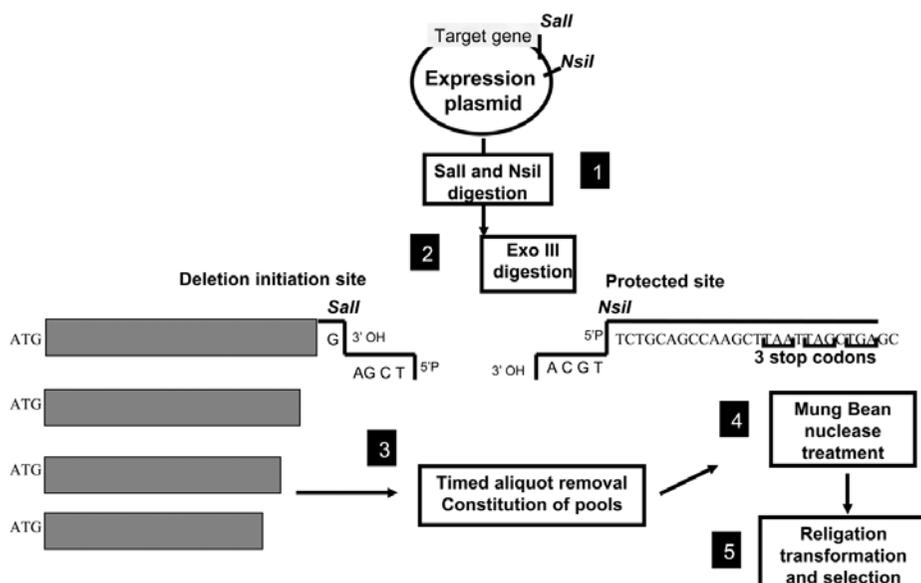
1. Before you can use POND<sup>R</sup>, you will need to create a new user account.
2. Paste your sequence in raw format without the first line that starts with a ">" sign.
3. Click on "submit."
4. The result is given as a gif file. The significance threshold above which residues are considered to be disordered is 0.5 and is indicated on the output file. Segments composed by more than 40 consecutive residues with threshold > 0.5 are highlighted with a thick black line.

#### 3.1.2.2.2. DISEMBL

1. Paste your sequence in raw format without the first line that starts with a ">" sign.
2. Click on "DisEMBL protein" button.
3. Follow the link to download the result file. (The result file is a postscript file, so you need a postscript previewer, such as Ghostscript.)

Each step should be performed on each divergent sequence of your alignment in order to cross-validate your results and to eliminate false-positives. Once disordered regions are clearly identified, one can:

1. Try to stabilize/order these regions by crystallizing the protein with protein-binding partner(s), substrates, or inhibitors (if already determined by biochemical experiments).
2. Reclone the fragment that is predicted to be ordered. As an example, such a bioinformatics analysis was used to generate information on the modular organization of phosphoproteins of the viral subfamily *Paramyxovirinae* and led to the resolution of the crystal structure of the extreme C-terminal domain of the measles virus phosphoprotein (20).



**Fig. 3.** General strategy for the obtention of nested 3'-deletion libraries.

### 3.2. Random Mutagenesis

Directed evolution methods, in which protein diversity libraries are screened for soluble variants, are a powerful tool for obtaining proteins suitable for structural studies. In the following sections we will describe some of the experimental protocols commonly used to generate protein diversity libraries (i.e., nested deletion libraries, error-prone PCR random mutagenesis, random amplification of cDNA fragments, and DNA shuffling). Other methods, such as yeast and phage display, will not be discussed in this chapter because of space limitations.

All these approaches need to be coupled with an efficient, rapid, and reproducible screening method that allows the identification of mutant forms with both increased solubility and stability. Recently, several new protein solubility screens have been developed that do not require structural and functional knowledge of the target (21). Among them, we mention the green fluorescent protein (GFP)-folding reporter method, in which a test protein (or protein domain) is expressed as an N-terminal fusion with GFP: the fluorescence yield of the GFP transduces information about the folding success or failure of the upstream fusion partner.

### 3.2.1. Nested Deletion Libraries

This system is based on the procedure developed by Henikoff (22) and allows the obtention of unidirectional deletions using Exonuclease III (*ExoIII*). This enzyme is used to specifically digest insert DNA from a unique 5'-protruding or blunt-end restriction site, whereas the adjacent unique restriction site is protected by a four base 3'-overhang end (see Fig. 3). The uniform rate of digestion of *ExoIII* allows deletion of predetermined lengths to be made simply by removing timed aliquots from the reaction. A collection of unidirectional deletions spanning several kilobases can be easily constructed in a few hours. The method described next is intended to provide the user with a pilot laboratory-scale protocol, which can of course be scaled-up for high-throughput approaches providing that appropriate (semi)-automated devices are available.

#### 3.2.1.1. RATIONALE FOR THE CHOICE OF THE EXPRESSION PLASMID

The gene of interest (with or without an N-terminal tag) is cloned under the control of a strong inducible promoter. The plasmid must contain a restriction site generating either a 5' overhanging or a blunt-end downstream of the stop codon, followed by a restriction site generating a 3' overhanging, protected end and by three stop codons in the three possible reading frames (see Fig. 3). The plasmid must also include an appropriate origin of replication and a selection marker.

A similar strategy can be used to generate a nested set of 5' deletion fragments. In that case, the protecting restriction site must be located immediately downstream of the ATG codon and followed by the restriction site sensitive to the *ExoIII* digestion. The GFP-encoding gene can be fused at the 3' end of the gene of interest, and increased solubility resulting from N-terminal truncations of the protein of interest can be assessed by following the yield of fluorescence. When generating N-terminal truncations however, only one-third of the deleted fragments will result in truncations in the desired reading frame and DNA sequencing is mandatory to assess the reading frame of selected clones.

#### 3.2.1.2. PREPARATION OF THE DNA TEMPLATE

1. If both restriction endonucleases are able to function effectively in the same buffer, perform a simultaneous double digest. Otherwise, perform the low ionic strength digest first and, after adjusting the reaction conditions, the second one.

As a general protocol combine:

- a. 10  $\mu$ g of DNA.
- b. 5  $\mu$ L of 10X restriction endonuclease buffer.
- c. 5  $\mu$ L of 10X bovine serum albumin (if required for the digestion conditions).
- d. 20 U of each enzyme.
- e. H<sub>2</sub>O to 50  $\mu$ L.

Incubate at the recommended temperature(s) for 1 h.

2. Check for complete linearization by agarose gel electrophoresis.
3. Add one-tenth volume of a solution of 1% SDS and 0.2 M Na<sub>2</sub>EDTA to the digest.
4. Purify the template by standard phenol/chloroform extraction followed by ethanol precipitation (23), resuspend it in 30  $\mu$ L of 1X *Exo*III reaction buffer, and estimate the DNA concentration by agarose gel electrophoresis.

### 3.2.1.3. GENERATION AND PURIFICATION OF NESTED DELETED DNA FRAGMENTS

The following protocol allows the generation of one set of deletions up to 1.5 kb in 200- to 300-bp increments. Six tubes are used, each containing pooled deletion fragments for 50 s intervals. Tube 1 will contain deletions produced in the first 50 s, tube 2 will contain deletions generated between 50 and 100 s, and so on. On average the first tube will contain 250 bases deletions, the second one 500 bases deletions, and so on.

1. Dilute a total of 5  $\mu$ g DNA in a total volume of 60  $\mu$ L in 1X *Exo*III reaction buffer. Place the DNA sample in a 37°C water bath.
2. Distribute 2  $\mu$ L of 10X Mung Bean nuclease buffer and 8  $\mu$ L of sterile distilled water to each of six 1.5-mL microcentrifuge tubes. Place the microcentrifuge tubes on ice. These will be referred to as Mung Bean-treated (MBT) tubes or pools.
3. To the 5  $\mu$ g DNA sample at 37°C, add 250 U of *Exo*III and gently mix.
4. Remove 2- $\mu$ L aliquots at 10-s intervals and place them in the appropriate MBT tubes, as described next. Once all five 2- $\mu$ L aliquots have been collected in one MBT tube, immediately place that MBT tube on dry ice.  
Add *Exo*III digested aliquots as follows:
  - a. 1–5 to MBT Tube 1.
  - b. 6–10 to MBT Tube 2.
  - c. 11–15 to MBT Tube 3.
  - d. 16–20 to MBT Tube 4.
  - e. 21–25 to MBT Tube 5.
  - f. 26–30 to MBT Tube 6.
5. After all aliquots have been taken, heat the six MBT tubes at 68°C for 15 min.
6. Place the tubes on ice for 5 min and then briefly centrifuge them.
7. Add 3 U of Mung Bean nuclease to each tube and incubate at 30°C for 30 min.
8. To each tube add 19  $\mu$ L of TE, pH 8.0.
9. Purify the deleted DNA by standard phenol/chloroform extraction followed by ethanol precipitation (23), resuspend each DNA pellet in 10  $\mu$ L of TE, pH 8.0, and verify the deletion sizes and DNA recoveries by agarose gel electrophoresis of a 5- $\mu$ L sample.
10. Ligate approx 100 ng of the DNA from each MBT tube. In a microcentrifuge tube combine the following:
  - a. 100 ng DNA.
  - b. 2.5  $\mu$ L 10X T4 DNA ligase buffer.
  - c. 150 U T4 DNA ligase.
  - d. H<sub>2</sub>O to 25  $\mu$ L vol.Incubate each of the six ligation reactions overnight at 16°C.

#### 3.2.1.4. TRANSFORMATION AND SELECTION OF CLONES

*E. coli*-competent cells (XL1 Blue; Stratagene) are transformed according to standard protocols (23) and the six pools are plated independently on six agar plates containing the appropriate antibiotic.

The following morning, select a few colonies from each plate and replica plate each onto a fresh agar plate containing the appropriate antibiotic. Incubate plates at 37°C overnight. Resulting colonies are then used to inoculate cultures for plasmid minipreparations. Plasmids are then purified and their size is checked by agarose gel electrophoresis. To ensure that an individual clone still retains its expected sequence integrity, it is recommended to digest the minipreparations with a restriction enzyme whose unique cleavage site lies on the vector and adjacent to the presumptive protected site. Linearization will prove that the *ExoIII* digestion has not proceeded beyond the protected site. This will also provide a more precise estimate of the size of the deleted clone.

Transform the selected plasmids (bearing the desired range of deletions) into an *E. coli* strain suitable for protein expression (BL21[DE3] or Rosetta; Novagen), and check for increased solubility of the target protein using a variety of methods as a function of the fusion tag reporter used. Finally, sequence promising candidates expressing higher amounts of soluble protein to ensure sequence integrity and proper reading frame.

#### 3.2.2. Error-Prone PCR Random Mutagenesis

PCR-based random mutagenesis is widely used for directed protein evolution and has gained popularity over chemical methods because it produces higher levels and a larger variety of mutations (24). The procedure involves performing a PCR reaction under conditions that reduce the fidelity of nucleotide incorporation, cloning the resulting PCR fragments, and then screening the resulting library for mutations affecting protein solubility and stability (25). Cloning of mutated PCR fragments can be carried out using either conventional cloning methods (i.e., methods based on the use of restriction endonucleases followed by ligation) or the recently developed Gateway Cloning Technology (Invitrogen), which relies on site-specific recombination between a target insert (or donor vector) and an expression destination vector. In that case, primers have to be designed so as to contain the specific recombination sites in their floating moiety. Mutations are deliberately introduced through the use of error-prone PCR polymerases and/or by modifying the reaction conditions. A number of commercially available kits have been developed, in which the desired mutation frequency is achieved by tuning either the buffer conditions (i.e., use of MnCl<sub>2</sub>-containing buffer and unbalanced dNTPs concentrations) or the initial amount of target DNA template. The method outlined next is that of the GeneMorph® random mutagenesis kit distributed by Stratagene.

**Table 1**  
**Quantity of DNA Required as a Function of the Expected Mutation Frequency**

Mutation frequency (mutations/kb)	Initial target quantity ( <i>see Note 1</i> )
0–3 (low range)	10–100 ng
3–7 (medium range)	10 pg to 10 ng
7–16 (high range)	Double or triple PCR, with 10–100 pg in each

### 3.2.2.1. HOW TO ACHIEVE THE DESIRED MUTATION FREQUENCY

Mutazyme DNA polymerase produces all possible nucleotide substitutions and can be used to mutagenize plasmid DNA targets up to 4.5 kb. It makes few insertion and deletion mutations, which create undesired frameshifts. Moreover, mutational hotspots have never been observed. Mutation frequency is the product of DNA polymerase error rate and the number of duplication events. With the GeneMorph random mutagenesis kit, low, medium, and high mutation frequencies are obtained by simply adjusting the initial target DNA amounts in the amplification reactions. The rationale is that, for the same PCR yield, targets amplified from low amounts of target DNA will undergo more duplication events than targets amplified from high concentrations of template. The more times a target is replicated, the more errors accumulate. Therefore, higher mutation frequencies are achieved simply by lowering DNA template concentration. Genomic DNA templates are not recommended owing to the low copy number of targets, only medium-to-high mutation frequencies are obtained. If genomic DNA is the only source of the target, then preliminary amplification of the target with a high-fidelity polymerase is recommended. **Table 1** provides the initial target amounts to be used as a function of the desired mutation frequency.

In directed evolution studies, mutation frequencies of one to four amino acid changes per protein (two to seven nucleotide changes per gene) are generally employed (26–28). High mutation frequencies (>8 mutations/kb) can be achieved by performing two or three sequential PCRs, in which the product of the first PCR serves as a template in the second PCR, and so on. The predicted mutation frequencies given in **Table 1** are calculated based on PCR yields ranging from 500 ng to 10  $\mu$ g.

Two useful equations to calculate the mutation frequency are given next:

$$d = [\text{Log}(\text{PCR yield}/\text{initial target amount})]/\text{Log } 2$$

$$\text{Mutation frequency (mutations/kb)} = 0.31 d + 0.41$$

### 3.2.2.2. HOW TO SET UP THE PCR

For optimal results, PCR primers should be designed so as to have similar melting temperatures ( $T_m$ ), ranging from 55 to 72°C. This reduces false priming

and ensures complete denaturation of primers at 94°C. PCR products generated by Mutazyme DNA polymerase are blunt-ended. However, this does not preclude the possibility of cohesive ligation. Indeed, restriction sites for endonucleases generating protruding ends can be introduced on the floating moiety of the primers, thereby allowing digestion of the PCR products prior to ligation into a digested destination vector. Likewise, the AttB1 and AttB2 sequences can be introduced in the forward and reverse primer, respectively, thus allowing cloning into Gateway destination vectors using the Gateway Cloning Technology (Invitrogen).

1. Prepare 50- $\mu$ L reactions as follows:
  - a. 41.5  $\mu$ L H<sub>2</sub>O.
  - b. 5  $\mu$ L 10X mutazyme reaction buffer.
  - c. 1  $\mu$ L 40 mM dNTP mix (200  $\mu$ M each final).
  - d. 0.5  $\mu$ L Primer mix (250 ng/ $\mu$ L of each primer).
  - e. 1  $\mu$ L Mutazyme DNA polymerase (2.5 U/ $\mu$ L).
  - f. 1  $\mu$ L Template (10 pg/ $\mu$ L to 100 ng/ $\mu$ L).

If the thermocycler is devoid of a heated lid, overlay each reaction with a few drops of mineral oil.

2. Perform the PCR as follows: heat at 95°C for 1 min and perform 30 cycles of 95°C 1 min, (primer T<sub>m</sub>, 5°C) 1 min and 72°C 1 min/kb, followed by one cycle at 72°C 10 min (for single-block thermocyclers reduce denaturation and annealing steps to 30 s).
3. Estimate the PCR yield by running 10  $\mu$ L of the PCR product onto agarose gel electrophoresis. For high mutations frequencies (>8 mutations/kb), either use less than 10 pg template and increase the number of cycles, or dilute the PCR product 1:1000 and use 1  $\mu$ L in a second amplification reaction.

### 3.2.3. Libraries of Random PCR cDNA Fragments

A novel approach to identify soluble protein domains has been developed (**11**), which combines tagged random primer PCR method (T-PCR) (**29**) and protein folding assay using GFP (**30**). This approach attempts to identify the boundaries of soluble domains experimentally by examining all possible protein fragments generated by amplification of random fragments from a template cDNA (see **Note 2**).

#### 3.2.3.1. T-PCR

Two primers, primer A (5'-GACCATGATTACGCCAAAGCTTN[15]-3') and primer B (5'-GACCATGATTACGCCAAAGCTT-3'), are used in two separate T-PCR reactions.

1. Perform the first T-PCR as follows: combine 50 ng DNA template and 100 pmol primer A in the presence of *Taq* DNA polymerase buffer and dNTPs in a 25- $\mu$ L total volume. If the thermocycler is devoid of a heated lid, overlay each reaction

with a few drops of mineral oil. Heat at 94°C for 2 min and perform two cycles of 94°C 1 min, 40°C 5 min, and 72°C 2 min.

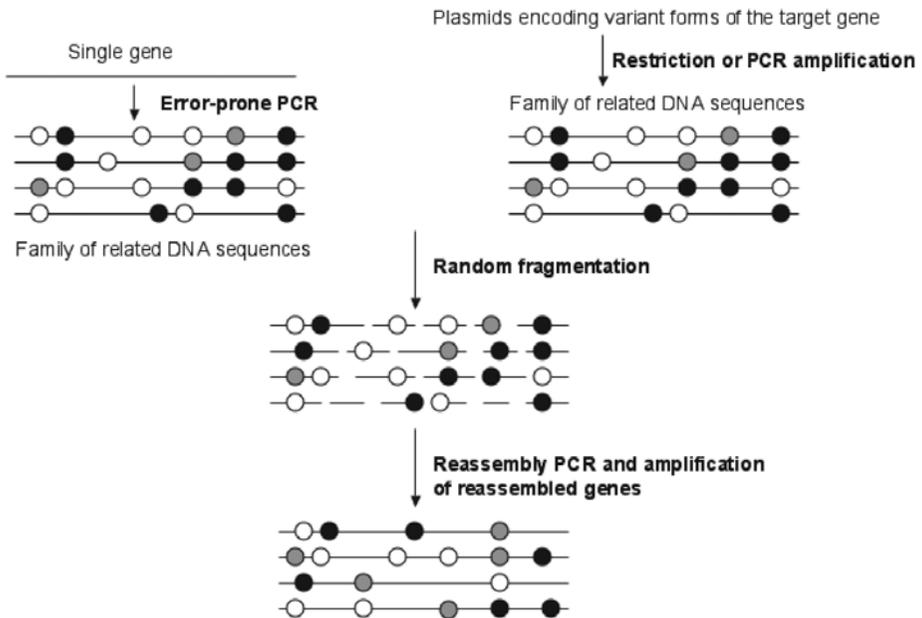
2. Purify the PCR products to remove the primer (QIAquick PCR purification kit; Qiagen, Chatsworth, CA).
3. Perform the second PCR using the first PCR-purified products and 100 pmol primer B in a 50- $\mu$ L total volume. Heat at 94°C for 2 min, and then perform 40 cycles of 94°C 1 min, 55°C 1 min, and 72°C 1 min.
4. Purify T-PCR products (spin columns, Qiagen).
5. Digest the purified T-PCR products with *Hind*III (corresponding site is underlined in the primers).
6. Separate the digested PCR products by agarose gel electrophoresis, and elute fragments corresponding to about 300 bp (Mini elute kit, Qiagen).
7. After purification of a *Hind*III-digested and dephosphorylated pPRO-GFP vector (obtained using standard protocols [23]), the ligation reaction is set up using a molar ratio of insert: vector of 10:1, a final amount of DNA of approx 100 ng in a 10- $\mu$ L total volume in the presence of 150 U of T4 DNA ligase in 1X T4 DNA ligase buffer.
8. Incubate overnight at 16°C and transform *E. coli* DH5 $\alpha$  competent cells (Novagen, Darmstadt, Germany) using standard protocols (23).

#### 3.2.3.2. SCREENING

1. Plate the transformed cells onto agar plates containing 100  $\mu$ g/mL ampicillin and incubate at 37°C for 18 h.
2. Observe colonies with a 366-nm UV transilluminator. Pick the brightest colonies, culture them in liquid medium, and analyze protein content of soluble fractions by SDS-PAGE.
3. Finally, sequence promising candidates expressing soluble proteins in the desired range of sizes to ensure sequence integrity and proper reading frame.
4. Prior to scale-up and purification by Ni Affinity Chromatography (for a detailed procedure see *The Protein Expression and Purification Handbook*, Qiagen), the selected plasmids are digested by *Nhe*I (to remove the GFP-encoding gene) and self-ligated.

#### 3.2.4. DNA Shuffling

The method of DNA shuffling (31), or “sexual PCR,” is used to recombine homologous DNA sequences during in vitro molecular evolution. In this system a pool of closely related sequences is fragmented randomly with DNase I, and these fragments (approx 50–200 bp) are reassembled into full-length genes via self-priming PCR and extension in a process called reassembly PCR. This procedure yields crossovers between related sequences owing to template switching (see Fig. 4 and Note 3). The PCR products are cloned in a suitable expression vector and the resulting library is then screened for increased solubility of the target protein using any of the most commonly used techniques (dot blot, use of fusion reporter tags, and others).



**Fig. 4.** Schematic illustration of the DNA-shuffling process.

The initial pool of related sequences can be obtained either by error-prone PCR random mutagenesis from a unique target gene template, or by restriction or PCR amplification from plasmids bearing variant forms of the gene of interest (see Fig. 4). When coupled with effective selection and applied reiteratively, DNA shuffling has been proven to be an efficient process for directed molecular evolution (32–34).

#### 3.2.4.1. PREPARATION OF GENES TO BE SHUFFLED

The initial pool of closely related genes can be obtained either by error-prone PCR random mutagenesis from a unique gene template (for the experimental procedure refer to **Subheading 3.2.2.**) or by using a number of plasmids harboring mutated versions of the target gene. In this case, the mutated genes can be obtained either by restriction (23) followed by separation on agarose gel electrophoresis and gel extraction (Mini elute kit, Qiagen), or by conventional PCR amplification with a high-fidelity enzyme, such as *Pfu* DNA polymerase (Promega, Madison, WI). Primers are designed to specifically recognize the 5' and 3' ends of the target gene and to bear either restriction sites for conventional subsequent restriction-mediated cloning, or site-specific recombination sites for Gateway Cloning Technology (Invitrogen).

1. In case PCR is used to obtain the initial set of related genes, prepare a 50- $\mu$ L reaction as follows:
  - a. 5  $\mu$ L 10X *Pfu* DNA polymerase buffer.
  - b. 4  $\mu$ L 10 mM dNTPs mix (200  $\mu$ M each final).
  - c. 15 pmol each primer.
  - d. 1  $\mu$ L *Pfu* DNA polymerase (2.5 U/ $\mu$ L).
  - e. 1  $\mu$ L template (10–50 ng/ $\mu$ L).
  - f. Water up to a total of 50  $\mu$ L.If the thermocycler is devoid of a heated lid, overlay each reaction with a few drops of mineral oil.
2. Perform the PCR as follows: heat at 95°C for 1 min and perform 30 cycles of 95°C 1 min, (primer  $T_m$ , 5°C) 1 min, and 72°C 2 min/kb, followed by 1 cycle at 72°C 8 min (for single-block thermocyclers reduce denaturation and annealing steps to 30 s).
3. Estimate the PCR yield by running 5  $\mu$ L of the PCR reaction onto an agarose gel electrophoresis.

#### 3.2.4.2. FRAGMENTATION WITH DNASE I

The PCR products are purified using the QIAquick PCR purification kit (Qiagen) and digested with DNase I.

Different reaction parameters, such as the incubation time and the enzyme-to-substrate ratio, have to be explored before optimal conditions (yielding random DNA fragments ranging from 50 to 200 bp in length) are found. As a first approach, the following conditions can be used: digest 2–4  $\mu$ g of DNA substrate with 0.15 U of DNase I in a 100- $\mu$ L total reaction volume in 50 mM Tris-HCl, pH 7.5, buffer containing 1 mM  $MnCl_2$  for 10–20 min at room temperature. The use of  $Mn^{2+}$  instead of  $Mg^{2+}$  during the DNase I fragmentation step has been reported to improve the fidelity of DNA shuffling (35), which otherwise would result in a point mutation rate of 0.7% (36). After assessing the length of the resulting fragments by agarose gel electrophoresis, the DNA fragments are purified by the QIAquick PCR purification kit (Qiagen) to remove residual  $Mn^{2+}$ .

#### 3.2.4.3. PCR REASSEMBLY

The purified DNA fragments are added to the reassembly PCR mixture at 20 ng/ $\mu$ L.

1. Prepare a 50- $\mu$ L reaction as follows:
  - a. Purified DNA fragments at a final concentration of 20 ng/ $\mu$ L.
  - b. 5  $\mu$ L 10X *Pfu* Turbo DNA polymerase buffer.
  - c. 4  $\mu$ L 10 mM dNTPs mix (200  $\mu$ M each final).
  - d. 1  $\mu$ L 2.5 U/ $\mu$ L *Pfu* DNA polymerase.
  - e. Water up to a total of 50  $\mu$ L.

Note that no primers are added at this point.

If the thermocycler is devoid of a heated lid, overlay each reaction with a few drops of mineral oil.

2. Perform the PCR as follows: heat at 94°C 1 min, and perform 30–45 cycles of 94°C 30 s, 50–55°C 30 s, and 72°C 30 s, followed by a final elongation step at 72°C 5 min.

#### 3.2.4.4. PCR AMPLIFICATION OF REASSEMBLED GENES (SEE FIG. 4)

The product of self-reassembling PCR is then diluted 40 times and used as template in a high-fidelity PCR amplification by using primers recognizing the 5' and 3' ends of the target gene.

1. Prepare a 50- $\mu$ L reaction as follows:
  - a. 1  $\mu$ L reassembly reaction.
  - b. 5  $\mu$ L 10X *Pfu* Turbo DNA polymerase buffer.
  - c. 4  $\mu$ L 10 mM dNTPs mix (200  $\mu$ M each final).
  - d. 40 pmol each primer.
  - e. 2  $\mu$ L 2.5 U/ $\mu$ L *Pfu* DNA polymerase.
  - f. Water up to a total of 50  $\mu$ L.If the thermocycler is devoid of a heated lid, overlay each reaction with a few drops of mineral oil.
2. Perform the PCR as follows: heat at 94°C 5 min, and perform 10–20 cycles of 94°C 30 s, 50°C 30 s, and 72°C 30 s.
3. Estimate the PCR yield by running 5  $\mu$ L of the PCR reaction onto an agarose gel electrophoresis and proceed to subsequent cloning. After screening, a subset of mutants with increased solubility may be selected for a further round of DNA shuffling. Repeating the whole procedure may lead to the isolation of mutants with increased solubility.

### 3.3. Protein Modifications

#### 3.3.1. Removal of Flexible Protein Regions

Structural heterogeneity resulting from protein flexibility often hinders crystallization. N- and C-terminal regions often prove to be flexible and truncation of part of these regions is currently a standard protocol in recombinant protein expression for crystallographic studies. More recently, it has been shown that insertions and deletions within a protein may also be well tolerated, without affecting its biological function, and internal deletions have even been employed to improve the diffracting quality of protein crystals (7).

1. Information on the protein regions, which should be deleted, can be provided by a bioinformatics approach. Sequence alignments with homologous proteins (see **Subheading 3.1.1.2.**) may clearly designate extra regions of the protein target, the removal of which may facilitate crystallization.
2. If the crystal structure of a homologous protein has been solved, the inspection of both the electron density map and the temperature factors after refinement will also pinpoint disordered regions that may be deleted.

3. Autoproteolysis and limited proteolysis are also useful tools, as these reactions are known to occur in flexible, exposed regions of proteins (*see Subheadings 3.3.2. and 3.3.3.*).

Once they are identified, these regions can be removed by deletion mutagenesis, using for instance the QuikChange site-directed mutagenesis kit (Stratagene) (*see Note 4*) according to the manufacturer's protocol.

### 3.3.2. Autoproteolysis as a Tool for Protein Engineering

It has been widely reported that some proteins can undergo spontaneous proteolysis, either when stored at 4°C in their purification buffer, or in the crystallization drop. This obviously generates heterogeneity, which is a major impediment to protein crystal formation, and which cannot be determined at once even when the protein homogeneity is checked by dynamic light scattering (*see Chapter 6*) at the end of the purification process.

#### 3.3.2.1. CHECKING FOR AUTOPROTEOLYSIS

We suggest here to determine whether and under which conditions the protein of interest is prone to autoproteolysis. Different samples of protein are treated as described next and analyzed by SDS-PAGE; the acrylamide concentration is chosen considering that the protein itself and smaller fragments should be detectable. Each sample consists of 9  $\mu\text{L}$  of SDS-loading buffer and 1  $\mu\text{L}$  of protein that has been subjected to the following procedure:

Lane 1: Just after purification, the protein sample was supplemented with 20–50% glycerol and stored at  $-80^{\circ}\text{C}$  for 1 wk. The volume of the protein required for preparing the SDS-PAGE sample is therefore corrected, depending on the glycerol concentration (2  $\mu\text{L}$  instead of 1  $\mu\text{L}$  if the storage was done in 50% glycerol).

Lane 2: The protein sample was stored at 4°C for 1 wk in its purification buffer.

Lane 3: The protein sample, supplemented with 10% glycerol (if not present in the last purification buffer), was stored at 4°C for 1 wk.

Lane 4: The protein sample, supplemented with 5 mM dithiothreitol (if not present in the last purification buffer) was stored at 4°C for 1 wk.

Lane 5: The protein sample was incubated for 1 wk at the same temperature than that used for crystallization trials.

Lane 6: The protein sample was incubated for 1 wk at 25°C.

Additional lanes of the gel can be used to check, after 1 wk, the profile of the protein sampled from several clear crystallization drops.

The last lane is reserved for molecular weight markers.

All samples are heated at 95°C for 5 min before loading on SDS-PAGE gel.

#### 3.3.2.2. IDENTIFICATION OF THE CLEAVAGE SITE

If a proteolytic product is observed after repeated experiments, the corresponding band on the SDS-PAGE gel is cut and identified by comparing its

N-terminal sequence (determined by Edman degradation) and its molecular mass (determined mass spectroscopy) with the full-length protein sequence and molecular mass. Once the cleavage site has been identified, the proteolytic agent can be found (using, for instance, the program PeptideCutter, <http://www.expasy.org/tools/peptidecutter/>), as well as the sequence features required for proteolysis (*see* **Table 2**).

#### 3.3.2.3. DELETING THE CLEAVAGE SITE

The first approach is often to try to crystallize the full-length protein. Mutation of one (or more) amino acids within the protease cleavage site can be performed (using the QuikChange kit, Stratagene; *see* **Note 4**) in order to remove the proteolytic site. If this site is not conserved among homologous proteins (as judged from sequence alignment, *see* **Subheading 3.1.1.2.**), the sequence of the homolog(s) can guide the choice of the amino acid mutation.

#### 3.3.2.4. WORKING WITH THE CLEAVED FORM OF THE PROTEIN

An alternative is to work with a homogeneous sample of the cleaved protein. A first approach consists in reproducing the proteolysis in a controlled manner (*see* **Subheading 3.3.3.**). The optimum conditions for performing the proteolytic reaction are reported in **Table 2** for a number of proteolytic enzymes. The protein of interest is subsequently purified to remove the proteolytic agent and it can be submitted to crystallization trials. One can also subclone the cleaved form of the protein, either according to the same procedure previously used for the full-length protein, or by PCR amplification of the region of interest followed by cloning a stop codon by site-directed mutagenesis (QuikChange kit, Stratagene; *see* **Note 4**). All these procedures have proven to be successful in obtaining diffracting crystals of previously “uncrystallizable” proteins (**37–39**).

#### 3.3.3. Limited Proteolysis as a Tool for Protein Engineering

Limited proteolysis is a classical approach to probe protein structure. As a matter of fact, even if most proteolytic enzymes actually recognize a particular primary sequence, the cleavage itself depends on the overall structural properties of the protein: cleavage will occur at flexible surface loops, linkers, or within unstructured regions. For most single domain proteins, the proteolytic reactions are expected to yield a nicked species that retains its overall folding under nondenaturing conditions, whereas multidomain proteins can potentially separate into individual structural domains. The use of such compact protein domain often leads to crystals more readily, or to better quality crystals (**37,40,41**).

No protocol that will lead to straightforward results can be provided here because this technique requires trial and error. The protease/protein ratio, the type and nature of the proteases employed, the number of aliquot analyzed, and

**Table 2**  
**Common Peptidases and Their Main Features**

Enzyme (type)	Primary specificity	MW (kDa)	Conditions for use Buffer	Inhibitors
$\alpha$ -Chymotrypsin (serine endopeptidase)	P1-P'1- (P1 = Y, F, W; P'1 $\neq$ P)	25	pH 7.5–8.5; 150 mM NaCl Ca <sup>2+</sup> activated	DFP, PMSF, TPCK, aprotinin
Dispase I	nonspecific	36	pH 7.5, 150 mM NaCl	EDTA, heavy atoms
Elastase	P1-P'1- (P1 = A,V,L, I,G,S,T; P'1 $\neq$ P)	25	pH 7.5, 150 mM NaCl	DFP, PMSF
Enteproteinase Glu-C V8	P1-P'1-(P1 = E; P'1 nonspecific)	30	100 mM NH <sub>4</sub> CO <sub>3</sub> pH 8.0–8.5, 150 mM NaCl	DFP, Hg <sup>2+</sup> , Cu <sup>2+</sup> , Zn <sup>2+</sup>
Papain (cysteine endopeptidase)	P2-P1-P'1-(P1 = R/K; P'1 $\neq$ V; P2 = hydrophobic)	23.4	pH 6.0–7.0, 150 mM NaCl	heavy metal ions, leupeptin, PMSF
Subtilisin (serine endopeptidase)	P1-P'1- (P1 nonspecific; P'1 nonspecific)	30	10 mM Tris pH 7.0–8.0, 150 mM NaCl, Ca <sup>2+</sup> activated	PMSF, DFP, aprotinin
Thermolysin (metalloendopeptidase)	P1-P'1-P'2 (P1; nonspecific; P'1 = L,F,I, V,F,Y,M,A; P'2 $\neq$ P)	37.5	10 mM Tris pH 7.0–8.5, 150 mM NaCl stabilized with two 2 mM Ca <sup>2+</sup>	EDTA, heavy atoms, citrate, Pi, oxalate
Trypsin (serine endopeptidase)	P1-P'1-(P1 = K,R; P'1 $\neq$ P)	23.5	pH 8.5–8.8, 150 mM NaCl stabilized with 20 mM Ca <sup>2+</sup>	PMSF, DFP, TLCK, leupeptin, aprotinin

the time-course of the experiment should be determined and further specifically improved for each protein of interest. The protease/protein ratios (w/w) typically used vary from 1/50 to 1/500 so that proteolysis is incomplete and intermediates may be observed accumulating in the course of time. Proteolysis reactions can be stopped by denaturation, addition of a specific protease inhibitor, or removal of the protease. Samples are generally analyzed by gel electrophoresis so that the band of interest can be cut and subsequently identified by Edman degradation sequencing chemistry and mass spectrometry. The reaction volume should therefore be designed to provide the appropriate number of gel sample aliquots. Time increments can be chosen by the experimentalist. Limitation of the proteolytic reaction can be achieved by modifying the amount of protease or by altering the optimal reaction conditions (pH, temperature, ionic strength). Once the experiment has been successfully done and results are analyzed, the fragment of interest can be either purified as a proteolytically stable fragment or recloned (*see Subheading 3.3.1.4*). The features of the enzymes that are mostly used in limited proteolysis experiments for structural analysis purposes are provided in **Table 2**, and an example of such an experiment is also given next.

1. Five similar control samples are prepared, each containing 1  $\mu\text{L}$  of the protein, 9  $\mu\text{L}$  of buffer A (10 mM Tris, pH 7.5 and 100 mM NaCl) and 10  $\mu\text{L}$  of SDS loading buffer. After heating 5 min at 95°C, the five samples are frozen at -20°C.
2. For each proteolytic enzyme, prepare three stock solutions at 5 mg/mL, 0.5 mg/mL, and 0.05 mg/mL in 1 mM HCl (to prevent self-digestion).
3. For each proteolytic enzyme, three Eppendorf vials are prepared, each containing 10  $\mu\text{L}$  of protein and 40  $\mu\text{L}$  of buffer A.
4. Three different ratios of protease/protein (1/10, 1/100, and 1/500) are tested for each protease. At  $t = 0$ , add 1  $\mu\text{L}$  of the 5 mg/mL protease stock to the first Eppendorf vial, 1  $\mu\text{L}$  of the 0.5 mg/mL stock to the second vial, and 2  $\mu\text{L}$  of the 0.05 mg/mL stock to the third vial. Repeat the experiment for each protease.
5. In each vial, 10- $\mu\text{L}$  aliquots are taken after 5, 15, 30, 60 min and placed in a new vial. The reaction is immediately stopped by mixing 10  $\mu\text{L}$  of each aliquot to 10  $\mu\text{L}$  of SDS-loading buffer and heating the mix during 5 min at 95°C. The samples are frozen to -20°C immediately.
6. All the different samples are thawed and the result of each of the five experiments is analyzed by SDS-PAGE.
7. The identification of the cleavage site is performed as described in **Subheading 3.3.2.2**. and a new protein fragment is obtained as described in **Subheading 3.3.2.4**.

#### 3.3.4. Modifying the Protein Carbohydrate Content of Glycoproteins

Many proteins of interest obtained from natural sources or eukaryotic expression systems do not crystallize or yield poor-quality crystals unsuitable

for crystallographic analyses. In fact, the N-linked glycans present on the surface of these molecules generate some conformational and compositional heterogeneity, which may be unfavorable to mediate crystal contacts. On the other hand, carbohydrates also greatly contribute to the protein solubility, and their systematic removal can lead to protein aggregation or precipitation; in addition, N-glycans can even be involved in the crystal packing (42). Crystallizing a glycoprotein, therefore, requires the identification of the sugar chains, which should be removed. The two following approaches have been successfully used to reduce/abolish the sugar content of glycoproteins and to crystallize them.

#### 3.3.4.1. SITE-DIRECTED MUTAGENESIS OF THE PUTATIVE N-GLYCOSYLATION SITES

Putative N-glycosylation sites are found in Asn-X-Ser/Thr motifs. Site-directed mutagenesis can be performed to generate several mutants (usually Asn into Asp or Gln). In the absence of previous information, individual mutations, as well as combinations of the different mutations, should be performed (43). Each mutant should then be purified and characterized in terms of expression level, solubility, folding, and activity. Mutants satisfying all the required criteria are subsequently submitted to crystallization trials. Mutations can be performed by the QuikChange kit (Stratagene) (see **Note 4**). This method was shown to be successful, for instance, for crystallizing the complex between IL-10 and IL-10R1 (44).

#### 3.3.4.2. PARTIAL OR TOTAL DEGLYCOSYLATION BY ENZYMATIC TREATMENT

As mentioned earlier, no rules exist concerning which glycan should be removed to promote crystallization or which enzyme should be used. Moreover, the required amount of enzyme necessary to deglycosylate a given protein depends on the actual assay conditions and the protein itself, and should therefore be determined empirically for each glycoprotein. Deglycosylation reactions are usually carried out overnight at 37°C, but depending on the sensitivity of the target protein to temperature, the time can be reduced and the amount of the glycosidase increased accordingly. We suggest the following protocol as a first attempt to establish the experimental conditions.

1. Label four Eppendorf vials containing 10  $\mu$ L of the protein (concentrated at 3 mg/mL) to be analyzed with “1,” “2,” “3,” and “-,” respectively.
2. Label two Eppendorf vials containing 10  $\mu$ L of the control glycoprotein (3 mg/mL of transferrin from human serum; Roche) with “C1,” “C2,” “C3,” and “C-,” respectively.
3. Add 10  $\mu$ L of the reaction buffer A to each vial and mix.
4. Add 2, 5, and 10  $\mu$ L of 1 U/ $\mu$ L N-glycosidase F (see **Note 5**) to all “+” vials, and adjust the volume to 30  $\mu$ L with the reaction buffer A to all vials. Incubate for 16 h at 37°C.

5. Mix a 10- $\mu$ L aliquot of each sample with 10  $\mu$ L of 2X SDS sample buffer and heat for 5 min to 95°C.
6. Analyze the results on SDS-PAGE comparing the apparent molecular weights for the sample and the control glycoproteins before and after treatment with N-glycosidase F. Analysis of the results may require the use of N-terminal sequencing and mass spectrometry. The mass changes resulting from typical posttranslational modifications of proteins can be found at: [http://www.sigmaaldrich.com/Area\\_of\\_Interest/Life\\_Science/Proteomics/Post\\_Translational\\_Modification/Mass\\_Changes.html](http://www.sigmaaldrich.com/Area_of_Interest/Life_Science/Proteomics/Post_Translational_Modification/Mass_Changes.html)). If required, run a new experiment modifying the amount of N-glycosidase F or the incubation time in order to empirically optimize the experimental parameters so as to perform a maximal deglycosylation.
7. Scale up your experiment to deglycosylate the amount of protein required to set up crystallization trials.
8. The same experiment can be run with Endoglycosidase H (Roche) to generate another type of deglycosylation (*see Note 5*), which will lead to a protein with different properties and probably a different behavior in crystallization trials. You can first estimate that 50–250 mU of Endoglycosidase H are usually sufficient to deglycosylate up to 1 mg glycoprotein per milliliter when incubated overnight at 37°C. The recommended buffer is 10–100 mM potassium or sodium acetate, pH 5.0–6.0.

### 3.3.5. Crystal Contact Engineering Via Surface Residues Mutagenesis

Point mutations of solvent-exposed amino acids have been used to turn proteins recalcitrant to crystallization into crystallizable proteins. This approach should not be neglected because it relies on straightforward experiments (directed mutagenesis) and has been demonstrated to be successful in the crystallization of many proteins (*8,45,46*). A point mutation alters only a small fraction of the protein surface area and how this mutation affects the protein behavior in crystallization assays is hardly predictable. However, we provide here some considerations useful in determining which residues are potential candidates for mutation.

1. If the structure of a protein homologous to the target protein has been solved, check the nature of the residues involved in the crystal contacts, and if they differ from the equivalent residues in the protein of interest (from ClustalW or T-coffee alignment, *see Subheading 3.1.1.2.*), then perform the corresponding mutations.
2. It has been shown that even if lysine and glutamate residues are found almost exclusively on the surface of proteins (*47,48*), they are very unlikely to participate in crystal contact, owing to their high-conformational entropy. Moreover, a survey conducted with a set of 223 unique protein structures showed that arginine and glutamine are the two most frequently observed amino acids found at crystal interfaces. In fact, an approach consisting of performing systematic mutations of surface lysine to arginine, glutamine, or alanine, and of glutamate to glutamine, aspartate, or alanine has proved many times to be successful (*46,49*).

If the surface exposure of such residues cannot be deduced from the structure of homologous proteins, some indications may be provided by the results of secondary structure predictions (*see Subheading 3.1.1.3.*; other secondary structure prediction programs, such as Jpred [<http://www.compbio.dundee.ac.uk/~www-jpred/>] or PredictProtein [<http://cubic.bioc.columbia.edu/predictprotein/>], can also be used). Alternatively, considering the results of the studies previously mentioned (47,48), lysine and glutamate residues can be arbitrarily mutated with an approx 90% chance of being surface exposed. The corresponding proteins should subsequently be purified (if soluble), checked for proper folding, and submitted to crystallization trials.

3. A similar approach can be performed concerning surface hydrophobic residues. Their mutation to polar residues can increase the protein solubility and eventually lead to crystallization (50).

### 3.3.6. Fusion Proteins

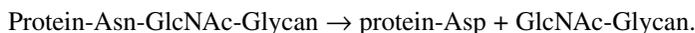
Fusion of a whole protein gene or a smaller tag to the target protein DNA sequence has become a standard tool used to improve the solubility or to facilitate the purification of the protein of interest. However, one cannot predict how the introduction of this new amino acid sequence will affect the crystallization. Concerning small tags like the most commonly used His-tag, literature reports many examples of proteins that have been tagged at their N- or C-terminus and for which crystals suitable for X-ray analysis have been obtained. On the other hand, very few examples of target proteins fused to another protein have been crystallized, one of the main reasons being the conformational heterogeneity induced by the flexible linker connecting the two proteins. Interestingly, a recent strategy consisting of rigidly fusing the protein of interest to MBP via a short three to five amino acid spacer led to the determination of three crystal structures (51,52).

However, crystallization of the target protein usually requires that the fused protein (MBP, GST, Thioredoxin, protein A,  $\beta$ -galactosidase, and others) be removed after purification by a potentially problematic cleavage step. When successful, this method generates a «tagless» pure protein likely to crystallize (*see Chapter 1*).

We mention here the IMPACT™ (intein-mediated purification with an affinity chitin-binding tag) system (New England Biolabs). IMPACT is a protein purification system that uses the inducible self-cleavage activity of a protein splicing element (termed intein) to separate the target protein from the affinity tag. It distinguishes itself from all other purification systems by its ability to purify, in a single chromatographic step, a native recombinant protein without the use of a protease. The cloning, expression, and purification are performed according to the manufacturer's instructions (<http://www.neb.com/nebecomm/products/productE6900.asp>).

#### 4. Notes

1. This quantity refers to the amount of the target and not to the total amount of plasmid DNA.
2. Only one-third of obtained clones will contain an in-frame fusion with GFP, and only a fraction of them will contain gene fragments of the template initial gene.
3. Once the initial pool of closely related sequences has been generated, recombination can also be achieved by using random amplification of PCR DNA fragments (refer to **Subheading 3.2.3.**).
4. Point mutations, as well as insertions or deletions of multiple amino acids, are performed with the QuikChange site-directed mutagenesis kit (Stratagene). This method allows the site-directed mutation in any double-stranded plasmid; it requires neither the presence, nor the engineering of unique restriction sites, nor the use of particular vectors or multiple transformation steps, nor the subcloning in M13-based bacteriophage vectors. Any other mutagenesis method can be used instead.
5. In theory, *N*-glycosidase F is able to release all common classes of *N*-glycans from the protein backbone. In practice, the different carbohydrate chains of a glycoprotein are not equally sensitive to *N*-glycosidase F and the reaction under native conditions can lead to partial deglycosylation. This enzyme is an amidase, which converts asparagines to aspartic acid in the following reaction:



#### References

1. Gerstein, M. (1998) Patterns of protein-fold usage in eight microbial genomes: a comprehensive structural census. *Proteins* **33**, 518–534.
2. Gerstein, M. (1998) How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold Des.* **3**, 497–512.
3. Sulzenbacher, G., Gruez, A., Roig-Zamboni, V., et al. (2002) A medium-throughput crystallization approach. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2109–2115.
4. Ding, H. T., Ren, H., Chen, Q., et al. (2002) Parallel cloning, expression, purification and crystallization of human proteins for structural genomics. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2102–2108.
5. Skinner, M. M. and Terwilliger, T. C. (1996) Potential use of additivity of mutational effects in simplifying protein engineering. *Proc. Natl. Acad. Sci. USA* **93**, 10,753–10,757.
6. Heinz, D. W., Baase, W. A., and Matthews, B. W. (1992) Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. *Proc. Natl. Acad. Sci. USA* **89**, 3751–3755.
7. Dale, G. E., Kostrewa, D., Gsell, B., Stieger, M., and D'Arcy, A. (1999) Crystal engineering: deletion mutagenesis of the 24 kDa fragment of the DNA gyrase B subunit from *Staphylococcus aureus*. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1626–1629.

8. Longhi, S., Nicolas, A., Creveld, L., et al. (1996) Dynamics of *Fusarium solani* cutinase investigated through structural comparison among different crystal forms of its variants. *Proteins* **26**, 442–458.
9. Martin, G., Keller, W., and Doublé, S. (2000) Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J.* **19**, 4193–4203.
10. Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L., and Canard, B. (2002) An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* **21**, 2757–2768.
11. Kawasaki, M. and Inagaki, F. (2001) Random PCR-based screening for soluble domains using green fluorescent protein. *Biochem. Biophys. Res. Commun.* **280**, 842–844.
12. Boone, M. and Upton, C. (2000) BLAST Search Updater: a notification system for new database matches. *Bioinformatics* **16**, 1054–1055.
13. Berman, H. M., Westbrook, J., Feng, Z., et al. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
14. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
15. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
16. Rost, B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Meth. Enzymol.* **266**, 525–539.
17. Gouet, P., Courcelle, E., Stuart, D. I., and Metz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308.
18. Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701–3708.
19. Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A., and Mornon, J. P. (1990) Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences. *Biochimie* **72**, 555–574.
20. Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B., and Longhi, S. (2003) Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J. Biol. Chem.* **278**, 44,567–44,573.
21. Waldo, G. S. (2003) Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7**, 33–38.
22. Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
23. Sambrook, J., Fritsch, E., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd ed.*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
24. Fromant, M., Blanquet, S., and Plateau, P. (1995) Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Anal. Biochem.* **224**, 347–353.

25. Melnikov, A. and Youngman, P. J. (1999) Random mutagenesis by recombination-al capture of PCR products in *Bacillus subtilis* and *Acinetobacter calcoaceticus*. *Nucleic Acids Res.* **27**, 1056–1062.
26. Cherry, J. R., Lamsa, M. H., Schneider, P., et al. (1999) Directed evolution of a fungal peroxidase. *Nat. Biotechnol.* **17**, 379–384.
27. Shafikhani, S., Siegel, R. A., Ferrari, E., and Schellenberger, V. (1997) Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *BioTechniques* **23**, 304–310.
28. You, L. and Arnold, F. H. (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* **9**, 77–83.
29. Grothues, D., Cantor, C. R., and Smith, C. L. (1993) PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nucleic Acids Res.* **21**, 1321–1322.
30. Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691–695.
31. Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391.
32. Cramer, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291.
33. Cramer, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14**, 315–319.
34. Salamone, P. R., Kavakli, I. H., Slattery, C. J., and Okita, T. W. (2002) Directed molecular evolution of ADP-glucose pyrophosphorylase. *Proc. Natl. Acad. Sci. USA* **99**, 1070–1075.
35. Lorimer, I. A. and Pastan, I. (1995) Random recombination of antibody single chain Fv sequences after fragmentation with DNaseI in the presence of Mn<sup>2+</sup>. *Nucleic Acids Res.* **23**, 3067–3068.
36. Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10,747–10,751.
37. Eakin, A. E., McGrath, M. E., McKerrow, J. H., Fletterick, R. J., and Craik, C. S. (1993) Production of crystallizable cruzain, the major cysteine protease from *Trypanosoma cruzi*. *J. Biol. Chem.* **268**, 6115–6118.
38. Kirchhofer, D., Guha, A., Nemerson, Y., et al. (1995) Activation of blood coagulation factor VIIa with cleaved tissue factor extracellular domain and crystallization of the active complex. *Proteins* **22**, 419–425.
39. Leppanen, V. M., Parast, C. V., Wong, K. K., Kozarich, J. W., and Goldman, A. (1999) Purification and crystallization of a proteolytic fragment of *Escherichia coli* pyruvate formate-lyase. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 531–533.
40. Lemieux, M. J., Song, J., Kim, M. J., et al. (2003) Three-dimensional crystallization of the *Escherichia coli* glycerol-3-phosphate transporter: a member of the major facilitator superfamily. *Protein Sci.* **12**, 2748–2756.

41. Gaudier, M., Gaudin, Y., and Knossow, M. (2001) Cleavage of vesicular stomatitis virus matrix protein prevents self- association and leads to crystallization. *Virology* **288**, 308–314.
42. Vallee, F., Lipari, F., Yip, P., Sleno, B., Herscovics, A., and Howell, P. L. (2000) Crystal structure of a class I alpha1,2-mannosidase involved in N-glycan processing and endoplasmic reticulum quality control. *EMBO J.* **19**, 581–588.
43. Nachon, F., Nicolet, Y., Viguie, N., Masson, P., Fontecilla-Camps, J. C., and Lockridge, O. (2002) Engineering of a monomeric and low-glycosylated form of human butyrylcholinesterase: expression, purification, characterization and crystallization. *Eur. J. Biochem.* **269**, 630–637.
44. Josephson, K., McPherson, D. T., and Walter, M. R. (2001) Purification, crystallization and preliminary X-ray diffraction of a complex between IL-10 and soluble IL-10R1. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1908–1911.
45. Zhang, F., Basinski, M. B., Beals, J. M., et al. (1997) Crystal structure of the obese protein leptin-E100. *Nature* **387**, 206–209.
46. Longenecker, K. L., Garrard, S. M., Sheffield, P. J., and Derewenda, Z. S. (2001) Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 679–688.
47. Baud, F. and Karlin, S. (1999) Measures of residue density in protein structures. *Proc. Natl. Acad. Sci. USA* **96**, 12,494–12,499.
48. Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E., and Bell, J. A. (1997) Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* **28**, 494–514.
49. Mateja, A., Devedjiev, Y., Krowarsch, D., et al. (2002) The impact of Glu—>Ala and Glu—>Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1983–1991.
50. Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R., and Craigie, R. (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl. Acad. Sci. USA* **92**, 6057–6061.
51. Liu, Y., Manna, A., Li, R., et al. (2001) Crystal structure of the SarR protein from *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* **98**, 6877–6882.
52. Center, R. J., Kobe, B., Wilson, K. A., et al. (1998) Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. *Protein Sci.* **7**, 1612–1619.



## Production of Selenomethionyl Proteins in Prokaryotic and Eukaryotic Expression Systems

Sylvie Doublé

### Summary

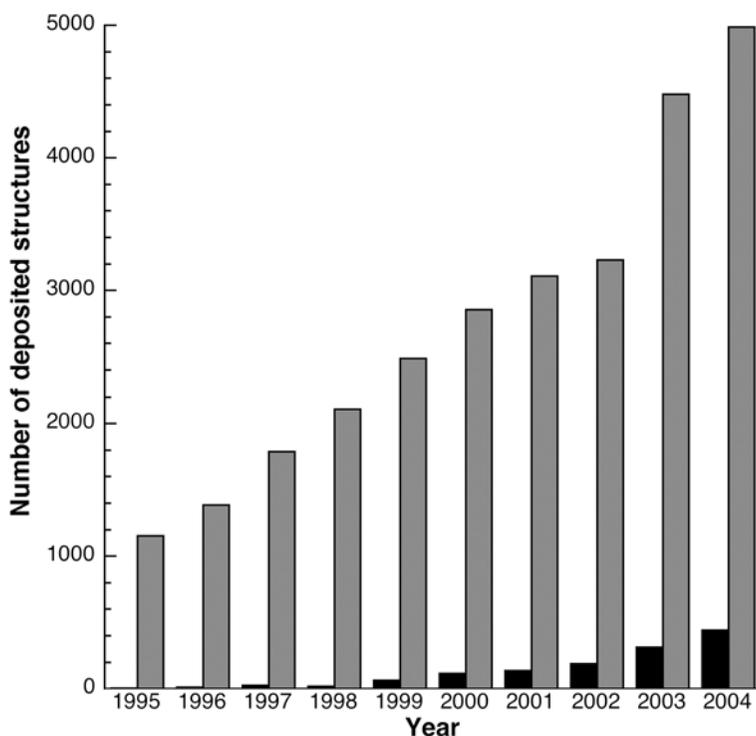
The use of selenomethionine as a phasing tool was first reported in 1990. Engineering of selenomethionyl proteins for structure determination is now routine. In fact, selenium is by far the most commonly used anomalous scatterer for multiwavelength anomalous diffraction studies. The past few years have seen new developments, which demonstrated the feasibility of expressing selenomethionyl protein in eukaryotic systems. In this chapter, the different methods available for producing selenomethionine-labeled proteins in bacteria, as well as in yeast and mammalian cells will be presented, along with tips for purifying and crystallizing selenomethionyl proteins.

**Key Words:** Selenomethionine; selenomethionyl protein expression; selenocysteine; multi-wavelength anomalous diffraction.

### 1. Introduction

The possibility of using selenomethionine as a phasing tool was first reported in 1990 by Hendrickson and collaborators (1), and even they probably did not foresee how prevailing this phasing technique would turn out to be. There are several reasons for this popularity: the substitution is rather straightforward, and now can be accomplished both in *Escherichia coli* and in eukaryotic cells. Substituting selenomethionine for methionine usually has benign effects on the labeled protein, although a change in kinetic properties after selenomethionine labeling can occur (2). In addition, substitution with selenomethionine was reported to enhance the stability of some methionine-rich proteins (3). And even in cases where the selenium-derived phases are not sufficient to solve a structure, they can be used in combination with molecular or isomorphous replacement phases, and at the very least, the selenium positions can help model building by pinpointing the location of methionine residues via the use of anomalous difference maps. The number of structures phased with selenomethionine steadily increased from

From: *Methods in Molecular Biology*, vol. 363: *Macromolecular Crystallography Protocols: Volume 1: Preparation and Crystallization of Macromolecules* Edited by: S. Doublé © Humana Press Inc., Totowa, NJ



**Fig. 1.** Growth of selenomethionine protein structures vs time. The number of structures of selenomethionyl proteins (black bars) deposited in the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) has increased steadily over the past few years. For comparison the total number of deposited structures per year is also shown (gray bars). Selenomethionyl protein structures accounted for almost 9% of the structures deposited in 2004.

fewer than 10 in the mid-1990s to more than 400 some 10 yr later (*see Fig. 1*). In fact, they represent about 80% of all structures solved via multiple anomalous diffraction (*see Note 1*). This surge is undoubtedly correlated with the increase in the number of synchrotron beamlines available for measuring anomalous diffraction data (*4*). It is also due, in part, to recent, exciting developments in the methods for producing selenomethionyl proteins, notably in eukaryotic systems. This chapter will review the expression of selenomethionyl proteins in *E. coli* and in the following eukaryotic systems: yeast, insect, and mammalian cells.

## 2. Materials

### 2.1. Stock Solutions for Bacterial Growth

1. 5X M9 salts: 64 g  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 15 g  $\text{KH}_2\text{PO}_4$ , 2.5 g NaCl, and 5.0 g  $\text{NH}_4\text{Cl}$ ; add deionized water to a final volume of 1 L, sterilize by autoclaving (*5*).

2. 1.0 M MgSO<sub>4</sub> (sterilize by autoclaving).
3. 1.0 M CaCl<sub>2</sub> (sterilize by autoclaving).
4. 20% (w/v) Glucose (filter through 0.2- $\mu$ m filter).
5. 2 mg/mL Biotin (sterile-filtered with 0.2- $\mu$ m filter).
6. 2 mg/mL Thiamine (sterile-filtered with 0.2- $\mu$ m filter).
7. All amino acids except methionine, phenylalanine, tryptophan, and tyrosine: mix amino acids, each at a concentration of 4 mg/mL. Filter-sterilize with a 0.2- $\mu$ m filter and store at 4°C.
8. Phenylalanine, tryptophan, and tyrosine: each at 4 mg/mL. Adjust mixture to pH 8.0 with NaOH. Sterile-filter with a 0.2- $\mu$ m filter. Store at 4°C.
9. 10 mg/mL L-Selenomethionine (Fisher/Acros [Fair Lawn, NJ] or Sigma [St. Louis, MO]) (sterile filtered).
10. DL-Selenocystine (Sigma/Aldrich) (sterile filtered).
11. Luria Broth (LB) medium (5): 10 g tryptone, 5 g yeast extract, 10 g NaCl, and 950 mL of deionized water. Adjust the pH to 7.0 with 5 N NaOH. Adjust the volume to 1 L with deionized water. Sterilize by autoclaving.
12. 0.2- $\mu$ m filters (Fisher; Nalgene, Rochester, NY).

### 2.1.1. Minimal Medium for Growing Methionine Auxotroph Strain

1. *E. coli* cells auxotrophic for methionine: B834(DE3) (Novagen/EMD Biosciences, San Diego, CA) (see Note 2).
2. M9 medium: 1X M9 salts, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 5 g/L (0.5%) glucose (dextrose) (see Note 3).
3. 40 mg/L of all amino acids except methionine.
4. 40–100 mg/L Selenomethionine.
5. 2 mg/L Thiamine.
6. 2 mg/L Biotin.
7. Antibiotic (plasmid dependent).

Make sure that pH of the medium is around 7.4.

### 2.1.2. Defined Medium for Inhibition of Methionine Biosynthesis

1. 1X M9 salts, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 5 g/L (0.5%) glucose (dextrose).
2. 100 mg/L each of lysine, phenylalanine, and threonine.
3. 50 mg/L each of isoleucine, leucine, and valine.
4. 40–100 mg/L Selenomethionine.
5. 2 mg/L Thiamine.
6. 2 mg/L Biotin.
7. Antibiotic (plasmid dependent).

Check that pH is 7.4.

### 2.1.3. Auto-Inducing Medium

#### 2.1.3.1. STOCK SOLUTIONS

1. 20X NPS solution: 66 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 136 g KH<sub>2</sub>PO<sub>4</sub>, and 142 g Na<sub>2</sub>HPO<sub>4</sub>. Add deionized water to 1 L. Sterilize by autoclaving.

2. 50X 5052 solution: 250 g glycerol, 730 mL water, 25 g glucose, and 100 g  $\alpha$ -lactose. Add reagents in this order, stir until everything is dissolved, then add deionized water to a final volume of 1 L. It is advisable to gently heat up the solution in the microwave for a couple of minutes, as  $\alpha$ -lactose dissolves slowly at room temperature (6). Sterilize by autoclaving. Keep the solution for 1 wk before discarding.
3. 50X Amino acid mixture: 10 g each of all amino acids, except cysteine, tyrosine, and methionine: Na glutamate, lysine-HCl, arginine-HCl, histidine-HCl, aspartic acid, alanine, proline, glycine, threonine, serine, glutamine, asparagine, valine, leucine, isoleucine, phenylalanine, and tryptophan in a final volume of 1 L. Filter-sterilize with a 0.2- $\mu$ m filter, wrap in aluminum foil, and store at room temperature until use.
4. 10,000X Trace metals solution: 100-mL aliquot made of 50 mL of 0.1 M FeCl<sub>3</sub>·6H<sub>2</sub>O dissolved in 0.1 M HCl, 2 mL 1 M CaCl<sub>2</sub>, 1 mL 1 M MnCl<sub>2</sub>·4H<sub>2</sub>O, 1 mL 1 M ZnSO<sub>4</sub>·7H<sub>2</sub>O, 1 mL 0.2 M CoCl<sub>2</sub>·6H<sub>2</sub>O, 2 mL 0.1 M CuCl<sub>2</sub>·2H<sub>2</sub>O, 1 mL 0.2 M NiCl<sub>2</sub>·6H<sub>2</sub>O, 2 mL 0.1 M Na<sub>2</sub>MoO<sub>4</sub>·5H<sub>2</sub>O, 2 mL 0.1 M Na<sub>2</sub>SeO<sub>3</sub>·5H<sub>2</sub>O, and 2 mL 0.1 M H<sub>3</sub>BO<sub>3</sub>. All stock solutions should be autoclaved, except for FeCl<sub>3</sub>·6H<sub>2</sub>O. Trace metal solution should be wrapped in aluminum foil and kept at room temperature.
5. 1000X Vitamin: 100-mL aliquot containing 2 mL each of 10 mM nicotinic acid, 10 mM pyridoxine-HCl, 10 mM thiamine-HCl, 10 mM *p*-aminobenzoic acid, 10 mM pantothenate, 5 mL of both 100  $\mu$ L folic acid and 100  $\mu$ M riboflavin, and 80 mL of sterile deionized water. Vitamin solution should be wrapped in aluminum foil and kept at room temperature.
6. 1000X Vitamin with vitamin B<sub>12</sub>. Same as in **item 5**, except add 4 mL of 5 mM vitamin B<sub>12</sub>.
7. 40% (w/v) Glucose solution: 40 g of glucose and sterile deionized water to a final volume of 100 mL. Autoclave or sterile filter with 0.2- $\mu$ m filter.
8. 80% (w/v) Glycerol solution: 80 g of glycerol and sterile deionized water to a final volume of 100 mL. Autoclave.
9. 25 mg/mL Methionine solution: 2.5 g methionine in 100 mL water, filter-sterilize with 0.2- $\mu$ m filter. Wrap in aluminum foil, and store at 4°C.
10. 25 mg/mL Selenomethionine solution: 2.5 g selenomethionine in 100 mL water, filter-sterilize with a 0.2- $\mu$ m filter. Wrap in aluminum foil, and store at 4°C.
11. 5 mM Vitamin B<sub>12</sub>.
12. Agar (Fisher).

#### 2.1.3.2. MEDIA

1. PA-0.5G medium: (for scale-up inoculum, 100 mL final volume); 92 mL water, 100  $\mu$ L 1 M MgSO<sub>4</sub>, 10  $\mu$ L 10,000X trace metal solution, 1.25 mL 40% glucose, 5.0 mL 20X NPS solution, 1.0 mL 50X amino acid mixture, 0.4 mL L-methionine solution, 100  $\mu$ L 1000X vitamin solution including vitamin B<sub>12</sub>, and appropriate antibiotics. Each component should be sterile.
2. PA-0.5G plates: add 10 g/L agar to the solution in **item 1**.

3. 1 L PASM-5052 medium: (for large-scale growth and auto induction): 900 mL water, 1 mL 1 M MgSO<sub>4</sub>, 100 µL 10,000X trace metal solution, 20 mL 50X 5052 solution, 50 mL 20X NPS solution, 20 mL 50X amino acid mixture, 0.4 mL L-methionine solution, 5 mL selenomethionine solution, 1 mL 1000X vitamin solution lacking vitamin B<sub>12</sub>, and appropriate antibiotics. Each component should be sterile (see **Note 4**).

## 2.2. Expression in Eukaryotic Cells

### 2.2.1. Expression in Yeast

1. Yeast strain expressing protein of interest.
2. Standard equipment, glassware, and plasticware for yeast expression.
3. Phosphate-buffered saline (PBS): in 800 mL deionized water dissolve: 8 g of NaCl, 0.2 g KCl, 1.44 g Na<sub>2</sub>HPO<sub>4</sub>, and 0.24 g KH<sub>2</sub>PO<sub>4</sub>. Adjust pH to 7.4 with HCl. Adjust volume to 1 L with additional deionized water. Sterilize by autoclaving.
4. YPD medium: 1% yeast extract, 2% peptone, and 2% dextrose. Add water to 1 L. Autoclave.
5. Synthetic complete medium: 0.09 mg/mL each of adenosine sulfate, uracil, L-tryptophan, L-histidine-HCl, L-arginine-HCl, L-tyrosine, L-leucine, L-isoleucine, L-lysine-HCl; 0.15 mg/mL L-phenylalanine, 0.3 mg/mL L-glutamic acid, 0.3 mg/mL L-aspartic acid, 0.45 mg/mL L-valine, 0.6 mg/mL L-threonine, 1.2 mg/mL L-serine, 0.34 mg/mL thiamine, 0.12 mg/mL L-cysteine, 0.3 mg/mL L-glutamine, 0.3 mg/mL succinic acid, 0.2 mg/mL L-proline, 0.2 mg/mL L-alanine, 0.01 mg/mL inositol, 1.34% (w/v) yeast nitrogen base without amino acids (Invitrogen, Carlsbad, CA), 3% (w/v) dextrose (*Saccharomyces cerevisiae*), and 0.1 mg/mL L-selenomethionine (Fisher/Acros or Sigma).
6. GF/B filter (Whatman, Florham Park, NJ).

### 2.2.2. Insect Cell Expression

1. Recombinant baculovirus-expressing protein of interest.
2. Standard equipment, glassware, and plasticware for insect cell expression.
3. Sf9, Sf21, or High Five™ cells (Invitrogen).
4. Serum-free insect cell culture media: Sf-900 II SFM for Sf9 and Sf21; Express Five® SFM for High Five cells (Invitrogen).
5. Methionine-free Grace's medium (Invitrogen).
6. Heat-inactivated, dialyzed fetal bovine serum (FBS) (Invitrogen or Sigma).
7. L-Selenomethionine (Fisher/Acros or Sigma).

### 2.2.3. Mammalian Cell Expression

1. Standard equipment, glassware, and plasticware for mammalian cell expression.
2. Chinese hamster ovary cell line stably expressing protein of interest.
3. Medium to grow transfected cells (e.g., Dulbecco's modified Eagle's medium/Ham's nutrient mixture F12; DMEM/F12) (JRH Biosciences, Lexena, KS).
4. Hank's buffered salt solution (Fisher).

5. Serum-free medium lacking methionine (JRH Biosciences).
6. L-Selenomethionine (Fisher/Acros or Sigma).

### 2.3. Protein Purification

1. Reducing agents: Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) (Pierce), dithiothreitol (DTT) (Pierce, Rockford, IL).

## 3. Methods

Selenomethionyl proteins for structural studies were first produced in *E. coli* (1). Hendrickson et al. observed that Cowie and Cohen had successfully grown an *E. coli* strain auxotrophic for methionine for 100 generations in a medium containing selenomethionine (7), which led them to establish an expression system for selenomethionyl proteins for structural studies (1). More recently selenomethionine labeling has been performed in yeast, insect, and mammalian cells. Methods for producing selenomethionine-labeled protein in these expression systems are described next.

### 3.1. Expression of Selenomethionyl Protein in *E. coli*

There are several ways to express a selenomethionine-labeled protein in *E. coli*. One can either use a bacterial strain auxotrophic for methionine or inhibit the methionine biosynthesis pathway, in which case any *E. coli* strain can be used (8–10). A recent development involves the use of an auto-inducing medium, which, it turns out, does not require the use of a strain auxotrophic for methionine. All three protocols are described next.

#### 3.1.1. Expression in a Methionine Auxotroph Strain

Expression of selenomethionyl proteins in *E. coli* was first performed in DL41, a methionine auxotroph strain originally designed by LeMaster (1). More recently B834 cells (Novagen) have found routine use in labeling proteins with <sup>35</sup>S-methionine for nuclear magnetic resonance studies or selenomethionine incorporation for crystallographic studies. In a survey of papers published in *Acta Crystallographica* volume D in 2003, B834 (DE3) cells were used in the vast majority of laboratories (see Notes 2 and 5).

Bacterial cells tend to grow more slowly in selenomethionine medium and have a longer lag time. They also reach a final cell density that is lower than that of cells grown in a medium containing methionine (see Note 6).

1. Isolate single colonies by streaking the strain of interest on an LB plate (5) supplemented with the appropriate antibiotics. Incubate at 37°C.
2. Inoculate 5 mL of LB medium supplemented with appropriate antibiotics with one colony and grow overnight at 37°C.
3. Use the 5-mL culture to inoculate 100 mL of defined medium supplemented with selenomethionine: minimal medium M9 with glucose at 5 g/L, with all amino acids

except methionine at a concentration of 40 mg/L, L-selenomethionine at 50–60 mg/L, thiamine and biotin at 2 mg/L (see **Note 7**), and the appropriate antibiotics.

4. Use the 100-mL culture to inoculate a 10-L fermenter or use 10 mL per 2-L baffled flask (see **Note 8**). Medium is the same as in **step 3** and should be prewarmed prior to inoculation.
5. Harvest cells before they reach the stationary phase, centrifuge at 5000g for 20 min, and freeze the cell pellet in liquid nitrogen. Store at  $-80^{\circ}\text{C}$  until purification (see **Note 9**).

### 3.1.2. Inhibition of the Methionine Biosynthesis Pathway

High concentrations of isoleucine, lysine, and threonine block methionine biosynthesis in *E. coli* by inhibiting an aspartokinase, the first enzyme in the methionine biosynthesis pathway. In addition, leucine and phenylalanine act in synergy with lysine. FKBP12 was the first selenomethionyl protein to be expressed by inhibiting methionine biosynthesis (**10**), where a nonauxotrophic prokaryotic strain was grown in a defined medium devoid of methionine but supplemented with selenomethionine and the amino acids known to inhibit methionine biosynthesis (isoleucine, lysine, threonine, leucine, and phenylalanine). This method is now routinely used and has yielded many structures (**8**). In fact, a survey of crystallization papers published in *Acta Crystallographica* in 2003 revealed that the inhibition of methionine biosynthesis was used in about 40% of cases of selenomethionyl protein production. In our laboratory, we favor this method over the use of a strain auxotrophic for methionine, as cells grow faster and to a higher density, giving rise to an increased protein yield (see **Note 10**).

1. Grow cells (1 mL) overnight in rich medium such as LB (**5**).
2. Spin down cells (a few minutes at 1300g in a microcentrifuge) and resuspend in 1 mL of M9 medium with carbon source at 4 g/L and then add to 1 L of the same, prewarmed medium.
3. Grow cells to mid-log phase before adding lysine, phenylalanine, and threonine at 100 mg/L, isoleucine, leucine, and valine at 50 mg/L, and L-selenomethionine at 60 mg/L. Amino acids can be added as powder.
4. Induce 15 min after addition of the amino acids.
5. Harvest cells in mid-to-late-log phase, centrifuge (5000g for 20 min), and flash freeze cell pellets in liquid nitrogen. Store at  $-80^{\circ}\text{C}$  while awaiting purification (see **Note 9**).

### 3.1.3. Auto-Induction Medium

An interesting observation by Bill Studier (Brookhaven National Lab, Upton, NY) has led to the formulation of novel auto-inducing media for the large-scale expression of wild-type and selenomethionyl proteins (**11**). Unintended induction of expression was traced back to small amounts of lactose in some commercially prepared complex growth media. That observation

and the fact that glucose and amino acids were known to inhibit induction by lactose led to the development of auto-inducing media. The yields of proteins are reported to be several-fold higher than those obtained by induction with IPTG (11). The media contain a blend of carbon sources that give rise to high-density cell growth and spontaneously induce protein production in *lac*-based expression systems. These media have been used in several laboratories at Brookhaven National labs and elsewhere, and have been adopted by the Center for Eukaryotic Structural genomics at the University of Wisconsin-Madison (6). When the auto-inducing media are used for the production of selenomethionyl proteins, the incorporation of selenomethionine is typically greater than 90% (6,11). The reported expression and incorporation levels should easily offset the time investment necessary to prepare the stock solutions for the defined media. The auto-induction media are also commercially available (Overnight Express Autoinduction System; EMD Biosciences). The following protocol was adapted from ref. 6.

1. Transform plasmid of interest into B834(DE3) or BL21(DE3) (see Note 11).
2. Plate transformed cells on PA-0.5G agar plate supplemented with appropriate antibiotics. Grow overnight at 37°C.
3. Pick a colony and transfer to a 10-mL test tube containing 3 mL of PA-0.5G medium supplemented with the appropriate antibiotics. Incubate at 37°C with shaking (300 rpm) for about 8 h.
4. Transfer the test tube to 25°C for 30 min, then inoculate a flask containing 100 mL PA-0.5G medium supplemented with the appropriate antibiotics at 25°C.
5. Incubate with shaking for 18 h at 25°C.
6. After 18 h, use 20-mL aliquots to inoculate four 2-L baffled flasks or PET (polyethylene terephthalate) bottles containing 500 mL of PASM-5052 medium supplemented with the appropriate antibiotics (see Note 12).
7. Incubate with shaking for approx 24 h at 25°C.
8. Centrifuge cultures at 5000g for 20 min. Freeze pellets and store at -80°C until purification.

### 3.2. Expression in Eukaryotes

Often eukaryotic proteins will not express well in *E. coli* either because they end up in inclusion bodies, are not properly folded, or lack the required post-translational modifications. When this happens, one commonly turns to eukaryotic expression systems using yeast, insect, or mammalian cells.

#### 3.2.1. Yeast

Incorporation of selenomethionine in proteins expressed in yeast has been performed in both *S. cerevisiae* (12) and *Pichia pastoris* (13,14). Substitution levels are reported to be less than 50% (Table 1), which, however, has sufficed

**Table 1**  
**Expression of Selenomethionyl Proteins in Eukaryotic Cells**

Expression system	Protein expressed	Molecular weight (kD)	Number of Met residues	Incorporation (%)	Reference
<b>Yeast</b>					
<i>Pichia pastoris</i>	$\beta$ -Mannanase	41	10	40	(14)
<i>P. pastoris</i>	Dextranase	60	12	50	(13)
<i>Saccharomyces cerevisiae</i>	RNA Polymerase II	520	119	50	(12)
<b>Insect cells</b>					
Sf9	Glycoprotein D	31	6	85	(51)
Sf9	Human CG	38	4	84	(52)
Sf9	PPT1	31	8	76	(19)
Sf21	TRAF2	22	18	40	(17)
Sf21	$\mu 1_3\sigma 3_3$	350	111	100	(53)
High Five™	H2-M	43	5	95	(54)
High Five	Dipeptidyl peptidase IV	85	14	30–40	(18)
High Five	Ro autoantigen	61	21	$\geq 40$	(48)
<b>Mammalian cells</b>					
CHO	EGF receptor	75	20	N.D. <sup>a</sup>	(20)
CHO	B7-1	22	6	60	(22)
CHO	vCCI	25	5	N.D. <sup>a</sup>	(23)
CHO	Human CG	38	4	92	(24,25)
CHO	Sialoadhesin	13	2	86	(26)
CHO	HER-1	19	2	N.D. <sup>a</sup>	(21)
BHK	Transferrin	80	9	89	(27)

<sup>a</sup>N.D. Not determined.

BHK, baby hamster kidney; CHO, Chinese hamster ovary.

for locating selenium sites and calculating phases for dextranase (15) and the large RNA polymerase II complex (16).

1. Grow yeast in 50–150 mL of YPD to an optical density of 4.5.
2. Centrifuge culture at 3000g for 10 min.
3. Wash cells three times with either 0.9% (w/v) NaCl or PBS.
4. Resuspend cells in 500 mL of synthetic complete medium.
5. For *P. pastoris*: induce by adding 1% methanol initially. After 24 h, 0.25–0.5% methanol is added each 12 h. Let expression proceed for 3–6 d. If protein is secreted

in the medium, harvest supernatant and pass through a GF/B filter (Whatman). Otherwise harvest cells and centrifuge.

6. For *S. cerevisiae*: monitor growth and harvest when it slows down. Centrifuge cells, then freeze in liquid nitrogen. After thawing, lyse the cells by shaking in the presence of glass beads.

### 3.2.2. Insect Cells

Selenomethionyl protein expression has been reported for *Spodoptera frugiperda* Sf9 and Sf21 cells and *Trichoplusia ni* High Five cells. Reported substitution ranges from approx 40 to 100% (see **Table 1**). Even when the substitution was less than 50%, the authors were able to combine phases from the selenomethionyl data with phases from other derivatives, which yielded interpretable electron density maps (**17,18**). The next protocol is based on a review by Bellizzi et al. (**19**).

1. Grow insect cells in methionine containing serum-free medium (e.g., Sf-900 II SFM or Express Five SFM) and infect with baculovirus. Use the same growth conditions as for wild-type protein (cell density, multiplicity of infection, temperature, and percentage air saturation).
2. After 8–36 h, gently spin down (300g) the cells at 20°C (see **Note 13**).
3. Resuspend cells in an equal volume of methionine-free Grace's medium supplemented with 5–10% (v/v) heat-inactivated dialyzed FBS (see **Note 14**).
4. Grow cells for 4–12 h to deplete intracellular pools of methionine (see **Note 15**).
5. Gently spin down the cells at 20°C.
6. Resuspend cells in an equal volume of methionine-free Grace's medium supplemented with 5–10% heat-inactivated dialyzed FBS and 50 mg/L SeMet and grow for an additional 36–48 h.
7. Centrifuge cells (800g) at 4°C for 10 min.
8. Harvest supernatant if protein is secreted, otherwise harvest cells.

### 3.2.3. Mammalian Cells

Selenomethionine-substituted proteins have been expressed in Chinese hamster ovary (**20–26**) and baby hamster kidney cells (**27**). Reported substitution rates in mammalian cells range from 60 to approx 90% (see **Table 1**).

The following protocol is adapted from a report by Hamaoka et al. (**21**).

1. Grow cells in the same medium as that used for native protein expression.
2. Wash cells once with Hank's buffered salt solution.
3. Incubate cells in methionine-free medium supplemented with 50 mg/L L-selenomethionine.
4. After 12–24 h, replace medium with fresh selenomethionine-containing medium.
5. After 3–7 d, harvest supernatant.

### 3.3. Cell-Free Expression

There are a handful of papers that report the expression of selenomethionyl proteins in cell-free expression systems. These could be a viable alternative if the expression systems previously described fail to produce any selenomethionyl protein (28,29).

### 3.4. What to Do When There are Too Few Methionines

A rule of thumb for a successful multiwavelength or single wavelength anomalous diffraction analysis is that the protein should contain one methionine residue per 100 amino acids (30). The average number of methionine residues naturally occurring in proteins is about 1 in 59 (1) so it is likely that a given protein will have enough methionines for structure determination. If this is not the case, leucines or isoleucines can be mutated into methionines. Another option is to use selenocysteine in combination with selenomethionine to doubly label the protein (31). Third, oxidizing the selenomethionyl protein in solution or in its crystalline form was shown to maximize the anomalous signal of selenium (32–34).

#### 3.4.1. Introduction of Additional Methionines

The isomorphous and anomalous contributions of selenium should be calculated (35) in order to find out whether additional methionines need to be introduced (36,37). If this is the case, isoleucine and leucine are choice amino acids for mutation into methionine (36) (see Note 16). The selection of isoleucine and leucine is based on the analysis of amino acid substitutions in related proteins encoded in the Dayhoff mutation probability matrix (38).

#### 3.4.2. Use of Selenomethionine and Selenocysteine in Double Labeling

Strub et al. reported the double labeling of proteins with selenocysteine and selenomethionine in *E. coli* for phasing purposes (31). This method requires the use of a specific strain auxotrophic for cysteine and a minimum medium supplemented with both L-selenomethionine and DL-selenocystine. The authors calculated that 60–80% of protein sequences have at least one methionine per 100 residues, which should be sufficient for a successful multiwavelength anomalous diffraction analysis (30), meaning that 20–40% of proteins are not amenable to phasing via selenomethionine substitution. This percentage goes down to 7–12% when the protein is doubly labeled with selenomethionine and selenocysteine.

#### 3.4.3. Oxidation

In 1998, Smith and Thompson reported that oxidized aqueous selenomethionine displayed an increased magnitude in its absorption edge when compared

with reduced selenomethionine (32). Soon after, other groups reported using selenium oxidation as a means to obtain a better anomalous signal (33,34). In the case of the TolC transporter (33) all traces of  $\beta$ -mercapoethanol were removed before treating the protein with 0.1% hydrogen peroxide for 10 s. The oxidized protein was then placed in a dialysis bag and dialyzed against peroxide-free buffer prior to crystallization. For the threonine synthase, oxidation occurred in the crystals without treatment with peroxide (34). It may seem counterintuitive to oxidize a selenomethionyl protein when most purification protocols emphasize the need to degas all buffers and add a fair amount of reducing agents (see **Subheading 3.5.**). What is important here is to have a homogeneous sample, be it completely reduced or completely oxidized. Oxidizing the protein, either in solution or in crystalline form, is certainly worth trying if the measured anomalous signal is insufficient for phase determination.

### 3.5. Purification and Crystallization

Selenomethionyl proteins are prone to oxidation, which requires that the following steps be taken. Buffers used for purification should be degassed (see **Note 17**). A reducing agent (0.2–2 mM TCEP or 5–20 mM DTT; see **Note 18**) and EDTA (0.2–1 mM) should be added, unless you are opting to fully oxidize the protein in an attempt to maximize the anomalous signal (see **Subheading 3.4.3.**). The purified protein should be used right away in crystallization experiments or stored at  $-80^{\circ}\text{C}$ . Alternatively, it can be frozen and stored in liquid nitrogen as frozen droplets. The level of selenomethionine incorporation can be checked with mass spectrometry (39). Another option is quantitative amino acid analysis, where it is actually the disappearance of methionine that is monitored.

Selenomethionyl-substituted proteins usually crystallize in conditions similar to those used for the wild-type protein. There are a few exceptions where the space group or cell parameters were altered (40–42). Selenomethionyl-substituted proteins tend to be more hydrophobic and somewhat less soluble than their natural counterparts, so one should expect to adjust the crystallization conditions accordingly (see **Note 19**). If crystals of the selenomethionyl protein do not appear *de novo*, it often helps to microseed with crystals of the native protein (see Chapter 7 for seeding protocols). Once crystals appear they should be harvested and flash cooled right away as we have routinely seen in our laboratory that selenomethionyl protein crystals tend to degrade more rapidly than native protein crystals (see Chapter 1 of volume 2 for cryocooling protocols).

Finally, a note of caution: selenomethionine has been shown to be teratogenic in animals. The precautions one usually takes while handling heavy metals apply: wear nitrile gloves and protective clothing (see **Note 20**).

#### 4. Notes

1. Data for 2004 were obtained from Protein Data Bank ([www.pdb.org](http://www.pdb.org)).
2. B834 (pLysS) is also available. pLysS cells carry a plasmid that encodes T7 lysozyme, a natural inhibitor of T7 RNA polymerase. This strain would be the one to use if there is a need to suppress basal expression of the protein of interest prior to induction.
3. Other groups report using 2X M9 (*see* medium composition at <http://alf1.mrc-lmb.cam.ac.uk/~ramak/madms/segrowth.html>).
4. MgSO<sub>4</sub> and the trace metal solution should be mixed with water before addition to 20X NPS to avoid precipitate formation.
5. B834 is a methionine auxotroph of *E. coli* B and BL21 is a Met<sup>+</sup> derivative of B834 obtained via P1 transduction (**11**). Both lack the lon and ompT proteases.
6. When trying to express a selenomethionyl protein under the control of the temperature-sensitive  $\lambda$  repressor it may be necessary to increase the induction temperature by a few degrees Celsius (44°C) (**43**).
7. Optionally, adenine, guanosine, thymine, and uracil can be added at a concentration of 0.5 g/L.
8. The final dilution of the initial methionine from the rich medium should be as high as possible because methionine is incorporated preferentially over selenomethionine. Ultimately one needs to find a compromise between optimal cell growth and maximal selenomethionine incorporation.
9. Cell lysis occurs shortly after late log phase when bacteria are grown in a selenomethionine-containing medium. Cells should not be allowed to reach stationary phase before inoculating a larger culture. Similarly, care must be taken to harvest cells during mid- or late log phase.
10. This method is potentially applicable to other prokaryotic strains. For example, incorporation of selenomethionine in *Pseudomonas fluorescens* has been reported (**44**).
11. It was found that the combination of 10  $\mu\text{g}/\text{mL}$  methionine and 125  $\mu\text{g}/\text{mL}$  SeMet in PASM-5052 medium can repress the endogenous synthesis of methionine (**11**). Using an *E. coli* strain auxotrophic for methionine, therefore, is not required and strains such as BL21(DE3) cells can be used.
12. Two-liter disposable PET beverage bottles can be used instead of the glass baffled flasks (**6,45,46**). Structural genomics groups have adopted these plastic bottles because they are inexpensive, can fit in standard shaker baskets, and do not need to be sterilized, provided that the appropriate antibiotics are added to the medium. Bottles are autoclaved and disposed of after use, which eliminates the risk of cross-contamination. The PET bottles are available through Continental Glass and Plastic or Ball Corporation (both in Chicago, IL).
13. One should pay attention to the time allowed for cell growth in a methionine-containing medium postinfection if the protein is expressed intracellularly. The synthesis of recombinant proteins in insect cells was shown to start between 12 and 24 h after infection (**47**). If the cells stay longer than 12 h in a Met<sup>+</sup> medium, one runs the risk of obtaining a mixture of Met and SeMet proteins. This is not an issue if

the protein is secreted because the protocol involves changing the medium, which would get rid of the nonlabeled recombinant protein.

14. Dialyzed FBS is commercially available. It is obtained via dialysis by tangential flow filtration using a 10,000 molecular-weight cutoff membrane. Dialysis substantially decreases the concentration of low molecular weight molecules in the serum, such as methionine. The heat inactivation is done at 56°C for 30 min.
15. The number of hours the cells stay in a medium deficient in methionine will affect the final selenomethionine substitution level. For example, it was reported that a lower selenomethionine substitution was seen when cells were in a methionine-free medium for 7 h instead of 12 (48). The presence of yeastolate in the medium can also affect the substitution.
16. Gassner and Matthews propose that the optimal amino acid to be mutated into methionine is leucine, followed by phenylalanine, isoleucine, and valine (49). The ranking, which is based on the analysis of isomorphous differences among T4 lysozyme protein variants, differs from that proposed by Leahy et al.: leucine, isoleucine, and to a lesser extent, valine (36). But it is reassuring that the two methods agree that leucine should be the amino acid of choice.
17. The use of argon during purification and crystallization can prevent oxidation and it was shown to be critical for obtaining diffracting crystals of mammalian CoA transferase (50).
18. TCEP-HCl is a good alternative to DTT or  $\beta$ -mercaptoethanol as a reducing agent in chromatographic buffers and crystallization experiments. It is more stable than DTT and  $\beta$ -mercaptoethanol, it is non-volatile, and it retains its reducing power at acidic and basic pH (greater than pH 7.5). TCEP-HCl should be used in molar excess over the protein sample, i.e., if the protein concentration is 0.1 mM, then the TCEP concentration should be 0.2–0.3 mM.
19. Usually one should reduce supersaturation, i.e., decrease the protein or precipitant concentration.
20. Latex gloves *should never be used* to handle toxic or hazardous material, as solutions can easily pass through latex. Nitrile gloves should be used instead. Nitrile also has the advantage that it is not allergenic. A list of recommended glove types can be seen at <http://www.ehs.washington.edu/ohs/updatestipsgloves.shtm>.

## Acknowledgments

This work is supported by grants from the National Institute of Health (GM62239) and the Human Frontier Science Program Organization (RG P22/2003). The author would like to thank Anne Mason (University of Vermont) for help with **Table 1**.

## References

1. Hendrickson, W. A., Horton, J. R. and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.

2. Bernard, A. R., Wells, T. N., Cleasby, A., Borlat, F., Payton, M. A., and Proudfoot, A. E. (1995). Selenomethionine labelling of phosphomannose isomerase changes its kinetic properties. *Eur. J. Biochem.* **230**, 111–118.
3. Gassner, N. C., Baase, W. A., Hausrath, A. C., and Matthews, B. W. (1999) Substitution with selenomethionine can enhance the stability of methionine-rich proteins. *J. Mol. Biol.* **294**, 17–20.
4. Jiang, J. and Sweet, R. M. (2004) Protein Data Bank depositions from synchrotron sources. *J. Synchrotron. Radiat.* **11**, 319–327.
5. Sambrook, J. and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
6. Sreenath, H. K., Bingman, C. A., Buchan, B. W., et al. (2005) Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr. Purif.* **40**, 256–267.
7. Cowie, D. B. and Cohen, G. N. (1957) Biosynthesis by *E. coli* of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* **26**, 252–261.
8. Doublé, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530.
9. Berne, P. F., Doublé, S., and Carter, C. W., Jr. (1999) Molecular biology for structural biology. In: *Crystallization of Nucleic Acids and Proteins, 2nd ed.*, (Ducruix, A. and Giegé, R., eds.), Oxford University Press, Oxford, UK.
10. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and Rapamycin. *J. Mol. Biol.* **229**, 105–124.
11. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234.
12. Bushnell, D. A., Cramer, P., and Kornberg, R. D. (2001) Selenomethionine incorporation in *Saccharomyces cerevisiae* RNA polymerase II. *Structure (Camb)* **9**, R11–R14.
13. Larsson, A. M., Stahlberg, J., and Jones, T. A. (2002) Preparation and crystallization of selenomethionyl dextranase from *Penicillium minioluteum* expressed in *Pichia pastoris*. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 346–348.
14. Xu, B., Munoz, I. I., Janson, J. C., and Stahlberg, J. (2002) Crystallization and X-ray analysis of native and selenomethionyl beta-mannanase Man5A from blue mussel, *Mytilus edulis*, expressed in *Pichia pastoris*. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 542–545.
15. Larsson, A. M., Andersson, R., Stahlberg, J., Kenne, L., and Jones, T. A. (2003) Dextranase from *Penicillium minioluteum*: reaction course, crystal structure, and product complex. *Structure (Camb)* **11**, 1111–1121.
16. Cramer, P., Bushnell, D. A., Fu, J., et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649.
17. McWhirter, S. M., Pullen, S. S., Holton, J. M., Crute, J. J., Kehry, M. R., and Alber, T. (1999). Crystallographic analysis of CD40 recognition and signaling by human TRAF2. *Proc. Natl. Acad. Sci. USA* **96**, 8408–8413.

18. Aertgeerts, K., Ye, S., Tennant, M. G., et al. (2004) Crystal structure of human dipeptidyl peptidase IV in complex with a decapeptide reveals details on substrate specificity and tetrahedral intermediate formation. *Protein Sci*, **13**, 412–421.
19. Bellizzi, J. J., Widom, J., Kemp, C. W., and Clardy, J. (1999) Producing selenomethionine-labeled proteins with a baculovirus expression vector system. *Structure Fold. Des.* **7**, R263–R267.
20. Ogiso, H., Ishitani, R., Nureki, O., et al. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787.
21. Hamaoka, B. Y., Dann, C. E., 3rd, Geisbrecht, B. V., and Leahy, D. J. (2004) Crystal structure of *Caenorhabditis elegans* HER-1 and characterization of the interaction between HER-1 and TRA-2A. *Proc. Natl. Acad. Sci. USA* **101**, 11,673–11,678.
22. Davis, S. J., Ikemizu, S., Collins, A. V., et al. (2001) Crystallization and functional analysis of a soluble deglycosylated form of the human costimulatory molecule B7-1. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 605–608.
23. Carfi, A., Smith, C. A., Smolak, P. J., McGrew, J., and Wiley, D. C. (1999) Structure of a soluble secreted chemokine inhibitor vCCI (p35) from cowpox virus. *Proc. Natl. Acad. Sci. USA* **96**, 12,379–12,383.
24. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994) Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein. *Structure* **2**, 545–558.
25. Lustbader, J. W., Wu, H., Birken, S., et al. (1995) The expression, characterization, and crystallization of wild-type and selenomethionyl human chorionic gonadotropin. *Endocrinology* **136**, 640–650.
26. May, A. P., Robinson, R. C., Aplin, R. T., Bradfield, P., Crocker, P. R., and Jones, E. Y. (1997) Expression, crystallization, and preliminary X-ray analysis of a sialic acid-binding fragment of sialoadhesin in the presence and absence of ligand. *Protein Sci.* **6**, 717–721.
27. Wally, J., Halbrook, P. J., Vornrhein, C., et al. (2006) The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. *J. Biol. Chem.*, in press.
28. Kigawa, T., Yamaguchi-Nunokawa, E., Kodama, K., et al. (2002) Selenomethionine incorporation into a protein by cell-free synthesis. *J. Struct. Funct. Genomics* **2**, 29–35.
29. Yokoyama, S. (2003) Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**, 39–43.
30. Hendrickson, W. A. and Ogata, C. M. (1997) Phase determination from multi-wavelength anomalous diffraction measurements. *Methods Enzymol.* **276**, 494–523.
31. Strub, M. P., Hoh, F., Sanchez, J. F., et al. (2003) Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure (Camb)* **11**, 1359–1367.
32. Smith, J. L. and Thompson, A. (1998) Reactivity of selenomethionine: dents in the magic bullet? *Structure* **6**, 815–819.

33. Sharff, A. J., Koronakis, E., Luisi, B., and Koronakis, V. (2000) Oxidation of selenomethionine: some MADness in the method! *Acta Crystallogr. D Biol. Crystallogr.* **56**, 785–788.
34. Thomazeau, K., Curien, G., Thompson, A., Dumas, R., and Biou, V. (2001) MAD on threonine synthase: the phasing power of oxidized selenomethionine. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1337–1340.
35. Boggon, T. J. and Shapiro, L. (2000) Screening for phasing atoms in protein crystallography. *Structure Fold. Des.* **8**, R143–R149.
36. Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P., and Hendrickson, W. A. (1994) Crystallization of a fragment of human fibronectin: introduction of methionine by site-directed mutagenesis to allow phasing via selenomethionine. *Proteins* **19**, 48–54.
37. Skinner, M. M., Zhang, H., Leschnitzer, D. H., et al. (1994) Structure of the gene V protein of bacteriophage f1 determined by multiwavelength X-ray diffraction on the selenomethionyl protein. *Proc. Natl. Acad. Sci. USA* **91**, 2071–2075.
38. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.
39. Cohen, S. L. and Chait, B. T. (2001) Mass spectrometry as a tool for protein crystallography. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 67–85.
40. Hagemeyer, C. H., Shima, S., Warkentin, E., Thauer, R. K., and Ermler, U. (2003) Coenzyme F420-dependent methylenetetrahydromethanopterin dehydrogenase from *Methanopyrus kandleri*: the selenomethionine-labelled and non-labelled enzyme crystallized in two different forms. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1653–1655.
41. McAuley, K. E., Cummins, I., Papiz, M., Edwards, R., and Fordham-Skelton, A. P. (2003). Purification, crystallization and preliminary X-ray diffraction analysis of S-formylglutathione hydrolase from *Arabidopsis thaliana*: effects of pressure and selenomethionine substitution on space-group changes. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2272–2274.
42. Poulsen, J. C., Harris, P., Jensen, K. F., and Larsen, S. (2001) Selenomethionine substitution of orotidine-5'-monophosphate decarboxylase causes a change in crystal contacts and space group. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1251–1259.
43. Witkowski, A. and Smith, S. (1998) Successful expression of a selenomethionyl protein under control of the temperature-sensitive lambda repressor requires higher than normal temperature. *Biotechniques* **24**, 934–936.
44. Dong, C., Kotzsch, A., Dorward, M., van Pee, K. H., and Naismith, J. H. (2004) Crystallization and X-ray diffraction of a halogenating enzyme, tryptophan 7-halogenase, from *Pseudomonas fluorescens*. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1438–1440.
45. Millard, C. S., Stols, L., Quartey, P., Kim, Y., Dementieva, I., and Donnelly, M. I. (2003) A less laborious approach to the high-throughput production of recombinant proteins in *Escherichia coli* using 2-liter plastic bottles. *Protein Expr. Purif.* **29**, 311–320.

46. Stols, L., Millard, C. S., Dementieva, I., and Donnelly, M. I. (2004) Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. *J. Struct. Funct. Genomics* **5**, 95–102.
47. O'Reilly, D. R. and Miller, L. K. (1988) Expression and complex formation of simian virus 40 large T antigen and mouse p53 in insect cells. *J. Virol.* **62**, 3109–3119.
48. Stein, A. J., Fuchs, G., Fu, C., Wolin, S. L., and Reinisch, K. M. (2005) Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. *Cell* **121**, 529–539.
49. Gassner, N. C. and Matthews, B. W. (1999) Use of differentially substituted selenomethionine proteins in X-ray structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1967–1970.
50. Bateman, K. S., Brownie, E. R., Wolodko, W. T., and Fraser, M. E. (2002) Structure of the mammalian CoA transferase from pig heart. *Biochemistry* **41**, 14,455–14,462.
51. Carfi, A., Gong, H., Lou, H., et al. (2002) Crystallization and preliminary diffraction studies of the ectodomain of the envelope glycoprotein D from herpes simplex virus 1 alone and in complex with the ectodomain of the human receptor HveA. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 836–838.
52. Chen, W. Y. and Bahl, O. P. (1991) Selenomethionyl analog of recombinant human choriogonadotropin. *J. Biol. Chem.* **266**, 9355–9358.
53. Liemann, S., Chandran, K., Baker, T. S., Nibert, M. L., and Harrison, S. C. (2002) Structure of the reovirus membrane-penetration protein, Mu1, in a complex with its protector protein, Sigma3. *Cell* **108**, 283–295.
54. Fremont, D. H., Crawford, F., Marrack, P., Hendrickson, W. A., and Kappler, J. (1998) Crystal structure of mouse H2-M. *Immunity* **9**, 385–393.

## How to Use Dynamic Light Scattering to Improve the Likelihood of Growing Macromolecular Crystals

Gloria E. O. Borgstahl

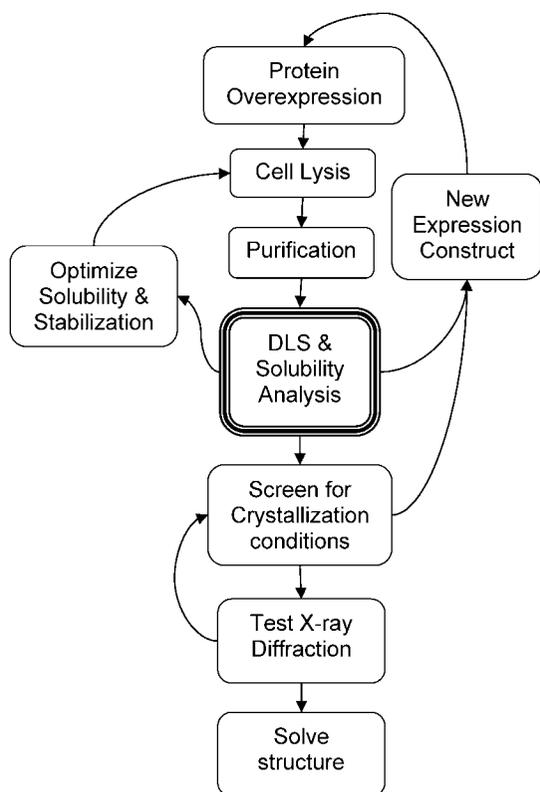
### Summary

Dynamic light scattering (DLS) has become one of the most useful diagnostic tools for crystallization. The main purpose of using DLS in crystal screening is to help the investigator understand the size distribution, stability, and aggregation state of macromolecules in solution. It can also be used to understand how experimental variables influence aggregation. With commercially available instruments, DLS is easy to perform, and most of the sample is recoverable. Most usefully, the homogeneity or monodispersity of a sample, as measured by DLS, can be predictive of crystallizability.

**Key Words:** Solubility; aggregation; hydrodynamic radius; monodispersity; polydispersity; dynamic light scattering; DLS.

### 1. Introduction

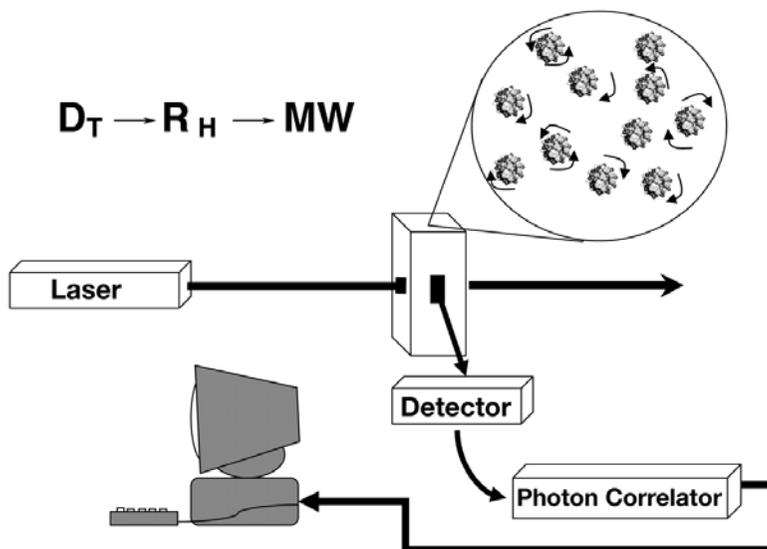
Along the road to atomic structure determination by X-ray crystallography, a major challenge is the growth of high-quality crystals of the macromolecule of interest. The level of interest in the macromolecule seems to correlate well with the level of difficulty in its crystallization. It is helpful to optimize the composition of the protein sample and the experimental conditions, such as buffer components and temperature, to increase the likelihood of crystallization. To optimize these parameters it is necessary to analyze the aggregation state and stability of a macromolecular sample. There are many methods that can be used to measure size or aggregation state, including sedimentation equilibrium, size exclusion chromatography, native gel electrophoresis, and light scattering. Of these methods, dynamic light scattering (DLS), otherwise known as quasielastic light scattering, is the easiest to implement, the quickest to perform, and the least destructive to the sample (**Fig. 1**). For complex cases a combination of these



**Fig. 1.** Flow diagram of methods involved in growing protein crystals. By placing dynamic light scattering and solubility analysis in the center of the process the chances of growing crystals are optimized

methods may be needed to interpret the data. Samples that are monodisperse in solution, as measured by DLS, are much more likely to crystallize (1,2). Along with this chapter, there are several other useful chapters written on the use of DLS analysis in crystallization (3–5).

The DLS instrument is easy to use and detailed knowledge of the underlying physics of molecular sizing is usually not needed. Therefore, only a brief explanation is included here. A microcuvet of protein solution is illuminated by laser light (Fig. 2). The molecules in solution are undergoing Brownian motion and cause fluctuations in the scattered light intensity. This change in light intensity is measured by a detector placed at a 90° angle to the incident laser light. The translational diffusion coefficient  $D_T$  is derived from these data using an auto-correlation function. In general, small particles diffuse “faster” than large particles. A hydrodynamic radius ( $R_H$ ) of the molecules in solution can be calculated from  $D_T$ . In general, particles must differ in  $R_H$  by 50% or more to be well

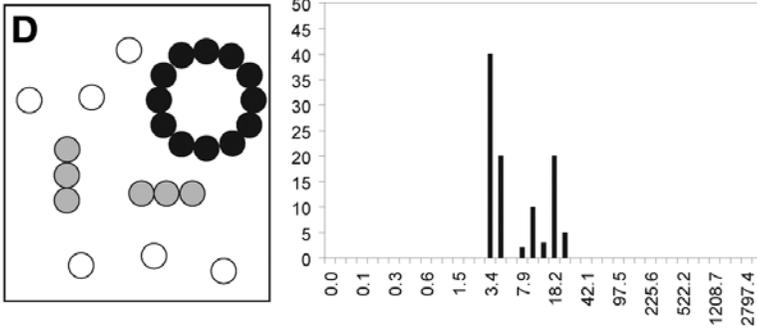
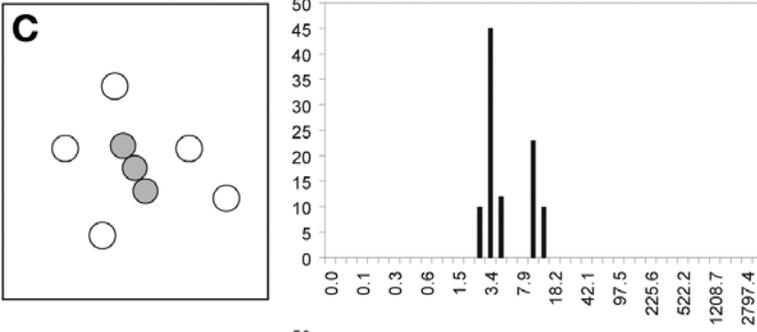
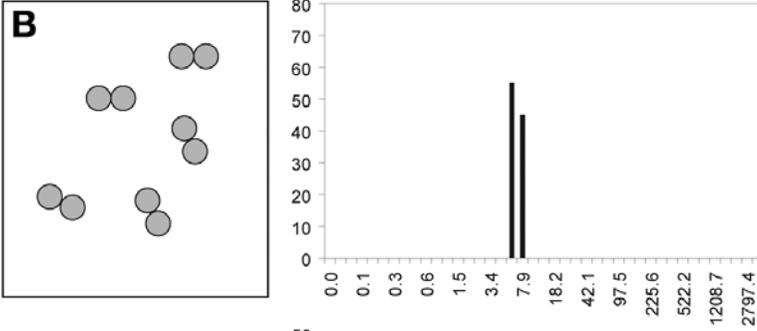
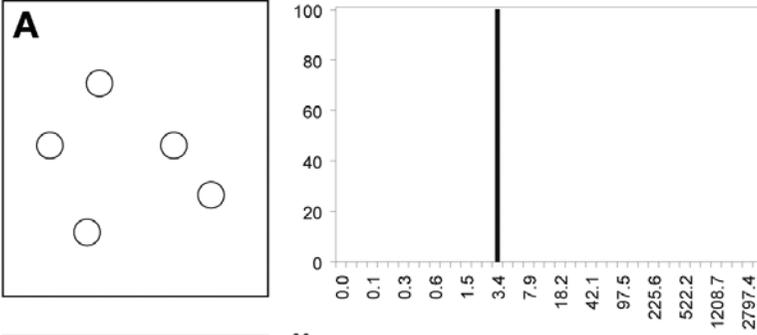


**Fig. 2.** Simplified schematic of the dynamic light scattering experiment.

separated by DLS (Fig. 3). Molecular weight (MW) can also be estimated, if the shape of the molecule is assumed, e.g., spherical or elongated. DLS as an estimator of MW is not recommended and must be used with caution. The shape definitions used to estimate MW may not accurately represent the particles in solution, and for polydisperse samples the  $R_H$  and MW will be based on a weighted average of more than one species. Multiple-angle static light scattering downstream from size-exclusion chromatography is the recommended light-scattering method to measure absolute MW of molecules in solution. In this chapter, DLS is primarily used to assess the aggregation state of a sample and to measure polydispersity, which is predictive of crystallizability.

## 2. Materials

1. Protein Solutions DynaPro MS/X instrument with temperature control (0–60°C) installed and correctly configured.
2. Dynamics software v6 installed on a compatible PC.
3. Protein Solutions 12- $\mu$ L quartz cuvetts.
4. Protein Solutions microfiltration system.
5. Syringe tip (0.2- $\mu$ m filters).
6. 20–30- $\mu$ L Protein sample.
7. Water, 1% Triton X-100, and a range of appropriate buffer and salt solutions, all 0.2- $\mu$ m filtered.
8. Compressed air, either house air or in a can.
9. Lens paper.
10. Small ultrasonic cleaner for cleaning cuvetts.



---

**Fig. 3.** Examples of solutions that differ in composition (left) and fake dynamic light scattering regularization histograms (right). The relative amount of light scattered by each bin, percentage of intensity, is plotted against the discrete particle sizes,  $R_H$ , in nanometers on a log scale. **(A)** A monodisperse, monomodal solution of monomers (with mean  $R_H$  of 3.4 nm, %Pd of 10%); very likely to crystallize. **(B)** A monodisperse, monomodal solution of dimers (with mean  $R_H$  of 6.8 nm, %Pd of 14%). The polydispersity is greater than the monomeric solution but it is still very likely to crystallize. **(C)** A bimodal solution of monomers contaminated by trimers (with mean  $R_H$  of 3.4 and 10.2 nm). Less likely to crystallize so put in less screening effort. **(D)** A multimodal solution of monomer, trimer, and dodecameric aggregates (with mean  $R_H$  of 3.4, 10.2, and 19.3 nm). Unlikely to crystallize but you might as well give it a little try.

Note: the mean  $R_H$  is defined by the weighted average of the number of bins comprising the peak. The polydispersity of each peak is indicated by the width. Species that differ in  $R_H$  by more than 50% are separable if their polydispersity is small. For example, the dimers and trimers differ in size by only 33%. Thus, if solution B was mixed with solution C there would be a single peak (monomodal) and it would be very broad ranging from 2.6 to 13.8 nm. This mixture would be monomodal with mean  $R_H$  of 5.9 nm, very polydisperse with %Pd of around 50%, and highly unlikely to crystallize.

11. Sterile plastic transfer pipets.
12. 100- $\mu$ L Pipetman with capillary pipet tips.
13. Table-top microcentrifuge and appropriate tubes.
14. Microconcentrators.
15. Good buffers (sodium salts) for use in Mueser's solubility screen, each at 100 mM: MES-NaOH, pH 5.8; PIPES-HCl, pH 6.5; HEPES-HCl, pH 7.5; and TAPS-HCl, pH 8.5.
16. Chloride salts for use in Mueser's solubility screen, each at 100 mM:  $\text{NH}_4\text{Cl}$ ; NaCl; KCl; LiCl;  $\text{MgCl}_2$ ; and  $\text{CaCl}_2$ .
17. Ammonium salts for use in Mueser's solubility screen, each at 100 mM:  $\text{NH}_4$  formate;  $\text{NH}_4$  acetate;  $\text{NH}_4$  cacodylate;  $(\text{NH}_4)_2 \text{SO}_4$ ;  $(\text{NH}_4)_3 \text{PO}_4$ ; and  $\text{NH}_4$  citrate.

### 3. Methods

The methods described next outline (1) how to prepare the sample, (2) how to take DLS measurements, and (3) how to interpret the results. The DynaPro instrument is run by the Dynamics software package (*see Note 1*). The methods and strategies described in this chapter can be applied to any DLS instrument, although the specific details of the instrument and software may be different.

#### 3.1. Preparation of Instrument

##### 3.1.1. Measuring DLS Data

While preparing for the experiment, many DLS measurements should be taken to check the state of the instrument. First, turn on the DynaPro MS/X

instrument, the temperature control unit, and start the Dynamics software. Place the quartz cuvet containing your sample or water in the sample holder on the optics block. The frosted side of the cuvet must point to the left side of the holder, as marked on the instrument. In the software, open a “New” experiment. On the left side of the Experiment window is the “Tree View.” The Tree View is used to select groups or categories of information for viewing in the display side of the Experiment window. There are three main nodes in the Tree View: Hardware, Parameters, and Measurements. Connect the instrument using the Hardware node. Set the temperature using the Parameters → Instrument node. Then proceed to take DLS measurements by clicking on the green “Start” button on the Experiment window tool bar. Subcategories in the Measurements node are the individual measurements (Meas no.), each of which is further broken down into Acquisitions (Acq no.) and Readings (Read no.). The display format for the information in the Measurement node is dependent on which view button is selected in the Experiment window tool bar (e.g., Datalog grid or Regularization graph). By default, each data acquisition is accumulated over a 10-s window of time. During the course of an experiment, each acquisition collected is displayed in the Measurements node along with the corresponding calculated data. Ten or more acquisitions are recommended per measurement. Each acquisition is the average of 10 readings. The Cumulants, displayed in the Datalog grid, and Regularization analysis data is calculated by the software and can be displayed at any level of detail, including the average over all the data, the average over each acquisition, or over each reading. To stop data collection, click on the red “Stop” button in the Experiment window tool bar.

### 3.1.2. Cuvet Cleaning and Clean Water Count

The DLS cuvet and apparatus must be very clean to ensure good data quality. Therefore, it is highly recommended that the first step of the DLS experiment is to measure a good clean water count. The clean water count also checks the condition of the instrument components.

1. Fill the cuvet with 0.2- $\mu\text{m}$  filtered deionized water. Avoid air bubbles.
2. Take DLS measurements to get a clean water count rate.
3. If the clean water count is reasonable and steady, you can start taking DLS measurements on your sample. For the DynaPro MS/X instrument counts less than 25,000 are good. Less sensitive instruments, such as the DynaPro 801, may have counts less than 10,000 for clean water. See **Example 1** for an example of acceptable clean water data. Reasonable clean water count rates are similar to those made when the cuvet and instrument were new. The clean water count rates should be noted in the instrument log book. If it is too high the cuvet and/or the microfilter kit need to be thoroughly cleaned or, in the worst case, should be replaced.

Remember, any large particles, such as dust, will scatter light intensely and will interfere with the DLS signal from the molecules of interest. Therefore, it

### Example 1 Acceptable Clean Water Count Data

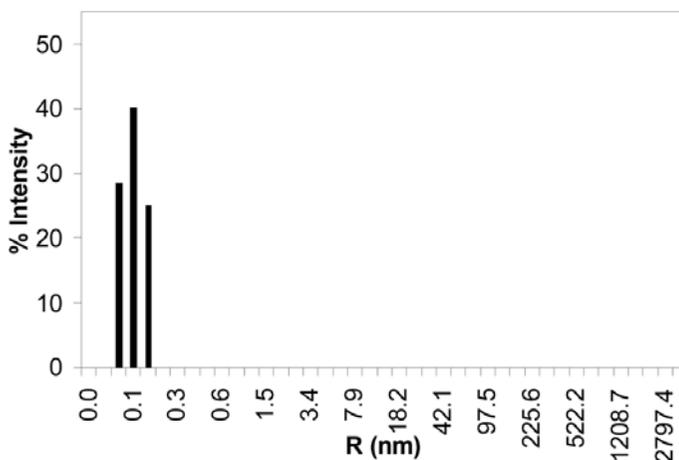
Data log grid

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	9.5	25	16340	0.3	0.1	0	0.026	1.002	165.138
Acq 2	19.5	25	18,240	0.1	0.1	0	0.174	1.002	117.940
Acq 3	29.5	25	18,195	0.1	0.1	0	0.159	1.001	100.560
Acq 4	39.5	25	18,559	0.2	0.1	0	0.100	1.001	71.775
Acq 5	49.5	25	19,285	0.6	0.1	1	0.013	1.001	118.706
Acq 6	59.5	25	17,181	0.5	0.2	1	0.025	1.001	150.398
Acq 7	69.5	25	17,748	0.1	0.1	0	0.297	1.002	149.658
Acq 8	79.6	25	19,150	0.1	0.1	0	0.232	1.000	99.076
Acq 9	89.6	25	16,615	0.5	0.1	1	0.057	1.000	124.772
Acq 10	99.7	25	15,131	0.1	0.1	0	0.762	1.000	138.819

Mean

Meas 1	99.7	25	17,644.4	0.26	0.1	0	0.185	1.001	123.684
--------	------	----	----------	------	-----	---	-------	-------	---------

Regularization graph



Regularization results summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	0.1	21.2	0	100.0	100.0

Interpretation: this is excellent water count data for the DynaPro MS/X. The regularization data at  $R_H < 0.5$  nm is owing to noise in the detector. Some people call it the solvent peak, but it is really because of after pulse noise in the photodiode. This noise surfaces from time to time in the data if the protein concentrations are too low. In this case, it can be ignored or filtered out of the regularization fit by clicking on the check mark for that peak in the Regularization Results Summary. Note that the SOS errors are high with water count data because there are no macromolecules present.

is necessary to thoroughly clean the cuvet before and after use. Be careful not to scratch the cuvettes. Clean all dust off the outer surface. The following cuvet-cleaning procedure is recommended (*see Note 2*):

1. Using a sterile plastic transfer pipet, flush the cuvet multiple times with a 1% Triton X-100 solution.
2. Rinse the cuvet three to five times with sterile-filtered, deionized water.
3. The interior of the cuvet can be dried using compressed air. Alternatively, if you have more time, invert the cuvet and allow it to dry.
4. Polish the exterior surface with lens paper and remove dust with compressed air. Note: tissues and other wipes should not be used as they can scratch the surface of the cuvet. The cuvet cap must also be dust free.
5. Repeat the clean water count and cleaning procedure until a reasonably low clean water count is obtained. The clean water count must be stable for 2–3 min.

## **3.2. Preparation of Sample**

### *3.2.1. Estimation of Minimal Concentration*

In order to make efficient use of your samples, it is helpful to know the minimum protein concentration needed for DLS measurements. Usually, it is safe to assume that DLS measurements made at low protein concentration represent fairly well the sample at the much higher concentrations typically used by crystal farmers. Also, if needed, DLS measurements can be repeated at higher concentrations. The minimal protein concentration needed is dependent on the MW of the macromolecule and the particular instrument. Smaller proteins will need to be at a higher concentration. For example, for the DynaPro MS/X the minimal concentration for a 10-kDa protein is approx 0.6 mg/mL and for a 100-kDa protein is 0.06 mg/mL. To estimate the minimum protein concentration that is needed for any DynaPro model use Tools→Calculations→Optimization in Dynamics v6. To obtain higher quality DLS data use a protein concentration two- to threefold higher than that recommended by the Calculator.

### *3.2.2. Sample Preparation*

Before DLS measurements can be taken the sample must also be cleaned of any dust or other particles. All buffers must be also cleaned by filtration. This can be done by using the microfilter kit provided by Protein Solutions or by centrifugation. Centrifugation is the easiest. The microfilter kit is more difficult to use but has the advantage of removing particles by pore size (*see Note 3*). The procedure used to clean the sample for DLS must also be used to prepare the sample before crystallization.

1. Centrifugation procedure:
  - a. Prepare the sample (e.g., thaw frozen stock, dilution of stock, or mixed components of a complex) in a suitable clean, dust-free microcentrifuge tube. The microcentrifuge tubes can be purged of dust with compressed air.

- b. Centrifuge 5–10 min at 15,000g in a table-top microcentrifuge.
  - c. Transfer supernatant with a 100- $\mu$ L pipetman with clean capillary tips to a dust-free, clean microcentrifuge tube or pipet the supernatant directly into the clean DLS cuvet. Do not disturb the pellet. Remember, only the top portion of the sample is dust free after centrifugation.
2. Microfilter kit procedure:
- a. Disassemble the microfilter system and syringe completely. Thoroughly clean all parts by rinsing/soaking in deionized water and then air-dry the parts. If needed, ultrasonication or 1% Triton X-100 can be used to clean the parts followed by thorough rinsing with deionized water.
  - b. Partially reassemble microfilter system by fitting the Teflon housing into the metal housing and seat the O-rings properly in each half.
  - c. Using the tweezers, place a filter disk into the “needle” half of the metal housing on top of the O-ring (*see Note 4*).
  - d. Tightly screw the two metal housing pieces together. The filter disk will be held in place by the two O-rings.
  - e. Reassemble the syringe and load with filtered water. Insert the syringe needle into the Teflon needle guide in the filter housing. Filter water through the microfilter system by pressing gently on the syringe plunger. Dispense into the cuvet and take clean water count data again to ensure that the microfilter system is clean. Keep passing water through the microfilter system until the clean water count test is passed.
  - f. Load the syringe with filtered buffer. Thoroughly and gently rinse the filter with buffer before filtering your protein sample.
  - g. Remove the syringe from the housing and dispense any remaining buffer to waste.
  - h. If you are using the 12- $\mu$ L cuvet, then load the syringe with 20  $\mu$ L or more of protein sample and reinsert the syringe needle into the filter housing. Approximately 5–8  $\mu$ L of sample will be lost to the filter system.
  - i. Gently depress the syringe plunger to dispense one to two drops onto a paper. These two drops are sufficient to displace any remaining buffer in the needle that was used to wet the filter disk.
  - j. The sample can now be directly loaded into the cuvet from the microfilter needle. Remove any air bubbles that are created, for example by sucking them back into the needle. Place the cap on the cuvet.
  - k. Disassemble the microfiltration system and thoroughly clean and dry all components before placing them back into the case.

### 3.3. DLS Data on the Sample

#### 3.3.1. Measuring DLS Data on the Sample

Place the quartz cuvet containing your sample in the sample holder on the optics block. Set the temperature (*see Note 5*) and proceed to take DLS measurements. At least 10–20 acquisitions should be taken for each solution condition. To check reproducibility, the measurements should be done in duplicate or triplicate if there is enough sample. After measurement, the sample can be recovered

using a pipetman and capillary-style pipet tips for use in crystallization screens or other experiments.

### 3.3.2. Interpretation of DLS Data

Before you start to analyze your DLS results make sure the solvent and sample conditions are properly entered into the Dynamics software. The viscosity and refractive index of the buffer you are studying can adversely affect the estimate of  $R_H$  and  $D_T$ , respectively. The Dynamics software has a pull-down menu of frequently used buffers under the Parameters → Solvent node. These are fairly accurate if your protein concentrations are low (*see Note 6*).

When interpreting the DLS data there are several things to keep in mind. First, know the limitations of the instrument. For example, the Protein Solutions MS/X has a lower limit of 0.5 nm and an upper limit of 1- $\mu$ m particle size. Be aware that you will not be able to deconvolute the intensity measurements coming from your sample of interest, the buffer system the protein is in, or impurities in the solution. If you are using a complicated buffer system, DLS data on the buffer alone can be helpful in interpreting the results (*see Example 2*). Data from other experiments such as native polyacrylamide gel electrophoresis, electron microscopy, and size-exclusion chromatography can be helpful in interpreting results (*6*).

In studying the results from Dynamics, first study the measurements statistics table to evaluate the quality of your data and the modality of your sample (*Fig. 3*). Outliers can be filtered out using the “Data Filter” or individually marked manually. Both methods are accessed by a right-click on the Measurements Datalog grid. Guidelines for marking outliers are given in the Help software. Guidelines for the interpretation of the statistics in the Measurements Datalog grid are outlined in *Table 1* and described next.

1. Use the baseline parameter to judge if your sample is monomodal, bimodal, or multimodal. The quality of the fit of the data to a given autocorrelation function is indicated by the baseline value. Monomodal distributions are defined by a baseline ranging from 0.997 to 1.002. Bimodal distributions have a baseline range of 1.003 to 1.005. Baselines greater than 1.005 are from multimodal samples, dust, or noise.
2. The Dynamics software determines the uniformity of sizes through a monomodal (single particle size with a Gaussian distribution) curve fit analysis called Cumulants. The quality of the data is represented in the sum of squares (SOS) error statistic reported for each sample acquisition (a single correlation curve) in the Datalog grid view of Dynamics v6. The SOS error is the SOS difference between the measured data and the Cumulants-calculated intensity correlation curves. The SOS errors less than 20.0 are good and errors less than 5.0 are considered negligible (*see Note 7*) and probably represent the best samples. We have noted that, with the higher sensitivity provided by the DynaPro MS/X, the SOS errors on polydisperse samples tend to be higher than on less sensitive instruments such as the DynaPro 801, nevertheless the rules in *Table 1* still hold true.

## Example 2

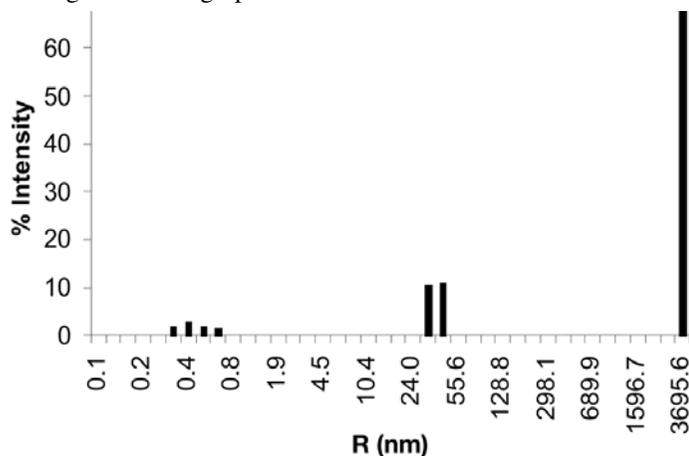
### Part A: DLS Data on Replication Protein A Buffer

Data log grid<sup>a</sup>

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	10	25	80,513	52.2	120.5	35238	0.24	1.003	302.977
Acq 2	20	25	73,262	51.6	120.1	34,243	0.26	1.01	508.992
Acq 3	30	25	59,220	32.1	146	11,281	0.19	1.003	518.801
Acq 4	40.1	25	71,035	47.1	118.3	27,704	0.25	1.006	478.842
Acq 5	50.1	25	59,948	36.6	159.8	15,337	0.19	1.003	461.535
Acq 6	60.1	25	66,565	41.9	154.4	20,973	0.19	1.009	399.663
Acq 7	70.1	25	91,962	83.1	144.2	104,370	0.3	1.067	492.541
Acq 8	80.1	25	94,199	113.1	208.5	214,887	0.32	1.131	609.86
<del>Acq 9</del>	<del>90.1</del>	<del>25</del>	<del>13,3398</del>	<del>326.3</del>	<del>244.7</del>	<del>2,562,320</del>	<del>0.534</del>	<del>1.344</del>	<del>1531.05</del>
<del>Acq 10</del>	<del>100.1</del>	<del>25</del>	<del>37,1576</del>	<del>1072.3</del>	<del>50.9</del>	<del>414,494,00</del>	<del>1.046</del>	<del>2.359</del>	<del>458.088</del>
Mean									
Meas 1	220.3	25	299261	103.7	188.9	175181	0.29	1.15	531.697

<sup>a</sup>For this sample, 20 data points were taken but only the first 10 are shown. The following data filter was applied before regularization of the data: minimum amplitude 0, maximum amplitude 1, baseline limit  $1 \pm 1$ , maximum SOS 1000. Thus, data for Acq 9 & 10 did not pass the filter and were not used as indicated by the strikethrough. For this data 13% did not pass the filter.

#### Regularization graph



#### Regularization results summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	0.4	30.6	0	8.4	100.0
<input checked="" type="checkbox"/> Peak 2	37.1	13.8	15,783	21.7	0.0
<input checked="" type="checkbox"/> Peak 3	3695.6	0.0	749,619,000	69.9	0.0

Interpretation: The SOS and baseline are very high because there is no macromolecule present and the buffer is polydisperse. Compared with **Example 1**, Peak 1 can be attributed to noise in the detector. Peaks 2 and 3 are from components in the buffer and are probably due to CHEGA10 detergent micelles.

(Example 2 continues)

**Example 2 (continued)**

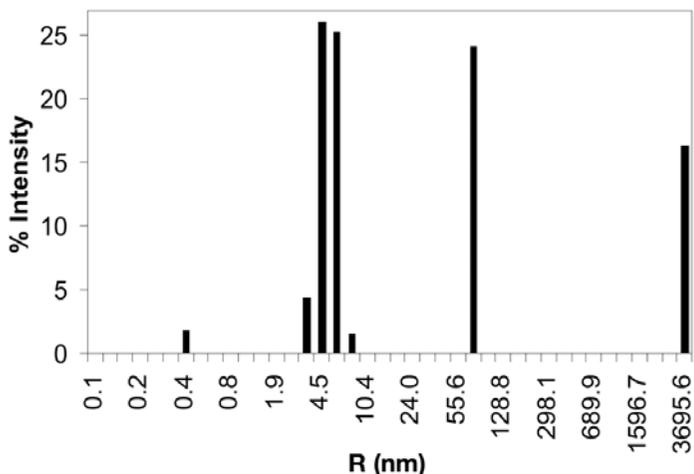
**Part B: DLS Data on RPA Heterotrimer at 0.9 mg/mL**

Data log grid<sup>a</sup>

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	9	25	464,697	11.9	104.1	1101	0.588	0.999	938.809
Acq 2	19	25	366,039	7.6	61.6	385	0.587	1.003	150.868
Acq 3	29.1	25	383,781	8	74.2	439	0.574	1.006	194.406
<del>Acq 4</del>	<del>39.1</del>	<del>25</del>	<del>1,295,610</del>	<del>148.0</del>	<del>241.4</del>	<del>402739</del>	<del>0.233</del>	<del>1.319</del>	<del>1136.620</del>
<del>Acq 5</del>	<del>49.1</del>	<del>25</del>	<del>1,255,210</del>	<del>147.0</del>	<del>191.2</del>	<del>396358</del>	<del>0.304</del>	<del>1.327</del>	<del>1558.530</del>
<del>Acq 6</del>	<del>59.1</del>	<del>25</del>	<del>610,538</del>	<del>19.3</del>	<del>115.9</del>	<del>3426</del>	<del>0.529</del>	<del>1.017</del>	<del>1780.420</del>
Acq 7	69.1	25	485,968	12.4	101.9	1228	0.594	1.004	803.41
Acq 8	79.1	25	648,094	15.9	103.8	2180	0.53	1.018	838.817
Acq 9	89.2	25	393,460	8.9	91.2	553	0.623	1.002	288.664
Acq 10	99.2	25	374,938	8.3	75.5	470	0.627	1	235.677
Mean									
Meas 1	249.4	25	517,431	10.1	95.3	761	0.574	1.013	529.

<sup>a</sup>Greater than 20 data points were taken, only the first 10 are shown. The following data filter was applied before regularization of the data: minimum amplitude 0, maximum amplitude 1, baseline limit  $1 \pm 1$ , maximum SOS 1000. Thus, data for Acq 4-6 did not pass the filter and were not used as indicated by the strikethrough. For this data 20% did not pass the filter.

Regularization graph



Regularization results summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	0.4	0.0	0	1.7	98.6
<input checked="" type="checkbox"/> Peak 2	5.1	18.5	154	55.3	1.4
<input checked="" type="checkbox"/> Peak 3	73.6	0.0	78,597	23.4	0.0
<input checked="" type="checkbox"/> Peak 4	3695.6	0.0	749,619,000	19.6	0.0

(Example 2 continues)

**Example 2 (continued)**

Interpretation: the data are very polydisperse as indicated by high baseline and high SOS error. Owing to the polydispersity of the buffer, comparison to the buffer alone DLS data is needed for interpretation of the protein sample DLS data. Peak 1, observed in the buffer only data too, is due to the low protein concentration used. Peaks 4 and 3 in buffer only data, is because the polydispersity of the buffer and probably represent CHEGA10 micelles. Identification of peak 3 components is ambiguous and may be due to the buffer components, protein or both. Peak 2 is owing to the protein, in this case RPA heterotrimer (~110 kDa). DLS could be used in this case to find the minimal concentration of CHEGA10 to solubilize the protein without micelles. Note that the peak attributed to detector noise is dominating the percentage mass calculation.

**Table 1**  
**Interpretation and Use of the Statistical Parameters Calculated by Dynamics v6<sup>a</sup>**

Parameter	Interpretation
Baseline	
0.997–1.002	Monomodal distribution
1.003–1.005	Bimodal distribution
>1.005	Multimodal distribution
Sum of squares (SOS)	
1.000–5.000	Low noise, negligible error
5.000–20.000	Background error because of noise, low protein concentration, or a small amount of polydispersity
>20.000	High noise/error owing to high polydispersity in size distribution (aggregation), irregular solvent
Normalized polydispersity	Note, this parameter should be used for monomodal distributions only.
%Pd < 15	Monodisperse solution, very likely to crystallize
%Pd < 30	A moderate amount of polydispersity, more likely to crystallize
%Pd > 30	A significant amount of polydispersity, less likely to crystallize

<sup>a</sup>Adapted from the DynaPro Operator Manual, Protein Solution, Inc. Note %Pd in Dynamics v6 was called  $C_p/R_H$  in older versions of the software.

3. The polydispersity statistic will tell you the likelihood of crystallizing your sample. The polydispersity (Pd or standard deviation) is indicative of the distribution in the peak or subpeak. By default, %Pd, or normalized polydispersity, is listed in the Datalog grid and the Regularization results summary in Dynamics v6.0. Here, %Pd is calculated by dividing Pd by  $R_H$  and reported as a percent. In older

versions of the Dynamics software, this statistic was called  $C_p/R_H$ . If your sample is monomodal, the mean %Pd of the sample can be read straight off the Datalog grid. If your sample is multimodal then mean  $R_H$  and %Pd for each peak can be obtained from the Regularization graph. If the sample is monomodal and the %Pd is less than 15% your sample is monodisperse and very likely to crystallize during screening (*see* **Note 8**). Go on to screen the sample for crystallization (**Fig. 1**), perhaps gauging your level of effort on the quality of the DLS analysis of that sample.

### 3.4. How to Use DLS Data to Find Conditions That Will Improve Crystallization Results

If your sample is multimodal, or monomodal but polydisperse (%Pd > 30%), or simply will not crystallize, the following experimental considerations can help improve your DLS and crystallization results.

1. Increasing the solubility of the sample will typically decrease the aggregation and polydispersity of the sample. Therefore, it is helpful to perform the Mueser solubility screen on your sample to optimize buffer conditions. The following protocol that has been adopted successfully several times (7,8). It is especially powerful when coupled to the DLS polydispersity analysis.
  - a. Dialyze 2–5 mg protein against deionized water (no buffer, no salt). Most proteins will precipitate under these conditions.
  - b. Resuspend the precipitated protein and aliquot the precipitate into 20 1.5-mL microcentrifuge tubes. Centrifuge at maximum speed in a table-top centrifuge for 2–5 min to repellet the protein. Remove the supernatant.
  - c. Each tube will be a separate experiment. Add either buffer, chloride salt, or ammonium salt (20 mL of 100 mM solutions described in **Subheading 2., steps 15–17**), and resuspend the pellet. Let stand at room temperature for 10 min, centrifuge to pellet the undissolved protein, and measure the protein concentration of the supernatant amount redissolved. If the pellets completely dissolve then use less volume or more protein.
  - d. When you have information about which buffer and/or salt are best, it is suggested that you try a cross-coupled experiment. For example, if LiCl and  $(NH_4)_2SO_4$  give good solubility, then perhaps  $(Li)_2SO_4$  is worth trying.
  - e. Test the final buffer condition for maximum solubility using mini concentrators. The best starting protein concentration for crystallization is one-half of the maximum solubility.
2. Perform a series of DLS experiments to test the effect of ionic strength, pH, protein concentration, organic solvents, detergents, and other additives on the polydispersity of your sample (*see* **Example 3**). Perform crystallization experiments at the solution condition where your sample is the most monodisperse and stable over time.
3. Test the effect of temperature on your sample. Find the temperature where your protein is the most monodisperse and then set the crystallization incubator to this temperature for crystal growth. The Event Scheduler can automate these experiments (right-click on Parameters node to access the Event Scheduler).

**Example 3 Part A:  
Monodisperse DLS Data on Crystallizable RPA14/32 Heterodimer,  
at a Concentration of 10 mg/mL**

Data log grid<sup>a</sup>

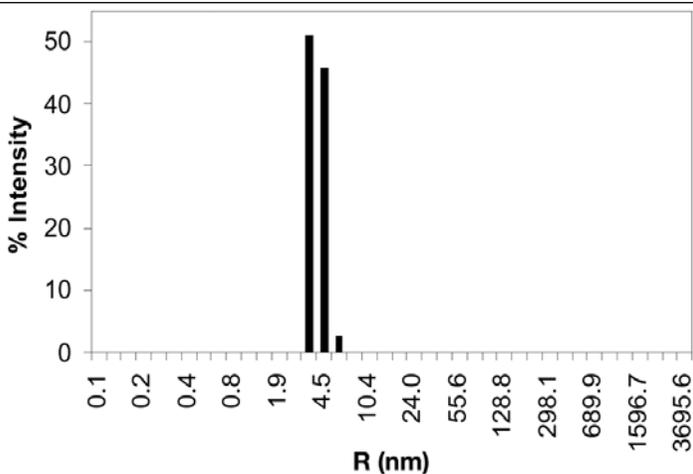
Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	10	25	1,872,940	3.9	22.5	80	0.496	1.000	0.812
Acq 2	20	25	1,872,590	3.9	14.7	83	0.552	1.000	1.145
Acq 3	30	25	1,888,920	3.9	17.2	81	0.544	1.000	0.761
Acq 4	40.1	25	1,894,110	3.9	24	82	0.521	1.000	1.124
Acq 5	50.1	25	1,908,650	3.9	29.1	84	0.51	1.000	1.83
Acq 6	60.1	25	1,892,830	3.9	23.5	82	0.504	1.000	1.239
Acq 7	70.1	25	1,891,400	3.9	21.5	82	0.495	1.000	0.862
Acq 8	80.1	25	1,888,290	3.9	23.6	83	0.489	1.000	1.355
Acq 9	90.1	25	1,882,720	3.9	10.9	83	0.49	1.000	1.385
Acq 10	100.1	25	1,876,270	3.9	29.7	82	0.49	1.000	1.071

Mean

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Meas 1	300.4	25	1,867,130	3.9	12.2	83	0.509	1.000	1.1375

<sup>a</sup>Greater than 20 data points were taken, only the first 10 are shown. The application of a data filter was not needed.

Regularization graph



Regularization results summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	4.0	15.8	84	100.0	100.0

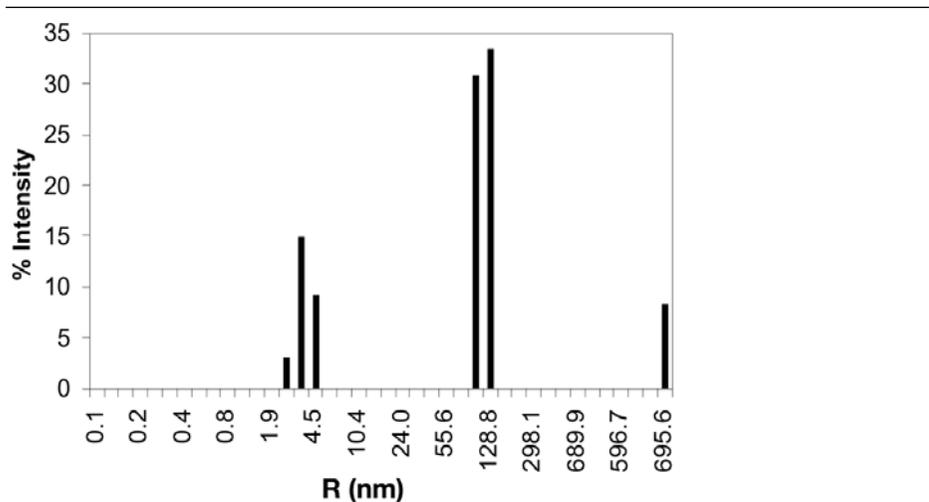
Interpretation: the sample gave a monomodal fit with a baseline of 1.000 and SOS of 1. RPA14/32 is a 46-kDa heterodimer with a predicted  $R_H$  of 3.1 nm. A dimer of heterodimers has a predicted  $R_H$  of 4.1 nm. Therefore, these data indicate primarily a dimer of heterodimers in solution. With a %Pd of 16%, this sample crystallized readily out of several crystallization conditions and several space groups (10).

**Example 3 Part B:****Multimodal DLS Data on RPA14/32 Heterodimer, at a Concentration of 5 mg/mL, diluted With 30 mM HEPES pH 7.8**Data log grid<sup>a</sup>

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	10	25	3,976,320	69.5	120.1	68,813	0.336	1.010	401.397
Acq 2	20	25	2,970,080	42.4	137	21568	0.377	1.003	837.184
Acq 3	30	25	2,968,450	43.6	118.7	23036	0.384	1.003	832.709
Acq 4	40.1	25	2,327,790	21.9	115.9	4613	0.386	1.003	958.650
Acq 5	50.1	25	2,388,140	22.8	118	5087	0.371	1.003	833.651
Acq 6	60.1	25	2,560,250	29.7	116.5	9434	0.388	1.003	984.212
Acq 7	70.1	25	2,680,860	32.8	116.8	11840	0.39	1.002	902.132
Acq 8	80.1	25	2,796,570	35.8	118	14553	0.368	1.000	849.627
Acq 9	90.1	25	2,777,060	36	114.7	14790	0.365	1.001	876.540
Acq 10	100.2	25	2,600,990	27.6	116.2	7896	0.374	1.001	914.389
Mean									
Meas 1	310.5	25	2925130	38.3	118.1	17040	0.362	1.002	819.282

<sup>a</sup>Greater than 20 data points were taken, only the first 10 are shown. The application of a data filter was not needed.

## Regularization Graph



## Regularization Results Summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	3.7	17.4	70	27.4	99.9
<input checked="" type="checkbox"/> Peak 2	113.7	13.8	217490	64.3	0.0
<input checked="" type="checkbox"/> Peak 3	3695.6	0.0	749619000	8.3	0.0

Interpretation: the increased baseline and very high SOS error indicate that dilution with this buffer introduces polydispersity into the sample. The regularization fit indicates three peaks. Peaks 2 and 3 are from aggregated protein. When many data acquisitions are taken this aggregation becomes worse over time (data not shown). Therefore, this is not a suitable buffer condition for this protein.

### Example 3 Part C:

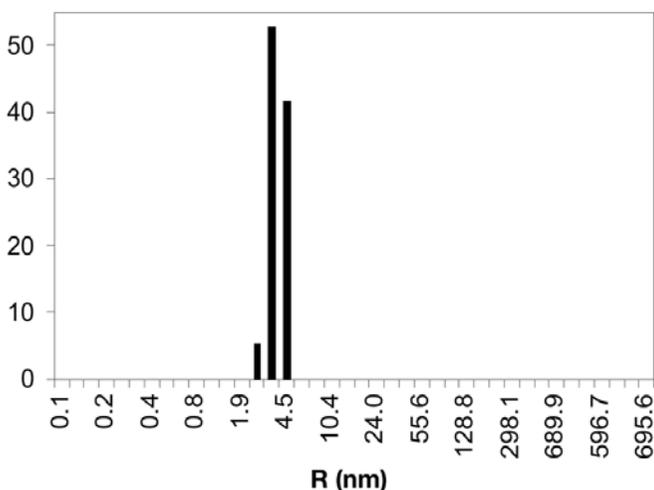
#### Monodisperse DLS Data on RPA14/32 Heterodimer at a Concentration of 5 mg/mL, Diluted with 30 mM HEPES pH 7.8, 200 mM KCl, and 10 mM DTT; Crystallizable in This Solution Condition

Data log grid<sup>a</sup>

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	9	25	1,047,070	3.6	14.6	68	0.531	1.000	1.244
Acq 2	19	25	1,053,610	3.8	30.6	77	0.540	1.000	2.441
Acq 3	29	25	1,053,860	3.7	35.1	73	0.528	1.000	1.797
Acq 4	39.1	25	1,061,240	3.8	26.2	76	0.509	1.000	2.154
Acq 5	49.1	25	1,067,120	3.7	21.3	73	0.515	1.000	0.983
Acq 6	59.1	25	1,029,520	3.8	21.4	77	0.593	1.000	2.747
Acq 7	69.1	25	1,026,860	3.9	28.5	80	0.599	1.000	2.787
Acq 8	79.1	25	1,034,900	3.8	19.3	77	0.592	1.000	1.178
Acq 9	89.1	25	1,054,730	3.8	30.4	77	0.566	1.000	2.361
Acq 10	99.1	25	1,074,230	3.8	32.9	79	0.545	1.000	2.580
Mean									
Meas 1	329.5	25	1056810	3.8	24.2	75	0.532	1.000	1.581

<sup>a</sup>Greater than 20 data points were taken, only the first 10 are shown. The application of a data filter was not needed.

Regularization graph



Regularization results summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	3.8	15.9	77	100.0	100.0

Interpretation: when compared with **Parts A and B of Example 3**, dilution of the protein into this buffer is good. The baseline of 1.000 and SOS of 1.6 shows that the sample is monomodal. The %Pd of 16 indicates that the protein is still monodisperse after dilution. Therefore, the protein is stable in this buffer and this is a good starting point for crystallization trials.

4. Test the effect of binding partners (protein, peptides, or oligonucleotides) or substrates on  $R_H$  and monodispersity. The effect of stoichiometry of mixing can also be tested (see **Example 4**). Anything that makes the molecule smaller and more compact may render it more crystallizable (9).
5. Use DLS analysis to help optimize the protein purification protocol, e.g., to avoid inappropriate disulfide bond formation (10).
6. The protein itself should be considered a crystallization variable (11). Use DLS analysis to help select the best construct. For example, perhaps three deletions of different lengths are made from the N-terminus of the protein. Then put your biggest crystallization effort into the sample with the best monodispersity.
7. DLS is also helpful to test the effects of storage procedures, e.g., freezing vs refrigeration and to assess shelf life (12–18).

#### 4. Notes

1. Protein Solutions has also written “Dynapro Data Interpretation Guide;” be sure to obtain a copy. In addition, several texts and articles have been written on the collection and analysis of DLS data (12–16).
2. If the protocol described does not clean the cuvet try placing it in a sonicating bath for 15–20 min and then clean it again. Do not use *concentrated* acids and bases to clean the cuvet as they can etch the surface. Do not use organic solvents (e.g., ethanol) to rapidly dry the cuvet as they can leave a thin residue on the surface. Always clean the cuvet thoroughly after use and before storage. If it is not stored

#### Example 4

##### DLS Data on a RPA:Rad52 Complex at a Concentration of 8 mg/mL

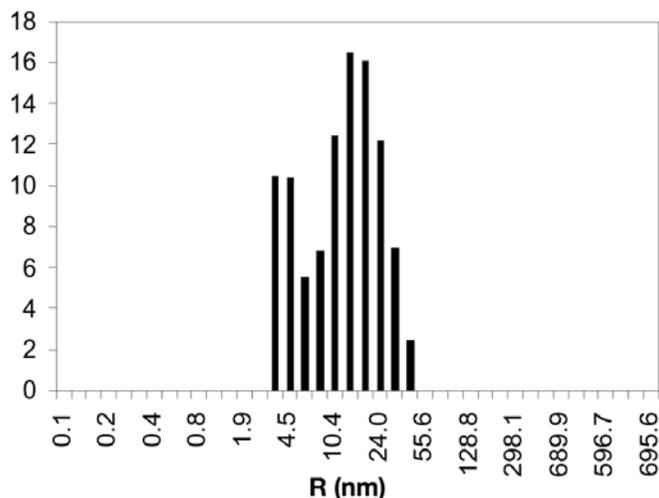
Data log grid<sup>a</sup>

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	%Pd	MW-R (kDa)	Amp	Baseline	SOS
Acq 1	9	25	4,495,800	10.8	64.5	872	0.311	1.001	14.876
Acq 2	39.1	25	5,018,720	10.8	54.1	875	0.332	1.000	20.763
Acq 3	49.1	25	4,969,060	10.7	61.9	870	0.328	1.000	19.873
Acq 4	69.1	25	4,941,200	11	61.2	921	0.325	0.999	21.940
Acq 5	89.2	25	4,953,440	11.2	64.7	969	0.334	1.000	20.287
Acq 6	99.2	25	4,894,110	10.9	69.8	903	0.327	0.999	18.665
Acq 7	109.2	25	4,878,090	10.9	72.9	894	0.327	0.999	19.605
Acq 8	119.3	25	4,881,350	11	72	925	0.330	1.000	21.353
Acq 9	129.3	25	4,869,450	10.9	61.5	901	0.330	1.001	20.228
Acq 10	139.3	25	4,835,920	10.8	65.7	879	0.334	1.000	18.834
Mean									
Meas 1	289.5	25	4,939,260	10.9	61	904	0.32	1.000	18.077

<sup>a</sup>Greater than 20 data points were taken, only the first 10 are shown. The application of a data filter was not needed.

**Example 4 (continued)**

Regularization graph



Regularization results summary

Item	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	4.4	21.8	106	26.4	92.0
<input checked="" type="checkbox"/> Peak 2	18.0	45.5	2901	73.6	8.0

Interpretation: the sample in this case is a complex of two proteins, RPA heterotrimer (110 kDa) and Rad52 heptameric ring (~350 kDa), mixed with an approximate one to one molar ratio. The data are polydisperse as indicated by the high SOS error. The regularization analysis shows that the sample is bimodal and polydisperse. This sample did not crystallize. Possible reasons are that the molar ratio was not exact or that a portion of the molecules is inactive. The next step is to vary the stoichiometry of RPA to Rad52 to try to obtain 100% monodisperse complex in solution. If this is not successful the complex will need to be separated from free RPA by size exclusion chromatography before crystallization trials.

in a clean state, you will cause the next user many headaches. If the cuvet window becomes scratched it will need to be replaced.

3. The sample may be unexpectedly lost during the filtration process. This could be because of aggregation, unexpected quaternary structure, or unusual binding to the filter. Also, it is important to take a protein reading or run a sodium dodecyl sulfate-polyacrylamide gel electrophoresis PHAST gel (Amersham Biosciences, Piscataway, NJ) on 1  $\mu$ L of sample before and after filtration to access how much, if any, is lost owing to filtration. Consider pore size and the MW of your protein. The Dynamics software includes a MW calculator that will estimate  $R_H$  for you. Use it to see which pore size to use. For example, do not use 0.020- $\mu$ m size pores if your protein is larger than 150 kDa. If needed, a filter with larger pore size can be used.

4. The Whatman Anotop filters used with the microfilter kit are very brittle. Care must be taken in handling them so that they do not crack. They must also be seated properly on the O-ring so that solution cannot pass around the filter.
5. Useful information on the effect of temperature on aggregation can be gained from starting the DLS measurements at 4°C and then stepping up the temperature in 5° increments and taking DLS measurements at each temperature from the same sample. The Event Scheduler node in Dynamics software can be used to automate these measurements. To access the Even Scheduler right-click on the Parameters node. The sample should be incubated 30 min at each temperature before DLS data (10–20 measurements) are taken. The maximum temperature for the DynaPro MS/X is to 60°C. Temperature can be used to control the aggregation of a protein (17). Ramping down from high temperature has been used in the crystallization of macromolecules, e.g., insulin (18).
6. The viscosity is influenced by protein concentration and buffer components, such as alcohols and glycerol. For precise measurements of  $R_H$ , viscosity and refractive index can be measured with a viscometer and refractive index detector, respectively, and then entered by hand into the software.
7. Regularization analysis will be able to give you some size information about your sample even if the SOS error is high and your baseline is high. Both of these statistics absorb error so the amount of trust you should place in the data in these cases should be proportionately low.
8. The %Pd is the statistic most useful for predicting crystallizability, if your sample is monomodal. A special exception is made for bi- or multimodal samples where one or more of the peaks result from something in the solvent (e.g., detergent micelles) that produces light scattering. If the protein peak can be identified from the noise through appropriate control experiments, the  $C_p/R_H$  for this peak off the regularization graph can be used to predict crystallizability. Also, the concentration of detergent can be optimized by DLS to eliminate the presence of micelles in the protein solution.

## Acknowledgments

The author would like to thank Tim Mueser and Doba Jackson for intellectual contributions and for critically reading drafts of this manuscript, and Gilson Baía for providing the DLS data used in the examples. Wasantha Ranatunga and Jeff Habel helped develop the laboratory's expertise in DLS. Grants from the Ohio Cancer Research Associates, Inc., the US Army Medical Research and Materiel Command DAMD17-98-1-8251, and the American Cancer Society RSG GMC-104160 supported this work.

## References

1. Zulauf, M. and D'Arcy, A. (1992) Light scattering of proteins as a criterion for crystallization. *J. Cryst. Growth* **122**, 102–106.
2. D'Arcy, A. (1994) Crystallizing proteins: a rational approach. *Acta Cryst.* **D50**, 469–471.

3. Wilson, W. W. (2003) Light scattering as a diagnostic for protein crystal growth: a practical approach. *J. Structural Biol.* **142**, 56–65.
4. Bergfors, T. M. (1999) *Protein Crystallization Techniques, Strategies and, Tips*. International University Line, La Jolla, CA.
5. Ferré-D'Amaré, A. R. and Burley, S. K. (1997) Dynamic light scattering in evaluating crystallizability of macromolecules. *Meth. Enzymol.* **276**, 157–166.
6. Ranatunga, W., Jackson, D., Lloyd, J. A., Forget, A. L., Knight, K. L., and Borgstahl, G. E. O. (2001) Human Rad52 exhibits two modes of self-association. *J. Biol. Chem.* **276**, 15,876–15,880.
7. Mueser, T. C., Rogers, P. H., and Arnone, A. (2000) Interface sliding as illustrated by the multiple quaternary structures of liganded hemoglobin. *Biochemistry* **39**, 15,353–15,364.
8. Collins, B. K., Tomanicek, S. J., Lyamicheva, N., Kaiser, M. W., and Mueser, T. C. (2004) A preliminary solubility screen used to improve crystallization trials. Crystallization and preliminary X-ray structure determination of Aeropyrum pernix Flap Endonuclease-1. *Acta Cryst.* **D60**, 1674–1678.
9. Jackson, D., Dhar, K., Wahl, J. K., Wold, M. S., and Borgstahl, G. E. (2002) Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA. *J. Mol. Biol.* **321**, 133–148.
10. Habel, J. E., Ohren, J. F., and Borgstahl, G. E. O. (2001) Dynamic light scattering analysis of full-length, human RPA14/32 dimer: purification, crystallization and self-association. *Acta Cryst. D* **D57**, 254–259.
11. Dale, G. E., Oefner, C., and D'Arcy, A. (2003) The protein as a variable in protein crystallization. *J. Struct. Biol.* **142**, 88–97.
12. Pusey, P. N., Koppel, D. E., Schaefer, D. W., Camerini-Otero, R. D., and Koenig, S. H. (1974) Intensity fluctuation spectroscopy of laser light scattered by solutions of spherical viruses: R17, Obeta, BSV, PM2 and T7. I. Light-scattering technique. *Biochemistry* **13**, 952–959.
13. Camerini-Otero, R. D., Pusey, P. N., Koppel, D. E., Schaefer, D. W., and Franklin, R. M. (1974) Intensity fluctuation spectroscopy of laser light scattered by solutions of spherical viruses: R17, Obeta, BSV, PM2 and T7. II. Diffusion coefficients, molecular weights, solvation, and particle dimensions. *Biochemistry* **13**, 960–970.
14. Pecora, R. (1985) *Dynamic Light Scattering: Applications of Photon Correlation Spectroscopy*. Plenum Press, New York, NY.
15. Brown, R. G. W. (1990) Miniature laser light scattering instrumentation for particle size analysis. *Applied Optics* **29**, 1.
16. Phillies, G. D. J. (1990) Quasielectric light scattering. *Anal. Chem.* **62**, 1049A–1057A.
17. Ranatunga, W., Jackson, D., II, R. A. F., and Borgstahl, G. E. O. (2001) Human Rad52 protein has extreme thermal stability. *Biochemistry* **40**, 8557–8562.
18. Long, M. M., Bishop, J. B., Nagabhushan, T. L., Reichert, P., Smith, G. D., and DeLucas, L. J. (1996) Protein crystal growth in microgravity: bovine insulin, human insulin and human alpha interferon. *J. Crystal Growth* **168**, 233–243.



## Screening and Optimization Methods for Nonautomated Crystallization Laboratories

Terese Bergfors

### Summary

Crystallization of biological macromolecules is becoming increasingly automated. However, for various reasons, many laboratories still perform at least some aspects of the work manually. A typical crystallization project entails two distinct steps: screening and optimization. The aim of the initial phase is to screen the many parameters affecting crystallization, and as broadly as possible. If any promising conditions are found, these are optimized with other protocols. This chapter describes procedures for manual screening by the vapor diffusion and microbatch methods in 96- and 24-well plate formats. For optimization, several protocols are presented, including grid and additive screens, seeding, and manipulation of the drop kinetics. The scoring of crystallization results and methods for distinguishing protein and salt crystals are also discussed in this chapter.

**Key Words:** Additives; crystallization; grid screens; hanging drop; microbatch; optimization; screening; seeding; sitting drop; sparse matrix; vapor-diffusion.

### 1. Introduction

The crystallization of biological macromolecules is usually a two-step process. The first of these is a broad sampling of the solution parameters known to affect crystallization. Useable crystals can sometimes be found by the initial search; more often however, further experiments are required to optimize the lead conditions. At present, no algorithms exist to predict which crystallization conditions will be successful for any particular macromolecule. As the crystallization databases expand with the input from structural proteomics, this situation may change. Data mining may reveal common trends for certain types of molecules (**1**) or find “hot spots” in the crystallization space. Currently though, the most used method for initial screening is a design known as the sparse matrix, which has its origin in the work of Jancarik and Kim (**2**). Nowadays,

many sparse-matrix crystallization kits are available commercially. **Table 1** lists sources for these as well as some other types of screens.

Kits, however, are of limited usefulness at the level of optimization because these experiments must be designed on a case-by-case basis, depending on which lead conditions are found in the screening phase. An often used optimization design is a systematic grid search varying two parameters at a time.

Structural genomics and proteomics have led to new, automated technologies for crystallization. Although many laboratories do not have the need or economic resources for automation, some of the high-throughput developments are applicable in nonautomated laboratories (**3**). This chapter will present manual methods of crystallization as practiced classically, but also incorporate some developments, e.g., 96-well crystallization plates from high-throughput operations (*see also* Chapter 9 for high-throughput crystallization and optimization techniques).

## 2. Materials

1. 96-Well crystallization plates for sitting-drop vapor-diffusion experiments, e.g., Crystal Quick™, (Greiner Bio-One, Frickenhausen, Germany, cat. no. 609171), Intelli-Plate, (Hampton Research, Aliso Viejo, CA, cat. no. HR3-297), or Corning plate (Hampton Research, cat. no. HR3-271).
2. 96-Well plates for microbatch experiments under oil (e.g., Imp@ct™, (Greiner, cat. no. 673101).
3. Optically clear sealing tape for 96-well plates (e.g., Greiner, cat. no. 676070).
4. Commercial kits (*see Table 1* for suppliers).
5. Paraffin oil (e.g., BDH, Poole, UK, cat. no. 29436 5H).
6. 25–250  $\mu\text{L}$  Eight-channel (electronic) pipet (e.g., Biohit Proline, Helsinki, Finland, cat. no. 710220).
7. 0.2–10  $\mu\text{L}$  Eight-channel (electronic) pipet (e.g., Biohit Proline, cat. no. 710300).
8. 0.2–10  $\mu\text{L}$  Single-channel (electronic) pipet (e.g., Biohit Proline, cat. no. 710520).
9. Racked pipet tips for the previously mentioned pipets.
10. 24-Well tissue culture plates, e.g., Linbro dishes (Hampton Research, cat. no. HR3-110) or XRL plates (Molecular Dimensions, Cambridgeshire, UK, cat. no. MD3-11).
11. Silanized glass cover slips, 0.2–0.3-mm thick, and 18–22 mm in diameter as appropriate to fit the 24-well plates.
12. High-vacuum grease.
13. Stock solution of 20% sodium azide or 0.22- $\mu\text{m}$  microcentrifuge tube filter (e.g., Whatman Anopore, Kent, UK, cat. no. 6830 0202).
14. Tweezers.
15. Canned air (available from photography supply companies, Hampton Research, and others).
16. Stereo-microscope.

### 2.1. Optional Materials

1. 1.2–2.2 mL 96-well storage blocks (e.g., Axygen, Union City, CA 997-O-DW-20-C).

**Table 1**  
**Portfolio of Screens<sup>a</sup>**

Name	Type of screen	Source or manufacturer
Crystal Screen HT <sup>b</sup> , MemFac	Sparse matrix	Hampton Research ( <a href="http://www.hamptonresearch.com">www.hamptonresearch.com</a> )
Wizard 1 and 2 <sup>b</sup> , Cryo 1 and 2 <sup>b</sup>	Sparse matrix	Emerald BioStructures ( <a href="http://www.decode.com/emeraldbiostructures">http://www.decode.com/emeraldbiostructures</a> )
Flexible Sparse Matrix	Sparse matrix	Zeelen, J. ( <b>13</b> )
Personal Structure Screen	Sparse matrix	Molecular Dimensions ( <a href="http://www.moleculardimensions.com">www.moleculardimensions.com</a> )
Crystallization Basic Kit for Proteins	Sparse matrix	Sigma Aldrich ( <a href="http://www.sigmaaldrich.com">www.sigmaaldrich.com</a> )
JBScreen HT 1L <sup>b</sup> JBScreen HT 2L <sup>b</sup>	Grid	Jena Bioscience ( <a href="http://www.jenabioscience.com">www.jenabioscience.com</a> )
Imperial College Screen	Grid	Haire, L. ( <b>14</b> )
Grid Screens <sup>TM</sup> , Quik Screen, SaltRX HT <sup>b</sup> Sodium Malonate Screen	Grid	Hampton Research
Footprint	A solubility footprint of the protein	Molecular Dimensions
Crystool	Random	<a href="http://www-structure.llnl.gov">http://www-structure.llnl.gov</a>
ZetaSol	Based on the net charge of the protein and the Hofmeister's series	Molecular Dimensions
Index HT <sup>TMb</sup>	Combines grid, incomplete factorial, and sparse matrix features all in one screen	Hampton Research
Clear Strategy Screens	Allows full control over screening pH; contains "built-in" anomalous scatterers and cryoprotectants	Molecular Dimensions
Incomplete Factorial	A statistically efficient design	Carter, C. W., Jr. and Carter, C. W. ( <b>15</b> )

<sup>a</sup>This list contains options available at the time of writing. New screens appear frequently.

<sup>b</sup>Available in a high-throughput format (96-well storage block).

2. Super-glue.
3. An animal hair or whisker.
4. Seed Bead™ (Hampton Research, cat. no. HR2-320).
5. IZIT™ dye (Hampton Research, cat. no. HR4-710).
6. Nextal (Montreal, Canada) crystallization support (cat. no. NCS-24-001).
7. Cryschem™, Q-plate, or sitting bridge inserts (Hampton Research).
8. Acupuncture needles.
9. Centrifuge with a swing-out rotor that can accept microtiter plates.
10. Pre-Crystallization Test, PCT™ (Hampton Research, cat. no. HR2-140).
11. Low-viscosity silicone oil (BDH, cat. no. 73002 4N) or Al's oil (Hampton Research, cat. no. HR3-413).
12. A silanizing solution, e.g., Repel-Silane (Amersham Biotech, Uppsala, Sweden, cat. no. 17-1332-01) or Aqua Sil (Hampton Research, cat. no. HR4-611).

### 3. Methods

#### 3.1. Preparation of the Sample for the Initial Screen

The protein or other biological macromolecule is the single most important component in the crystallization trials. The sample purity, homogeneity, and stability are major determinants of the outcome of the crystallization experiment.

1. Assay the purity of the sample by gel electrophoresis or other methods. At a minimum this means a SDS gel, stained with Coomassie blue. The protein should appear at least 90% pure. Run native and isoelectrofocusing gels if possible. Archive the gel(s) for comparison with future batches of the protein and to monitor possible changes in the protein over time. For similar reasons, archive 5–10  $\mu\text{L}$  of the protein by flash-freezing it in liquid nitrogen. Store this aliquot at  $-80^\circ\text{C}$ .
2. Exchange the purification buffer for 10 mM HEPES, pH 7.0 or 10 mM Tris-HCl, pH 8.0. Substitute other buffers as necessary if HEPES or Tris are inappropriate for a particular protein (*see Note 1*). The ionic strength of a 10 mM buffer may be too low to keep the protein in solution. NaCl should then be included (begin with 25–150 mM). Reducing agents, cofactors, or detergents may also be required by certain proteins.
3. Determine the protein concentration by standard procedures. The concentration of protein to use in the crystallization trial will depend on how soluble the molecule is. For a poorly soluble protein, this may be as low as 2–4 mg/mL (*see Note 2*). For highly soluble proteins, begin screening at 20–40 mg/mL. These values should be considered as rough guidelines only (*see Note 3*). As a rule of thumb, the more soluble the protein, the more concentrated it should be for the purpose of initial screening.
4. The protein solution should be protected from bacterial growth. Add 0.02% sodium azide or pass the protein through a 0.22- $\mu\text{m}$  filter (*see Note 4*). Buffers and precipitants used in the crystallization trials should be filtered.
5. Use freshly purified protein when possible. Otherwise, avoid repeated freezing and thawing of the entire protein stock. Divide the protein solution into 100- $\mu\text{L}$

aliquots. Store them at a temperature appropriate for the protein, usually  $-80^{\circ}\text{C}$ . Before setup of the crystallization trials, thaw one aliquot and centrifuge it at  $16,000g$  for 5 min to pellet dust, aggregated molecules, and so on. The supernatant should be free of any turbidity before use in the crystallization drops.

### 3.2. Principles of the Vapor-Diffusion Method

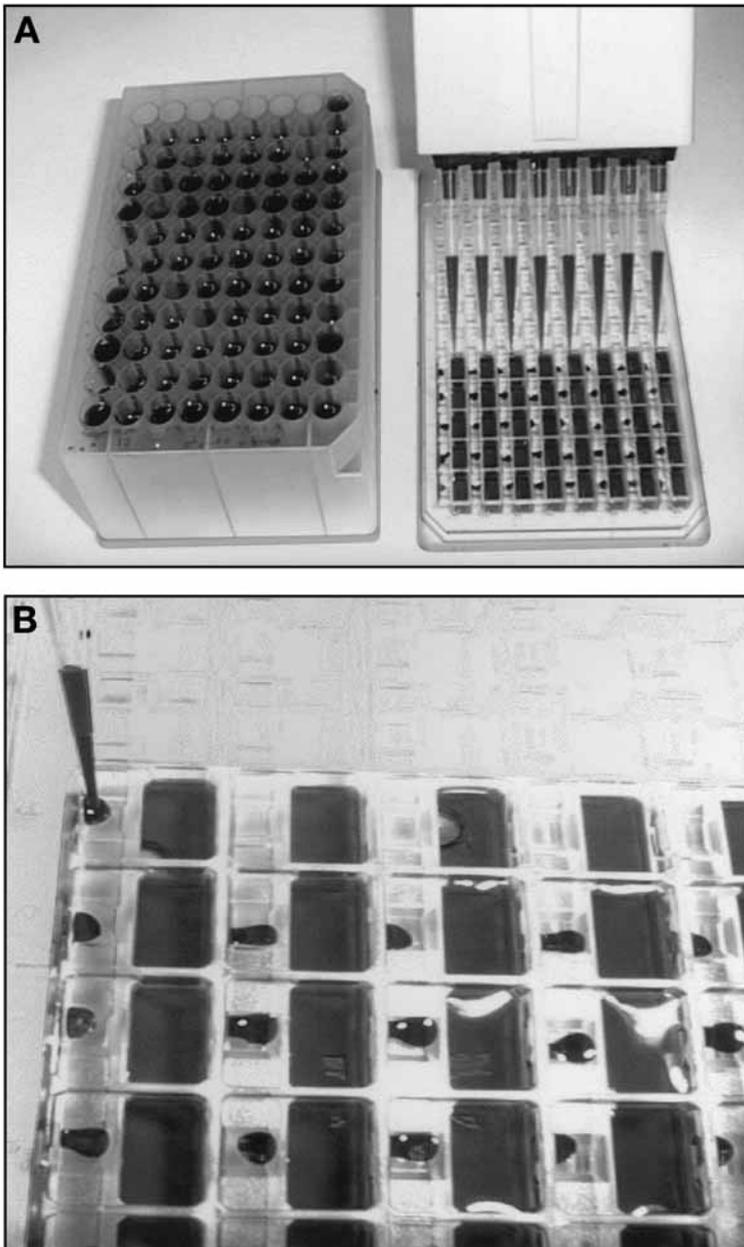
Crystals can only grow from supersaturated solutions; therefore, the protein must be brought to supersaturation. This chapter describes two methods for doing this: vapor-diffusion and microbatch experiments. In these approaches, a solution containing buffer and precipitating agent is prepared. The precipitant can be a salt, polymer, organic solvent, or combinations thereof. Additional components, e.g., dithiothreitol, azide, detergents, and so on, may also be included; collectively all these ingredients constitute what is called the mother liquor, i.e., the solution from which the crystals grow.

In the vapor-diffusion experiment, a droplet of the undersaturated protein solution is mixed with a droplet of the mother liquor. This mixture is then equilibrated against a much larger reservoir containing only mother liquor. The concentration difference between the reservoir and the droplet causes water to leave the droplet; it diffuses as a vapor into the reservoir. As this happens, the protein and mother liquor both become more concentrated. This leads to a supersaturated state of protein while at the same time, the increased precipitant concentration lowers the protein's solubility. Under favorable conditions, this combination of events drives the protein out of solution as a highly ordered solid phase, namely a crystal.

The geometry of the vapor-diffusion experiment has several variations: the two most common ones are sitting drops and hanging drops. Examples of these two will be described in the next sections.

#### 3.2.1. Procedure for Sitting-Drop Vapor-Diffusion Setup in 96-Well Plates

1. Select one of the commercially available screens sold in a 96-well storage block format. Alternatively, prepare your own screen in a similar storage block.
2. Using the larger volume (25–250  $\mu\text{L}$ ) multichannel pipet, transfer 50–100  $\mu\text{L}$  of each of the 96 solutions from the storage block to the reservoirs of the 96-well crystallization plate for sitting drops (**Fig. 1A**).
3. Set the single-channel pipet on its multidispensing mode to aspirate 8  $\mu\text{L}$  of protein solution from a microcentrifuge tube (**Fig. 1B**). Dispense the 8  $\mu\text{L}$  as 1- $\mu\text{L}$  aliquots into the first row of sitting drop wells. Repeat for the remaining rows (*see Note 5*).
4. With the smaller volume multichannel pipet (0.2–10  $\mu\text{L}$ ), transfer 1  $\mu\text{L}$  from the plate reservoirs to the sitting drop wells corresponding to them. This droplet should be delivered so that it merges with the 1- $\mu\text{L}$  protein droplet. Stirring or mixing of the combined droplets is optional (*see Note 6*).



**Fig. 1.** Setup in a 96-well plate for sitting-drop vapor-diffusion experiments. **(A)** A multichannel pipet transfers aliquots of the screening solutions (left) to the reservoirs of a Crystal Quick™ plate (right). **(B)** Delivery of the protein solution with a repeating, single-channel pipet to the shallow depressions (close-up view). The solutions have been dyed to enhance visibility in the photograph.

5. Seal the experiment with a transparent tape or optically clear film, such as that used for PCR plates.
6. Examine the droplets under the microscope. If there are air bubbles or the two 1- $\mu\text{L}$  droplets have not coalesced, centrifuge the plate at 300g for 5 min. (Any rotor for microtiter plates will be able to accept the crystallization dishes.)
7. Store the plate at a constant, controlled temperature.
8. Repeat with a different screen, temperature, or method, e.g., microbatch instead of vapor diffusion, if there is sufficient protein.

### 3.2.2. Procedure for Hanging-Drop Vapor Diffusion in 24-Well Plates

The setup with 96-well plates using multichannel pipets is compact, quick, and reduces the need for reagents to a minimum. However, the most frequently used item of plasticware for vapor diffusion in the last decade has been the 24-well disposable tissue-culture plate. Next is a description of how to perform vapor-diffusion setups with hanging drops in these plates (*see* [Fig. 2](#) and [Note 7](#)).

1. Choose a 24-well plate type with thick rims around the reservoirs, e.g., Linbro or XRL plates.
2. Grease the raised rims with high-vacuum grease. The grease can be smeared onto the rim with a fingertip or a small painting brush, or dispensed from a 10-cc syringe filled with grease.
3. Pipet 0.5–1.0 mL of each kit reagent (bought commercially or prepared in-house) into a well of the plate.
4. Place a cover slip on the bench-top. Remove any dust on it with a puff of canned air.
5. Carefully pipet 1–5  $\mu\text{L}$  of protein solution onto the cover slip. Try to form a drop that is as spherical as possible (*see* [Note 8](#)).
6. Add 1–5  $\mu\text{L}$  of the reservoir solution from the first well to the protein droplet on the cover slip. Avoid the formation of air bubbles in the two droplets. Remove any such bubbles by a quick poke with an acupuncture needle.
7. With a pair of tweezers, carefully invert the cover slip without disturbing the droplet.
8. Suspend the hanging drop over the reservoir by placing the cover slip onto the greased rim. Press gently to ensure a tight seal between cover slip and rim.
9. Prepare the rest of the plate in the same manner.
10. Inspect the seals at low magnification under the microscope for air gaps between the rim and cover slip. Press out any air bubbles or, failing this, pry up the cover slip and recoat the rim with a thicker layer of grease. Gaps will lead to dehydration of the experiment.
11. Gently transfer the plate to the crystallization room without shaking, bumping, or splashing the drops (*see* [Note 9](#)).

### 3.3. Principles of Crystallization by Microbatch

Microbatch can be used instead of vapor diffusion or in parallel with it (*see* [Note 10](#)). In microbatch, the protein and mother liquor components are mixed at high concentrations to achieve supersaturation directly. By contrast, the protein



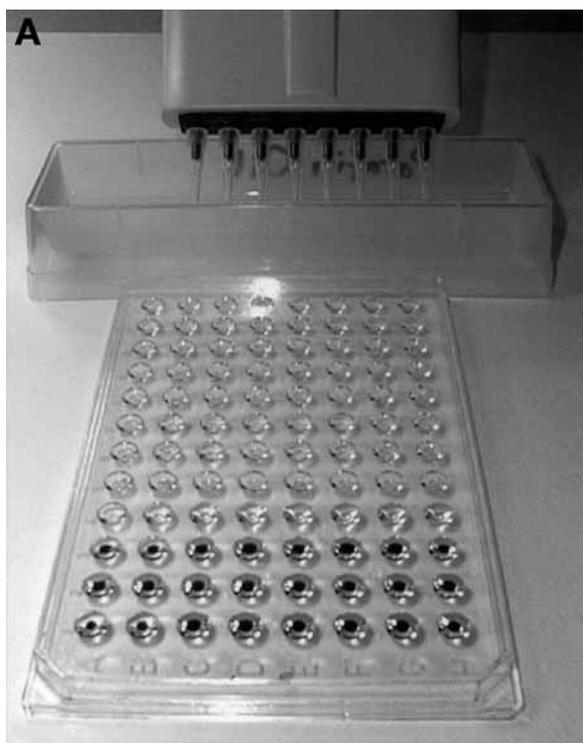
**Fig. 2.** Hanging-drop vapor-diffusion setup in 24-well tissue culture plates.

The hanging drop, suspended on a cover slip, is about to be sealed on the greased rim. The well contains 1 mL of the screening solution. Solutions have been dyed to enhance visibility in the photograph.

solution in a vapor-diffusion experiment is initially undersaturated, but becomes supersaturated during the course of the trial. The microbatch experiment is done under oil to prevent evaporation of the drop components. There is no equilibration with a reservoir.

### 3.3.1. Procedure for Microbatch Under Oil in 96-Well Plates

1. Dispense 25–30 mL of paraffin oil (*see Note 11*) into a reagent basin (**Fig. 3A**).
2. With a multichannel pipet, transfer 10–20  $\mu\text{L}$  of the oil from the reagent basin into the depressions of the microbatch plate.
3. Using the single-channel pipet, dispense the protein solution into the oil-filled wells. The protein will sink because it is heavier than the oil.



**Fig. 3.** Microbatch in 96-well plates. **(A)** Transfer of oil from the reagent basin to the Imp@ct™ microbatch plate with a multichannel pipet. The first three rows at the bottom of the photograph show coalesced droplets of the protein and screening solutions under the oil. **(B)** Storage of the plate in a plastic sandwich box with a wetted sponge towel.

4. Select or prepare a 96-well storage block of screening solutions.
5. With the smaller volume multichannel pipet (0.2–10  $\mu\text{L}$ ), transfer 1  $\mu\text{L}$  of each of the 96 solutions from the storage block to the microbatch dish. The droplet of screening reagent will also eventually sink to the bottom of the oil-filled well and coalesce with the protein droplet. To speed up the coalescence, centrifuge the plates for 5 min at 300g (optional).
6. Place the microbatch experiment inside a plastic sandwich or freezer box, e.g., Tupperware™ with some wetted paper towels at the bottom. This will supplement the effect of the oil layer in preventing drying-out of the experiment. Close the box with its close-fitting plastic lid (**Fig. 3B**).

### 3.4. Examination of the Crystallization Drop Results

1. Examine the drops under a stereomicroscope immediately after setup, the next day, then after 1 wk.
2. If the overwhelming majority of the drops in the screen have remained clear after 1 wk, repeat the screen with at least double the protein concentration.
3. If more than 80% of the drops have precipitated after 1 wk, the protein concentration may be too high. Reduce the protein concentration by half and repeat the screen (*see Note 12*).
4. Give each drop a score (*see Table 2*). Look for any trends that emerge with respect to the mother liquor components, e.g., pH or precipitant type.
5. Optimize the promising conditions. However, first verify that any crystals or crystalline precipitates are indeed protein and not salt (*see Table 3*).
6. Continue to monitor the experiment on a weekly basis for 1 mo and thereafter monthly or until the experiment dries out.

### 3.5. Optimization

Optimization can be carried out in many ways, e.g., by:

1. Fine-tuning the crystallization conditions found by the initial screen. This can include adjustments or changes in the pH, temperature, or protein/precipitant concentrations.
2. The introduction of low-molecular weight additives, substrates, or ligands.
3. The use of macro- or microseeding.
4. Varying the kinetics of the experiment.
5. Combinations of **steps 1–4**.

#### 3.5.1. Optimization Protocol With Grid Screens: Three Examples

Because optimization is a different kind of sampling problem than initial screening, the search design is also different. There are many sophisticated experimental designs for efficient optimization, e.g., the orthogonal array, Box-Behnken, central composite, and Hardin–Sloane approaches. However, a systematic grid screen of two parameters is the simplest way to begin optimization because the design is intuitively obvious.

**Table 2**  
**Scoring System for Results in Crystallization<sup>a</sup>**

Score	Description
0	Clear drop
1.	Denatured protein. Sometimes accompanied by formation of a skin over the drop surface.
2.	Heavy amorphous precipitate. Most of the protein in the drop has precipitated and the drop is full of grayish, grainy, sand-like particles.
3.	Light amorphous precipitate.
4.	Sporadic precipitation. Nonamorphous.
5.	Crystalline precipitate. Glassy, scintillating, more transparent than amorphous precipitate. Crystalline precipitates will exhibit birefringence under polarized light whereas amorphous ones do not.
6.	Spherulites. Round and chunky, semi-ordered. Sometimes needle-shaped crystals begin to grow out from these.
7.	Oils. This a semi-liquid phase of highly concentrated protein. Oils can be distinguished from spherulites by probing them with a needle.
8.	Gels. Gelatinous protein appears as irregular, transparent patches.
9.	Phase separation. Often appears as hundreds of small droplets or a mixture of droplets and a few big ones.
10.	Crystals: sea-urchins or needle clusters. Thin needle-like crystals originating from a single site or sometimes from spherulites.
11.	Crystals: needles growing singly.
12.	Crystals: plates. Thin, two-dimensional.
13.	Crystals: three-dimensional but growing on top of each other or in inseparable clusters.
14.	Crystals: single, three-dimensional, distinct edges. Check diffraction quality: do not assume that there is any relationship between optical appearance and internal crystalline order.

<sup>a</sup>Pictures of the different phenomena can be found at <http://xray.bmc.uu.se/terese/crystallization/library.html>.

The protocols listed next will use the following hypothetical example as the case to be optimized. The initial screen performed at 20°C has produced hundreds of promising, but too small, needle-shaped crystals. The reservoir consisted of 20% polyethylene glycol (PEG) 4000, 0.2 M unbuffered sodium acetate, and 0.1 M Tris-HCl buffer, pH 8.0. The protein concentration was 20 mg/mL and 1 μL of it was mixed with 1 μL of the reservoir solution in a sitting drop, then equilibrated against the reservoir in a vapor-diffusion setup.

1. Perform at least one of the following setups at a temperature different from that used in the initial screen, e.g., instead of 20°C try 4 or 8°C.
2. The protocols are described for vapor-diffusion setups; microbatch can also be used.

**Table 3**  
**Salt or Protein? Methods for Testing the Crystals<sup>a</sup>**

Method	Description
X-ray diffraction	This is the definitive method for determining if a crystal is salt or protein. If the crystal can be mounted, check the diffraction pattern on an in-house X-ray source. The answer will be immediately obvious because the diffraction spots for salt are at high resolution and far apart, whereas they are close together for macromolecular crystals. As a result, collect a 3–5° oscillation picture with the detector close to the crystal.
Snap test	With the drop at low magnification under the microscope, use an acupuncture needle or other sharp object to break the crystal. Salt crystals are extremely hard and can even be heard to snap upon breaking. Most protein crystals crumble and smash easily upon probing them.
IZIT™ dye	Hampton Research sells a proprietary blue dye, IZIT, which binds to most proteins. The dye will concentrate in the crystals if they are protein and turn them blue, although there are reports of “false-negatives,” i.e., protein crystals that failed to turn blue. The IZIT dye itself can also crystallize, usually appearing as extremely long thin needles that form within minutes. Instructions for using the dye accompany the product.
Run an SDS gel on the crystals.	<ol style="list-style-type: none"> <li>1. Collect the crystals in an excess of mother liquor (e.g., 50 <math>\mu\text{L}</math>) and place them in a microcentrifuge tube. A single large crystal (0.5 <math>\text{mm}^3</math>) usually contains enough protein to be visible on a SDS gel developed with silver staining. Otherwise, pool together as many small crystals as possible.</li> <li>2. Centrifuge 5 min at 16,000g to pellet the crystal(s).</li> <li>3. Remove the supernatant (save it for the gel) and resuspend the crystals in 50 <math>\mu\text{L}</math> fresh mother liquor; centrifuge again. Repeat twice. The washing steps are necessary to remove traces of protein still in solution, i.e., not incorporated into the crystals. At the same time, it is important that the mother liquor itself does not cause the crystals to dissolve. Check that the crystals are visible at the bottom of the tube after the final wash step.</li> <li>4. Remove the supernatant after the last wash and dissolve the crystals directly in 5–20 <math>\mu\text{L}</math> of SDS gel-loading buffer.</li> <li>5. Run the washing supernatants on the same gel as the crystals along with a sample of the protein solution as a control. If the washing has been done properly, the gel may show some protein in the first washes, but there should be none visible in the final wash.</li> </ol>

<sup>a</sup>Many of the components commonly used for crystallizing protein can also give rise to salt crystals. Therefore no crystal should be celebrated until it is known if it is salt or protein.

## 3.5.1.1. GRID SCREEN 1: PEG AND PROTEIN CONCENTRATIONS

1. Set up three rows of PEG 4000 at concentrations of 10, 15, 20, and 25% in the reservoirs. Include the buffer previously listed (0.1 M Tris-HCl, pH 8.0) and the 0.2 M sodium acetate. Use a reservoir volume appropriate for the plate type, i.e., 100  $\mu$ L in a 96-well plate and 500–1000  $\mu$ L in a 24-well plate.
2. Prepare dilutions of the protein solution at 10, 15, and 20 mg/mL.
3. Make a droplet (1–5  $\mu$ L protein and 1–5  $\mu$ L reservoir solution) as described in the screening section and equilibrate it against its corresponding reservoir.
4. Repeat for the four PEG concentrations at all three protein concentrations to give a total of 12 drops.

## 3.5.1.2. GRID SCREEN 2: pH

1. In this optimization, prepare a reservoir solution consisting of 20% PEG 4000 and 0.2 M sodium acetate but substitute different buffers for the 0.1 M Tris-HCl, pH 8.0. A broad scan of pH using only four points could be made with:
  - a. 0.1 M sodium acetate or citrate buffer, pH 4.5.
  - b. 0.1 M cacodylate or MES, pH 6.0.
  - c. 0.1 M HEPES or Tris-HCl, pH 7.5.
  - d. 0.1 M Bis-Tris propane or glycine-NaOH, pH 9.0.
2. The range of pH values should be selected based on what is already known about the protein's solubility with respect to pH. The pH can be screened more finely, e.g., every 0.5–1.0 pH units, if there is sufficient protein.
3. Material permitting, the grid can be enlarged to screen pH while simultaneously varying the concentrations of protein and PEG.

## 3.5.1.3. GRID SCREEN 3: DIFFERENT PEGS

Substitute other molecular weight PEGs for the PEG 4000. For example, try PEG 1000, monomethyl ether PEG 5000, and PEG 10,000 at four concentrations each, e.g., 5, 10, 15, and 20%.

## 3.5.2. Optimization With an Additive Screen

Small molecules such as chaotropes, cosmotropes, cations, detergents, amphiphiles, polyamines, chelators, linkers, polyamines, and others are often used as additives in crystallization.

1. Select one of the additive/detergent kits from suppliers like Hampton Research and Molecular Dimensions or refer to the literature for suggestions.
2. For the optimization previously mentioned example, prepare 25 mL of 20% PEG 4000, 0.1 M sodium acetate, and 0.1 M Tris-HCl, pH 8.0. Preparing all 25 mL at once will ensure that the reservoirs are identical.
3. Pipet 1 mL into each of the wells in a 24-well tissue culture plate. (There will be 1 mL left over.)
4. Prepare droplets consisting of 2.5  $\mu$ L of the protein and 2.5  $\mu$ L of reservoir solution.

5. Add 0.5  $\mu\text{L}$  of the kit additive to this 5  $\mu\text{L}$  droplet. (It is not necessary to include the additive in the 1-mL reservoir.)
6. As a control, substitute 0.5  $\mu\text{L}$  of buffer for additive.
7. Equilibrate against the reservoir as usual.
8. Rate the results with a simple scoring system:
  - a. Same as the control, the additive has no effect (=).
  - b. Crystals have improved, e.g., they are bigger (+).
  - c. Worse than the control, the crystal morphology has deteriorated, the protein has precipitated, or the drop remains clear (-).
9. Usually one or more additives will cause some improvement to the crystal quality (at least visually, this does not necessarily mean the diffraction quality has improved). However, even additives that have negative effects are of interest because they are obviously perturbing the system. Repeat the experiments for these additives at one-tenth of their original concentrations.
10. Additives can dramatically change the rates of nucleation and growth, so be patient.

### 3.5.3. Protocol for Optimization by Seeding

Nucleation is the formation of the first ordered aggregates of molecules. These ordered aggregates are the templates on which more of the molecules preferentially accumulate, eventually building the crystal. The probability that some molecules will meet and form an ordered aggregate is greater the more there are of them; thus, the higher the level of supersaturation, the more likely it is that nucleation will occur. These high levels of supersaturation, however, tend to lead to the formation of too many crystals.

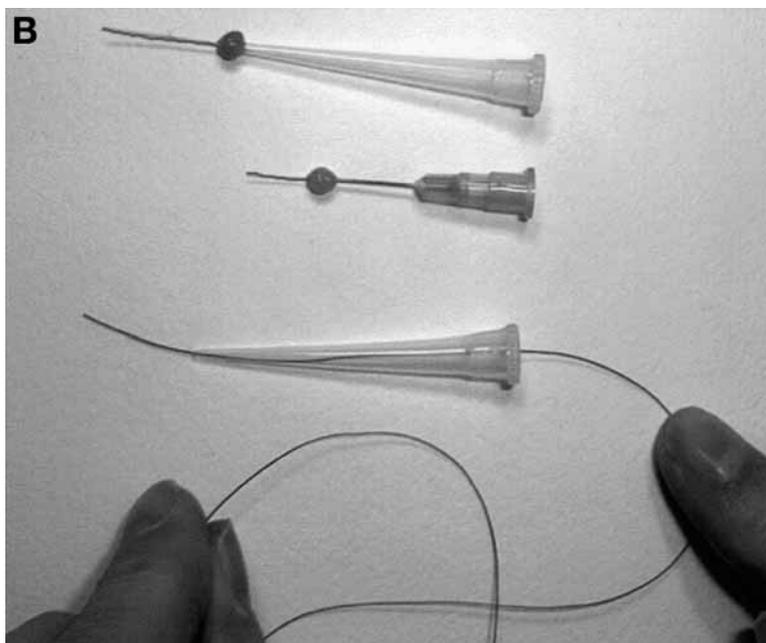
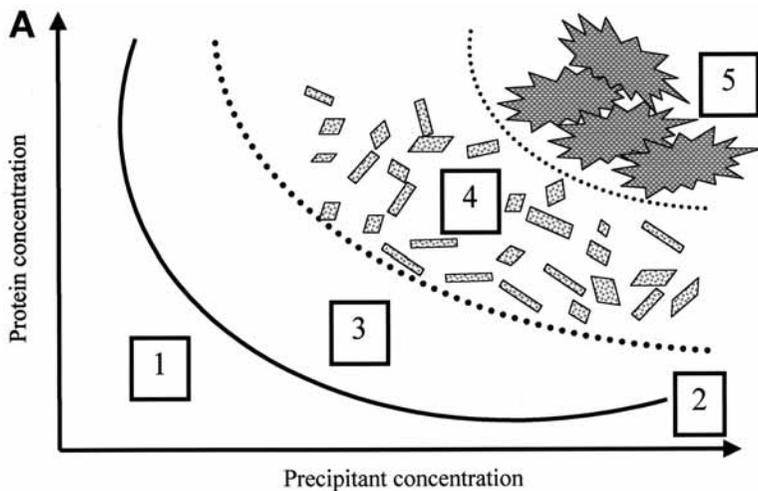
Nonetheless, these small crystals can be used as seeds. They are put into a new experiment set up at a lower level of supersaturation, i.e., the metastable zone in the phase diagram (**Fig. 4A**). This procedure, known as seeding, bypasses the need for spontaneous nucleation because the seed serves as the ready-made nucleus for growth. By limiting the number of seeds that are introduced, ordered growth of just a few, and, therefore, larger, crystals is possible.

The drawback of this method is that a phase diagram for the protein is rarely available. The concentrations of protein and precipitant corresponding to the metastable zone must be determined empirically by lowering one or the other, or both.

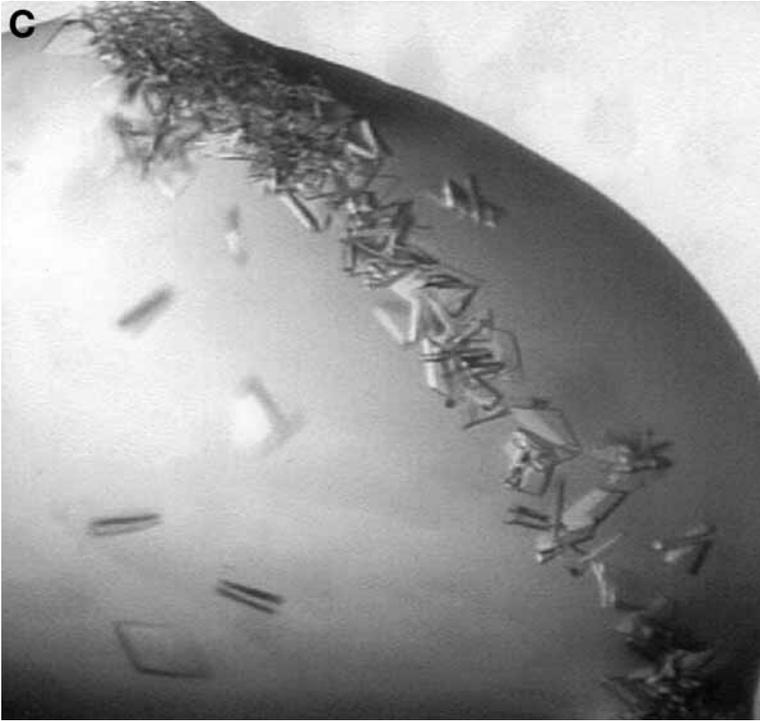
Many seeding protocols for macro- and microseeding exist. One easy and fast variant of microseeding is streak seeding (**4**).

#### 3.5.3.1. PROTOCOL FOR STREAK SEEDING

1. Either microbatch or vapor diffusion can be used (*see Note 13*). The microbatch oil will not affect the seeds. Equilibrate the drops overnight vs the reservoirs if using vapor diffusion.



**Fig. 4.** Seeding. **(A)** Hypothetical phase diagram. (1) Undersaturated zone; no solid phase of the protein is possible. (2) The dark thick line indicates the limit of solubility. (3) Metastable region; the level of supersaturation will support crystal growth but is not high enough to initiate spontaneous nucleation. This is the best region in which to place a seed crystal. (4) Labile zone; spontaneous nucleation occurs in this region and the crystals form. (5) At high levels of supersaturation the protein precipitates. **(B)** Streak-seeding wands made with horsetail hair. The horsetail hair is threaded through a pipet tip or hypodermic needle and fixed into place with a drop of super-glue or wax. A scalpel blade should be used to trim the hair to a working length of 1–2 cm. (*Figure continues*)



**Fig. 4.** (continued) (C) Result of a streak-seeding experiment showing crystals growing along the streak line.

2. With the previous example as the model again, set up a  $3 \times 4$  grid screen as previously described. The reservoirs can be 5, 10, 15, and 20% PEG 4000, 0.1 M Tris-HCl, pH 8.0, and 0.2 M unbuffered sodium acetate. The protein concentration can be 5, 10, and 15 mg/mL. Equilibrate overnight.
3. Make a seeding wand with horsetail hair or an animal whisker, e.g., from a cat or rabbit (**Fig. 4B**).
4. To pick up the seeds for transfer to the new drops made in **step 2**, touch or scratch the surface of the parent crystal with the seeding wand (*see Note 14*).
5. Draw a line with the seeding wand through all the new drops in one row. There are enough microcrystals trapped on the wand to inoculate four to six drops without retouching the parent crystal to pick up more.
6. Wait 2–7 d and then examine the drops. In conditions that are supersaturated, the seed crystals will grow along the streak line (**Fig. 4C**). Drops that remain clear indicate that the conditions are undersaturated, which has caused the seeds to dissolve. Try to find the concentrations where the protein appears to be only slightly supersaturated, i.e., no spontaneous nucleation occurs but inserted seeds grow.

### 3.5.3.2. PROTOCOL FOR SEEDING WITH A DILUTION SERIES OF SEED STOCK

Further refinements can be introduced to provide better control over the number of seed nuclei that are transferred.

1. Use the streak-seeding protocol in **Subheading 3.5.3.1.** to find the concentrations of protein and precipitant that correspond to the metastable zone of the phase diagram (**Fig. 4A**).
2. Set up a row of five identical drops within this metastable zone. Equilibrate overnight if using the vapor-diffusion method.
3. Crystals grown in the previous streak-seeding experiments or from the initial screen can be used as seeds. Crush or pulverize the crystal(s) in the parent drop with a needle, tissue homogenizer, glass rod, Seed Bead, or similar tool.
4. Flood the drop of crushed crystals with 50–100  $\mu\text{L}$  of mother liquor to make recovery of the seeds easier.
5. Transfer the seeds to a microcentrifuge tube. Vortex to ensure an even dispersion of the seeds, then immediately take 10  $\mu\text{L}$  of seed stock and add it to a new tube containing 90  $\mu\text{L}$  mother liquor (= a 10X dilution). Repeat in this manner to make the series of five dilutions (i.e.,  $10^2$ ,  $10^3$ ,  $10^4$ , and  $10^5$ ) of the seed stock.
6. From the first of these tubes (the 10X dilution), remove a 0.1- to 0.5- $\mu\text{L}$  aliquot and add it to one of the new drops made in **step 2**. Repeat with the second drop for the second (100X) dilution and so on. Use a fresh pipet tip for each drop (*see Note 15*).
7. Save the seed dilutions at the temperature at which they grew. Wait 2–7 d, then examine the drops to find which dilution gave the optimum number of seeds.
8. Having once determined the optimum dilution, the same seed stock can be used many times (*see Note 16*).

### 3.5.4. Optimization by Varying the Kinetics of the Experiment

The rate at which supersaturation occurs can greatly affect the outcome of the crystallization experiment and even determine if crystals appear or not. There are many ways of manipulating the kinetics and some possibilities are given here:

1. Vary the mixing ratio of protein and reservoir solution in the droplets. Instead of 1  $\mu\text{L}$  protein and 1  $\mu\text{L}$  of the reservoir solution, try ratios of 3:1 and 1:3.
2. Increase the size of the droplet. Try 10  $\mu\text{L}$  protein plus 10  $\mu\text{L}$  reservoir instead of 1  $\mu\text{L}$  plus 1  $\mu\text{L}$ .
3. If the initial screen used hanging drops, try sitting drops and vice versa.
4. Try a 96-well plate instead of a 24-well plate, or vice versa.
5. Change the method, e.g., use microbatch instead of vapor diffusion and vice versa (*see Note 17*).
6. To slow down vapor-diffusion experiments in 24-well plates, cover the 1000- $\mu\text{L}$  solution in the reservoir with 100–500  $\mu\text{L}$  of paraffin oil (**5**).
7. To speed up microbatch experiments, exchange the 100% paraffin oil for a mixture of 50% paraffin:50% silicone oil (**6**).

### 3.6. What to Do if There are No Crystals to Optimize or the Optimization is Unsuccessful

1. Recheck the drops in the original screen. Crystals have been known to appear after many months. Have a colleague also check the drops; they may see things that have been missed.
2. Is the protein concentration set high enough? (*See Note 18.*)
3. Is the protein pure enough? Consider a further purification step or a different fraction from the current purification. Rescreen.
4. Check the stability of the protein. Is it degrading during the course of the crystallization trials? Run a gel on the protein from the drop and compare it with the archived material.
5. If there are no crystals, look for the next best results, e.g., spherulites, crystalline precipitates, or phase separation, and optimize around these conditions.
6. Add a ligand, substrate, or cofactor to the protein and rescreen. Binding of a ligand, etc., can completely change the conformation of the molecule, making it more amenable to crystallization.
7. Vary the type of search strategy for initial screening. Try a design other than the sparse matrix (**Table 1**).
8. Consider modifying the molecule if extensive screening is unsuccessful. Examples of possible modifications are removing or moving affinity tags (e.g., from the N- to the C-terminus), making truncations, chemically or genetically modifying residues, deglycosylating the protein, or expressing it in a different system.
9. Some researchers report that dynamic light scattering is a useful diagnostic method for determining if a protein solution is likely to crystallize or not (**7,8**) (*see* Chapter 6).

### 4. Notes

1. Many proteins have been successfully crystallized in phosphate buffer but it will easily give rise to inorganic salt crystals, e.g., calcium phosphate. Citrate is frequently used as a buffer but keep in mind that it chelates metal ions. Cacodylate is another common crystallization buffer, handle it with caution because it is an arsenic compound.
2. Glycerol (try 10–30%) will increase protein solubility. If hydrophobic interactions between molecules are the cause of aggregation, neutral detergents can improve the solvation properties.  $\beta$ -octyl glucoside (0.25–0.5%) and CHAPS (0.1–0.3%) are frequently used.
3. Hampton Research sells a product called PCT<sup>TM</sup>, Pre-Crystallization Test, for finding an appropriate protein concentration for use in their Crystal Screen.
4. Microcentrifuge tube filters, such as Whatman's Anopore, have an advantage over syringe filters because there is no hold-up loss. Anopore is an inert substance with low protein-binding properties. Nonetheless the careful researcher will confirm the protein concentration again after filtration. Sodium azide can be used to prevent bacterial growth but beware of this compound's toxicity. Moreover, it is a ligand for some proteins and may later appear in electron-density maps.

5. Volumes larger or smaller than 1  $\mu\text{L}$  can also be used.
6. Stirring or mixing of the droplet will increase any nucleation that occurs. This effect may or may not be desirable, depending on whether the protein nucleates too easily (giving rise to too many crystals) or hardly at all (drops remain clear).
7. The 24-well plates can be converted for sitting-drop setups by the insertion of commercially available plastic bridges that are placed into the reservoirs. There are also 24-well plates specifically made for sitting drops, e.g., Cryschem™ and Q Plate.
8. Protein wets glass and spreads all over the cover slip, especially if detergents are present. To avoid this, cover slips should be made of plastic or silanized if made of glass. Already silanized cover slips can be bought from the suppliers given in [Table 1](#) or they can be silanized in-house with a silanizing solution, e.g., Repel-Silane (Amersham Biotech) or AquaSil (Hampton Research).
9. For hanging-drop vapor-diffusion experiments at 4°C, place the crystallization plates inside a Styrofoam box. This is to minimize local fluctuations in temperature that can give rise to condensation on the cover slips. Sitting drops and microbatch are less vulnerable to condensation problems.
10. Vapor diffusion and microbatch create supersaturation of the protein in quite different ways. In a comparison of the two methods on six proteins, 30% of the successful conditions were unique for each technique, respectively ([9](#)).
11. Paraffin oil is a highly effective barrier, permitting little evaporation. The experiment will not dry out for at least 1 mo. Diluting the paraffin with silicone oil will greatly increase the rate of evaporation. This can be an advantage during screening because results will be obtained more rapidly ([10](#)). However, the experiments dry out in a matter of days. A 50:50 mixture of paraffin:silicone oil is commercially available (Al's oil, Hampton Research) or can be made in-house. 100% silicone oil will lead to dryness overnight in the Imp@ct plates and therefore is not suitable in this low-volume type of microbatch plate.
12. To determine if the precipitated protein is denatured or simply too highly supersaturated, a simple test can be performed. Try to redissolve the precipitate by flooding the drop with mother liquor from the reservoir, buffer, or just water. Denatured protein will not redissolve.
13. A plate designed for seeding into hanging drops is manufactured by Nextal ([www.nextalbiotech.com](http://www.nextalbiotech.com)). It consists of greaseless crystallization supports that can easily be opened and resealed again.
14. The same seeding wand can be reused a few (5–10) times.
15. The seeds can be transferred by streak seeding instead of as aliquots. Wipe the seeding wand clean between each drop with a Kimwipe.
16. The seeds will settle rapidly to the bottom of the tube. It is therefore important to vortex the tube immediately before use. Under the microscope, check an aliquot of the seed stock to confirm that the seeds are still present. Seeds can dissolve because of temperature changes or bacterial contamination of the storage solution.
17. Changing the method from vapor diffusion to microbatch, or vice versa, will have a greater effect on the outcome than variations of the same method, like substituting sitting drops for hanging drops ([11](#)).

18. Although there are proteins that have been crystallized at only 1 mg/mL, in general, poorly soluble proteins are poor candidates for crystallization. Nucleation is favored by high levels of supersaturation, which in turn requires high protein concentrations. If simpler methods (*see Note 2*) do not improve solubility, mutations can be introduced. One of many examples where this has worked is HIV-1 integrase (*12*). Substitution of a single amino acid made it possible to reach a concentration of 25 mg/mL and eliminated the detergent needed to keep the wild-type protein soluble.

## Acknowledgments

I thank Joanne Yeh and Matti Nikkola for helpful discussions regarding the chapter and Mark Harris for his help with the photographs.

## References

1. Hennessy, D., Buchanan, B., Subramanian, D., Wilkosz, P. A., and Rosenberg, J. M. (2000) Statistical methods for the objective design of screening procedures for macromolecular crystallization. *Acta Crystallogr. D. Biol. Crystallogr.* **56**, 817–827.
2. Jancarik, J. and Kim, S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409–411.
3. Villaseñor, A., Sha, M., Thana, P., and Browner, M. (2002) Fast drops: a high-throughput approach for setting up protein crystal screens. *BioTechniques* **32**, 184–189.
4. Stura, E. and Wilson, I. (1991) Applications of the streak seeding technique in protein crystallization. *J. Cryst. Growth.* **110**, 270–282.
5. Chayen, N. (1997) A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Cryst.* **30**, 198–202.
6. Chayen, N. (1997) The role of oil in macromolecular crystallization. *Structure* **5**, 1269–1274.
7. Bergfors, T. (1999) Dynamic Light Scattering. In: *Protein Crystallization: Techniques, Strategies, and Tips*, (Bergfors, T., ed.), International University Line, La Jolla, CA, pp. 29–38.
8. Zulauf, M. and D'Arcy, A. (1992) Light scattering of proteins as a criterion for crystallization. *J. Cryst. Growth* **122**, 102–106.
9. Baldock, P., Mills, V., and Shaw Stewart, P. (1996) A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. *J. Crystal Growth* **168**, 170–174.
10. D'Arcy, A., MacSweeney, A., Stihle, M., and Haber, A. (2003) The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 396–399.
11. Rupp, B. (2003) Maximum-likelihood crystallization. *J. Structural Biol.* **142**, 162–169.
12. Dyda, F., Hickman, A., Jenkins, T., Engelman, A., Craigie, R., and Davies, D. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1985.

13. Zeelen, J. (1999) A flexible sparse matrix screen. In: *Protein Crystallization: Techniques, Strategies, and Tips*, (Bergfors, T., ed.), International University Line, La Jolla, CA, pp. 77–90.
14. Haire, L. (1999) Imperial College grid screen. In: *Protein Crystallization: Techniques, Strategies, and Tips*, (Bergfors, T., ed.), International University Line, La Jolla, CA, pp. 125–129.
15. Carter, C. W., Jr. and Carter, C. W. (1979) Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12,219–12,223.



## Improving Marginal Crystals

Charles W. Carter, Jr. and Madeleine Riès-Kautt

### Summary

The physical chemistry of crystal growth can help to identify directions in which to look for improved crystal properties. In this chapter, we summarize how crystal growth depends on parameters that can be controlled experimentally, and relate them to the tools available for optimizing a particular crystal form for crystal shape, volume, and diffraction quality. Our purpose is to sketch the conceptual basis of optimization and to provide sample protocols derived from those foundations. We hope to assist even those who chose not to use systematic methods by enabling them to carry out rudimentary optimization searches armed with a better understanding of how the underlying physical chemistry operates.

**Key Words:** Protein crystallization; optimization; crystallogensis; solubility; supersaturation; nucleation; crystal growth; Hofmeister series; anions; cations; salts; polymers; response surface; stationary points.

### 1. Introduction

During the search for crystals of a new protein suitable for structural studies, one will likely wish to improve on initial results. In general, sparse searches sample different regions of the experimental crystallization space, and it is statistically unlikely that any sampling points will correspond precisely to optimal combinations of all variables. Many try to improve “hits” from an initial screen simply by sampling more finely a range surrounding the original conditions. We will not describe such straightforward protocols, but rather focus on alternatives for situations where they fail and attempt to exploit what is known of crystallogensis.

Crystals may produce less than the desired diffraction effects in one or both of two major respects: size and/or internal order. Diffracted intensity varies approximately with the square of the number of unit cells present in the X-ray beam, ( $I$ ) and, hence, with the crystal thickness in the X-ray beam,

as well as relative molecular alignment. We first consider these two diffraction determinants.

### 1.1. Crystal Volume

An individual crystal volume,  $V_x$ , is proportional to its mass, whose average value is roughly equal to the mass of supersaturated protein in the drop divided by the number of crystals,  $n_x$ . The supply or *reserve* of supersaturated protein,  $R$ , is determined by the product,  $(C - C_{sol}) * v_d$ , of the drop volume  $v_d$  and the difference between the initial protein concentration  $C$  and the solubility  $C_{sol}$ ;  $n_x$ , the number of seeds in the crystallizing mixture is proportional to the nucleation rate,  $J$ . Both  $R$  and  $J$  can be controlled experimentally and either systematically or randomly optimized, resulting in crystals of more optimal volume.

### 1.2. Diffraction Quality

Both local and long-range disorder in a crystal reduce diffraction quality (2,3). Disorder results from a variety of sources. Optimization of disordered crystals is, therefore, a much more difficult process that is exacerbated whenever the source of impurity is unknown. Nevertheless, an understanding of the various sources of disorder can suggest ways toward improvement. In some cases, growth kinetics may be involved and modification of the rates of crystal growth may lead to improved diffraction. Growth kinetics can be modified in several distinct ways, including the use of gels (4), variation of the rates of equilibration (5), and in some, but not all cases, microgravity (6).

## 2. Materials

1. Protein purification equipment.
2. Polyacrylamide gel electrophoresis equipment.
3. Ultraviolet/Vis spectrophotometer.
4. Tabletop centrifuge.
5. Bradford assay (e.g., Bio-Rad, Hercules, CA).
6. Crystallization plates (Hampton Research, Aliso Viejo, CA).
7. Microscope.

## 3. Methods

### 3.1. Physical Chemistry of Crystallogenesis

Different experimental variables alter the production of crystals in ways that can be understood in terms of the physicochemical properties summarized in Table 1. Few of these effects are linearly related to a single factor, and most are significantly interrelated, so it is important to understand how nucleation and growth depend on the important experimental variables, in order to decide how best to adjust them.

As the initial concentrations of macromolecule and crystallizing agents are always known, the unknowns in **Table 1** are the macromolecular solubility,  $C_{sol}$ , the solubility constants,  $Ks_i$ , and the number of molecules in the critical nucleus,  $n$ . These are discussed in more detail in the next sections with a view toward making useful approximations for experimental design.

### 3.1.1. Solubility and Supersaturation

The central notion about which all of crystallogenesis revolves is that of the supersaturation *ratio*,  $S = C/C_{sol}$  or the actual macromolecule concentration divided by the solubility. Solubility is defined as the concentration of soluble protein in equilibrium with the crystalline form, at given temperature and pH values, and in the presence of a given concentration of solvent compounds (i.e., water, buffer, crystallizing agents, stabilizers, additives). It corresponds to the equilibrium after crystal growth ceases: additional crystalline protein does not dissolve, but adding reservoir solution without protein leads to the dissolution of the protein crystals. Crystallization invariably involves using salts, polymers, or pH and/or temperature changes to reduce solubility, bringing the protein solution to a supersaturated state.

Much protein and time are required to accurately define equilibrium solubility values. However, it is very helpful to estimate the residual protein concentration equilibrated with crystals, at least within an order of magnitude. This can be done as follows:

1. Withdraw 1–2  $\mu\text{L}$  or the whole sample from a crystallization drop, where crystals have remained unchanged for at least 1 or 2 wk. Centrifuge and take an accurately measured volume (0.5–1.5  $\mu\text{L}$ ) of the clear supernatant.
2. Determine the residual protein concentration in the supernatant by an OD measurement at 280 nm or the Bradford method.

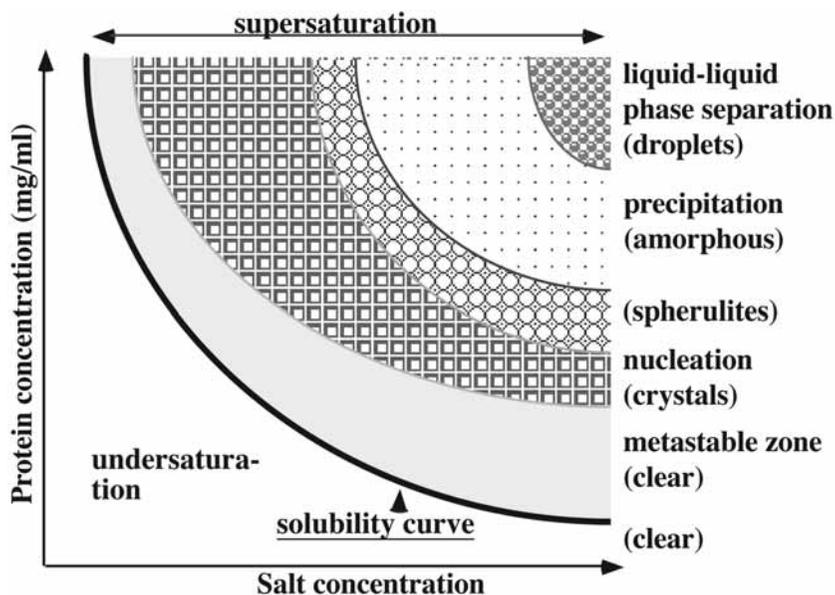
Measuring the residual protein concentration gives:

1. An estimate of  $C_{sol}$ . Growing few and large crystals is difficult to achieve under conditions of  $C_{sol} \leq 1 \text{ mg/mL}$ .
2. The approximate supersaturation,  $S$ , of the starting conditions.
3. The supply of supersaturated protein,  $R$ . For a given  $S$ , the higher the solubility, the more protein is available to feed crystals.
4. Measurements at *different values* of a given variable, e.g., ionic strength, temperature, pH, and others, enables one to estimate the *slope*,  $Ks$ , and intercept,  $A$ , of the  $\ln(C_{sol})$  curve. These values help for extrapolating to the nucleation zone at lower or higher values of the variable depending on how steep the solubility variation is.
5. A guide for the preparation of seeding experiments.
6. The amount of protein needed in solutions used for transferring and mounting crystals (see **Note 1**).

**Table 1**  
**Principal Experimental Determinants of Crystallogensis**

Experimental variable	Principal downstream effects	Functional dependence
C (macromolecule); μ chemical potential	Supersaturation ratio association/dissociation equilibria	(see below) $\mu = RT \cdot \ln C$
C <sub>ag</sub> (crystallizing agent);	Solubility; ionic strength, I	$I = (1/2) (\sum_i [\text{ion}]_i \times Z_i^2)_i$
Σ effective total concentration <sup>a</sup> ; K <sub>s</sub> empirical solubility constant	Total supersaturation induced by all agents	$\Sigma = \sum_{\text{agents}, i} K_{S_i} \cdot C_{\text{ag}}$
C <sub>sol</sub> solubility of the protein	Supersaturation ratio	$C_{\text{sol}} = A \exp(-\Sigma)$ (7); $A = C_{\Sigma \rightarrow 0}$
S supersaturation ratio	(Homogeneous) nucleation rate	$S = C/C_{\text{sol}}$
J nucleation rate	Number of crystals produced	$J = \text{const} \cdot S^n$ (8,13) n = the number of molecules in a critical nucleus
M “neutral” moment	Solubility	$M = K_{S_{\text{PEG}}} / \Sigma$
R macromolecular supply	Mass of supersaturated protein (i.e., available to feed the crystals)	$R = (C - C_{\text{sol}}) \cdot v_d$
V <sub>x</sub> Crystal volume; v <sub>d</sub> = volume of drop or dialysis cell	Irradiated volume; integrated intensity	$V_x \propto R/J \propto (C - C_{\text{sol}}) \cdot v_d (C/C_{\text{sol}})^n$

<sup>a</sup>Effective total concentration is a term introduced to approximate the total effects on solubility of all crystallizing reagents.



**Fig. 1.** Schematic phase diagram showing the different zones and types of solid phase. The different types of solid phase shown here are rarely encountered in one single set of conditions: the purpose here is to show where each type of solid is located in relation to the others.

### 3.1.2. The Solubility Diagram

The solubility curve (**Fig. 1**) delineates the boundary between under- and supersaturated solutions as a function of the parameter(s) used to change the solubility. Below the solubility curve the solution is *undersaturated* and the biological macromolecule will not crystallize. A solution above the solubility curve is said to be *supersaturated*. The initial macromolecular concentration must be greater than the equilibrium concentration, ( $S > 1.0$ ), in order for nucleation to occur. If the supersaturation ratio is only marginally greater than 1.0, nuclei do not form, but crystal growth of a seed introduced in the solution is possible.

The two axes of solubility diagrams represent the experimental dimensions within which optimization of crystal volume must be sought. The y-axis, macromolecule concentration, is a simple matter to control, as it is determined by a single variable. The x-axis, however, is less obvious in general. For a protein that is crystallized by increasing salt concentration, pH, or temperature, this axis can also represent a single variable. Complications arise, however, when multiple parameters are changed simultaneously or when several crystallizing agents are used. Other factors, notably the net charge and, hence, the pH relative to the isoelectric point, pI, bring additional dimensionality to the solubility

diagram. Even for well-characterized systems like hen egg white (HEW) lysozyme, multiple regimes occur within which the solubility decreases exponentially at different rates with increasing salt concentration and with different lysozyme net charges (9). Integral membrane proteins represent a further example: their crystallization typically requires optimal concentrations of salts to saturate the soluble domains and polymeric volume exclusion to saturate the micellar portions (see Chapter 10). In such cases, the solubility curve becomes a multidimensional *surface*, none of whose parameters are known, *a priori*.

A final complication in making explicit use of solubility diagrams arises because its precise shape is almost never determined in practice. Thus, it remains a conceptual heuristic rather than an analytical tool. Nonetheless, it is a centrally useful heuristic.

### 3.1.3. Nucleation

Crystals originate from small aggregates called nuclei, which form spontaneously (*homogeneous* nucleation) at a rate,  $J$ , which rises increasingly rapidly as  $S$  increases.  $J$  can be altered either by changing  $C$  or  $C_{sol}$ . *Heterogeneous* nucleation resulting from dust or other foreign particles, including seeds, can be intentionally introduced to overcome the nucleation barrier. Heterogeneous nucleation and seeding are far less reproducible than homogeneous nucleation, but have been used successfully to increase crystal volume and, occasionally, the success rate for crystallization itself (10).

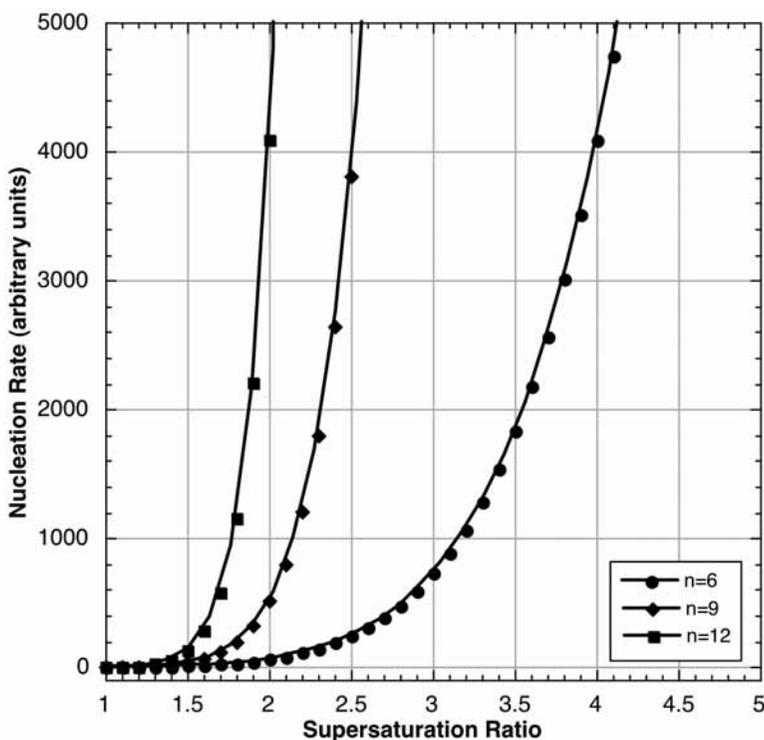
An important problem in optimization is to find a combination of experimental factors that produces a small number of nuclei within an appropriate timeframe, i.e., reducing  $J$ . Selecting experiments in the neighborhood of a “hit” must ensure that  $J$  is sampled at approximately equally spaced multiples. An effective way to do this is to make explicit use of the power law dependence of  $J$  on  $S$ . In turn, this relates  $J$  to  $C$  via the chemical potential of the macromolecule,  $\mu = \ln C$ , and  $C_{ag}$ , given approximately by  $\Sigma$  (Table 1).

For modest values of  $C_{sol}$ ,  $J$  is, to a good approximation, proportional to  $S$ , raised to a power,  $n$ , approximately equal to the number of molecules in the least stable aggregate along the crystallization path (11)

$$J = \text{const} * S^n \quad (1)$$

**Equation 1** can be understood from the fact that nucleation is a stochastic process, linked to the probability of  $n$  molecules joining to form a nucleus.  $J$  is, approximately, the reciprocal of the time taken to produce the first crystals, and can be determined by following the time course of screening experiments. The slope of  $\log(J)$  vs  $\log[C]$  at constant  $C_{sol}$  provides an experimental measurement of  $n$  (12,13).

The inverse dependence of  $J = (C/C_{sol})^n$  on  $C_{sol}$ , has important practical consequences. First, supersaturation ratios required to crystallize are lower at higher



**Fig. 2.** Power-law dependence of the nucleation rate (and by implication crystal growth rate) on the supersaturation. Examples are shown for crystallization with critical nuclei sizes of 6 (similar to lysozyme, for which  $n = 5$ ), 9, and 12 monomers.

$C_{sol}$ , i.e., at low  $C_{ag}$ . The observed nucleation of very high lysozyme concentrations at high solubility (9,14) suggests that the nucleation barrier decreases at very high  $C_{sol}$ . On the other hand, at high  $C_{ag}$ , the low  $C_{sol}$  implies that  $J$  will increase faster with both  $C$  and  $C_{ag}$ , making it difficult to strike an appropriate balance. **Figure 2** illustrates this concept, showing the narrowing that occurs between the precipitation zone and the solubility at low solubility. Moreover, the power law dependence also narrows the range of useful  $S$  for increasing critical nucleus sizes (**Fig. 1**).

The excess of macromolecule in solution,  $R = (C - C_{sol}) \cdot v_d$ , will eventually appear as a solid phase. Thus, in addition to the notion of supersaturation, the amount of supersaturated solute also needs to be considered. At low  $S$  only a small protein supply can nourish a seed, and this effect is much more pronounced at low  $C_{sol}$ . Taking an example of nucleation occurring at  $S$  of 5, there will be 4  $\mu\text{g}/\mu\text{L}$  (i.e.,  $5 - 1$ ) protein available if  $C_{sol} = 1 \text{ mg/mL}$ , but as few as 0.4  $\mu\text{g}/\mu\text{L}$  (i.e.,  $0.5 - 0.1$ ) if  $C_{sol} = 0.1 \text{ mg/mL}$ . For these reasons, a one-dimen-

sional search normal to the solubility isoclines, first, produces a small number of small crystals, then a small number of larger crystals, then a larger number of smaller crystals, then a very large number of microcrystals, spherulites, precipitates, and, finally, liquid–liquid phase separation or “oils” (Fig. 1).

### 3.1.4. Effectiveness of Different Parameters on Solubility

Crystallization parameters that can most easily be changed to increase or decrease protein solubility are:

1. Temperature.
2. Protein net charge, via the pH of the crystallizing solution.
3.  $C_{ag}$  (i.e., ionic strength and neutral polymers like polyethylene glycols [PEG]), represented approximately by  $\Sigma$ .
4. The nature of the crystallizing agent and the buffer, when they act as counterions and coions and/or as ligands of the protein.
5. The dielectric constant, i.e., addition of organic solvents or heavy water.

#### 3.1.4.1. TEMPERATURE

As is well known for small molecules, solubility changes with temperature. For proteins, opposite behaviors have been observed;  $C_{sol}$  can either increase or decrease with increasing temperature. This direct, respectively inverse, solubility behavior with temperature, which indicates the sign of the enthalpy change on crystallization,  $\Delta H_{cryst}$ , is characteristic neither of a given protein, nor of a crystallizing agent, but is a coupled effect of both. Indeed bovine trypsin inhibitor (BPTI) solubility increases with temperature when it is crystallized with KSCN, but decreases with temperature in the presence of NaCl or ammonium sulfate (15).

Moreover, the sensitivity of solubility toward temperature changes depending on the solubility value itself:

$$dC_{sol}/dt = -(\Delta H_{cryst}/RT_0^2) C_{sol}(t) \quad (2)$$

where  $R$  is the molar gas constant,  $t$  the temperature in degree Celsius, and  $T_0$  is 273 K. **Equation 2** shows that the higher the solubility  $C_{sol}$ , the larger the magnitude of  $dC_{sol}/dt$ , when  $\Delta H_{cryst}$  is independent of the salt concentration (16) (see **Note 2**).

#### 3.1.4.2. pH

pH changes modify the protein net charge:

1.  $C_{sol}$  varies most rapidly near the pKa values of the most abundantly charged residues (i.e., at pH 4.5 for Asp and Glu, 6.2 for His, 9.3 for free Cys, 9.5 for Tyr, 10.4 for Lys, and 12.0 for Arg); outside the range of these values, the solubility changes smoothly.
2. Solubility is *minimal* at the *pI* of the protein. Conversely solubility increases, when the net charge increases.

These effects can be exploited by sampling net charge intervals rather than pH values. It is advisable to screen conditions close to the pI and on each side of it for a *specific* protein net charge of about 0.5–1.0 charge/kDa, provided the protein is stable within the screened pH range. Typically this means screening a positively charged protein at pH values about 4.5 and negatively charged ones greater than pH 8.5.

### 3.1.4.3. $\Sigma$ ; IONIC STRENGTH AND POLYMER CONCENTRATIONS

Within limits,  $C_{sol}$  decreases exponentially with  $C_{ag}$ , according to  $C_{sol} = A \exp(-\Sigma)$  (17). This behavior is observed for both types of crystallizing agents: neutral polymers, like PEG (18), and for salts (19,20). Hence, the logarithm of  $C_{sol}$  is related linearly to  $C_{ag}$ , both in salt- and volume exclusion-induced crystallization, with proportionalities known as “solubility constants,”  $K_s$ . As the individual effects of salts and polymers are both exponential, they should be multiplicative (see Note 3). This is the rationale for the approximation  $\Sigma = \Sigma_i K_{s_i} * C_{agi}$  (Table 1). The solubility coefficients,  $K_s$ , can be estimated from the slopes of  $\ln(C_{sol})$  vs  $\ln C_{ag}$ , and are generally between 2 and 10 for salts (17).

PEG polymers, especially those of higher molecular weight, have considerably larger  $K_s$  values, and because the effect on solubility is almost entirely an excluded volume effect, it is more pronounced for combinations involving high molecular weight macromolecules and high molecular weight PEGs (21). A survey of published solubility constants for different PEG molecular weights and different proteins (17,21) reveals that the variations in  $K_s$  values for different PEGs are almost entirely related to the size of the protein. Molar  $K_s$  values for an unknown problem can be estimated roughly from the following relation (Eq. 3; unpublished), which reproduces 45 published values for 13 different proteins with an  $R^2 = 0.75$ :

$$K_{sPEG}(PEG_{MW}) = (-0.00723 + 0.000824 * M_w^{1/3}) * MW_{PEG} \quad (3)$$

where the cube root of the protein molecular weight is an approximation of the Stokes radius.

### 3.1.4.4. M; THE NEUTRAL MOMENT

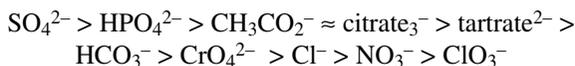
Mixtures of salts and neutral polymers are becoming increasingly useful in producing macromolecular crystals. For this reason, it is useful to be able to approximate the overall reduction in solubility because of the combined effects of the two kinds of crystallizing agents. In the simplest case, the effects of salt and PEG can be considered independent and treated linearly,  $\Sigma = K_{s,salt} C_{salt} + K_{s,PEG} C_{PEG}$ , so solubility can be represented by  $C_{sol} = A * \exp^{-\Sigma}$ . Given the total reduction in solubility, the proportion arising from volume exclusion and/or dielectric effects can be represented by a second parameter,  $\Sigma = K_{s,PEG}[PEG]/\Sigma$ ,

which might be called the “organic” or “neutral” moment (22).  $M$  simply estimates the fraction of  $\Sigma$  owing to added polymer or alcohol. Thus, two empirical parameters,  $\Sigma$  and  $M$ , represent the total reduction in solubility and the proportion of that total resulting from each type of component (see Note 3).

### 3.2. The Hofmeister Series

Solvent constituents play an important role if they interact with a protein. The protein net charge is then changed not only by a pH variation, but also by adsorption of charged species to the macromolecule as it would be by mutation of charged residues. It is important to consider the protein and adsorbed ions as a whole: different salts of a protein exhibit different solubility behaviors. This concerns not only ions of the crystallizing agent but also the buffer or other additives. It is common for a protein to dissolve to higher concentration at the same pH with a given buffer rather than in another one. When the pH is varied, solvent compounds, like buffer, crystallizing agent, or additives may also change their charge.

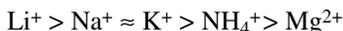
Anions differ greatly in their ability to salt out protein solutions (23), and the order of these effects depends on the net charge of the protein (20,24). For the acidic protein collagenase from *Hypoderma lineatum* (24), with a pI of 4.6 and a negative net charge at the crystallization pH of 7.2, the effectiveness of ions to decrease solubility follows the original Hofmeister series:



For the basic protein HEW lysozyme (pI = 11) crystallized with a positive net charge (pH 4.5), the order of the anions is reversed (19).

When a crystallizing condition has to be optimized, a qualitative approach is the stepwise replacement of the original salt by another among the Hofmeister series in order to increase, or decrease, the solubility depending on what is expected to improve crystal growth. In addition, optimization may imply testing salts of the same chemical class, namely among halides (NaCl, NaBr, NaI, NaF), or carboxylic acids (acetate, citrate, propionate, tartrate, malate) (25,26).

Cations can also modulate solubility. Initially, Hofmeister tested only a few monovalent cations and observed that:



Lysozyme solubility has since been investigated with a larger range of cations (16). Multivalent cations, like, Mn, Co, and Yb, have shown to increase the solubility of the protein, probably by adsorption of co-ions. Addition of 5–10 mM Co, Cu, Ni, Zn, or lanthanides has also shown to be effective (27).

### 3.3. Navigating the Solubility Diagram

A schematic phase diagram, as shown in **Fig. 1**, is a useful tool for understanding qualitatively the major trends of crystallization trial outcomes. However preparing an optimization protocol, as described here, involves numerical values of  $C_{sol}$ ,  $J$ , and  $n$ , which are usually not accessible for a new protein. Similarly, one also needs to know how sensitive nucleation is to the supersaturation ratio (**Eq. 1**), because the solubility diagram varies considerably from protein to protein and, for a given protein, with the net charge and the choice of crystallizing agents.

One way to solve the problem is to use approximations summarized in **Table 1**, which reflects the limited work done to rationalize how the factors form the coordinate system of the solubility diagram for complex crystallizing regimes. An initial insight arose from the optimization of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase (TrpRS) crystals: using the product of the protein and salt concentrations as  $x$ -axis worked better than using only salt concentration. Later, a heuristic connection between the product of the two concentrations and  $J$  suggested that natural coordinates for optimization are the supersaturation ratio,  $S$ , and the supply,  $R$  (**28**).

Moreover, because of the power law dependence of the *nucleation rate* on  $S$  (**Eq. 1**), values of  $\Sigma$  should be chosen that sample approximately equal multiples of  $J$ , centered, if possible, in the region of greatest curvature in **Fig. 1**. Thus, higher  $S$  values should be more closely spaced. If the solubility and  $Ks$  for the relevant system are known, uniform sampling can be achieved as follows (from **Table 1**). Assume that the current conditions represent something close to the best conditions, in which  $J$  is one nucleus/drop and per unit of time. Use  $J = \text{const} * \{C/C_{sol}\}^n = \text{const} * (C/A)^n \exp^{n * \Sigma_0} = 1$  to determine the scaling constant, const. For two different solutions at constant protein concentrations, the ratio of  $J_2/J_1$  will be  $S_2^{*n} / S_1^{*n} = \exp(n[\Sigma_2 - \Sigma_1]) = \exp(n\Delta\Sigma)$ . Equal multiples of the  $J$  are, therefore, achieved by using equal increments of  $\Sigma$ , and in turn by equal increments of  $C_{ag}$ . The actual increments  $\Delta\Sigma$  are given by  $\ln(k)/n$  for  $k$ -fold increases in nucleation rate, i.e.,  $J_2/J_1 = k$ . Interestingly, this development underscores the appropriateness of grid screens to optimize nucleation. As an example, which we have verified in practice for HEW lysozyme ( $n = 5$ ), increasing the nucleation rate twofold requires a  $\Delta\Sigma = \ln(2)/5 = 0.138$ . As  $Ks$  in this case is approx 3.16, this implies increments of 0.038  $M$  NaCl. More generally when  $n$  is unknown and may lie between 5 and 20, several values of  $\Delta\Sigma$  should be tested and finer grid spacings,  $\ln(2)/9$ ,  $\ln(2)/14$ , and so on, may give better results.

### 3.4. Biochemical Purity

Impurities invariably corrupt crystals, so there is little doubt that the single most important way to improve the quality and size of crystals is to ensure that

the starting material is scrupulously homogeneous. Incorporating any nonidentical or misoriented material into the lattice will degrade the final, macroscopic crystal. The disproportionate deterioration caused by molecular fragments and aggregates arises from the fact that their inclusion into the lattice propagates over many unit cells, leading to strain and strain-induced disorder. Impurities can be contamination of the initial starting material by foreign proteins, unintended oligomers of the same protein, alternate conformations of the desired protein, and/or misdirected incorporation of correct macromolecules owing to inappropriate rates of growth (29). Oligomers of the macromolecule to be crystallized often fit well into parts of the lattice, while disrupting continuity between unit cells (30). For similar reasons, proteolytic degradation products of the primary constituent can seriously corrupt crystal order. Furthermore, although the starting material may be of high purity, degradation can occur from trace amounts of proteases over the time-scale of a crystallization experiment (days to weeks). These types of impurities can affect crystal growth at less than a percent of total mass.

Structurally unrelated contaminants can also become trapped into crystal lattices by waves of multiple growth layers. These develop when both diffusion and attachment kinetics influence the rate-limiting steps (6). Moreover the proportion of impurities in the remaining crystallization solution increases as crystals of the molecule of interest grow. Indeed, cessation of growth can result from excessive impurity concentrations at the crystal surface, even though the solution may still be supersaturated.

The gain in crystal quality can be substantial, even once crystallization conditions are known. Although additional chromatographic steps invariably incur losses of material, the prerequisite to those who would improve their crystals is to revisit the question of purity seriously. Any contaminant bands that appear in overloaded Coomassie-stained sodium dodecyl sulfate-polyacrylamide gels of crystals and/or supernatant should encourage further efforts for purification. The following suggestions might prove useful at least for large subsets of problems.

1. Introduce a new step in purification or repeat more selectively a previous step that involved making a too generous cut.
2. Remove higher oligomers of proteins by gel-filtration chromatography. Generally, proteolysis products cannot be removed, because interactions frequently remain strong enough to sustain a native-like molecular weight on the time-scale of the chromatography.
3. Eliminate proteolytic fragments by preventing degradation in the first place. Use appropriate protease inhibitors (which can, in turn, become molecular contaminants in their own right), or genetically modify the material to be purified (*see Note 4.*)
4. Crystallization itself is basically a purification method. Batch crystallization can reduce the presence of contaminants if crystals are already known to grow from a particular salt or combination of crystallizing agents. This can be done either on a

preparative scale using standard crystallization conditions or by trituration of precipitated phases on a chromatography support, like Sepharose 4B, eluting with solutions of decreasing salt or PEG concentrations (see **Note 5**). Selective dissolution of the desired molecule (31,32) from the solid phase in equilibrium with solution on an inert support is capable of quite substantial purification (33).

### 3.5. Optimization Methods

This section focuses on several available crystal optimization methods: response surface modeling, alteration of growth kinetics via microgravity or gel acupuncture, and crystal annealing.

#### 3.5.1. Response Surface Modeling

Mathematical modeling of what are called “response surfaces” (34–36) provides an analytic solution to the problem of locating the optimal combination of multiple variables affecting crystal properties. This quantitative approach has three stages:

1. Design. Systematically change all the experimental conditions simultaneously in a statistical way, using the approximate relationships outlined in **Table 1**, to induce variations in crystal growth.
2. Experiments and scores. Use quantitative metrics, such as crystal dimensions, resolution limits, and/or mosaicity to estimate how the system behaves at the corresponding point in the experimental space.
3. Fitting and testing models. Use mathematical models to examine these estimates together as a group. Modeling provides a way to describe the effects of all experimental variables. Moreover, models provide accurate *interpolation* within the range of experimental variables originally sampled; occasionally a good model may predict behavior *outside* it (34). Quadratic polynomial models are particularly useful local approximations because they can possess “stationary points,” where their gradient vanishes, and which may represent optima (22,37).

Response surface experiments require a nonuniform sampling strategy, highlighting differences between better and worse behaviors. For a single variable,  $x$ , the goal is to distinguish as accurately as possible between a parabola,  $y = ax^2 + bx + c$  and the corresponding linear relation,  $y = x + b$ . Three groups of experimental points have maximal impact on this distinction: those near the suspected maximum value of the parabola and those at upper and lower limits of  $x$ . Averaging over multiple measurements of the three groups of experimental points gives the most accurate and significant quadratic coefficient of  $ax^2$ . A similar strategy applies for response surface models involving more than one variable. Experimental points are also selected near the center of the design and evenly distributed around the *perimeter* of the experimental space. Designs that minimize errors in parameter estimation are owing to Hardin and Sloane, and are called “minimum-prediction variance” designs (38).

## 3.5.1.1. A DESIGN MATRIX

The Hardin–Sloane design matrix in **Table 2** is for determining the stationary point for HEW lysozyme while testing three variables. The first three columns in **Table 2** represent a statistical distribution of experiments, such that the range of each of the three chosen variables,  $V1$ ,  $V2$ , and  $V3$ , in the experiment is within the range  $-1$  to  $1$ . The remaining columns make use of these values, together with the ranges intended for the experiment. Here the three variables are used for  $V1 = \text{pH}$ ,  $V2 = \ln C$ , and  $V3 = J$ .

The starting  $\{0;0;0\}$  condition to be optimized in this example is:  $V1 = \text{pH} = 4.8$ ;  $C = 12.68 \text{ mg/mL}$  (i.e.,  $V2 = \ln C = 2.54$ ), and  $C_{\text{NaCl}} = 0.6 \text{ M}$  (i.e.,  $V3 = J = 1$  nucleus/drop per day). These conditions correspond to those determined by response surface analysis of HEW lysozyme using both dialysis and vapor-diffusion experiments and represent those that might be obtained for a screening “hit.” An Excel spreadsheet in which the calculations are encoded for use in adapting it to new problems is available from [carter@med.unc.edu](mailto:carter@med.unc.edu).

Independently varied factors are the lysozyme concentration,  $C$ , the solution pH, and the salt concentration  $C_{\text{NaCl}}$  and refer to the initial reservoir and drop or dialysis cell (lysozyme) concentrations. The pH centered on 4.8 is aimed to range from 4.1 to 5.5,  $\ln C$  is centered on 2.54 and ranges from 2.39 to 2.69. The salt concentration is determined indirectly from the intended changes in the nucleation rate,  $J$ , using relations for  $S$  and  $\Sigma$  from **Table 1**, which together implicitly define the protein and salt concentrations. The nucleation rate at  $\{0;0;0\}$  is presumed to be 1 on an arbitrary scale (approx 1 nucleus per drop per day) and is varied over a relative range from 0.14 to 7, the former being slower by a factor of approx 7, the latter being faster by the same factor. In other words, the design samples nucleation rates over a 50-fold range centered at the starting  $\{0;0;0\}$  condition. An intermediate in the generation of  $C_{\text{NaCl}}$  is the supersaturation value  $C/A$ , where it is the  $y$ -intercept of the low-salt region of the experimental solubility curve or its extrapolation for the measurement of residual protein concentration (§2.1). Salt concentrations are then calculated from the corresponding values of  $J$ . In this way  $J$ , which is normally an experimental consequence, becomes an experimentally controlled variable. The value of  $A$  at  $14^\circ\text{C}$  is 32, and solubility corrections for 4 and  $21^\circ\text{C}$  were estimated by an algorithm provided by [ref. 39](#).

This design should produce sufficient variation in the numbers and sizes of crystals that optical measurement of their dimensions will allow fitting them to a trivariate quadratic polynomial of the form:

$$\begin{aligned}
 Q = & \beta_{0c+} (\beta_{\text{ph}} * \text{pH}) + (\beta_{\text{lyso}} * C) + (\beta_{\text{NaCl}} * C_{\text{NaCl}}) + (\beta_{\text{pH-lyso}} * \text{pH} * C) \\
 & + (\beta_{\text{pH-NaCl}} * \text{pH} * C_{\text{NaCl}}) + (\beta_{\text{lyso-NaCl}} * C * C_{\text{NaCl}}) + \beta_{\text{pH}}^2 * \text{pH}^2 \\
 & + \beta_{\text{lyso}}^2 * C^2 + \beta_{\text{NaCl}}^2 * C_{\text{NaCl}}^2
 \end{aligned}
 \tag{4}$$

**Table 2**  
**Harden–Sloane Experiment to Optimize Three Variables Associated With HEW Lysozyme Crystal Growth at Three Temperatures<sup>a</sup>**

	<i>V1</i>	<i>V2</i>	<i>V3</i>	pH	<i>C</i>		<i>J</i>	<i>C</i> <sub>NaCl</sub> ( <i>M</i> ) (at 14°C)	<i>C</i> <sub>NaCl</sub> ( <i>M</i> ) (at 4°C)	<i>C</i> <sub>NaCl</sub> ( <i>M</i> ) (at 21°C)
					ln <i>C</i>	(mg/mL)				
1.	-0.0032	-0.0032	-0.0032	<b>4.8</b>	2.54	<b>12.67</b>	0.99	<b>0.600</b>	<b>0.283</b>	<b>0.821</b>
2.	-0.8787	1	1	<b>4.2</b>	2.69	<b>14.73</b>	7.06	<b>0.676</b>	<b>0.360</b>	<b>0.898</b>
3.	0.3937	1	1	<b>5.1</b>	2.69	<b>14.73</b>	7.06	<b>0.676</b>	<b>0.360</b>	<b>0.898</b>
4.	-0.0032	-0.0032	-0.0032	<b>4.8</b>	2.54	<b>12.67</b>	0.99	<b>0.600</b>	<b>0.283</b>	<b>0.821</b>
5.	1	-0.8787	1	<b>5.5</b>	2.41	<b>11.11</b>	7.06	<b>0.765</b>	<b>0.449</b>	<b>0.987</b>
6.	-1	1	0.0528	<b>4.1</b>	2.69	<b>14.73</b>	1.11	<b>0.559</b>	<b>0.243</b>	<b>0.781</b>
7.	0.0082	-1	-1	<b>4.8</b>	2.39	<b>10.91</b>	0.14	<b>0.524</b>	<b>0.207</b>	<b>0.745</b>
8.	1	-1	-1	<b>5.5</b>	2.39	<b>10.91</b>	0.14	<b>0.524</b>	<b>0.207</b>	<b>0.745</b>
9.	-1	1	-1	<b>4.1</b>	2.69	<b>14.73</b>	0.14	<b>0.429</b>	<b>0.112</b>	<b>0.650</b>
10.	1	0.0528	-1	<b>5.5</b>	2.55	<b>12.78</b>	0.14	<b>0.474</b>	<b>0.157</b>	<b>0.695</b>
11.	1	1	0.3937	<b>5.5</b>	2.69	<b>14.73</b>	2.16	<b>0.601</b>	<b>0.285</b>	<b>0.823</b>
12.	-1	-1	1	<b>4.1</b>	2.39	<b>10.91</b>	7.06	<b>0.771</b>	<b>0.455</b>	<b>0.993</b>
13.	1	0.3937	1	<b>5.5</b>	2.60	<b>13.45</b>	7.06	<b>0.705</b>	<b>0.389</b>	<b>0.927</b>
14.	1	-1	0.0528	<b>5.5</b>	2.39	<b>10.91</b>	1.11	<b>0.654</b>	<b>0.338</b>	<b>0.876</b>
15.	1	1	-0.8787	<b>5.5</b>	2.69	<b>14.73</b>	0.18	<b>0.444</b>	<b>0.127</b>	<b>0.665</b>
16.	-0.0032	-0.0032	-0.0032	<b>4.8</b>	2.54	<b>12.67</b>	0.99	<b>0.600</b>	<b>0.283</b>	<b>0.821</b>
17.	-1	-1	0.0082	<b>4.1</b>	2.39	<b>10.91</b>	1.02	<b>0.648</b>	<b>0.332</b>	<b>0.870</b>
18.	-0.0032	-0.0032	-0.0032	<b>4.8</b>	2.54	<b>12.67</b>	0.99	<b>0.600</b>	<b>0.283</b>	<b>0.821</b>
19.	-1	0.0528	1	<b>4.1</b>	2.55	<b>12.78</b>	7.06	<b>0.721</b>	<b>0.405</b>	<b>0.943</b>
20.	-1	0.0082	-1	<b>4.1</b>	2.54	<b>12.70</b>	0.14	<b>0.476</b>	<b>0.159</b>	<b>0.697</b>
21.	-0.0032	-0.0032	-0.0032	<b>4.8</b>	2.54	<b>12.67</b>	0.99	<b>0.600</b>	<b>0.283</b>	<b>0.821</b>
22.	-1	-1	-1	<b>4.1</b>	2.39	<b>10.91</b>	0.14	<b>0.524</b>	<b>0.207</b>	<b>0.745</b>
23.	0.0528	-1	1	<b>4.8</b>	2.39	<b>10.91</b>	7.06	<b>0.771</b>	<b>0.455</b>	<b>0.993</b>
24.	0.0528	1	-1	<b>4.8</b>	2.69	<b>14.73</b>	0.14	<b>0.429</b>	<b>0.112</b>	<b>0.650</b>

Conversion of the variables *V1*, *V2*, and *V3* of the first columns to the corresponding experimental values within the second box is achieved as follows: pH = 4.8 + (*V1*\* 0.7); ln*C* = 2.54 + (*V2*\*0.15), which gives  $C_{\text{lyso}} = \exp(\ln C)$ ;  $J = V3 * (3.2^{1.68} * V3)$ . This converts the range -1 to 1 of *V3* into *J* values of approx 0.14 to approx 7, providing a range of 7/0.14 = 50-fold. *C*<sub>NaCl</sub> is then extracted through  $C_{\text{NaCl}} = \Sigma_0 + \ln(kp)/n + \ln C_0/[C]$ , where the subscript 0 refers to the center of the design. An intermediate in the generation of *C*<sub>NaCl</sub> is the supersaturation value *C/A*, where *A* = 32 mg/mL is the *y*-intercept of the low-salt region of the experimental 14°C solubility curve or its extrapolation for the measurement of residual protein concentration (§2.1). Alternate salt concentrations in the final two columns for 4 and 21°C were generated using temperature dependence corrections from [ref. 39](#).

### 3.5.1.2. SCORING

Scoring must reflect variations in the properties to be optimized. Crystal dimensions can be scored directly by microscopic inspection. At least two of the three spatial dimensions of a crystal can usually be determined using a microscopic reticle. Often the third dimension can be estimated, albeit with less precision. These measurements provide estimates for the sizes and shapes of crystals and constitute the most useful scores for the purpose of optimization (40). Experience suggests that the ratio of the smallest to the largest dimension (width/length) is often the most sensitive score, and leads to better response–surface models than do estimates of crystal volume.

As a first step of analyzing the scores of the experiment, they may be plotted on a two- or three-dimensional plot in order to visualize the correlation.

### 3.5.1.3. ANALYSIS OF MODELS BY MULTIPLE REGRESSION AND THE ANALYSIS OF VARIANCE

Multiple regression provides  $\beta^i$  values that minimize the sum of squared differences between observed and calculated scores. This predictive model estimates,  $Q_{\text{calc}}$ , for the experimental result, based on contributions from the different factors. If there are  $K$  adjustable parameters in the model, they can be estimated by minimizing the sum of the squares of differences between  $Q_{\text{obs}}$  and  $Q_{\text{calc}}$  over all the experiments in a design of  $N > K$  experiments. Statistics programs like JMP (41) and SYSTAT (42) are natural tools for this kind of fitting. Details of fitting and interpreting the fit for such an equation and locating its stationary point by differentiation and setting the gradient equal to zero have been described (22,37).

### 3.5.1.4. STATIONARY POINTS

Stationary points are solutions to the simultaneous equations obtained by equating to zero the partial derivatives of the response surface. They represent the coordinates of the functional optimum of the model and, hence, of the experimental behavior of the system. Using these coordinates can be expected to produce optimal results.

Stationary points may be determined for any desirable crystalline property which can be “scored” precisely enough for optimization, including volume, shape, diffraction limits, stability, and relative freedom from secondary nucleation (34). Finding and using conditions close to stationary points of analytical response surfaces has several important advantages:

1. Optimization. Conditions at a maximum produce crystals that are in some sense optimized. Two different crystal forms of TrpRS had given trouble with reproducibility, inadequate volume, and/or unsuitable morphology. Replicated 20-experiment

Hardin–Sloane response–surface designs in four variables showed that these difficulties arose because we were growing at insufficient supersaturation (34). We avoided having to use repeated seeding (43,44), a difficult procedure that also produced unwanted growth of satellite crystals.

2. Reproducibility. Even the most carefully performed experiments can suffer from a frustrating level of irreproducibility. One source of variability is that the partial derivatives of the underlying, multidimensional response surface are large if the experimental conditions actually used are far from an optimum. Working instead at stationary points, where they are zero, helps insulate crystal growth from experimental errors in pH determination, pipetting errors, temperature fluctuations, and so on.
3. Insight. Empirical response surfaces provide scientific documentation about crystallogensis that is otherwise difficult to achieve, including conclusions with important and interesting biochemical relevance (40).

### 3.5.1.5. ANALYSIS AND VERIFICATION

Identification of a stationary point does not guarantee optimality. One must first examine the behavior nearby to determine whether it corresponds to a maximum, a minimum, or to a saddlepoint. This can be done by evaluating the second partial derivatives: negative curvature in all variables implies a local maximum, whereas positive curvature implies a minimum. Mixed second partial derivatives imply a saddlepoint. An accessory strategy in such cases is to examine two-dimensional level surface plots in all subspaces. Frequently, the dominant feature of a response surface is not a stationary point, but a “ridge” along which the value of the function increases, but normal to which it decreases. Optimum behavior also must be verified with replicated experiments at the experimental conditions specified by the stationary point.

### 3.5.1.6. STEEPEST ASCENT

Response surfaces may not turn out to be homogeneous quadratic polynomial functions. Often, when the initial conditions are far from a stationary point, they correspond to planes or ridges. In such cases, a steepest ascent algorithm provides a path to the optimal region. Steepest ascent, or line searching, is a most intuitive algorithm. Sensing that something improves, one marches toward that direction which gives the greatest improvement. This single search direction may, however, be a composite of several variables. The defining characteristic of a line search is that one can specify the gradient of the function one tries to optimize. Far from a stationary point, the available experiments may fit a plane or ridge function.

### 3.5.2. Altering Growth Kinetics

Crystal growth in microgravity can enhance the diffraction limits of some macromolecular crystals (45,46). The rationale for such results is complex but

important. In the absence of convective fluid movements, recruitment of molecules to the lattice from the near surface of the growing crystal can deplete a layer of solution nearest the crystal. This depletion layer slows the growth, altering the incorporation of impurities and allowing orderly addition of correctly oriented molecules to the lattice. Two different processes compete to determine the rate of addition to a growing crystal: diffusive recruitment of molecules from the bulk solution and the specific rate of attachment of molecules that have arrived at the surface. Either rate can dominate the actual growth kinetics, or neither of them may dominate. In the latter case, instabilities produce multiple layers of new growth, which also lead to incorporation of a variety of imperfections. Altering the rate of diffusion can therefore either minimize or accentuate these “step bunches.” In the former case, crystal order may improve, whereas in the latter it may actually deteriorate (6).

The physical advantages of microgravity can be approximated in terrestrial situations by limiting the diffusion of crystallizing agent and protein systematically and by allowing the crystal growth to take place in gels, which limit convection and may help in retaining impurities (4). Gel acupuncture samples large ranges of the phase diagram over time, increasing the chance to test optimal supersaturation conditions.

### 3.5.3. Annealing

Thorne (47,48) has established a firm theoretical and experimental basis for a phenomenon called annealing. Cryoprotection introduces strains that increase mosaicity and degrade diffraction resolution. The critical phenomenon appears to be differential thermal expansions of the protein lattice contacts and the bulk aqueous solvent. Annealing can restore the pre-existing order by allowing compensation for these differential effects. Careful, systematic determination of cryoprotection conditions (49) is, therefore, a highly justifiable investment, as the annealing appears only to restore pre-existing order that is degraded by cryoprotection. Moreover, the insight that can be obtained by such systematic study of cryoprotection conditions provides useful data with which to design optimal annealing conditions. (See ref. 50 and Chapter 3 of volume 2 for more detailed descriptions and protocols.)

## 4. Notes

1. Often reservoir solution is used to transfer the crystal. This can be done safely only if the protein concentration in the transfer solution matches that of the drop. Often solubility is high in the drop, and the crystal will start to dissolve if the reservoir contains no protein.
2. It is worth testing the temperature effect by introducing a crystallization plate into a thermally regulated box to observe whether the solid protein phase, crystals, or precipitate, dissolves or develops further.

3. The effects of salt and PEGs are not additive in general, and may even be of opposite sign (51). Whenever it is known or suspected that they will interact, an interaction term proportional to the product of (salt) and (PEG) should be included in  $\Sigma$ .
4. Usually, this is a two-step procedure. Regions likely to be disordered are first identified and then removed by mutagenesis (52).
5. Trituration works because most impurities are present at low concentrations. Even high precipitant concentrations will not reduce their solubility below their concentrations, so they remain in the soluble phase, whereas most of the molecule of interest is in the precipitated/crystallized phase. Washing a slurry of microcrystals, therefore, removes essentially all contaminating proteins that are not actually incorporated into the lattices of the solid phase. As this procedure generally has a high yield and can be done rapidly, it can be iterated to good effect.

## References

1. James, R. W. (1967) *The Optical Principles of The Diffraction of X-Rays*. G. Bell and Sons LTD, London, UK.
2. Dobrianov, I., Finkelstein, K. D., Lemay, S. G., and Thorne, R. E. (1998) X-ray topographic studies of protein crystal perfection and growth. *Acta Cryst.* **D54**, 922–937.
3. Snell, E. H., Bellamy, H. D., and Borgstahl, G. E. O. (2003) Macromolecular crystal quality. *Meth. Enzymol.* **368**, 268–287.
4. Garcia-Ruiz, J. M. (2003) Counter-diffusion methods for macromolecular crystallization. *Meth. Enzymol.* **368**, 130–153.
5. Luft, J. R. and DeTitta, G. T. (1996) Kinetic aspects of macromolecular crystallization. *Meth. Enzymol.* **276**, 110–131.
6. Vekilov, P. G., Alexander, J. I., and Rosenberger, F. (1996) Nonlinear response of layer growth dynamics in the mixed kinetics-bulk transport regime. *Phys. Rev. B* **E54**, 6650–6660.
7. Green, A. A. (1932) Studies on the physical solutions of the chlorides and sulfates of varying concentration. *J. Biol. Chem.* **95**, 47–66.
8. Ataka, M. and Tanaka, S. (1986) The growth of large, single crystals of lysozyme. *Biopolymers* **25**, 337–350.
9. Retailleau, P., Riès-Kautt, M., and Ducruix, A. (1997) No salting-in of lysozyme chloride observed at low ionic strength over a large range of pH. *Biophys. J.* **73**, 2156–2163.
10. Stura, E. (1999) Seeding techniques. In: *Crystallization of Nucleic Acids and Proteins, A Practical Approach*, (Ducruix, A., Giegé, R., eds.), IRL Press, Oxford, UK, pp. 177–207.
11. Ataka, M. and Michihiko, A. (1988) Systematic studies on the crystallization of lysozyme: determination and use of phase diagrams. *J. Cryst. Growth* **90**, 86–93.
12. Hofrichter, J., Ross, P. D., and Eaton, W. A. (1974) Kinetics and mechanism of deoxyhemoglobin S gelation: a new approach to understanding sickle cell disease. *Proc. Nat. Acad. Sci. USA* **71**, 4864–4868.

13. Hofrichter, J., Ross, P. D., and Eaton, W. A. (1976) Supersaturation in sickle cell hemoglobin solutions. *Proc. Nat. Acad. Sci. USA* **73**, 3035–3039.
14. Retailleau, P. (1997) Cristallologénèse de sels de lysozyme: étude des interactions en solution et de la solubilité d'un polyélectrolyte à faible force ionique (Ph.D.), Université Paris XI Orsay, December 16, 1996.
15. LaFont, S., Veesler, S., Astier, J. P., and Boistelle, R. (1997) Comparison of solubilities and molecular interactions of BPTI molecules giving different polymorph. *J. Crystal Growth* **173**, 132–140.
16. Bénas, P., Legrand, L., and Riès-Kautt, M. (2002) Strong and particular effects of cations on lysozyme chloride solubility. *Acta Cryst.* **D58**, 1582–1587.
17. Arakawa, T. and Timasheff, S. N. (1986) Theory of protein solubility. *Meth. Enzymol.* **114**, 49–77.
18. Gaucher, J. -F., Riès-Kautt, M., Reiss-Husson, F., and Ducruix, A. (1997) Solubility diagram of the *Rhodobacter sphaeroides* reaction center as a function of PEG concentration. *FEBS Lett.* **401**, 113–116.
19. Riès-Kautt, M. and Ducruix A. (1992) Phase diagrams. In: *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, (Ducruix, A. and Giegé, R., eds.), IRL Press, Oxford, UK, pp. 195–218.
20. Riès-Kautt, M. and Ducruix, A. F. (1989) Relative effectiveness of various ions on the solubility and crystal growth of lysozyme. *J. Biol. Chem.* **264**, 745–753.
21. Ingham, K. C. (1990) Precipitation of proteins with polyethylene glycol. *Meth. Enzymol.* **182**, 301–306.
22. Carter, C. W., Jr. (1999) Experimental design, quantitative analysis, and the cartography of crystal growth. In: *Crystallization of Nucleic Acids and Proteins, 2nd ed.* (Giegé, R. and Ducruix, A., eds.), IRL Press, Oxford, UK, pp. 75–120.
23. Riès-Kautt, M. and Ducruix, A. (1997) Inferences drawn from physicochemical studies of crystallogenesis and precrystalline state. *Meth. Enzymol.* **276**, 23–59.
24. Carbonnaux, C., Riès-Kautt, M., and Ducruix, A. (1995) Relative effectiveness of various anions on the solubility of acidic *Hypoderma lineatum* collagenase at pH 7.2. *Protein Sci.* **4**, 2123–2128.
25. Riès-Kautt, M. (1999) Strategy 2: an alternative to sparse matrix screens. In: *Crystallization of Proteins: Techniques, Strategies, and Tips. A Laboratory Manual*, (Bergfors, T., ed.), International University Line, La Jolla, CA.
26. McPherson, A. (2001) A comparison of salts for the crystallization of macromolecules. *Protein Sci.* **10**, 418–422.
27. Trakhanov, S. and Quiocho, F. A. (1995) Influence of divalent cations in protein crystallization. *Protein Sci.* **4**, 1914–1919.
28. Carter, C. W., Jr. (1996) A local approximation to supersaturation affords a useful coordinate transformation for the study of crystal growth. *Acta Cryst.* **53**, 647–654.
29. Vekilov, P. G. (2003) Molecular mechanisms of defect formation. *Meth. Enzymol.* **368**, 170–187.
30. Carter, D. C., Lim, K., Ho, J. X., et al. (1999) Lower dimer impurity incorporation may result in higher perfection of HEWL crystals grown in microgravity: case study. *J. Cryst. Growth* **196**, 623–637.

31. Holmes, W. M., Hurd, R. E., Reid, B. R., Rimerman, R. A., and Hatfield, G. W. (1975) Separation of transfer ribonucleic acid by sepharose chromatography using reverse salt gradients. *Proc. Nat. Acad. Sci. USA* **72**, 1068–1071.
32. Gulewicz, K., Adamiak, D., and Sprinzl, M. (1985) A new approach to the crystallization of proteins. *FEBS Lett.* **189**, 179–182.
33. Schekman, R., Weiner, J. H., Weiner, A., and Kornberg, A. (1975) Ten proteins required for conversion of fX174 single-stranded DNA to duplex form *in vitro*. *J. Biol. Chem.* **250**, 5859–5865.
34. Carter, C. W., Jr, Yin, Y. (1994) Quantitative analysis in the characterization and optimization of protein crystal growth. *Acta Cryst.* **D50**, 572–590.
35. Chester, A., Weinreb, V., Carter, C. W., Jr., and Navaratnam, N. (2004) Optimization of apolipoprotein B mRNA editing by APOBEC1 apoenzyme and the role of its auxiliary factor, ACF. *RNA* **10**, 1399–1411.
36. Yin, Y. and Carter, C. W., Jr. (1996) Incomplete factorial and response surface methods in experimental design: yield optimization of tRNA<sup>Trp</sup> from *in vitro* T7 RNA polymerase transcription. *Nucleic Acids Res.* **24**, 1279–1286.
37. Carter, C. W., Jr. (1997) Response surface methods for the optimization and improving reproducibility of crystal growth. *Meth. Enzymol.* **276**, 74–99.
38. Hardin, R. H. and Sloane, N. J. A. (1993) A new approach to the construction of optimal designs. *J. Stat. Plan. Inference* **37**, 339–369.
39. Muschol, M. and Rosenberger, F. (1997) Liquid-liquid phase separation in super-saturated lysozyme solutions and associated precipitate formation/crystallization. *J. Chem. Phys.* **107**, 1953–1962.
40. Carter, C. W., Jr., Doublé, S., and Coleman, D. E. (1994) Quantitative analysis of crystal growth tryptophanyl-tRNA synthetase crystal polymorphism and its relationship to catalysis. *J. Mol. Biol.* **238**, 346–365.
41. SAS JMP: Scientific discovery software. (1999) In. 3.2.2 ed. Cary, NC: SAS Institute, Inc.
42. Wilkinson L. (1987) SYSTAT, The System for Statistics. In. 5.2.1 ed. Evanston, IL 60601: SYSTAT, Inc.
43. Thaller, C., Weaver, L. H., Eichele, G, Wilson, E., Karlsson, R., and Jansonius, J. (1981) Repeated seeding technique for growing large single crystals of proteins. *J. Mol. Biol.* **147**, 465–469.
44. Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R., and Jansonius, J. (1985) Seed enlargement and repeated seeding. *Meth. Enzymol.* **114**, 132–135.
45. Symersky, J., Devedjiev, Y., Moore, K., Brouillette, C., and DeLucas, L. (2002) NH<sub>3</sub>-dependent NAD<sup>+</sup> synthetase from *Bacillus subtilis* at 1 Å resolution. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1138–1146.
46. Borgstahl, G., Vahedi-Faridi, E. O. A., Lovelace, J., Bellamy, H. D., and Snell, E. H. (2001) A test of macromolecular crystallization in microgravity: large well ordered insulin crystals. *Acta Crystallogr. D. Biol. Crystallogr.* **57**, 1204–1207.
47. Kriminski, S., Caylor, C. L., Nonato, M. C., Finkelstein, K. D., and Thorne, R. E. (2002) Flash-cooling and annealing of protein crystals. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 459–471.

48. Kriminski, S., Kazmierczak, M., and Thorne, R. E. (2003) Heat transfer from protein crystals: implications for flash-cooling and X-ray beam heating. *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 697–708.
49. Garman, E. F. and Doublíé, S. (2003) Cryocooling of macromolecular crystals: optimisation methods. *Meth. Enzymol.* **368**, 188–216.
50. Hanson, B. L., Harp, J. M., and Bunick, G. J. (2003) The well-tempered protein crystal: annealing macromolecular crystals. *Meth. Enzymol.* **368**, 217–238.
51. Papanikolaou, Y. and Kokkinidis, M. (1997) Solubility, crystallization and chromatographic properties of macromolecules strongly depend on substances that reduce the ionic strength of the solution. *Protein Eng.* **10**, 847–850.
52. Koth, C. M., Orlicky, S. M., Larsen, S. M., and Edwards, A. M. (2003) Use of limited proteolysis to identify protein domains suitable for structural analysis. *Meth. Enzymol.* **368**, 77–84.

## Optimization Techniques for Automation and High Throughput

Naomi E. Chayen

### Summary

The main effort in the area of crystallization for structural genomics is currently being invested in automation of high-throughput screening procedures to identify potential crystallization conditions. However, screening in itself, even in massive quantities, is not enough; it has to be complemented by an equally important procedure in crystal production, namely crystal optimization. This chapter describes optimization methods for turning low-quality crystals into useful diffracting ones and presents practical ways of automating such methods and adapting them to high throughput. The methods enable the control of the crystallization environment as the trial takes place. They include the use of oils, gels, and the uncoupling of the nucleation and growth phases of crystallization.

**Key Words:** Gels; microbatch; nucleation; protein crystallization; structural genomics; proteomics; high throughput; vapor diffusion; robotics.

### 1. Introduction

Protein crystallization has gained a new strategic and commercial relevance in structural genomics where X-ray crystallography plays a major role. Production of high-quality crystals has always posed a problem for X-ray crystallography, and with the advent of structural genomics this problem has become even more acute. The abundance of screening protocols combined with progress in automation has resulted in production of numerous crystals. The next crucial step is to design optimization procedures in order to turn these crystals into useful diffracting ones.

The most common way to optimize crystallization experiments is to fine-tune the crystallization conditions by varying parameters, such as the concentration of protein, precipitants, pH, temperature, and others. This, in fact, is extended

screening, which focuses on a more defined range of conditions. A second means of optimization is to actively influence and control the crystallization environment while the trial takes place in order to lead crystal growth in the direction that will yield the best results.

Screening procedures can readily be automated and adapted to high throughput (e.g., **refs. 1–6**). These procedures are very valuable as they are essential in order to find initial conditions for crystallization and, in some cases, can also produce high-quality crystals. However, when screening and subsequent optimizing by fine tuning of the initial screening conditions fail, it is necessary to implement the second approach—that of optimizing the crystallization conditions by actively controlling the crystallization process. Such methods are currently conducted manually, as they are complicated and do not easily lend themselves to automation and high throughput. There is, however, an urgent need to find ways to automate and miniaturize these techniques in order to cope with the vast numbers of “leads” resulting from screening procedures (7). To date only few such optimization methods have been automated.

This chapter presents detailed protocols for optimization by controlling the crystallization environment. These techniques have been automated and can thus be easily applied as high-throughput experiments. The protocols described are to be used for turning low-quality crystals (e.g., showers, needles, haystacks, and others) into useful diffracting crystals.

## **2. Materials**

### **2.1. Crystallization Plates**

1. Microbatch plates (Douglas Instruments, Hampton Research, Molecular Dimensions, Greiner).
2. 24-Well plates, e.g., Linbro, VDX (Hampton Research), or XRL (Molecular Dimensions).
3. Cryschem plates (for sitting and sandwich drops) and sealing tape (Hampton Research).
4. When used with robotics, plates suitable for the particular robots are recommended, e.g., 96-well plates (Hampton Research), or 1536 well plates (Nunc).

### **2.2. Tools and Chemicals**

1. Siliconized cover slips with a diameter to fit the 24-well plates (Hampton Research, Molecular Dimensions).
2. Clear or white Vaseline (any pharmacy or drugstore).
3. Silicone oil (Hampton Research, Molecular Dimensions, Douglas Instruments).
4. Paraffin oil (Hampton Research, Molecular Dimensions, Douglas Instruments).
5. Plasticine, vacuum oil, or grease (for sealing the cover slips) (Hampton Research, Molecular Dimensions).

6. Pipets of all sizes, pipet tips.
7. Loops of various sizes for harvesting crystals (Hampton Research).
8. Tetramethyl orthosilane (TMOS) solution (Fluka, cat. no. 87682). Make fresh as required.
9. Sodium metasilicate solution (Sigma, cat. no. 33844-3). Make fresh as required.
10. 1 M Acetic acid.

### 2.3. *Optional*

1. Nextal plates (Nextal Biotechnologies, Montreal, Canada).
2. Al's oil (Hampton Research).
3. Three-dimensional (3D) screen (Molecular Dimensions).
4. Gelled surface kit (Molecular Dimensions).

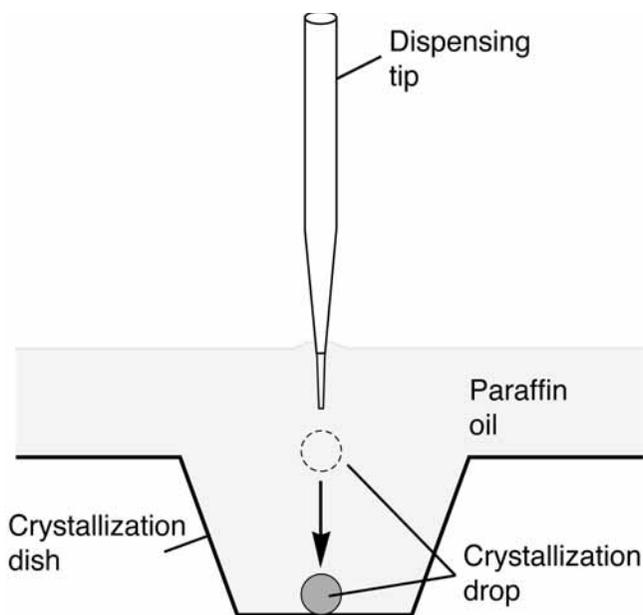
## 3. Methods

This chapter concentrates on optimization techniques by microbatch and vapor diffusion, which are the most common crystallization methods. The inherent differences between these methods dictate when to choose microbatch rather than vapor diffusion and vice versa. The protocols describe the performance of the techniques when performed manually and automatically using robotics.

### 3.1. *The Microbatch Method*

The microbatch method was designed to obtain maximum information on the molecule to be crystallized while using minimal amounts of sample (less than 1  $\mu\text{L}$  per trial) and to save experimenter's time (8). Microbatch is the simplest crystallization method because the molecule to be crystallized and the crystallizing agents are mixed at their final concentrations at the start of the experiment. Supersaturation is achieved upon mixing, thus, no equilibration process takes place. The trials are dispensed and incubated under paraffin oil in order to prevent evaporation of such small volumes. Microbatch trials can be performed either manually or automatically. When dispensed manually the minimum volume that can accurately be dispensed is 0.5  $\mu\text{L}$ . Using a robot, thousands of experiments can be generated per day using nanoliter quantities of sample. Microbatch is used for screening, optimization by fine tuning of conditions, and is especially useful for optimization by controlling the crystallization environment. This is because in batch, conditions are known and constant and it is, therefore, straightforward to follow the history of the sample and to achieve reproducibility (9).

The procedure for setting up microbatch experiments and harvesting crystals under oil is given next. These procedures form the basis for performing all the microbatch experiments described in this chapter (*see Note 1*).



**Fig. 1.** A crystallization drop dispensed under oil. Once dispensed the drop, which is heavier than the paraffin oil, migrates to the bottom of the dish (based on [Fig. 1](#) of [ref. 20](#)).

### 3.1.1. Setting Up Manually

#### 3.1.1.1. MIXING OF THE PROTEIN SOLUTION WITH ONE SOLUTION CONTAINING CRYSTALLIZING AGENTS

1. Pipet or dispense 6 mL of paraffin oil into a microbatch plate (*see Note 2*). The oil will spread over the plate and cover the wells.
2. Using a Gilson P2 or similar pipet, withdraw 1  $\mu\text{L}$  of the precipitant solution from its container. You can also use the automatic pipet described in Chapter 7.
3. Insert the tip into the well under the surface of the oil and dispense the 1- $\mu\text{L}$  drop (*see Note 3*). If you find it difficult to hold the tip in mid-oil, you can rest the edge of the tip on the floor of the plate as you dispense. As you withdraw the tip from the oil, the drop will detach from it and fall to the bottom of the well ([Fig. 1](#)).
4. Add 1  $\mu\text{L}$  of protein solution to that well in the same way. The two (separate) 1- $\mu\text{L}$  drops join and become a 2- $\mu\text{L}$  drop. If the drops do not coalesce, mix them gently with the pipet tip (*see Note 4*).
5. Incubate at the temperature of your choice.
6. Observe crystals under a light microscope (*see Note 5*).

### 3.1.2. Setting Up by a Robot

When the experiments are performed by a robot, the precipitant solutions are dispensed simultaneously under oil by 1–384 syringes depending on the robotic

system. Protein is then added to the precipitant drops using a dedicated syringe for the protein solution. Some robots have a routine for mixing the drops. Alternatively the crystallization dish can be centrifuged after dispensing.

### 3.1.3. *Mixing of the Protein Solution With Several Solutions Containing Crystallizing Agents*

In most optimization cases, four or five different ingredients may be needed in the drop. It is difficult to manually pipet all the ingredients into one drop directly under the oil. Therefore, once promising conditions are found, mix the protein solution and the crystallizing agents in an Eppendorf tube or, if quantities are very small, on a cover slip. Once mixed, draw the drop with a pipet tip and dispense under the oil as described in **Subheading 3.1.1**.

When performed by a robot, the different ingredients are placed in different channels/syringes of a dispensing system and dispensed simultaneously under the oil by the action of motorized syringes. All robots have routines whereby they pick up chosen stock solutions and dispense them into a well to which protein is added simultaneously or later on.

## 3.2. *Crystallization of Membrane Proteins in Microbatch*

The prospect of crystallizing membrane proteins under oil had initially been received with scepticism owing to doubts about the suitability of an oil-based method for crystallizing lipophilic compounds. However, in the last 5 yr an increasing number of proteins in a variety of different detergents have been crystallized in microbatch under oil (**10**). Some of these could only be crystallized in microbatch and not by other crystallization methods. Crystals of these proteins were produced in 1.2–2- $\mu$ L drops using a robot. The drops in oil do not spread out as they do in vapor diffusion over the siliconized cover slips, dispensing is quick and simple, and the presence of detergents does not present any difficulties (**7,10**).

The procedure for dispensing microbatch trials containing membrane proteins is identical to that described in **Subheadings 3.1.1– 3.1.3**.

## 3.3. *Harvesting Crystals From Microbatch*

1. Add cryo-protectant solution to the drop.
2. Make sure that the crystals are “happy” (i.e., not cracked or dissolved) by looking at them under the microscope.
3. Using a loop, take the crystals out of the oil and freeze.

If this proves difficult try this alternative method:

1. Add harvest solution into the well containing the crystal (*see Note 6*).
2. After several minutes withdraw the enlarged drop using a standard 20- to 200- $\mu$ L micropipet, which had its tip cut off with a scalpel in order to widen its bore.
3. Transfer the drop into a depression well containing more harvest solution.
4. From this stage onwards the handling proceeds as it would from any diffusion trial.

If crystals stick to the supporting surface, loosen them gently inside the drop using microtools (Hampton Research) or with an animal whisker (9). Alternatively, set up the drops in the container described in **Subheading 3.5**.

### **3.4. Optimization by Controlling the Crystallization Environment**

#### *3.4.1. Crystallization in Gels*

Growth of crystals in a gel medium can improve the quality of crystals in comparison with solution media (11). This is because the presence of gel in the crystallization drop reduces convection and sedimentation, in some way mimicking crystallization in microgravity.

Using vapor diffusion and counter diffusion techniques, setting up gelled drops is more demanding and time consuming than standard solution trials. In contrast, setting up gelled drops in microbatch, described next, is as easy and simple as setting the trials in standard microbatch (12,13).

#### *3.4.2. Preparation of the Gel Stock Solutions*

Prepare a 2-mL stock solution of TMOS at 5% (v/v) as follows:

1. Add 0.1 mL TMOS solution to 1 mL deionized water in a glass beaker.
2. Stir the solution vigorously using a stirrer at high speed (*see Note 7*).
3. Top up the solution to 2 mL with deionized water.
4. Stir vigorously for an additional 10–15 min keeping the beaker covered.

To prepare a 2-mL stock solution of sodium metasilicate at 5% (v/v):

1. Dilute a sodium metasilicate solution, which has an initial pH of 11.6, with distilled water to bring it close to the desired concentration.
2. Adjust the pH to 6.5 by addition of 1 M acetic acid while stirring.

The experiments are performed in **Subheading 3.4.3**.

#### *3.4.3. Manual Dispensing*

1. Mix the protein to be crystallized, the crystallizing agents, and one of the freshly made gel solutions in an Eppendorf tube. The gel solution should be at a final concentration of 0.2%.
2. Once mixed, draw a drop (0.5–5  $\mu\text{L}$ ) with a pipet tip and dispense under the oil as described in **Subheading 3.1**.

#### *3.4.4. Dispensing by Robot*

1. Choose a computer-controlled dispensing system consisting of several channels/syringes in which precipitant, buffer, protein, and additives can be put into different channels and dispensed simultaneously by the action of motorized syringes.
2. Fill a crystallization tray with paraffin oil as done for standard microbatch trials.
3. Load the gel solution while it is a low-viscosity liquid into the dispensing system in the same way as the other components of the crystallization trial.

4. The gel solution is dispensed under oil simultaneously with all the other components to form one drop (*see Note 8*).
5. Incubate the trials at the temperature of your choice.
6. After a given time (this will depend on the type of gel. It will take 12–16 h in the case of TMOS) polymerization occurs and the drops gel.
7. Harvesting crystals from the gelled drops is done in the same way as with standard microbatch trials.

### 3.5. “Containerless” Crystallization

Heterogeneous nucleation, which is often detrimental to the production of diffraction-quality crystals, can be induced by the contact of a crystallization trial with the walls of its supporting vessel (14). Crystallization in a “containerless” setup, in which a crystallization drop is suspended between two oils of different densities, results in reduction of heterogeneous nucleation (15,16), thus, leading to the production of a smaller number of high-quality crystals.

The protocol for setting up “containerless” crystallization trials is described in **Subheading 3.5.1**.

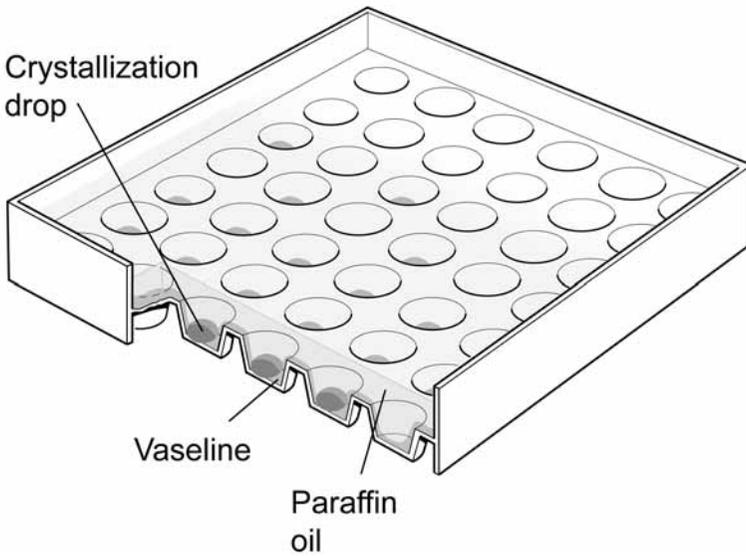
#### 3.5.1. Preparation of the Crystallization Dish

1. Squeeze 3–5 cm of Vaseline (*see Note 9*) from a tube into wells of a crystallization plate or onto a flat dish which has a lid.
2. Place the dish in a 40–60°C oven for a few minutes until the Vaseline becomes liquid. It will then spread and form a layer that covers the bottom of the crystallization plate or dish.
3. Allow to cool at room temperature for 15–20 min. The Vaseline will harden and form a film at the bottom of the dish.

#### 3.5.2. Dispensing of the Drops

1. Dispense 50  $\mu\text{L}$  to 2 mL of paraffin oil depending on the size of the dish (*see Note 10*) on top of the Vaseline surface.
2. Mix the protein and precipitant in a small Eppendorf tube.
3. Dispense the crystallization drop into the oil by inserting the pipet tip into the paraffin oil with the pipet tip just touching the Vaseline surface. Drop size can be any size you choose.
4. The drop will situate itself between the Vaseline surface and the oil (**Fig. 2**).
5. Cover the dish with a lid or sealing tape.
6. Incubate at the temperature of your choice.
7. When crystals appear they can be lifted directly out of the drop with a loop or a spatula. The greased surface provides a stable interface to the upper layer. This prevents crystals from migrating to the walls, making them much easier to harvest compared with standard microbatch trials.

Any robot that dispenses microbatch trials can be used for this procedure.



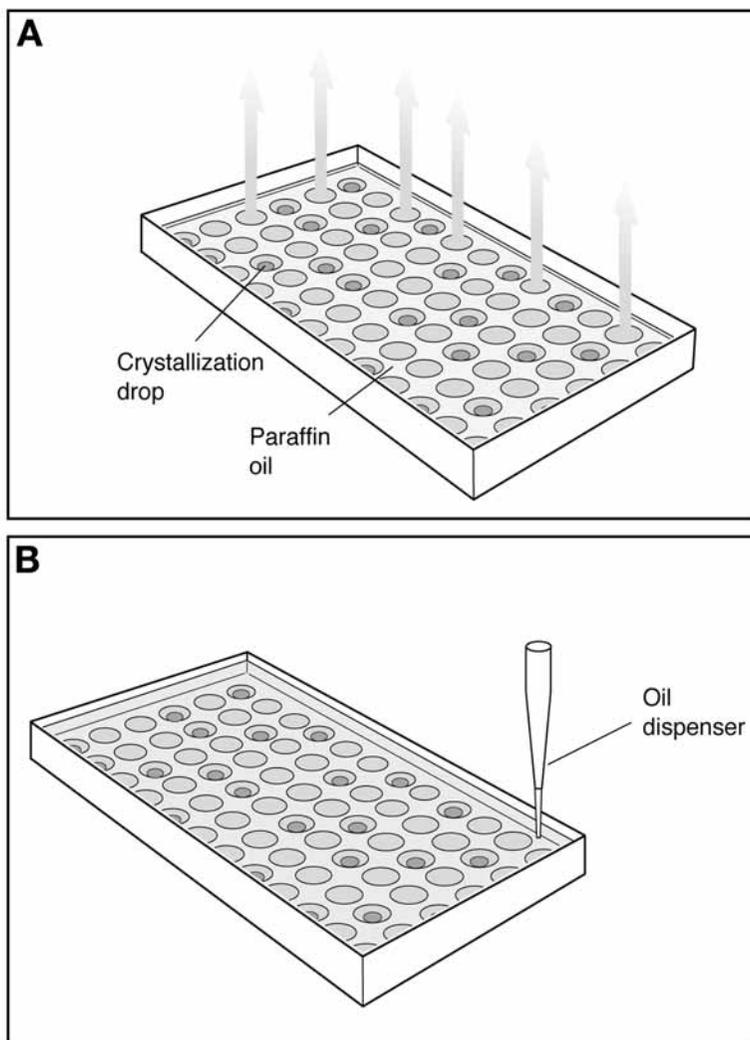
**Fig. 2.** Containerless crystallization (based on [Fig. 3](#) of [ref. 13](#)).

### 3.6. Control of Evaporation Kinetics

Nucleation is a prerequisite and the first step in crystal growth, yet excess nucleation yields a large number of small crystals instead of a small number of useful ones. A means of controlling nucleation in microbatch by inducing nucleation and then stopping it before it becomes excessive is described next. This is achieved by controlled evaporation, and therefore concentration, of the drops through a thin oil layer. Evaporation is later arrested by increasing the thickness of the oil layer ([13](#)). If trials were allowed to evaporate without arresting, the drops would dry out. Arrest of evaporation at the early stages of nucleation will result in the formation of fewer crystals of better quality.

1. Set up several trays containing microbatch trials under a layer of paraffin oil so that the oil just covers the trials ([Fig. 3A](#)) (*see Note 11*).
2. Allow incubation for a given time (*see Note 12* for guidance on how to select the time).
3. Top up the oil (at different times for the different trays) so that it fills the dish ([Fig. 3B](#)).
4. Continue to incubate at the temperature of your choice.
5. Observe every few days.

A robot will first dispense the trials under the lower volume of oil. The robot is programmed to add oil at various time intervals after setting up the trials.



**Fig. 3.** Controlled evaporation by altering the thickness of the oil layer covering the trials. **(A)** A thin layer of paraffin that enables evaporation. **(B)** Arrest of evaporation by addition of more paraffin oil. (Based on [Fig. 4](#) of [ref. 13](#).)

### 3.7. Decoupling Nucleation and Growth

Nucleation requires conditions that are different from those that promote growth. The most common way of decoupling nucleation and growth is by seeding ([17](#)). However, quenching of nucleation using dilution is more amenable to high-throughput processing. Dilution can be achieved in both microbatch ([18](#)) and hanging-drop methods ([19](#)). The aim of dilution is to start

the trial at nucleation conditions and after a given time “back off” to conditions of growth.

### 3.7.1. *In Microbatch*

1. Set up microbatch trials under conditions that yield low-quality crystals.
2. Dilute the trials at given times after setup (*see Note 13* for guidance on how to select the times) by adding either buffer or protein in buffer at a volume that is 5–10% of the total drop volume.

If performing manually use a pipet of your choice. If performed by a robot, the robotic system is programmed to revisit the drops at given times in order to add the diluent.

### 3.7.2. *In Hanging Drops*

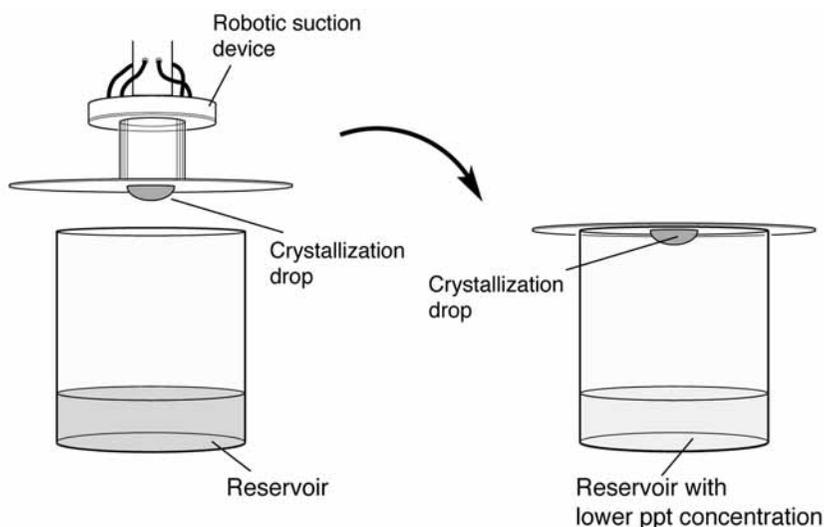
1. Use Linbro-type plates.
2. Set up trials with reservoirs containing progressively lower concentrations of precipitant solutions until you determine the conditions that yield clear drops.
3. Fill six to nine reservoirs with a solution containing lower precipitant concentration (determined in **step 1**) that would result in producing a clear drop if crystallization drops were set up under these conditions.
4. Grease the rim of the plate with oil (not vacuum grease, *see Note 14*) and cover the reservoirs with cover slips or a lid (*see Note 15*).
5. Set up 6–10 trials under conditions that are known to generate low-quality crystals.
6. At a given time after setup (*see Note 16* for guidance on how to select the time) transfer one of the cover slips onto a reservoir containing the lower precipitant concentration.
7. After a further time interval transfer a second cover slip, then the third, fourth, and so on.
8. Leave one or two drops without transferring them and one or two drops under the low-precipitant concentration to act as controls.
9. The time required for crystal formation will be longer in the transferred drop compared with the control experiment, which has not been transferred, but the crystals should be fewer and better ordered.

If performing these experiments manually, transfer the drops by hand. The transfer takes 1–2 s (*see Note 14*). If using a robot, program it to perform the transfers. This, of course, will increase the number of experiments that can be done.

A robot which dispenses hanging-drop trials can be used to perform such trials automatically (**Fig. 4**).

## 3.8. *Control of the Speed of Vapor-Diffusion Trials*

Often, numerous tiny crystals are formed in protein crystallization trials. In some cases this occurs because the process of crystallization takes place too rapidly. A way to approach supersaturation more slowly in order to avoid the



**Fig. 4.** Robotic transfer of hanging drops from nucleation to growth conditions using a suction device (based on [Fig. 5](#) of [ref. 13](#)).

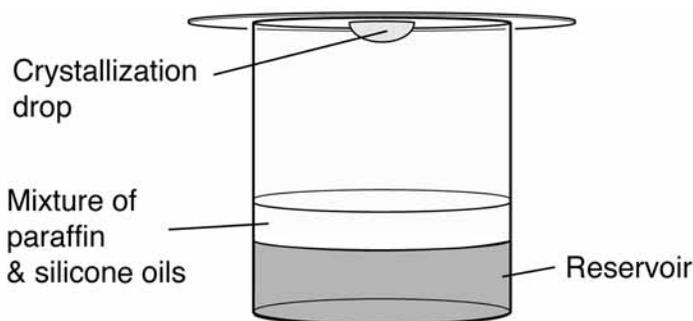
formation of small crystals, twinned crystals, or precipitate is by slowing down the equilibration rate. This is achieved by placing a paraffin/silicone oil mixture as a barrier over the reservoir of a hanging or sitting drop trial ([20](#)). The mix of paraffin and silicone oil can be varied as needed. It was found that volumes of 250–500  $\mu\text{L}$  placed over 0.6- to 1-mL reservoirs in standard Linbro plates (corresponding to a layer thickness of 1.25–2.5 mm) were most efficient. The type of oil and the thickness of the oil layer situated above the reservoir control the speed of crystallization. In trials containing an oil barrier, crystals required longer periods (e.g., 8 d compared to 24 h) to grow to full size, yet their number was reduced and their quality was much improved (e.g., [refs. 20,21](#)).

The advantage of this technique is that no change is required to the crystallization conditions or to the method used. It can be applied in Linbro, VDX, Cryschem, Nextal, or any other vessel, and it can easily be automated by adding one extra step to the procedure used by existing crystallization robots.

The protocol for inserting an oil barrier to slow down vapor-diffusion trials is described next.

### 3.8.1. Preparation of the Oils

1. Mix paraffin and silicone oils at different ratios, e.g., equal volumes, or 60% paraffin and 40% silicone, and others (see **Note 2**). The oils are totally miscible. Al's oil is a ready-made mix of equal volumes of paraffin and silicone oils.
2. Shake well and let stand for several minutes until the bubbles disappear.



**Fig. 5.** A hanging-drop trial with an oil barrier above the reservoir.

### 3.8.2. Set Up of Hanging Drops

1. Use a Linbro-type plate type for hanging drops.
2. Coat the lips of the reservoirs with grease or oil (unless your plates are pregreased).
3. Pipet into each well 0.6–1.0 mL of the reservoir solution, which is known to produce your best crystals (*see Note 17* for choice of reservoir solution).
4. Pipet a measured volume of a mixture of paraffin and silicone (*see Note 18*) over all the reservoirs except for one reservoir. The oil will form a layer above the reservoirs (**Fig. 5**).
5. Dispense the hanging drops on the cover slips as usual by mixing the protein solution with the reservoir solution. Use the reservoir without oil as your source of precipitant for all the drops.
6. Place the cover slips over the wells containing the oil layer and cover.
7. Place the last drop over the reservoir without the oil. This drop will act as your control.
8. Incubate at the temperature of your choice.

### 3.8.3. Sitting and Sandwich Drop

In the case of sitting and sandwich drops, set up the trials as you would normally and place the layer of oil above the reservoir before sealing the plates with tape.

## 4. Notes

1. Microbatch can be used for almost all known precipitants, buffers, and additives. The oils do not interfere with the common precipitants such as salts, PEG, MPD, jeffamine, glycerol, and ethanol. Microbatch, though, should not be used for crystallization trials containing small, volatile organic molecules such as phenol, dioxane, or thymol because these molecules dissolve into the oil (**7,9,10**).
2. The microbatch method involves using a layer of paraffin oil thick enough (4 mm or 6 mL), in standard microbatch plates, or a ratio of 1:50 between the drop and the

oil, to render evaporation through it negligible within the time-scale of a crystallization experiment (typically 1 wk to 1 mo). A thinner layer of paraffin oil will allow evaporation and drying of the drops. Other oils, such as silicone oil or mixtures of paraffin and silicone oils, also allow evaporation of the trials and are used in microbatch only for screening purposes (22) or in optimization using vapor diffusion as described in **Subheading 3.8**.

3. Pipets usually have two stops when pressing them. Dispense the drop into the oil while holding the pipet on the first stop, otherwise it will introduce air bubbles into the drop.
4. A great advantage of microbatch is that it provides a robust environment in which you need not be too careful when setting up experiments. You can easily mix the drops and shake the trays once the drops are setup and even once crystals have grown. This is because the oil buoys the crystallization drops and protects them from physical shock, as well as from airborne contamination. This makes the experiments easily transportable.
5. Crystals under oil can be seen clearly under the microscope. Some crystallization plates are birefringent making it difficult to distinguish between the plate and the crystals. Improved plates are currently being designed by Greiner, Douglas instruments, and other companies selling crystallization plates.
6. As for vapor diffusion, in microbatch the harvest solution contains a slightly higher (~5%) concentration of precipitant than that in the drop.
7. When water is first added to the gel, a phase separation occurs, which looks like oil drops in the solution. When shaking well or stirring vigorously these drops disperse. By addition of more water and further vigorous stirring it becomes a solution.
8. Gelled drops can be dispensed in final drop volumes of 0.2  $\mu\text{L}$  and above but 0.7  $\mu\text{L}$  is the minimum volume at which the gel exercises an effect on crystal quality.
9. The Vaseline has to be white or transparent, not yellow as usually known. This allows good visibility of the crystals under the microscope.
10. MicroWell modules and Linbro-type plates can be used, but most other crystallization plates are suitable. Alternatively use the gelled surface kit from Molecular Dimensions. A low-density oil, usually paraffin, is applied as the top layer. The oil must provide a thick layer (as explained in **Note 2**) to cover the trials in order to prevent their evaporation. Many drops can be dispensed within a small, flat area coated with the grease, doing away with the need for wells.
11. The paraffin oil generally used in standard microbatch trials is not completely impermeable to the aqueous solution, which constitutes the crystallization drops. The conventional microbatch method, therefore, involves using a thick layer of oil as described in **Note 2**. However, a thinner layer allows evaporation to take place at a much faster rate.
12. The timing of arrest is selected by reference to the time it took for the first crystals to appear in the initial screens. For example, if crystals grew within 24 h, nucleation would have occurred anytime between setting up the experiments to several hours before the crystals appeared. Hence, arrest should be done at intervals of 1–2 h after setup. Trials that had been arrested too soon will produce clear drops,

whereas those that were arrested too late will yield low-quality crystals. Using robotics, the arrest can be done with larger numbers of trials at shorter intervals making it easier to pinpoint the right time.

13. The consideration for the time of dilution is the same as that for arrest as explained in **Note 12**.
14. In Linbro-type plates the reservoirs are generally sealed with grease, which makes it difficult to remove the cover slip after a period of time. Moreover, it is hard to ascertain the effectiveness of the seal over the new reservoir. Greasing with oil allows easier transfer of the cover slips from one reservoir to another especially when using robotics, which need to work very quickly. For manual use, Nextal plates, where each well is sealed by a screw cap that incorporates the cover slide, are ideal for such experiments.
15. To avoid shock to the drop owing to transfer, the low-concentration reservoir solutions are dispensed at the beginning of the experiment and sealed with cover slips, a lid, or with self-adhesive tape. This allows a humid atmosphere to develop above the well on which the cover slip will be transferred. The transfer itself should last 1–2 s.
16. The consideration for the time of dilution is the same as that for arrest and dilution as explained in **Note 12**.

These transfer experiments have led to the development and use of a “3D screen,” which has a built-in first optimization step. The screen works in the following way: hanging drops were transferred at various times (the transfer time is selected by reference to the time it took to see the first crystals in the preliminary screens) from a standardized set of screening solutions at high concentrations to screens at lower concentrations. For a decoupling experiment to be successful, it must be ensured that the solution is diluted to the metastable zone of conditions after incubation at spontaneous nucleation conditions. It was, however, found that when these conditions are not known beforehand, as is the case in a high-throughput environment, the best results were obtained when standard screening kits solutions were diluted to between 60 and 80%. These experiments can now be performed automatically with robots for setting up hanging drops, such as the Gilson Workstation or others. The screen consists of 48 solutions, 24 contain sparse matrix screening conditions; the other 24 are a 70% dilution of the precipitants in those (buffer and additive concentrations are kept constant). Using this screen, all hanging drops were first incubated for 3–6 h over the solutions at high concentrations. The cover slips holding the drops are then transferred over the reservoirs at 70% dilution (**13**).

17. This method works very well in cases where the precipitant is salt. It also works well with up to 12.5% PEG (various molecular weights) and MPD. If PEG concentration is higher this technique will not work. Try the experiments in **Subheading 3.7.2**. instead.
18. You can control the speed of crystallization by varying the thickness of your oil barrier by pipetting different volumes (from 100 to 700  $\mu\text{L}$ ) over the reservoir or by varying the oil ratios.

## References

1. Luft, J. R., Wolfley, J., Jurisica, I., Glasgow, J., Fortier, S., and DeTitta, G. T. (2001) Macromolecular crystallization in a high throughput laboratory: the search phase. *J. Cryst. Growth* **232**, 591–595.
2. Walter, T. S., Brown, D. J., Pickford, M., Owens, R. J., Stuart, D. I., and Harlos, K. (2003) A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. *J. Appl. Cryst.* **36**, 308–314.
3. Santarsiero, B. D., Yegian D. T., Lee, C. C., et al. (2002) An approach to rapid protein crystallization using nanodroplets. *J. Appl. Cryst.* **35**, 278–281.
4. Mueller, U., Nyarsik, L., Horn, M., et al. (2001) Development of a technology for automation and miniaturization of protein crystallization. *J. Biotechnol.* **85**, 7–14.
5. Kuhn, P., Wilson, K., Patch, M. G., and Stevens, R. C. (2002) The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr. Opin. Chem. Biol.* **6**, 704–710.
6. Hansen, C. L., Skordalakes, E., Berger, J. M., and Quake, S. R. (2002) A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *PNAS* **99**, 16,531–16,536.
7. Chayen, N. E. (2003) Protein crystallization for genomics: throughput versus output. *J. Struct. Funct. Genomics* **4**, 115–120.
8. Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L., and Blow, D. M. (1990) An automated system for microbatch protein crystallization and screening. *J. Appl. Cryst.* **23**, 297–302.
9. Chayen, N. E. (1998) Comparative studies of protein crystallization by vapor diffusion and microbatch. *Acta Cryst.* **D54**, 8–15.
10. Chayen, N. E. (2003) Crystallization of membrane proteins in oils. In: *Methods and Results in Crystallization of Membrane Proteins*, (Iwata, S., ed.), International University Line, La Jolla, CA, pp. 131–139.
11. Robert, M. -C., Vidal, O., Garcia-Ruiz, J. -M., and Otálora, F. (1999) Crystallization in gels and related methods. In: *Crystallization of Nucleic Acids and Proteins*, (Ducruix, A. and Giegé, R., eds.), Oxford University Press, Oxford, UK, pp. 149–175.
12. Moreno, A. Saridakis, E., and Chayen, N. E. (2002) Combination of oils and gels for enhancing the growth of protein crystals. *J. Appl. Cryst.* **35**, 140–142.
13. Chayen, N. E. and Saridakis, E. (2002) Protein crystallization for genomics: towards high-throughput optimization techniques. *Acta Cryst.* **D58**, 921–927.
14. Yonath, A., Mussig, J., and Wittman, H. G. (1982) Parameters for crystal growth of ribosomal subunits. *J. Cell Biochem.* **19**, 145–155.
15. Chayen, N. E. (1996) A novel technique for containerless protein crystallization. *Protein Engineering* **9**, 927–929.
16. Lorber, B. and Giegé, R. (1996) Containerless protein crystallization in floating drops: application to crystal growth monitoring under reduced nucleation conditions. *J. Cryst. Growth* **168**, 204–215.
17. Stura, E. A. (1999) Seeding. In: *Crystallization of Proteins: Techniques, Strategies and Tips*, (Bergfors, T., ed.), International University Line, La Jolla, CA, pp. 139–153.

18. Saridakis, E., Shaw Stewart, P. D., Lloyd, L. F., and Blow, D. M. (1994) Phase diagram and dilution experiments in the crystallization of carboxypeptidase G2. *Acta Cryst.* **D50**, 293–297.
19. Saridakis, E. and Chayen, N. E. (2000) Improving protein crystal quality by decoupling nucleation and growth in vapor diffusion. *Protein Sci.* **9**, 755–757.
20. Chayen, N. E. (1997) The role of oil in macromolecular crystallization. *Structure* **5**, 1269–1274.
21. Mayans, O., van der Ven, P. F. M., Wilm, M., et al. (1998) Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* **395**, 863–869.
22. D'Arcy, A., MacSweeny, A., Stihle, M., and Haber, A (2003) The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Acta Cryst.* **D59**, 396–399.

## Three-Dimensional Crystallization of Membrane Proteins

James Féthière

### Summary

Although the examination of the protein data bank reveals an important backlog in the number of three-dimensional structures of membrane proteins, several recent successes are serving as preludes to what will become a very prosperous decade in this field. Systematic investigations of various factors affecting the stability of membrane proteins, as well as their potential to crystallize three dimensionally, have paved the way for such achievements. The importance of the role of detergents both at the level of purification and crystallization is now well established. In addition, the recognition of the protein–detergent complex as the entity to crystallize, as well as the understanding of its physical–chemical properties and discovery of factors affecting these, have permitted the design of better crystallization strategies. As a consequence of the various efforts in the field, new crystallization methods for membrane proteins are being implemented. These have already been successful and are expected to contribute significantly to the future successes. This chapter will review some basic principles in membrane protein crystallization and give an overview of the current state of the art in the field. Some practical guidelines to help the novice approach the problem of membrane protein crystallization from the initial step of protein purification to crystallogenesis will also be given.

**Key Words:** Membrane proteins; crystallization; detergent; cloud point; surfactant; protein purification.

### 1. Introduction

Membrane proteins are a class of proteins that play specific roles in the communication between biological compartments and allow these to take on specialized functions. Their importance is recognized by the fact that greater than 20–35% of all open reading frames of currently known genomes encode for membrane proteins and more than 70% of all known pharmacological targets are membrane proteins (*1,2*). In order to understand the molecular mechanisms

involved in regulating the function of this class of proteins and be able to interfere with the latter for therapeutic interests, high-resolution structures at near atomic scale are needed. This can be achieved by either of three complementary methods: electron microscopy (EM), X-ray crystallography, and nuclear magnetic resonance (NMR). Although the latter allows structure determination of larger and larger membrane proteins in their solubilized state, its strength is best demonstrated in the determination of structures of selected transmembrane sections of membrane proteins rather than their complete structure. The former two require the obtainment of well-ordered two-dimensional (2D) and three-dimensional (3D) crystals, respectively. Determination of membrane protein structure from 2D crystals by EM has the advantage of providing structural information on the protein in its native environment and requires less sample. Although reasonably high resolution can be achieved with this method, it is still limited by the lack of efficiency of data acquisition and processing methods (3). On the other hand, X-ray crystallography is a well-established and routine method that can provide 3D structures of membrane proteins at atomic-scale resolution. However, obtaining well-ordered 3D crystals of membrane proteins is quite a milestone. Whereas this task is becoming expeditious and routine for soluble proteins, as demonstrated by the several thousand structures available in the Protein Data Bank (PDB; [www.rcsb.org](http://www.rcsb.org)), it is still quite a challenging endeavor for membrane proteins. Indeed, according to the most recent survey maintained by the membrane protein structure group at the Max Planck Institute in Frankfurt, Germany (<http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html>), less than 1% of the structures determined in the PDB are membrane proteins. Two main bottlenecks hamper regular and continuous success in the determination of membrane protein structure: quantity of protein and crystallization. The former has been extensively addressed and important progress in the overexpression of recombinant prokaryotic membrane proteins has been made, which led to successful structure determination (4-6). However, progress is still limited for eukaryotic membrane proteins and several systems often need to be investigated before satisfactory levels of overexpression can be achieved to support the crystallization efforts (7). The difficulties associated with membrane protein crystallization are due to their intrinsic properties that render the crystallization process slightly different from that of their soluble counterparts. These proteins reside in the lipidic membrane with polar heads and feet facing the surrounding solvent and interacting with the polar head groups of the lipids. These hydrophilic extensions cap a hydrophobic domain traversing the membrane and making contact with the alkyl chains of the lipids. The latter accounts for the main barrier associated with structural studies of these proteins. The amphipathic character of these proteins makes them insoluble in aqueous solvent, and detergents have to be used to extract

them from the membrane and keep them soluble in aqueous solvent. Because detergents interact directly with the protein, they can strongly affect both the stability and the homogeneity of membrane protein preparations, providing an additional variable to be optimized, and requiring additional quality controls. Furthermore, whereas most 2D crystals for EM are obtained from a reconstituted lipidic bilayer by detergent removal (3), 3D crystals are grown from detergent micelles (8). This accounts for another of the main difficulties associated with structural studies. Two types of these 3D crystals can be obtained (9): type I crystals are stacks of 2D crystals that are stabilized by protein–protein and protein–lipid hydrophobic interactions in the plane of the membrane and by polar interactions between hydrophilic surfaces of the proteins, perpendicular to this plane. Type II crystals grow from micelle-incorporated proteins, and are built from the establishment of contacts strictly between the polar surfaces of the proteins that are not shielded by the detergent micelle. From the standpoint of type II crystallization, instead of working with a single entity, the protein, we are now manipulating a protein–detergent complex (PDC) that we need to keep stable and homogeneous to force it into a well-ordered 3D crystal array that will diffract X-rays to high resolution. In such a crystallization system, the usual five-parameter array for soluble protein crystallization (protein, precipitant, buffer, additive, water) is complicated by the addition of a sixth variable in membrane protein crystallization: the detergent. Several excellent reviews and methodology articles are regularly being written on the subject and follow the progress being made in the field (10–13). However, we still have a long way to go to reach the smooth pace at which soluble protein are being crystallized, and many other guides like this one will be needed to remind us of the basic principles in membrane protein crystallization and provide new tricks to facilitate the task.

## 2. Materials

1. Hi-trap ion exchange and metal chelating columns (Amersham Biosciences, Piscataway, NJ).
2. Centriprep and microcon protein concentrators (Amicon, Millipore, Billerica, MA).
3. Microdialyzer (Pierce, Rockford, IL).
4. Detergent family kits (Anatrace, Maumee, OH).
5. A series of lipids from Avanti Polar Lipids (Alabaster, AL) or Sigma (St. Louis, MO).
6. Membrane protein crystallization screens from Molecular Dimensions (Cambridshire, UK) and Hampton Research (Aliso Viejo, CA) (see Table 1 in Chapter 7 for additional protein crystallization screens).
7. 96-Well crystal quick, and low-profile crystallization plates (Greiner Bio-One, Essen, Germany).
8. 24-Well Limbro cell culture plates (Hampton Research).
9. 96-Well microbatch plates (Hampton Research).
10. Clear seal film for 96-well plate (Greiner Bio-One).

11. Micro-bridge from Hampton Research for sitting-drop setups.
12. Siliconized cover slides (Hampton Research).
13. Paraffin and silicon oil (Hampton Research).

### 3. Methods

#### 3.1. System Components

##### 3.1.1. Detergents

Detergents are small, amphiphilic molecules classified as surfactants. In analogy to lipids, which they replace in membrane protein solubilization, they are composed of a hydrophilic head group and a hydrophobic alkyl chain. Both lipids and detergents are known to self-associate to form different types of multimolecular assemblies. Whereas lipids will spontaneously associate to form multilamellar structures (mono- or bilayers) in water, detergents will assemble as micelles with their aliphatic chains oriented toward the interior of the micelle, shielded from the aqueous environment, and their polar heads in contact with the solvent. Detergents are primarily classified according to the nature of the hydrophilic head group and can be ionic, nonionic, or zwitterionic. Ionic detergents are rather denaturing. They bind strongly to membrane proteins and because of the charged nature of their head groups, the micellar structure and behavior will be strongly dependent on solution parameters such as ionic strength and pH. The potential large micelle size and charge repulsion effects in these detergents make them unattractive for membrane protein crystallization. On the other hand, nonionic detergents are less aggressive and allow the isolation of membrane proteins in a native state. Formation of the micellar phase is a concentration-dependent process that is driven by hydrophobic interactions. At a certain monomer concentration, known as the critical micelle concentration (CMC), detergent monomers assemble as micelles and start to effectively solubilize membrane proteins. The length of the alkyl chain and the solvent environment are the two main parameters that will affect the CMC of a particular detergent (14). Micelles are very dynamic structures that can adopt different size and shapes, which are dependent on the type, structure, concentration, and packing of the monomers, as well as on the solvent environment (15–17). In membrane protein crystallization we are working with a PDC, and the detergent portion of this complex can account for greater than 50% of the mass of the particle. This ring of detergent molecules covering the hydrophobic domain of the protein in a PDC can interfere with crystal lattice formation by preventing close approach of the particles for establishment of polar contacts. Therefore, because of this large contribution of the detergent to the properties of the particle, several parameters related to detergent structure and behavior should be considered during crystallization trials. Table 1 shows some characteristics of some commonly used detergents.

**Table 1**  
**Selection of Detergents Used in the Crystallization of Membrane Proteins**

Alkyl chain length (code name)	Detergent name	Mol. mass	CMC (mM, %)	Aggregation number	Micelle size MW (kDa), stokes radius(Å)
C <sub>6</sub> (HxOG)	<i>n</i> -hexyl-β-D-glucopyranoside	264.4	250, 6.6	–	–
C <sub>7</sub> (HpOG)	<i>n</i> -heptyl-β-D-glucopyranoside	278.4	70, 1.9	–	–
C <sub>8</sub> (β-OG)	<i>n</i> -octyl-β-D-glucopyranoside	272.4	18–20, 0.53	27–100	~8–25 <sup>a</sup> , 15
C <sub>8</sub> (β-OTG)	<i>n</i> -octyl-β-D-thiogluco-pyranoside	308.4	9, 0.28	–	–
C <sub>9</sub> (NG)	<i>n</i> -nonyl-β-D-glucopyranoside	306.4	6.5, 0.20	–	–
C <sub>9</sub> (NTG)	<i>n</i> -nonyl-β-D-thiogluco-pyranoside	322.4	2.9, 0.093	–	–
C <sub>12</sub> (DDG)	<i>n</i> -dodecyl-β-D-glucopyranoside	348.5	0.19, 0.0066	200 <sup>a</sup>	~70 <sup>a</sup>
C <sub>8</sub> (OM)	<i>n</i> -octyl-β-D-maltopyranoside	454.4	19.5, 0.89	47	~21, 15.5–21
C <sub>8</sub> (OTM)	<i>n</i> -octyl-β-D-thiomaltopyranoside	470.6	8.5, 0.40	–	–
C <sub>9</sub> (NM)	<i>n</i> -nonyl-β-D-maltopyranoside	468.5	6.0, 0.28	55	26
C <sub>9</sub> (NTM)	<i>n</i> -nonyl-β-D-thiomaltopyranoside	484.6	3.2, 0.15	–	–
C <sub>10</sub> (DM)	<i>n</i> -decyl-β-D-maltopyranoside	482.6	1.8, 0.087	69	33
C <sub>10</sub> (DTM)	<i>n</i> -decyl-β-D-thiomaltopyranoside	498.6	0.9, 0.045	–	–
C <sub>11</sub> (UDM)	<i>n</i> -undecyl-β-D-maltopyranoside	496.6	0.59, 0.029	74	37
C <sub>11</sub> (UTM)	<i>n</i> -undecyl-β-D-thiomaltopyranoside	512.7	0.21, 0.011	–	–
C <sub>12</sub> (DDM)	<i>n</i> -dodecyl-β-D-maltopyranoside	510.6	0.17, 0.0087	78–92	40–47, 29
C <sub>12</sub> (DDTM)	<i>n</i> -dodecyl-β-D-thiomaltopyranoside	526.6	0.05, 0.026	–	–
C <sub>6</sub> DAO	<i>n</i> -hexyl- <i>N,N</i> -dimethylamine- <i>N</i> -oxide	145.4	–	–	–
C <sub>10</sub> DAO	<i>n</i> -decyl- <i>N,N</i> -dimethylamine- <i>N</i> -oxide	201.4	10.4 <sup>b</sup> , 0.21	–	–
C <sub>12</sub> DAO, (LDAO)	<i>n</i> -dodecyl- <i>N,N</i> -dimethylamine- <i>N</i> -oxide	229.4	1–2, 0.023	76	17
MEGA-9	nonanoyl- <i>N</i> -methylglucamide	335.5	25, 0.84	–	–
MEGA-10	decanoyl- <i>N</i> -methylglucamide	349.5	6–7, 0.21	–	–

Table continues

**Table 1 (continued)**

Alkyl chain length (code name)	Detergent name	Mol. mass	CMC (mM, %)	Aggregation number	Micelle size MW (kDa), stokes radius(Å)
Cymal5	5-cyclohexyl-1-pentyl-β-D-maltoside	494.5	2.4–5, 0.12	66	33
Cymal-6	6-cyclohexyl-1-hexyl-β-D-maltoside	508.5	0.56, 0.028	63	~32
Cymal-7	7-cyclohexyl-1-heptyl-β-D-maltoside	522.5	0.19, 0.0099	–	–
C <sub>8</sub> E <sub>4</sub>	Polyoxyethylene(4) octylether	306	7–8.5 <sup>b</sup> , 0.22	82	~25
C <sub>8</sub> E <sub>5</sub>	Polyoxyethylene(5) octyl ether	350	4.3–9.2 <sup>c</sup> , 0.15	–	–
C <sub>8</sub> E <sub>6</sub>	Polyoxyethylene(6) octylether	394	9.9, 0.39	32	~13
C <sub>10</sub> E <sub>5</sub>	Polyoxyethylene(5) decylether	378	–	–	–
C <sub>12</sub> E <sub>8</sub>	Polyoxyethylene(8) dodecyl ether	539	0.09, 0.0048	90–120	~66
C <sub>12</sub> E <sub>9</sub> or Thesit	Polyoxyethylene(9) dodecyl ether	583	0.05, 0.003	–	~64
C <sub>8</sub> HESO	<i>n</i> -octyl-2-hydroxyethyl sulfoxide	206	30, 0.62	–	–
Zwittergent 3–12	<i>n</i> -dodecyl- <i>N,N</i> -dimethyl-3-ammonio- 1-propanesulfonate	335	2.8, 0.094	–	–

<sup>a</sup>Value taken from the Calbiochem detergent booklet

<sup>b</sup>Value from **ref. 44**.

<sup>c</sup>Value from **ref. 70**.

Most values are obtained from the Anatrace catalog unless otherwise stated.

The length of the alkyl chain determines the size of the micelle, and a detergent with short alkyl chains will form small micelles that should be optimal for membrane protein crystallization. In addition, because they interact with the hydrophilic head and feet of the protein, polar head groups of the detergents can also interfere with crystal lattice formation in a size-dependent fashion. Although these detergents with short alkyl chains may be optimal for crystallization, it should be noted that they also tend to be more denaturing. Therefore, when screening for the best detergent, one should take into consideration the effect on the stability of the protein and on the formation of the crystal lattice, and aim for an optimal alkyl chain length and size of a polar head group that will satisfy these constraints. Interestingly, to date, the highest number of membrane protein crystals has been obtained with *n*-octyl- $\beta$ -D-glucopyranoside (**Table 2**), an eight carbon alkyl chain length detergent that forms small micelles.

Detergents have distinct phase behavior, which have some relevancy to the crystallization process. Above the CMC, monomers and micelles exist in equilibrium and a clear isotropic phase indicates absence of interaction between micelles. However, it is possible to modify this phase behavior by forcing micellar interactions to the point where micelles will start to aggregate and the solution becomes turbid, eventually leading to two separate immiscible phases: one detergent-rich and one detergent-poor with the PDC concentrating in the former with precipitated detergent micelles (**11,16,18**). This point, termed the “cloud point,” is specific to each detergent and is temperature dependent. The characteristic boundary in the phase transition of a detergent is called the consolute phase boundary. An upper consolute phase boundary is observed when a lowering of the temperature causes the phase transition, and a lower consolute phase boundary when the temperature needs to be raised (**19**). The most common detergents used in membrane protein crystallization do not exhibit this characteristic phase transition in pure water and for the ones that do, it does not occur at a useful temperature. However, modifications of the same solution parameters that influence the crystallization process such as salts, precipitant, additives, and impurities will induce this clouding behavior and shift the phase boundary to a different temperature range (**18–20**). This phase separation is detrimental to crystal growth, and the fact that many membrane protein crystals appear close to these boundaries is indicative of the role played by micellar interactions in the crystallization process (**21–24**). Therefore, one should try to optimize the experimental conditions to approach this phase boundary delicately and avoid the phase separation. A systematic investigation of the role of this phase behavior in the crystallization of the P-type ATPase indicates that one should find conditions that satisfy both micelle–micelle and protein–protein interactions for successful crystallization (**24**).

**Table 2**  
**Crystallization Conditions for Selected Membrane Proteins in Detergent Micelles**

Protein	Source	Precipitant	Salt	pH	Additives	Detergents	xCMC	T°C	Method	Ref. <sup>f</sup>
Reaction center	<i>Rhodopseudomonas viridis</i>	AS 2–3 M	—	6	HT 3%,	LDAO 0.5% TEA 3%	22	RT	VD	(26)
Reaction center	<i>Thermochromatium tepidum</i>	PEG 4K 25–45%	NaCl 360 mM	7	NaAzide 0.1%, EDTA 0.1 mM	β-OG 15 mM	1.3	4	VD	(71)
Reaction center	<i>Rhodobacter sphaeroides</i>	KP 1.6 M	—	7.5	Dioxane 4.2%, HT 7.35%	LDAO 0.09%	3.9	25	VD	(72)
Reaction center	<i>R. sphaeroides</i>	KP 1.4 M	—	7.0	Dioxane 1%, HT 3%	LDAO 0.09%	3.9	18	VD	(73)
LH2	<i>Rhodopseudomonas acidophilia</i>	AS 2.2 M	Phosphate 0.9 M	9.5	Benzamidine hydrochl. 2.5%	β-OG 1%	1.9	20	VD	(74)
LH2	<i>Rhodospirillum molischianum</i>	AS 3.0– 3.3 M	KP 150 mM	6.5	HT 3.2%, 0.1% NaAzide	UDAO 0.2%	—	20	VD	(75)
PS-I	<i>Synedococcus</i> sp.	Salting in	MgSO <sub>4</sub> 0.1 M to 6 mM	6.4	—	DDM 0.02%	2.3	4	Dial	(76,77)
PS-II	<i>Synechococcus vulcanus</i>	PEG 1450 6–7%	MgSO <sub>4</sub> 40 mM, CaCl <sub>2</sub> , NaCl 20 mM	6.5	—	DDM 0.02%	2.3	20	Dial	(78)
Cytochrome c oxidase	Bovine heart mitochondria	PEG 4K	—	6.8	—	DM 0.2%	2.3	4	Batch	(79–80)
Cytochrome c oxidase <sup>a</sup>	<i>Paracoccus denitrificans</i>	PEGMME 2K, 12%	NH <sub>4</sub> Ac 400 mM	8	NaAzide 0.1%	DDM	—	20	VD	(58,59)
ba3-cytochrome c oxidase	<i>Thermus thermophilus</i>	PEG 2K 6%	—	7	—	NG 0.4%	2	20	Batch	(81)
Cytochrome bc1 complex	Bovine heart mitochondria	PEG 4K 18%	KCl 0.5 M	7.0	Glycerol 10–20%	MEGA-10/ SPC 0.1%	0.5/	0–4	Batch	(82,83)

Cytochrome <i>bc1</i> complex	Bovine heart mitochondria	PEG 4K 6–8%	NaCl 100 mM	6.8	—	DDM 0.015%	1.7	4	VD	(84)
Cytochrome <i>bc1</i> complex <sup>a</sup>	<i>Saccharomyces cerevisiae</i>	PEG 4K 5–6%	—	8	—	UDM 0.05%	1.7	—	VD	(85)
Ubiquinol oxidase <i>bo3</i>	<i>Escherichia coli</i>	PEG 1500 9–10%	NaCl 100 mM, MgCl <sub>2</sub> 100 mM	7.5	—	β-OG 1%, 5% EtOH	1.9	4	VD	(86,87)
Fumarate reductase	<i>E. coli</i>	PEG 10K 10%	MgAc 85 mM	5.8	EDTA 0.1 mM, 0.001% DTT	C <sub>12</sub> E <sub>9</sub> 0.7%	233	—	VD	(88)
Fumarate reductase	<i>Wolinella succinogenes</i>	PEG 3350 10%	NaCl 150 mM	6.4	K <sub>3</sub> Fe(CN) <sub>6</sub> benzamidine 2.4% DMF 5% DMN 1 mM Malonate 1 mM	DDM 0.05% DM 0.2%	6, 2.3	—	VD	(89)
COX2, prostaglandin H2 synthase <sup>b</sup>	Human	PEG 4K 17–22%	NaCl 300–650 mM	6.7	—	C <sub>8</sub> E <sub>5</sub> 0.5%	3.3	—	VD	(90)
COX2, prostaglandin H2 synthase <sup>b</sup>	Murine	PEGMME 550 20–34%	NaCl 100 mM MgCl <sub>2</sub> 10–240 mM	8.0	—	β-OG 0.6%	1.1	18	VD	(91)
COX 1 prostaglandin H2 synthase <sup>1b</sup>	Sheep	PEG 4K 4–8%	NaCl 100–200 mM	6.7	—	β-OG 0.6%	1.1	—	VD	(92)
Squalene cyclase	<i>Alicyclobacillus acidocaldarius</i>	Na Citrate 0.1 M		4.8	0.01% LDAO	C <sub>8</sub> E <sub>4</sub> 0.3%	1.4	—	VD	(93,94)
Porin ompF	<i>E. coli</i>	PEG 2K 10.5%	MgCl <sub>2</sub> 700 mM	9.8	—	C <sub>8</sub> HESO 0.6% C <sub>8</sub> E <sub>4</sub> 0.1%	1/0.45	—	Dial	(95)
Maltoporin ( <i>lamb</i> )	<i>Salmonella typhimurium</i>	PEG 1500 28–32%	—	7.5	—	C <sub>8</sub> E <sub>4</sub> 0.3 % 0.8% C <sub>6</sub> DAO	1.4	18	VD	(96)
Maltoporin ( <i>lamb</i> )	<i>E. coli</i>	PEG 2K 15–18%	MgCl <sub>2</sub> 100 mM	7	—	C <sub>12</sub> E <sub>9</sub> 0.1% DM 0.4%	33/4.6	—	RT	Dial (97)
Porin	<i>Rhodobacter capsulatus</i>	PEG 600 23–30%	LiCl 300 mM	—	—	C <sub>8</sub> E <sub>4</sub> 0.6%	2.7	20	VD	(98)

(Table continues)

**Table 2 (continued)**

Protein	Source	Precipitant	Salt	pH	Additives	Detergents	xCMC	T°C	Method	Ref. <sup>f</sup>
Porin ompX	<i>E. coli</i>	2-propanol 30%	CaCl <sub>2</sub> 200 mM	4.6	Glycerol 20%	C <sub>8</sub> E <sub>4</sub> 0.6%	2.7	RT	VD	(99)
Porin ompA	<i>E. coli</i>	PEG 8K 12%		5.1	MPD 10%	C <sub>8</sub> E <sub>4</sub> 0.6%	2.7	20	VD	(100)
Porin ompK36	<i>Klebsiella pneumoniae</i>	PEG 2K 15%	MgCl <sub>2</sub> 500 mM	9.8	—	C <sub>8</sub> HESO 0.6% C <sub>8</sub> E <sub>4</sub> 0.1%	1/0.45		Dial	(101)
Porin omp32	<i>Comamonas acidovorans</i>	Li <sub>2</sub> SO <sub>4</sub> 1.3–1.4 M		7.5	—	β-OG 2%	3.7	17	VD	(102)
Porin ompT	<i>E. coli</i>	MPD 28%	NaCl 0.5 M	5.5	—	β-OG 1%	1.9	4	VD	(103)
OpcA	<i>Neisseria meningitidis</i>	PEG 4K 10–20%	ZnAc 150 mM ZnCl <sub>2</sub> 50 mM	7.5	HpOG 0.5%	C <sub>10</sub> E <sub>5</sub> 1%	—	—	VD	(104)
Sucrose porin ScrY	<i>S. typhimurium</i>	PEG 2K 12–15%	MgSO <sub>4</sub> 20 mM, LiCl 100 mM	7.7	—	β-OG 1.2% C <sub>6</sub> DAO 1% HpOG 1%	2.3/ 0.52	17	VD	(105, 106)
FhuA	<i>E. coli</i>	PEGMME 2K 11%	—	—	PEG 200 3%, cis-inositol 1% Glycerol 20%	C <sub>10</sub> DAO 0.8%	3.8	18	VD	(107, 108)
FepA	<i>E. coli</i>	PEG 1000 28%	NaCl 350 mM	8	Glycerol 10%, NaN <sub>3</sub> 2 mM, HT 1.75%	LDAO 0.06%	2.6	21	VD	(109, 110)
MalFGK <sub>2</sub>	<i>Thermococcus litoralis</i>	PEG 13–14%	Ca-acetate 400 mM	6.5	Glycerol 20%	DDM 0.05%	5.7	18	VD	(111)
Type III ADH <sup>b</sup>	<i>Gluconobacter suboxydans</i>	PEG 3350 6%	AS 150 mM	4.5	—	DDM 0.34 mM	2	4	VD	(112)
Fdh-N	<i>E. coli</i>	PEG 1500 6–12%	NaCl 10 mM	7.5–8.2	EtOH 5%	β-OG 1%	1.9	4	VD	(113)

GlpF	<i>E. coli</i>	PEG 2K 28%	MgCl <sub>2</sub> 300 mM	8.9	Glycerol 15% DTT 5 mM	β-OG 35 mM	1.8	—	VD	(114)
APQ1 <sup>c</sup>	Bovine RBC	PEGMME 550 20%		7.5	—	NG 13 mM	2	4	VD	(115, 116)
MscL	<i>Mycobacterium tuberculosis</i>	TEG 23–27%	AS 100–120 mM	3.6–3.8	GdCl <sub>3</sub> 1–3 mM or SmCl <sub>3</sub> (in D <sub>2</sub> O)	DDM 0.05%	5.7	4	VD	(5)
SERCA1a <sup>d</sup>	Rabbit sarcoplasmic reticulum	Na butyrate 0.8 M	CaCl <sub>2</sub> 10 mM, MgCl <sub>2</sub> 3 mM	6.1	NaN <sub>3</sub> 2.5 mM DTT 0.2 mM, 2.75 M glycerol	—	—	—	—	(46)
TolC	<i>E. coli</i>	PEGMME 2K 12.5%	NaCl 400 mM MgCl <sub>2</sub> 20 mM	7.4	PEG 400 10% HT 1.5%	β-OG, DDG, HxOG, HpOG 0.6%	—	25	VD	(117)
AcrB	<i>E. coli</i>	PEG 4K 7%	KNO <sub>3</sub> 20– 50 mM	5.6–6.5	Glycerol 10% DTT 20 mM	DDM 0.1%	11.4	25	VD	(118)
KcsA	<i>Streptomyces lividans</i>	PEG 400 48%	CaCl <sub>2</sub> 200 mM KCl 150 mM	7.5	DTT 2 mM	LDAO 5 mM	3.3	20	VD	(6)
ClC chloride channel	<i>E. coli</i>	PEG 400 34%	Na <sub>2</sub> SO <sub>4</sub> 50 mM Li <sub>2</sub> SO <sub>4</sub> 50 mM	8.5	—	OM 45 mM	2.4	20	VD	(119)
ClC chloride channel	<i>S. typhimurium</i>	PEG 400 26–31%	Na <sub>2</sub> SO <sub>4</sub> 75–100 mM Li <sub>2</sub> SO <sub>4</sub> 75–100 mM	4.6	—	OM 45 mM	2.4	20	VD	(119)
KvAP	<i>Aeropyrum pernix</i>	PEG 400 16–20% or MME 350	CdCl <sub>2</sub> 150– 200 mM KCl 100 mM	5.0	—	β-OG 30 mM	1.6	20	VD	(120)
MthK Ca-gated K channel	<i>Methanobacterium thermoautotrophicum</i>	PEG 350- MME 23–26%	CaCl <sub>2</sub> 200 mM, KCl 100 mM	6.5	—	LDAO 5 mM	3.3	20	VD	(121)
BR	<i>Halobacterium halobium</i>	AS 2.4–3.1 M	AP 750 mM	5.6	Benzamidine 1% D,L-pipecolic ac. 1%	β-OG 1%	1.9	4	VD	(122, 123)

(Table continues)

**Table 2 (continued)**

Protein	Source	Precipitant	Salt	pH	Additives	Detergents	xCMC	T°C	Method	Ref. <sup>f</sup>
Rhodopsin	Bovine	AS 3–3.4 M	ZnAc 65–90 mM	6.0–6.4	$\beta$ -ME 5–7 mM HT 0.55–0.75%	NG 0.45–0.55%	2.2–2.7	—	VD	(47)
OMPLA	<i>E. coli</i>	MPD 25–29%	CaCl <sub>2</sub> 1 mM	5.9	—	$\beta$ -OG 1.5%	2.8	22	VD	(124)
MAO-B <sup>e</sup>	Human	PEG 4K 12%	Li <sub>2</sub> SO <sub>4</sub> 70 mM, KP 25 mM	6.5	—	LDAO 2.6 mM or Zwittergent 3–12 8.5 mM	1.7/3	4	VD	(125)
BtuCD	<i>E. coli</i>	PEG 2K 21%	Mg-nitrate 0.3 M	8.0	MDP 0.8% (D <sub>2</sub> O)	LDAO 1%	43	—	VD	(4)
$\alpha$ -hemolysin	<i>Staphylococcus aureus</i>	AS 2.5 M PEGMME– 5K 0.25%	Na-cacodylate 75 mM	7.4	—	$\beta$ -OG 25 mM	1.3	RT	VD	(126)
LacY	<i>E. coli</i>	PEG 400 27–30%	CaCl <sub>2</sub> 200 mM	7.0	1,6-hexanediol 3%	CHAPS 0.8 mM DDM 0.01%	0.1/1.1	20	VD	(127)
GlpT	<i>E. coli</i>	PEGMME 2K 25–27%	NaCl 5– 100 mM	8.5– 8.9	SrCl <sub>2</sub> or MgCl <sub>2</sub> 5 mM MPD 5%, glycerol 20%	DDM 0.1– 0.25%, C <sub>12</sub> E <sub>9</sub> 0.04–0.1%	11–29/ 13–33	15–20	—	(48)
FAAH <sup>b</sup>	Rat	PEG 6K 7%	Li <sub>2</sub> SO <sub>4</sub> 100 mM	5.0	MPD 5%, DTT 0–20 mM	LDAO 2 mM	1.3	—	VD	(128)
PSII	<i>Thermosynechococcus vulcanus</i>	PEG 1450 6–7%	MgSO <sub>4</sub> 40 mM, CaCl <sub>2</sub> 20 mM NaCl 20 mM	6.0	—	DDM 0.02%	2.3	20	Dial	(78)

Fdn-N	<i>E. coli</i>	PEG 1500 6–12%	NaCl 100 mM, MgCl <sub>2</sub> 100 mM	7.5– 8.2	EtOH 5%	β-OG 1%	1.9	4	VD	(129)
c10F <sub>0</sub> F <sub>1</sub> -ATP synthase	<i>S. cerevisiae</i>	PEG 6K 12%	NaCl 150 mM, MgCl <sub>2</sub> 2 mM	8.0	AMP-PNP 1 mM, ADP 40 μM, DTT 1 mM, trehalose 50 mM, glycerol 10%, EDTA 1 mM	DDM 0.1%	11	4	MB	(130)
NarGHI	<i>E. coli</i>	PEG 3K 20%	NaAc 200 mM KCl 200 mM	7.0	EDTA 5 mM	Thesit 0.7 mM	14	—	VD	(131)

<sup>a</sup>Antibody complex crystals.

<sup>b</sup>Monotopic membrane proteins crossing the membrane only once.

<sup>c</sup>Crystals obtained from deglycosylated protein.

<sup>d</sup>Type I crystals in the absence of detergents.

<sup>e</sup>Not clearly classified as a membrane protein.

<sup>f</sup>References given correspond to papers where the crystallization conditions have been reported.

ADP, adenosine diphosphate; AMP-PNP, 5'-adenylyl-imidodiphosphate; AP, ammonium phosphate; AS, ammonium sulfate; ATP, adenosine triphosphate; β-OG, *n*-octyl-β-D-glucopyranoside; C<sub>10</sub>DAO, *n*-decyl-*N,N*-dimethylamine-*N*-oxide; C<sub>6</sub>DAO, *n*-hexyl-*N,N*-dimethylamine-*N*-oxide; C<sub>x</sub>E<sub>x</sub>, Polyoxyethylene(E), (C)Ether; C<sub>8</sub>HESO, *n*-octyl-2-hydroxyethyl sulfoxide; DDG, *n*-dodecyl-β-D-glucopyranoside; DDM, *n*-dodecyl-β-D-maltopyranoside; dial, dialysis; DM, *n*-decyl-β-D-maltopyranoside; DMF, dimethylformamide; DMN, dimethylnaphthoquinone; DTT, dithiothreitol; EtOH, ethanol; HT, 1,2,3-heptanetriol; HxOG, *n*-hexyl-β-D-glucopyranoside; HpOG, *n*-heptyl-β-D-glucopyranoside; KP, potassium phosphate; MEGA-10, decanoyl-*N*-methylglucamide; MME, monomethyl ether; MPD, 2-methyl-2,4,-pentanediol; NG, *n*-nonyl-β-D-glucopyranoside; OM, *n*-octyl-β-D-maltopyranoside; PEG, polyethylene glycol; SPC, short chain phosphatidyl choline; TEG, triethylene glycol; UDAO, *n*-undecyl-*N,N*-dimethylamine-*N*-oxide (C<sub>11</sub>DAO); UDM, *n*-undecyl-β-D-maltopyranoside; VD, vapor diffusion.

### 3.1.2. Special Additives

Knowing that micelle size and shape influence the close approach of the polar surfaces of the proteins in a PDC, manipulation of these parameters by the addition of special additives is expected to affect crystal lattice formation. The small amphiphile concept was proposed on the basis that detergent micelles might be too large to fit in the protein's crystal lattice (9). It was shown that the use of these small polar additives significantly reduced the size of certain detergent micelles (25), and their use was essential in obtaining crystals of a bacterial photosynthetic reaction center (26). A large number of these compounds were tested on bacteriorhodopsin crystals and the most effective was 1,2,3-heptanetriol. This small amphiphile (like many others) has limited effect on the CMC of several detergents used in membrane protein crystallization, but significantly affects the cloud point temperature, indicating a predominant effect on micellar interactions (9,19), which will eventually reflect on the establishment of polar interactions at the surface of the proteins for crystal lattice formation. This effect on the cloud point is not general, as these compounds have limited effect on the crystallization of membrane proteins in the presence of alkyl maltoside detergents (9,19). Attempts to generate new types of amphiphiles for solubilization and/or crystallization of membrane proteins have been limited. Recently however, new classes of compounds have been created and were shown to be efficient in stabilizing membrane proteins in detergent-free aqueous solution. Amphipols are synthetic polymers able to exchange for detergent and stabilize several membrane proteins by wrapping their hydrophobic side chains around the hydrophobic regions of the protein leaving their strong hydrophilic backbone to interact with the solvent (27,28). Tripod amphiphiles are conformationally rigid amphiphilic compounds built from a rigid quaternary carbon center that are able to solubilize and stabilize membrane proteins (29,30). Lipopeptide detergents are amphiphilic compounds built on a peptide scaffold supporting two alkyl chains (31). A proof of concept for the efficacy of these compounds came from the EM reconstruction of the  $F_0F_1$ -ATP synthase in the presence of amphipols (32), crystallization of the  $K^+$  channel (30) and stabilization of the chloroplast  $F_0F_1$ -ATP synthase (unpublished results) in the presence of tripod amphiphiles, and the stabilization of *lac*-permease by lipopeptides against aggregation (31).

## 3.2. Sample Preparation

### 3.2.1. Protein Purification and Characterization

As for soluble proteins, the initial step and primary requirement toward structure determination of a membrane protein is to obtain sufficient amounts of pure and well-characterized protein. If recombinant technologies are avail-

able for the protein being investigated, the use of affinity tags is a good advantage for purification and downstream processing, such as detergent exchange and refolding. These tags also allow rapid purification, minimizing the steps required to get pure protein. In addition, it offers the possibility of engineering the target to make it more stable and amenable to crystallization (*see Note 1*). The methods used should provide highly pure and homogeneous protein and be designed to minimize appearance of microheterogeneities created by either proteolysis, local denaturation, complex dissociation, variable stoichiometry of bound lipids, and posttranslational modifications. As membrane proteins require detergents for solubilization, one has to screen for the best detergent that will result in high yields of soluble and active protein. Apart from its efficiency in the extraction process, the main considerations in choosing a detergent for solubilization should be effect on bioactivity of the protein, compatibility with downstream purification schemes, ease of removal, and stability and solubility in the working conditions (**33**). An activity assay will help probe the integrity of the membrane protein in different detergents. Usually, the best detergents for solubilization are nonionic detergents with low CMCs. An excess of detergent, several folds above the CMC, is desirable for complete solubilization: in the range of 0.1 to 5% (v/v) for nonionic detergents. The concentration of membrane proteins should be kept above 1 mg/mL and the detergent:protein ratio around 10:1 to insure complete extraction of the lipids and formation of single protein–detergent complex per micelle. Once soluble and active protein has been obtained, classical chromatographic supports are used for further purification. The protein–detergent complex can be purified to homogeneity by density-gradient centrifugation, ion exchange, chromatofocusing, size exclusion, and affinity chromatography. Each of these methods have their pros and cons, and a combination is often preferable and necessary to get pure protein (*see Notes 2–4*). Detailed descriptions and usefulness of several protocols for membrane protein purification are available in **refs. 34** and **35**. Speed of purification is another important parameter for bioactivity. The advent of perfusion chromatography significantly reduces the time required for every purification steps, and enables one to rapidly get the protein in a more stable environment (**36**). Affinity purification with a specific ligand also guarantees that the active conformation of the protein has been purified and is especially handy if refolding has been performed (*see Note 5*). Resulting from their intrinsic nature, membrane proteins tend to form aggregates. Monodispersity being a stringent requirement for crystallization, size-exclusion chromatography should be performed to guarantee a monodisperse protein solution prior to crystallization. For this polishing step, care must be taken to choose the right analytical column with a good separation range, and the highest possible resolution (*see Note 6*). At every purification step, the level of purity should be monitored by

SDS-gel electrophoresis. Native gels and isoelectric focusing are extremely helpful in monitoring the homogeneity of the preparation and should be performed at the end of the procedure, after detergent exchange and after prolonged storage.

### 3.2.2. Protein Concentration and Storage

Once purified, the protein has to be concentrated prior to crystallization trials. At this stage, the solution should be exchanged for one with low buffer concentration (5–10 mM), no salt when possible (*see Note 7*), and a detergent concentrations one- to twofold above the CMC (*see Note 8*). One should aim for a protein concentration of approx 10 mg/mL or 10–100  $\mu$ M (**10**). Chromatography and ultrafiltration are the most common methods used for the concentration step. Many types of ultrafiltration devices exist. When starting from large volumes, centripreps (Amicon Inc.) are preferred to others (stir cells, centricon) because their geometry prevents concentration of the protein on the membrane, which might lead to irreversible aggregation. For volumes up to 500  $\mu$ L, use microcon filter devices (Amicon Inc.). The main drawback with these ultrafiltration devices is the risk of concentrating the detergent micelle. The use of filters with the highest molecular weight cut-off allowed by the protein being concentrated, or the exchange of the detergent for one with smaller micelle size or higher CMC, should help prevent this. To completely avoid this effect of the ultrafiltration devices, it is advisable, when possible, to use chromatographic techniques (affinity or ion exchange) followed by dialysis (*see Notes 9 and 10*). Be aware that additional delipidation can occur in a second chromatographic run and will alter the lipid–protein–detergent stoichiometry introducing some sample heterogeneity that can undermine the subsequent crystallization efforts (*see Note 11*).

Not every protein tolerates long-term storage at 4 or  $-20^{\circ}\text{C}$ . When storage is necessary, it is recommended to flash-freeze the sample in liquid nitrogen and store at  $-80^{\circ}\text{C}$ . Small aliquots of the concentrated protein in Eppendorf tubes are plunged into liquid nitrogen or alternatively, as suggested by Bergfors (**37**), drops of concentrated protein can be directly added to liquid nitrogen solution and transferred to  $-80^{\circ}\text{C}$ . In some cases, this procedure will affect the stability of the protein and addition of glycerol may be necessary. Store the protein at high concentration and avoid freeze–thaw cycles. Activity of the protein and reproducibility of crystals should be verified after thawing.

### 3.2.3. Detergent Exchange

As discussed earlier, the detergent type is probably the most important variable in the crystallization of membrane proteins. Stabilization of membrane proteins in a PDC is affected by the length of the alkyl chain of the detergent (**8**) and for certain membrane proteins, a one-carbon difference in the alkyl

chain length of a particular family of detergent can make the difference between crystal and no crystals (38,39). Therefore, thorough screening of detergents should be performed during crystallization trials. The best way to achieve complete exchange is by ion exchange or affinity chromatography. Dialysis and size-exclusion chromatography are rather inefficient because they often do not permit complete detergent exchange, especially for large micelles. On the other hand, repeated cycles of dilution/concentration with ultrafiltration devices suffer from the same limitations as the one mentioned at the concentration step, namely concentration of the detergent (*see Note 12*). With the appropriate chromatographic support, thorough washing of the protein bound to the column will warrant complete exchange. However, the extent of this washing step should take into consideration the delipidation occurring while the protein is bound to the column, and one should limit the volume to the minimum required for complete exchange. Usually, 5- to 10-column volume should suffice. A one-step elution with high-eluant concentration should be performed to avoid diluting the protein.

### 3.3. Crystallization

#### 3.3.1. Classical

Crystallization of membrane proteins is a challenging and time-consuming endeavor (11). However, it is not an insurmountable one, and similar crystallization techniques applied to soluble proteins can be used. When crystallizing a membrane protein, multiple physicochemical parameters have to be screened. However, the difficulty in this field comes from the additional level of parameterization introduced by the detergent. Each of the traditional parameters has to be screened against a whole array of detergents. Although choosing the best detergent for crystallization remains a trial-and-error procedure, the best initial candidates should correlate with the ones in which the protein was still active at the solubilization test stage and have moderate-to-high CMCs because these will form small micelles. The protein should stay soluble during the length of the crystallization process, and therefore sufficient detergent should be present. Working at concentrations one- to twofold the CMC is a good starting point. With the appropriate precipitant, crystallization will occur by driving the protein-detergent complex/solvent system into a state of reduced solubility where a certain degree of supersaturation will be attained and polar contacts between neighboring molecules will be established for lattice formation to occur. Additional considerations in membrane protein crystallization are related to the physical properties of the detergent used and the effect of solvent environment on these properties relevant to their role in the crystallization process (such as CMC and cloud point). If some phase separation is observed in certain drops, a more systematic investigation around this condition aiming at controlling more

tightly the micellar attraction should be performed. Solvent components and physical conditions (precipitating agents, temperature, small amphiphiles, and others) that will affect the consolute-phase boundary should be investigated (18,22,40). Precise investigation of the phase behavior of a particular detergent in different crystallization conditions will help to better rationalize the screening process. The aim is to bring the PDC into an area that will promote optimal micelle–micelle and protein–protein interaction for crystallization to occur. This strategy was successfully applied to the crystallization of  $\alpha$ -hemolysin (41), and the P-type ATPase (24). Another consideration is the crystallization method. As mentioned in the crystallization chapter by T. Bergfors (*see* Chapter 7), crystallization drops can be set with different methods: vapor diffusion in the sitting- or hanging-drop configuration, microbatch under different type of oils, and dialysis. The hanging-drop method suffers from the drawback of the reduced surface tension on the cover slide due to the detergent, which limits the volume of the drop; therefore, it is better to perform vapor diffusion in the sitting-drop configuration. In the microbatch under oil and dialysis method, equilibration between drop and reservoir is very fast and might be a problem for crystal growth. However, the advantage of these methods is that the concentration of the components does not vary, and detrimental effects of high concentration of detergent can be ruled out during the crystallization process. Detergent cost might be an issue in the dialysis method because, in contrast to the vapor diffusion technique, it has to be present both in the drop and the reservoir. It has been suggested that membrane protein crystallization in oil might be driven by a slow diffusion of the detergent from the protein solution to the surrounding oil, thereby reducing the solubility of the protein (42). However, it was recently demonstrated that no significant loss occurs from dissolution of the detergent into the oil. It is thus safe to assume that the detergent concentration remains constant during the crystallization process (43).

According to the reported data in the literature (*see* Table 2), most membrane proteins have been crystallized in PEG solutions with or without the addition of salts. However, in the initial stages of screening, it is advisable to get a broad idea of the behavior of the protein in different precipitants (“know your protein, know your PDC”). Using the grid-screening strategy will provide valuable information that will permit a better use of the crystallization data we will obtain from the larger scale commercial screens. Set up PEG, PEG/salt, or salt screens vs pH using different molecular weight PEGs and different salts and a pH interval of 0.5 unit. Perform these screens with a drop size of 1 + 1  $\mu$ L (protein + reservoir), and keep a thorough record of the results for future reference. Thereafter, set up the commercial screens. For starters, membrane protein-specific screens (Jena Biosciences, Hampton Research, Molecular Dimensions) should be used because they have been specifically designed for these proteins.

However, even these do not completely cover the whole crystallization space, and a more systematic screening based on knowledge about the properties of the protein (pI, stabilization factors, temperature sensitivity, and so on) and the detergent should be performed. A strategy that provides a direct comparison of different detergents for the same crystallization condition is presented in the methodology section. Incubating crystallization plates at different temperatures will provide some information on the effect of the different crystallization conditions on the consolute boundary of the detergents and help us rationalize the design of further crystallization experiments. A single of these membrane protein screens can generate a few others by investigating different temperatures for several detergents with or without amphiphiles added to the protein stock. If sample availability is not a problem, the other additional commercial screens could also be set up (Jena Biosciences, Hampton Research, deCODE Genetics, Molecular Dimensions). However, one should be aware that many of these screens contain conditions that are not membrane protein-friendly, such as high PEG concentrations and organic solvents, and one should consider diluting the precipitant in these.

Very often, other components are needed to induce crystallization. Additives are used to manipulate protein–solvent and protein–protein interactions, and also to stabilize the protein (specific ligand). In the membrane protein field, however, some have an additional role in the modification of the micelle shape and size. Additives that have a specific effect in membrane protein crystallization include other detergents with different micelle size (to form mixed micelles), small amphiphiles, and organic solvents. They often need to be added to the protein stock solution before setting up the crystallization experiment. One additive that has been particularly successful for membrane proteins is 1,2,3-heptanetriol. Considering their effect on the cloud point, these small amphiphilic compounds should also be screened at different temperatures. Commercial additive screens, as well as detergent screens, are available and should be used during the search for crystallization conditions and the optimization stage. One type of additive that has not been extensively investigated is the lipids. It is difficult to control the amount of lipids that remain bound to the protein after purification, and it was previously believed that total delipidation was preferable for successful crystallization (44,45). However, in some instances, their presence was necessary to obtain crystals (46–49). For certain cases, it may be necessary to have lipids present during the whole process of solubilization and purification as a stabilizing factor (50). Therefore, it is considered a good practice to verify the lipid content of the purified protein using thin layer chromatography and an enzymatic phosphorous assay for quantification. The results of these measurements should be correlated with activity assays and the homogeneity and monodispersity of the sample. During the

screening for different additives, different lipids should also be investigated. The source of lipids added need not to be from the original membrane as shown for the cytochrome *b6f* (51).

### 3.3.2. *New Development*

Considering the importance of lipids in the stabilization and crystallization of membrane proteins, the group of Rosenbusch in Basel has proposed a crystallization method in which the protein would be kept in a membranous environment throughout crystallization (52). It was observed that the lipidic cubic phase could reproduce the lateral pressure that the protein experiences in the native membrane in contrast to the reduced constraints experienced in a detergent micelle, resulting in less conformational freedom of the protein. This cubic phase is a continuous 3D curved lipid bilayer in which the membrane protein incorporates and diffuses laterally to establish contacts between apolar domains of neighboring proteins; polar contacts in the third dimension are also established to create this ordered stack characteristic of type I 3D-crystals. High-resolution diffracting crystals of bacteriorhodopsin (53,54) and halorhodopsin (55) have been obtained with this technique. In order to form this cubic phase, the dry lipid (most commonly, mono-olein [see Note 13]) is mixed with aqueous buffer containing the protein detergent solution by centrifugation or with an extruder (12) until formation of the cubic phase (see Note 14), as indicated by the transparent nonbirefringent gel-like appearance of the material. Crystallization is initiated by addition of the precipitant to the preformed cubic phase. Alternatively, all components necessary for crystallization can be mixed with the lipid prior to formation of the cubic phase. One should be aware that the components of the crystallization mixture (precipitant, detergent, buffer, and so on) can affect the formation of the cubic phase, and their effects should be investigated beforehand. A systematic investigation of the effect of certain commercial crystallization screens has started (56), and showed varying compatibility with the cubic phase at different temperatures. Crystals obtained from the cubic phase can be harvested with cryoloops after mechanical disruption of the cubic phase, or after enzymatic or detergent treatment of the cubic phase (57) and flash-frozen for data collection (see Note 15). Currently, several membrane proteins have been crystallized with this method, and continuous efforts in improving the experimental setup, investigation of new lipids, and the refinement of our understanding of how membrane protein crystals grow in this system will soon make this new methodology more attractive for membrane protein crystallization.

Most membrane protein crystals are of type II, with crystal contacts provided by the polar surfaces of the protein exposed to the aqueous solvent. Membrane proteins have various ratios of hydrophobic/hydrophilic surface

areas, and for many, this ratio favors the hydrophobic transmembrane domains. The presence of such small hydrophilic area is a barrier to the productive establishment of contacts for crystal lattice formation. In addition, the detergent ring shielding the transmembrane domains can be seen as a steric wall preventing the close approach of these hydrophilic groups. Therefore, strategies that will expand the hydrophilic surface area should facilitate the establishment of these desired contacts for crystallization. One strategy that is being employed is to use an intermediary such as a specific high-affinity antibody. This strategy pioneered by the group of Michel in Frankfurt, Germany (58), relies on generating Fab fragments from monoclonal antibodies recognizing discontinuous epitopes on the target protein, with high affinity. It has proven very successful in getting crystals and improving diffraction limits of various membrane proteins (59–62). In addition, because of the high specificity of the antigen–antibody interaction, these antibodies are valuable purification tools providing often a single-step immunoaffinity purification protocol. Another way to increase the polar surface of a membrane protein for crystallization purposes is to form complexes with partner soluble proteins. This can be done after purification of both partners or by expression of a fusion protein. Similar attempts have been investigated for the  $\beta$ -adrenergic receptor (63), the *lac* permease (64), and cytochrome *bo3* ubiquinol oxidase (65).

When crystals are obtained, their diffraction quality will be assessed by X-ray diffraction analysis, and lead our search for improved crystallization conditions. It is routine these days to collect X-ray diffraction data from flash-frozen crystals previously cryoprotected by some modification of the mother liquor (66). However, finding the right cryoprotection protocol for membrane proteins can be difficult, and diffraction quality should be determined at room temperature or 4°C in a capillary before cryogenic data collection. It is also a good practice to verify the content of the crystal by gel electrophoresis or mass spectrometry as detergent crystals can sometime be mistaken for protein (*see Note 16*).

### 3.4. General Methodology

#### 3.4.1. Purification

The purification protocol will be protein dependent, and one should optimize the solubilization procedure for high yield and purity of active protein.

1. Test different detergents for optimal solubilization and/or stabilization. An example of detergent screening for solubilization is provided in the chapter on the glutamate receptor by Madden and Safferling (*see Chapter 3*).
2. In certain cases, a combination of detergents might be better suited for optimal solubilization.
3. Insure that the protein is homogeneous and monodisperse using analytical size-exclusion chromatography or dynamic light scattering (*see Chapter 6*). This is

especially important when screening the different detergents. An alternative method for this has been proposed and consists in verifying sample homogeneity by negative-stain electron microscopy (67).

4. Keep in mind that in particular cases, additives used during solubilization can improve the yields of active protein substantially. Addition of lipids, small amphiphiles, specific ligands, or inhibitors should be tested.
5. In preparation for the screening of different detergents for crystallization, at the end of the purification protocol, prior to concentration, split the protein in different aliquots. If only ion exchange is used (i.e., no tags are present for metal or immunoaffinity chromatography), dilute the aliquots to decrease salt concentration (see **Note 17**). To perform concentration and detergent exchange simultaneously, pre-equilibrate a hi-trap column (the type will depend on the binding mode; a good selection is available from Amersham Biosciences) with the target detergent in the appropriate buffer. Load the protein and wash with 5- to 10-column volumes of buffer containing the new detergent. Elute with four to five 500- $\mu$ L pulses of eluent containing the target detergent at the desired concentration and collect in separate tubes. The concentrated protein in the new detergent should be in fractions two and three for a 1-mL hi-trap column.
6. It is a good idea to run a native gel to verify the aggregation state of the protein in the new detergent before crystallization.
7. Dialyze fractions containing protein against buffer of low ionic strength (5–10 mM) containing detergent at one- to twofold the CMC and any additive required to stabilize the protein.

### 3.4.2. Crystallization of a PDC

1. Prior to setup, centrifuge the samples in an ultracentrifuge for 30 min at 160,000g and transfer the protein in a new tube, taking care not to disturb any pellet that might have formed from aggregated protein.
2. With the protein available in three different detergents, start by setting up grid screens of PEGs, PEG + salt, and salt vs pH. Then, set the commercial membrane protein screens (MembFac from Hampton Research, MemStart and MemSys from Molecular Dimensions Ltd., JBS membrane screens from Jena Biosciences) in the 96-well format plates. In vapor diffusion, use the crystal-quick plates to allow direct comparison of the different detergents.
3. Try to set up all the screens with the same protein preparation and keep an aliquot frozen. Different protein stocks may behave differently. Therefore, if you obtain crystals from a particular preparation, make sure that you have an aliquot stored away in case the crystals cannot be reproduced from another purification.
4. Setups should be performed with 1  $\mu$ L protein and 1  $\mu$ L reservoir against a 150  $\mu$ L reservoir solution. It is not necessary to add detergent to the reservoir solution. Three wells are available per reservoir and allow the simultaneous screening of three different detergents (e.g., same head group and different chain length). Prepare plates in triplicate, seal with the clear seal film, and store at 4, 16, and 22°C.
5. For microbatch crystallization under oil, fill the reservoirs of a 96-well microbatch plate with 200  $\mu$ L paraffin oil. Dispense 1  $\mu$ L protein solution and 1  $\mu$ L crystal-

lization screen solution at the bottom of the well in the oil. Make sure the drops coalesce. Cover the plates and incubate at controlled temperatures.

6. After 1 d, observe the plates and record conditions where phase separation has occurred, as well as conditions where precipitate and/or crystals are visible. Categorize these conditions according to precipitant, salt, pH, and temperature.
7. For wells where crystals appeared, reproduce the condition, and screen different additives. The additive screens from Hampton Research and Molecular Dimensions are good starting points but are somewhat incomplete for membrane proteins. One should also screen for the effect of polar additives to evaluate the small amphiphile concept. In addition to the new ones mentioned in **Subheading 3.1.2.**, a good list of these can be found in **refs. 9 and 41**. Nonvolatile additives can be added to the crystallization drop only, whereas volatile additives should also be present in the reservoir. The effect of amphiphilic additives should be investigated at different temperatures because they will affect the phase boundary of the detergent. Do not forget to use lipids as additives.
8. In addition, the detergent screen from Hampton Research should be used as an additive screen to verify the influence of mixed micelles on the quality of the crystals.
9. For conditions in which phase separation occurred, one should screen pure detergent solutions against varying concentrations of salt and precipitant, in order to identify the phase separation boundary conditions.
10. The result of this latter screen are then used to design a new crystallization screen for the PDC with concentrations of precipitant and/or salt slightly lower than the one that induced phase separation of the pure detergent micelle. These conditions should be optimal for crystallization of the PDC away from the cloud point.

#### 4. Notes

1. If one were having difficulties crystallizing a particular membrane protein, a surrogate to mutagenesis would be to look for a different isoform or a homolog in a different species. You might improve not only the expression levels, but also the ease of crystallization (**4–6**).
2. In the purification procedure, a good initial step is to perform a stringent wash of the membrane fraction to remove tightly bound peripheral proteins. Use urea-alkaline treatment, high salt, and EDTA washes (**68**).
3. Ion exchange and chromatofocusing have been shown to significantly reduce the amount of lipids resulting in destabilization of membrane proteins.
4. When working with large membrane protein complexes inserted in detergent micelles, there is a size-exclusion limit of some ion-exchange matrices not initially intended for this type of separation. These large protein–detergent complexes will not have access to the entire volume of the particles for binding and, therefore, will bind only to the surface of the beads, resulting in a reduced capacity of the columns.
5. When purifying by affinity with specific ligands, binding is often very tight and extreme conditions (very low pH), not very protein-friendly, are required to elute the protein from the column. This might have some irreversible effect on the active conformation of the protein.

6. Size-exclusion chromatography is a useful method to separate protein aggregates that membrane proteins tend to form. One should be aware of the low resolution of this method when working with protein–detergent micelles. It is preferable to work with detergent having a small micelle size (the protein should be stable in such detergents).
7. During concentration many proteins precipitate and various amounts of salt might be necessary to maintain solubility. In this case include a small amount of NaCl (25 mM or higher, using the minimum required).
8. At the final protein concentration step, the amount of detergent present should take into consideration the dilution factor that will be introduced when forming the crystallization drops.
9. When concentrating by chromatography, use the smallest quantity of resin that will allow binding of the whole amount of protein to achieve the highest concentration possible. Elution usually requires one- to two-column volume. For dialysis, use the small microdialyzers from Pierce. They have a maximal capacity of 100–200  $\mu$ L and come in different molecular weight cutoffs. In addition, if one needs to perform an additional concentration, these devices can be placed in a dish containing a dry powder, such as substituted carboxymethyl cellulose (e.g., Aquacide<sup>®</sup> from Calbiochem) or PEG 20K as an alternative. These compounds will absorb the water through the dialysis membrane, and concentrate the sample further. This procedure requires no extra manipulation of the sample and is rapid.
10. If one has to screen for detergent at the crystallization stage and milligram quantities of proteins are available, the pure material should be divided prior to protein concentration and each aliquot concentrated on a small ion-exchange column pre-equilibrated with the target detergent.
11. Because of the many possibilities of modifying the protein during purification, it is advisable to always apply the same methods of purification, concentration, and detergent exchange between preparations. This is especially important when the initial crystallization setup has been successful.
12. One should be careful not to have too high a detergent concentration in the stock protein because of the risk of having too many empty micelles in the solution, and the risk of denaturation. In addition, too high concentrations increase the risk of getting detergent crystals.
13. Monoglycerides, such as mono-olein, are hygroscopic and light sensitive. Upon preparation, the bottle should be slowly brought up to room temperature to avoid overcondensation.
14. When centrifuging for forming the cubic phase, several centrifugation steps with intermittent reversal of the tube should be performed to improve the mixing of the lipid/aqueous components (69).
15. Because the lipids in the cubic phase will produce some powder diffraction rings, the less lipid that is frozen with the crystal, the less interference with protein diffraction data there will be.
16. Detergent crystals have also been shown to form at low temperature and could be mistaken for protein crystals being fragile and birefringent. For a nice gallery of

DDM crystals go to: <http://www2.mrc-lmb.cam.ac.uk/groups/stock/F1c10/yfc.htm>.

17. One should be aware that decreasing the salt concentration too much might cause precipitation of the protein. Therefore, only dilute to a level where the remaining salt concentration will not prevent binding to the ion-exchange column. Dialysis may also be used to decrease the salt concentration. Basically the aliquots should be conditioned for binding to the next column.

## Acknowledgment

The author would like to thank Dean Madden and Eckhard Hofmann for their comments on the chapter.

## References

1. Gerstein, M. (1998) How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold. Des.* **3**, 497–512.
2. Boyd D., Schierle C., and Beckwith J. (1998) How many membrane proteins are there? *Protein Sci.* **7**, 201–205.
3. Stahlberg, H., Fotiadis, D., Scheuring, S., et al. (2001) Two-dimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins. *FEBS Lett.* **504**, 166–172.
4. Locher, K. P., Lee, A. T., and Rees, D. C. (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091–1098.
5. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T., and Rees, D. C. (1998) Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive Ion Channel. *Science* **282**, 2220–2226.
6. Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., et al. (1998) The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**, 69–77.
7. Tate, C. G. (2001) Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett.* **504**, 94–98.
8. Michel, H. (2003) Crystallization of membrane proteins. In: *Macromolecular Crystallography*, vol. F, (Rossmann, M. G. and Arnold, E., eds.), Kluwer, Dordrecht, Germany, pp. 9499.
9. Michel, H. (1983) Crystallization of membrane proteins. *TIBS* **8**, 56–59.
10. Kühlbrandt, W. (1988) Three-dimensional crystallization of membrane proteins. *Q. Rev. Biophys.* **21**, 429–477.
11. Garavito, R. M., Picot, D., and Loll, P. J. (1996) Strategies for crystallizing membrane proteins. *J. Bioenerg. Biomembr.* **28**, 13–27.
12. Caffrey, M. (2003) Membrane protein crystallization. *J. Struct. Biol.* **142**, 108–132.
13. Abramson, J. and Iwata, S. (1999) Crystallization of membrane proteins. In: *Protein Crystallization: A Laboratory Manual*, (Bergfors, T., ed.), International University Line, La Jolla, CA pp. 199–210.
14. Neugebauer, J. M. (1990) Detergents: an overview. *Meth. Enzymol.* **182**, 239–253.

15. Garavito, R. M. and Ferguson-Miller, S. (2001) Detergents as Tools in membrane biochemistry. *J. Biol. Chem.* **276**, 32,403–32,406.
16. Zulauf, M., Furstenberger, U., Grabo, M., Jaggi, P., Regenass, M., and Rosenbusch, J. P. (1989) Critical micellar concentrations of detergents. *Meth. Enzymol.* **172**, 528–538.
17. Timmins, P. A., Leonhard, M., Weltzien, H. U., Wacker, T., and Welte, W. (1988) A physical characterization of some detergents of potential use for membrane protein crystallization. *FEBS Lett.* **238**, 361–368.
18. Zulauf, M. (1991) Detergent phenomena in membrane protein crystallization. In: *Crystallization of Membrane Proteins*, (Michel, H., ed.), CRC Press, Boca Raton, FL, pp. 53–72.
19. Rosenow, M. A., Williams, J. C., and Allen, J. P. (2001) Amphiphiles modify the properties of detergent solutions used in crystallization of membrane proteins. *Acta Crystallogr. D.* **57**, 925–927.
20. Garavito, R. M. (1991) Crystallizing membrane proteins: experiments on different systems. In: *Crystallization of Membrane Proteins*, (Michel, H., ed.), CRC Press, Boca Raton, FL, pp. 89–106.
21. Hitscherich, C., Jr., Kaplan, J., Allaman, M., Wiencek, J., and Loll, P. J. (2000) Static light scattering studies of OmpF porin: implications for integral membrane protein crystallization. *Protein Sci.* **9**, 1559–1566.
22. Hitscherich, C., Jr., Aseyev, V., Wiencek, J., and Loll, P. J. (2001) Effects of PEG on detergent micelles: implications for the crystallization of integral membrane proteins. *Acta Crystallogr. D.* **57**, 1020–1029.
23. Welte, W., Wacker, T., Leis, M., et al. (1985) Crystallization of the photosynthetic light-harvesting pigment-protein complex B800-850 of *Rhodospseudomonas capsulata*. *FEBS Lett.* **182**, 260–264.
24. Scarborough, G. (1994) Large single crystals of the *Neurospora crassa* plasma membrane H<sup>+</sup>-ATPase: an approach to the crystallization of membrane proteins. *Acta Crystallogr. D.* **50**, 643–649.
25. Timmins, P. A., Hauk, J., Wacker, T., and Welte, W. (1991) The influence of heptane-1,2,3-triol on the size and shape of LDAO micelles Implications for the crystallisation of membrane proteins. *FEBS Lett.* **280**, 115–120.
26. Michel, H. (1982) Three-dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from *Rhodospseudomonas viridis*. *J. Mol. Biol.* **158**, 567–572.
27. Tribet, C., Audebert, R., and Popot, J. L. (1996) Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci. USA* **93**, 15,047–15,050.
28. Prata, C., Giusti, F., Gohon, Y., Pucci, B., Popot, J. L., and Tribet, C. (2000) Nonionic amphiphilic polymers derived from Tris(hydroxymethyl)-acrylamidomethane keep membrane proteins soluble and native in the absence of detergent. *Biopolymers* **56**, 77–84.
29. Yu, S. M., McQuade, D. T., Quinn, M. A., et al. (2000) An improved tripod amphiphile for membrane protein solubilization. *Prot. Sci.* **9**, 2518–2527.

30. McQuade, D. T., Quinn, M. A., Yu, S. M., Polans, A. S., Krebs, M. P., and Gellman, S. H. (2000) Rigid amphiphiles for membrane protein manipulation. *Angew. Chem. Int. Ed Engl.* **39**, 758–761.
31. McGregor, C. L., Chen, L., Pomroy, N. C., et al. (2003) Lipopeptide detergents designed for the structural study of membrane proteins. *Nat. Biotechnol.* **21**, 171–176.
32. Wilkens, S., Zhou, J., Nakayama, R., Dunn, S. D., and Capaldi, R. A. (2000) Localization of the delta subunit in the *Escherichia coli* F1F0-ATP synthase by immuno electron microscopy: the delta subunit binds on top of the F1. *J. Mol. Biol.* **295**, 387–391.
33. Helenius, A., McCaslin, D. R., Fries, E., and Tanford, C. (1979) Properties of detergents. *Meth. Enzymol.* **56**, 734–749.
34. von Jagow, G., Link, T. A., and Schägger, H. (2003) Purification strategies for membrane proteins. In: *Membrane Protein Purification and Crystallization*, (Hunte, C., von Jagow, G, and Schägger, H., eds.), Academic Press, Amsterdam, The Netherlands, pp. 1–18.
35. Schägger, H. (2003) Techniques and basic operations in membrane protein purification. In: *Membrane Protein Purification and Crystallization*, (Hunte, C., von Jagow, G, and Schägger, H., eds.), Academic Press, Amsterdam, The Netherlands, pp. 19–53.
36. Roobol-Boza, M. and Andersson, B. (1996) Isolation of hydrophobic membrane proteins by perfusion chromatography-purification of photosystem II reaction centers from spinach chloroplasts. *Anal. Biochem.* **235**, 127–133.
37. Bergfors, T. (1999) Protein samples. In: *Protein Crystallization: A Laboratory Manual*, (Bergfors, T. M., ed.), International University Line, La Jolla, CA, pp. 17–25.
38. Palma, A. M., Thiyagarajan, P., Wagner, A. M., and Tiede, D. M. (1999) Effect of detergent alkyl chain length on crystallization of a detergent-solubilized membrane protein: correlation of protein-detergent particle size and particle-particle interaction with crystallization of the photosynthetic reaction center from *Rhodobacter sphaeroides*. *J. Crystal Growth* **207**, 214–225.
39. Shinzawa-Itoh, K., Ueda, H., Yoshikawa, S., Aoyama, H., Yamashita, E., and Tsukihara, T. (1995) Effects of ethyleneglycol chain length of dodecyl polyethyleneglycol monoether on the crystallization of bovine heart cytochrome c oxidase. *J. Mol. Biol.* **246**, 572–575.
40. Loll, P. J., Allaman, M., and Wiencek, J. (2001) Assessing the role of detergent-detergent interactions in membrane protein crystallization. *J. Crystal Growth* **232**, 432–438.
41. Song, L. and Gouaux, E. (1997) Membrane protein crystallization: application of sparse matrices to the alpha-hemolysin heptamer. *Meth. Enzymol.* **276**, 60–74.
42. Chayen, N. E. (1997) The role of oil in macromolecular crystallization. *Structure* **5**, 1269–1274.
43. Loll, P. J., Tretiakova, A., and Soderblom, E. (2003) Compatibility of detergents with the microbatch-under-oil crystallization method. *Acta Crystallogr. D.* **59**, 1114–1116.

44. Reiss-Husson, R. (1992) Crystallization of membrane proteins. In: *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, (Ducruix, A. and Giegé, R., eds.), IRL Press, Oxford, UK, pp. 175–193.
45. Garavito, R. M., Markovic-Housley, Z., and Jenkins, J. A. (1986) The growth and characterization of membrane protein crystals. *J. Crystal Growth* **76**, 701–709.
46. Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**, 647–655.
47. Okada, T., Le Trong, I., Fox, B. A., Behnke, C. A., Stenkamp, R. E., and Palczewski, K. (2000) X-ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. *J. Struct. Biol.* **130**, 73–80.
48. Lemieux, M. J., Song, J., Kim, M. J., et al. (2003) Three-dimensional crystallization of the *Escherichia coli* glycerol-3-phosphate transporter: a member of the major facilitator superfamily. *Prot. Sci.* **12**, 2748–2756.
49. Joanne Lemieux, M., Reithmeier, R. A. F., and Wang, D. N. (2002) Importance of detergent and phospholipid in the crystallization of the human erythrocyte anion-exchanger membrane domain. *J. Struct. Biol.* **137**, 322–332.
50. Pierre, Y., Breyton, C., Kramer, D., and Popot, J. L. (1995) Purification and Characterization of the Cytochrome b(6)f Complex from *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **270**, 29,342–29,349.
51. Zhang, H., Kurisu, G., Smith, J. L., and Cramer, W. A. (2003) A defined protein-detergent-lipid complex for crystallization of integral membrane proteins: The cytochrome b6f complex of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. USA* **100**, 5160–5163.
52. Landau, E. and Rosenbusch, J. P. (1996) Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. USA* **93**, 14,532–14,535.
53. Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P., and Lanyi, J. K. (1999) Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911.
54. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P., and Landau, E. (1997) X-ray structure of bacteriorhodopsin at 2.5 Å from microcrystals grown in lipidic cubic phase. *Science* **277**, 1676–1681.
55. Kolbe, M., Besir, H., Essen, L. O., and Oesterhelt, D. (2000) Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science* **288**, 1390–1396.
56. Cherezov, V., Fersi, H., and Caffrey, M. (2001) Crystallization screens: compatibility with the lipidic cubic phase for in meso crystallization of membrane proteins. *Biophys. J.* **81**, 225–242.
57. Landau, E. (2003) In cubo crystallization of membrane proteins. In: *Membrane Protein Purification and Crystallization: A Practical Guide*, (Hunte, C., von Jagow, G., and Schägger, H., eds.), Academic Press, Amsterdam, The Netherlands, pp. 285–302.
58. Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. (1995) Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842–846.

59. Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. (1995) Structure at 2.8 Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* **376**, 660–669.
60. Hunte, C. (2001) Insights from the structure of the yeast cytochrome bc<sub>1</sub> complex: crystallization of membrane proteins with antibody fragments. *FEBS Lett.* **504**, 126–132.
61. Dutzler, R., Campbell, E. B., and MacKinnon, R. (2003) Gating the selectivity filter in CIC chloride channels. *Science* **300**, 108–112.
62. Zhou, Y., Morais-Cabral, J. H., Kaufman, A., and MacKinnon, R. (2001) Chemistry of ion coordination and hydration revealed by a K<sup>+</sup> channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48.
63. Carpentier, E., Lebesgue, D., Kamen, A. A., Hogue, M., Bouvier, M., and Durocher, Y. (2001) Increased production of active human beta(2)-adrenergic/G(alpha)s fusion receptor in Sf-9 cells using nutrient limiting conditions. *Prot. Express. Purif.* **23**, 66–74.
64. Prive, G. G., Verner, G. E., Weitzman, C., Zen, K. H., Eisenberg, D., and Kaback, R. H. (1994) Fusion proteins as tools for crystallization: the Lactose permease from *Escherichia coli*. *Acta Crystallogr. D.* **50**, 375–379.
65. Byrne, B., Abramson, J., Jansson, M., Holmgren, E., and Iwata, S. (2000) Fusion protein approach to improve the crystal quality of cytochrome bo<sub>3</sub> ubiquinol oxidase from *Escherichia coli*. *Biochim. Biophys. Acta* **1459**, 449–455.
66. Garman, E. F. and Schneider, T. R. (1997) Macromolecular cryocrystallography. *J. Appl. Cryst.* **30**, 211–237.
67. Horsefield, R., Yankovskaya, V., Tornroth, S., et al. (2003) Using rational screening and electron microscopy to optimize the crystallization of succinate:ubiquinone oxidoreductase from *Escherichia coli*. *Acta Crystallogr. D.* **59**, 600–602.
68. Wertén, P. J. L., Hasler, L., Koenderink, J. B., et al. (2001) Large-scale purification of functional recombinant human aquaporin-2. *FEBS Lett.* **504**, 200–205.
69. Rummel, G., Hardmeyer, A., Widmer, C., et al. (1998) Lipidic cubic phases: new matrices for the three-dimensional crystallization of membrane proteins. *J. Struct. Biol.* **121**, 82–91.
70. le Maire, M., Champeil, P., and Møller, J. V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta* **1508**, 86–111.
71. Katayama, N., Kobayashi, M., Motojima, F., Inaka, K., Nozawa, T., and Miki, K. (1994) Preliminary X-ray crystallographic studies of photosynthetic reaction center from a thermophilic sulfur bacterium, *Chromatium tepidum*. *FEBS Lett.* **348**, 158–160.
72. Pokkuluri, P. R., Laible, P. D., Deng, Y. L., Wong, T. N., Hanson, D. K., and Schiffer, M. (2002) The structure of a mutant photosynthetic reaction center shows unexpected changes in main chain orientations and quinone position. *Biochemistry* **41**, 5998–6007.
73. Buchanan, S. K., Fritzsche, G., Ermler, U., and Michel, H. (1993) New crystal form of the photosynthetic reaction centre from *Rhodobacter sphaeroides* of improved diffraction quality. *J. Mol. Biol.* **230**, 1311–1314.

74. Papiz, M. Z., Hawthornthwaite, A. M., Cogdell, R. J., et al. (1989) Crystallization and characterization of two crystal forms of the B800-850 light-harvesting complex from *Rhodospseudomonas acidophila* strain 10050. *J. Mol. Biol.* **209**, 833–835.
75. Koepke, J., Hu, X., Muenke, C., Schulten, K., and Michel, H. (1996) The crystal structure of the light-harvesting complex II (B800-850) from *Rhodospirillum molischianum*. *Structure* **4**, 581–597.
76. Krauss, N., Schubert, W. D., Klukas, O., Fromme, P., Witt, H. T., and Saenger, W. (1996) Photosystem I at 4 Å resolution represents the first structural model of a joint photosynthetic reaction centre and core antenna system. *Nat. Struct. Biol.* **3**, 965–973.
77. Fromme, P. and Witt, H. T. (1998) Improved isolation and crystallization of photosystem I for structural analysis. *Biochim. Biophys. Acta* **1365**, 175–184.
78. Shen, J. R. and Kamiya, N. (2000) Crystallization and the crystal properties of the oxygen-evolving photosystem II from *Synechococcus vulcanus*. *Biochemistry* **39**, 14,739–14,744.
79. Tsukihara, T., Aoyama, H., Yamashita, E., et al. (1995) Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 Å. *Science* **269**, 1069–1074.
80. Tsukihara, T., Aoyama, H., Yamashita, E., et al. (1996) The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* **272**, 1136–1144.
81. Soulimane, T., Buse, G., Bourenkov, G. P., Bartunik, H. D., Huber, R., and Than, M. E. (2000) Structure and mechanism of the aberrant ba<sub>3</sub>-cytochrome c oxidase from *Thermus thermophilus*. *EMBO J.* **19**, 1766–1776.
82. Xia, D., Yu, C. A., Kim, H., et al. (1997) Crystal structure of the cytochrome bc<sub>1</sub> complex from bovine heart mitochondria. *Science* **277**, 60–66.
83. Yu, C. A., Xia, J. Z., Kachurin, A. M., et al. (1996) Crystallization and preliminary structure of beef heart mitochondrial cytochrome-bc<sub>1</sub> complex. *Biochim. Biophys. Acta* **1275**, 47–53.
84. Iwata, S., Lee, J. W., Okada, K., et al. (1998) Complete structure of the 11-subunit bovine mitochondrial cytochrome bc<sub>1</sub> complex. *Science* **281**, 64–71.
85. Hunte, C., Koepke, J., Lange, C., Manith, T., and Michel, H. (2000) Structure at 2.3 Å resolution of the cytochrome bc<sub>1</sub> complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Struct. Fold. and Design* **8**, 669–684.
86. Abramson, J., Larsson, G., Byrne, B., Puustinen, A., Garcia-Horsman, A., and Iwata, S. (2000) Purification, crystallization and preliminary crystallographic studies of an integral membrane protein, cytochrome bo<sub>3</sub> ubiquinol oxidase from *Escherichia coli*. *Acta Crystallogr. D.* **56**, 1076–1078.
87. Abramson, J., Riistama, S., Larsson, G., et al. (2000) The structure of the ubiquinol oxidase from *Escherichia coli* and its ubiquinone binding site. *Nat. Struct. Biol.* **7**, 910–917.
88. Iverson, T. M., Luna-Chavez, C., Cecchini, G., and Rees, D. C. (1999) Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* **284**, 1961–1966.

89. Lancaster, C. R., Kroger, A., Auer, M., and Michel, H. (1999) Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. *Nature* **402**, 377–385.
90. Luong, C., Miller, A., Barnett, J., Chow, J., Ramesha, C., and Browner, M. F. (1996) Flexibility of the NSAID binding site in the structure of human cyclooxygenase-2. *Nat. Struct. Biol.* **3**, 927–933.
91. Kurumbail, R. G., Stevens, A. M., Gierse, J. K., et al. (1996) Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* **384**, 644–648.
92. Picot, D., Loll, P. J., and Garavito, R. M. (1994) The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1. *Nature* **367**, 243–249.
93. Wendt, K. U., Feil, C., Lenhart, A., Poralla, K., and Schulz, G. E. (1997) Crystallization and preliminary X-ray crystallographic analysis of squalene-hopene cyclase from *Alicyclobacillus acidocaldarius*. *Prot. Sci.* **6**, 722–724.
94. Wendt, K. U., Poralla, K., and Schulz, G. E. (1997) Structure and function of a squalene cyclase. *Science* **277**, 1811–1815.
95. Pauptit, R. A., Zhang, H., Rummel, G., Schirmer, T., Jansonius, J. N., and Rosenbusch, J. P. (1991) Trigonal crystals of porin from *Escherichia coli*. *J. Mol. Biol.* **218**, 505–507.
96. Meyer, J. E. W., Hofnung, M., and Schulz, G. E. (1997) Structure of maltoporin from *Salmonella typhimurium* ligated with a nitrophenyl-maltotrioxide. *J. Mol. Biol.* **266**, 761–775.
97. Stauffer, K. A., Page, M. G., Hardmeyer, A., Keller, T. A., and Pauptit, R. A. (1990) Crystallization and preliminary X-ray characterization of maltoporin from *Escherichia coli*. *J. Mol. Biol.* **211**, 297–299.
98. Kreuzsch, A., Weiss, M. S., Welte, W., Weckesser, J., and Schulz, G. E. (1991) Crystals of an integral membrane protein diffracting to 1.8 Å resolution. *J. Mol. Biol.* **217**, 9–10.
99. Vogt, J. and Schulz, G. E. (1999) The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Struct. Fold. and Design* **7**, 1301–1309.
100. Pautsch, A., Vogt, J., Model, K., Siebold, C., and Schulz, G. E. (1999) Strategy for membrane protein crystallization exemplified with OmpA and OmpX. *Prot. Struct. Funct. Gen.* **34**, 167–172.
101. Dutzler, R., Rummel, G., Alberti, S., et al. (1999) Crystal structure and functional characterization of OmpK36, the osmoporin of *Klebsiella pneumoniae*. *Struct. Fold. and Design* **7**, 425–434.
102. Zeth, K., Schnaible, V., Przybylski, M., Welte, W., Diederichs, K., and Engelhardt, H. (1998) Crystallization and preliminary X-ray crystallographic studies of the native and chemically modified anion-selective porin from *Comamonas acidovorans*. *Acta Crystallogr. D* **54**, 650–653.
103. Vandeputte-Rutten, L., Kramer, R. A., Kroon, J., Dekker, N., Egmond, M. R., and Gros, P. (2001) Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *EMBO J.* **20**, 5033–5039.

104. Prince, S. M., Feron, C., Janssens, D., et al. (2001) Expression, refolding and crystallization of the OpcA invasin from *Neisseria meningitidis*. *Acta Crystallogr. D* **57**, 1164–1166.
105. Forst, D., Welte, W., Wacker, T., and Diederichs, K. (1998) Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat. Struct. Biol.* **5**, 37–46.
106. Forst, D., Schulein, K., Wacker, T., et al. (1993) Crystallization and preliminary X-ray diffraction analysis of ScrY, a specific bacterial outer membrane porin. *J. Mol. Biol.* **229**, 258–262.
107. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (1998) Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **282**, 2215–2220.
108. Ferguson, A. D., Breed, J., Diederichs, K., Welte, W., and Coulton, J. W. (1998) An internal affinity-tag for purification and crystallization of the siderophore receptor FhuA, integral outer membrane protein from *Escherichia coli* K-12. *Prot. Sci.* **7**, 1636–1638.
109. Smith, B. S., Kobe, B., Kurumbail, R., et al. (1998) Crystallization and preliminary X-ray analysis of ferric enterobactin receptor FepA, an integral membrane protein from *Escherichia coli*. *Acta Crystallogr. D.* **54**, 697–699.
110. Buchanan, S. K., Smith, B. S., Venkatramani, L., et al. (1999) Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nat. Struct. Biol.* **6**, 56–63.
111. Schiefner, A., Diederichs, K., Hashimoto, K., Boos, W., and Welte, W. (2002) Crystallization and preliminary X-ray analysis of the trehalose/maltose ABC transporter MalFGK2 from *Thermococcus litoralis*. *Acta Crystallogr. D* **58**, 2147–2149.
112. Chen, Z., Baruch, P., Mathews, F. S., et al. (1999) Crystallization and preliminary diffraction studies of two quinoprotein alcohol dehydrogenases (ADHs): a soluble monomeric ADH from *Pseudomonas putida* HK5 (ADH-IIB) and a heterotrimeric membrane-bound ADH from *Gluconobacter suboxydans* (ADH-GS). *Acta Crystallogr. D.* **55**, 1933–1936.
113. Jormakka, M., Tornroth, S., Abramson, J., Byrne, B., and Iwata, S. (2002) Purification and crystallization of the respiratory complex formate dehydrogenase-N from *Escherichia coli*. *Acta Crystallogr. D* **58**, 160–162.
114. Fu, D., Libson, A., Miercke, L. J. W., et al. (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486.
115. Sui, H., Walian, P. J., Tang, G., Oh, A., and Jap, B. K. (2000) Crystallization and preliminary X-ray crystallographic analysis of water channel AQP1. *Acta Crystallogr. D* **56**, 1198–1200.
116. Sui, H., Han, B. G., Lee, J. K., Walian, P., and Jap, B. K. (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878.
117. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919.

118. Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H., and Koshland, D. E., Jr. (2003) Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* **300**, 976–980.
119. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T., and MacKinnon, R. (2002) X-ray structure of a CIC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294.
120. Jiang, Y., Lee, A., Chen, J., et al. (2003) X-ray structure of a voltage-dependent K<sup>+</sup> channel. *Nature* **423**, 33–41.
121. Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B. T., and MacKinnon, R. (2002) Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522.
122. Schertler, G. F., Lozier, R., Michel, H., and Oesterhelt, D. (1991) Chromophore motion during the bacteriorhodopsin photocycle: polarized absorption spectroscopy of bacteriorhodopsin and its M-state in bacteriorhodopsin crystals. *EMBO J.* **10**, 2353–2361.
123. Essen, L. O., Siegart, R., Lehmann, W. D., and Oesterhelt, D. (1998) Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Natl. Acad. Sci. USA* **95**, 11,673–11,678.
124. Snijder, H. J., Ubarretxena-Belandia, I., Blaauw, M., et al. (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **401**, 717–721.
125. Binda, C., Newton-Vinson, P., Hubalek, F., Edmondson, D. E., and Mattevi, A. (2002) Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nat. Struct. Biol.* **9**, 22–26.
126. Gouaux, J. E., Braha, O., Hobaugh, M. R., et al. (1994) Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proc. Natl. Acad. Sci. USA* **91**, 12,828–12,831.
127. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615.
128. Bracey, M. H., Hanson, M. A., Masuda, K. R., Stevens, R. C., and Cravatt, B. F. (2002) Structural adaptations in a membrane enzyme that terminates endocannabinoid signaling. *Science* **298**, 1793–1796.
129. Jormakka, M., Byrne, B., and Iwata, S. (2003) Formate dehydrogenase: a versatile enzyme in changing environments. *Curr. Opin. Struct. Biol.* **13**, 418–423.
130. Stock, D., Leslie, A. G., and Walker, J. E. (1999) Molecular architecture of the rotary motor in ATP synthase. *Science* **286**, 1700–1705.
131. Bertero, M. G., Rothery, R. A., Palak, M., et al. (2003) Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. *Nat. Struct. Biol.* **10**, 681–687.



## Crystallization of Protein–DNA Complexes

Thomas Hollis

### Summary

Determining the crystal structure of a protein–DNA complex can provide a wealth of information regarding protein function and mechanism. The foundation for all successful X-ray structure determination is the ability to produce diffraction-quality crystals. Crystallization of protein–DNA complexes often presents unique challenges because of the additional parameters involved. This chapter will outline many of those challenges, including choice of DNA and formation of a stable protein–DNA complex and provide guidance in preparing for crystallization experiments. Additionally, techniques for oligonucleotide purification, sample preparation, and crystallization methods are provided. Careful thought and initial analysis of the protein–DNA complex prior to crystallization experiments followed by optimization of crystallization parameters can greatly increase the likelihood of producing well-diffracting crystals.

**Key Words:** Crystallization; DNA purification; protein–DNA complex; DNA; oligonucleotides; protein–DNA crystallization; X-ray crystallography.

### 1. Introduction

Many basic biological questions involve the structural basis for site-specific recognition of nucleic acids by proteins. Crystal structures of protein–DNA complexes can provide insight into protein function and DNA-binding activities, as well as guide mutational and biochemical assays. In order to determine the crystal structure of a protein–DNA complex, however, well-ordered crystals of the complex must be obtained. With advances in computational techniques and phasing methods, crystallization of biomolecules is often the rate-limiting step in X-ray crystal structure determination. Unfortunately, crystallization is still largely a trial and error process for each new molecule or complex to be crystallized. This chapter describes techniques for crystallizing protein–DNA complexes, including choice and purification of DNA oligonucleotides, preparing the complex, and crystallization conditions.

Crystallization of protein–DNA complexes can be considered a distinct problem to that of crystallizing the individual components. Although the general techniques and methods of crystallization trials are similar to those employed for crystallization of proteins or DNA alone, crystallization of a protein–DNA complex adds another dimension of crystallization space that often must be explored. In addition to purity and concentration of individual components, the protein:DNA ratio, as well as the length and composition of DNA, must be considered. Like crystallization of most macromolecules this is an empirical process. Broad screening using an array of DNAs under diverse conditions is usually required, followed by more directed screens and optimization once initial crystals are identified.

## 2. Materials

1. 1 M Triethylammonium bicarbonate (TEAB) buffer: 140 mL (101 g) triethylamine (Aldrich, St. Louis, MO), bring to 1 L with dH<sub>2</sub>O. Transfer to 2-L Erlenmeyer flask and add crushed dry ice while stirring until pH 7.0. Filter solution through 0.45- $\mu$  filter. Buffer can be aliquoted and frozen at  $-20^{\circ}\text{C}$  until needed.
2. Anion exchange buffer A: 0.01 N NaOH, 0.1 M NaCl, and 5% acetonitrile.
3. Anion exchange buffer B: 0.01 N NaOH, 1.0 M NaCl, and 5% acetonitrile.
4. Elution buffer: 0.1 M TEAB and 30% acetonitrile.
5. Wash buffer: 25 mM TEAB.
6. 10X Tris-boric acid-EDTA (TBE) buffer: 108 g Tris base, 55 g boric acid, 40 mL 0.5 M EDTA, pH 8.0; bring to 1 L with dH<sub>2</sub>O.
7. 20% Denaturing acrylamide stock solution: 250 mL 40% acrylamide (19:1 acrylamide:*bis*-acrylamide) (National Diagnostics, Atlanta, GA), 240 g urea; bring to 500 mL with dH<sub>2</sub>O. Store at 4°C until needed.
8. 10X DNA-annealing buffer: 25 mM MES pH 6.5, 200 mM NaCl, and 25 mM MgCl<sub>2</sub>.
9. Formamide-loading buffer: 1 mL 10X TBE buffer, 9 mL deionized formamide (Fluka, Basel, Switzerland), and 0.5% bromophenol blue.
10. 1.0 M Acetic acid solution.
11. Gel-fixing solution 1: 10% glacial acetic acid and 25% methanol.
12. Gel-fixing solution 2: 10% glacial acetic acid and 10% isopropanol.
13. 50  $\mu\text{M}$  Dithiothreitol (DTT) solution.
14. Silver nitrate solution: 0.1% silver nitrate.
15. Developing solution: 3% sodium carbonate and 0.02% formaldehyde.
16. Stop solution: 2.3 M citric acid.
17. Chromatographic instrument with ultraviolet (UV) detection system and fraction collector.
18. POROS<sup>®</sup> HQ10 4.6 mmD/100 mmL anion exchange column (Applied Biosystems, Foster City, CA).
19. Electrophoresis equipment.
20. Thin-layer chromatography plate (TLC) with 366-nm fluorescent indicator (EM Sciences, Darmstadt, Germany).

21. 0.2- $\mu\text{m}$  Spin filter for DNA (Corning, Inc., Corning, NY).
22. Sep-Pak<sup>®</sup> C18 cartridge (Waters, Milford, MA).
23. Dialysis tubing (3350 mw cutoff) (Pierce Chemical, Rockford, IL).
24. Crystallization supplies: crystallization plates, cover slips, and crystallization screens.
25. Hand-held long wavelength UV light (350- to 375-nm range).

### 3. Methods

#### 3.1. DNA Sequence Choice and Length

In designing DNA to be used in crystallization trials, several important factors need to be considered, primarily the length, sequence, and composition of the ends of the oligonucleotides. The idea is to create a structurally homogeneous population of the protein–DNA complex. Even seemingly small disturbances, such as a single base pair shift of the protein along the DNA in the complex could result in detrimental changes to the diffraction quality of the crystals and/or quality of the final electron density maps. In crystallizing protein–DNA complexes, experience has shown that often times it is more fruitful to experiment with a variety of DNA fragments under a relatively limited set of crystallization conditions rather than exploring more potential conditions with a limited number of DNA sequences (1).

Choosing the proper length of DNA to use in crystallization trials is an empirical process and must be determined for each new protein of interest. A general rule is to choose a sequence of minimal length that binds tightly to the protein and contains all necessary sequence requirements. The underlying reason for this is to minimize the number of positions the protein can bind on the DNA, as well as to reduce excess portions that may inhibit crystallization. Electromobility shift assays (EMSA) can be a useful tool in determining the minimum length of oligonucleotide necessary to form the complex. Often times, it is necessary to create a library of DNAs of varying lengths and screen each one with the protein in crystallization trials in order to successfully crystallize the protein–DNA complex. Beginning with a minimal binding sequence one could increase the length of DNA until a spectrum of lengths has been explored. Additionally, double-stranded DNA (dsDNA) oligonucleotides often tend to pack end-to-end within a crystal, forming a pseudocontinuous helix. Screening blunt-ended DNA in crystallization trials, as well as single- or double-base overhangs, which could facilitate this sort of packing, is often productive when crystallizing complexes containing dsDNA (1,2). Although both 3' and 5' ends have proven successful, the overhanging bases on one strand should be complementary to the overhanging bases on the opposite strand to allow the best end-to-end packing of the DNA.

The sequence of the DNA to be used is often dictated by the biological function of the protein, but in the case of a protein that binds sequence independently the

choice can be more difficult. It is important to choose a DNA sequence that allows protein binding in a single position in order to reduce structurally heterogeneous populations of the complex. Careful selection of the DNA sequence can help minimize mixed binding conformations that can ultimately decrease the quality of the crystals. If single-stranded DNA is to be used often times an oligonucleotide of uniform composition, such as poly-dT, is useful. This allows identical protein–DNA interactions regardless of where the protein binds on the DNA. Similarly, when using dsDNA choosing a palindromic sequence allows the protein to bind to either strand and maintain a structurally homogenous complex.

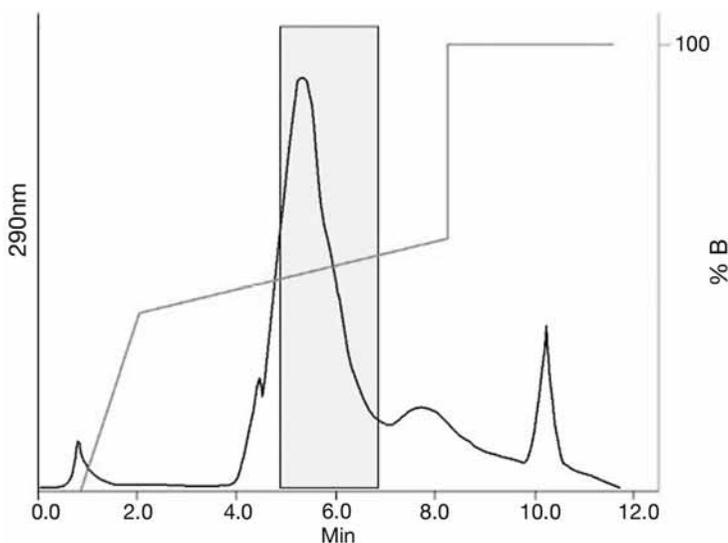
### 3.2. Sample Preparation

One of the most important considerations in crystallization experiments, whether protein or nucleic acid, is sample purity. There is a wealth of information and general knowledge available on purification of proteins (3–5), so little effort will be devoted here to explaining those methods. For the purpose of crystallization, however, it is generally desirable to have at least several milligrams of highly pure protein (>99%). The protein should be buffered in 20–50 mM buffer with any other components necessary for protein solubility such as salt, glycerol, or reducing agent (*see Note 1*). The underlying idea is to let the crystallization conditions influence the pH and salt concentration rather than be overwhelmed by the protein buffer.

Likewise, having highly purified DNA, whether synthesized oligonucleotides or biologically isolated DNA fragments, can be critical to obtaining well-ordered crystals of a protein–DNA complex. Typically, 300–500 nmol of purified DNA is needed to begin crystallization trials. This is easily obtained from purification of a 1  $\mu$ mol DNA synthesis reaction. The most common methods for purifying synthesized oligonucleotides for use in crystallization experiments are anion-exchange chromatography, or polyacrylamide gel electrophoresis (PAGE). The DNA should be deprotected (protecting groups removed after synthesis), which is normally the case for commercially ordered DNA. If necessary, this is achieved by incubation of the DNA in 1-mL neat ammonium hydroxide in a screw-top Eppendorf tube at 54°C for 12 h (5). The solution can then be lyophilized and the DNA resuspended in water. (**Caution:** oligonucleotides with unusual or modified bases incorporated into them may require a special deprotection procedure.) Both anion exchange and PAGE methods described here are suitable for purification of up to a 1- $\mu$ mol synthesis reaction.

### 3.3. Purification of DNA Oligonucleotides by Anion-Exchange Chromatography

It is recommended that the following procedure be used for the optimization of oligonucleotide purification on an analytical scale, until the optimal elution



**Fig. 1.** Elution profile of DNA oligomer from HQ10 column. After injection, the column is developed with a short, steep gradient followed by a longer shallow gradient (gray line) during which elution occurs. The lagging two-thirds of the peak are kept (gray shade) (**Subheading 3.3.**).

for a particular oligonucleotide is known. The parameters of sample elution will be dependent on the length of DNA oligomer. To achieve maximum resolution of separation during purification, a two-step gradient protocol is used. First, a short steep gradient raises the salt concentration to near the elution condition. This is followed by a long shallow gradient during which the oligonucleotide is eluted (**Fig. 1**) (*see Note 2*). All buffers and solutions for anion exchange should be filtered through a 0.2- $\mu\text{m}$  filter before use.

1. Attach POROS HQ10 column or other anion-exchange column to the chromatographic system and equilibrate with buffer A. The entire purification procedure will be run at a flow rate of 4.0 mL/min. Set UV detection at 290 nm.
2. Resuspend deprotected, dried DNA oligomer in 500  $\mu\text{L}$  dH<sub>2</sub>O. Filter through 0.2- $\mu\text{m}$  spin filter.
3. Dilute 5  $\mu\text{L}$  of this oligonucleotide solution into 300  $\mu\text{L}$  dH<sub>2</sub>O.
4. Inject diluted oligonucleotide onto column. Wash column with buffer A for 1 min.
5. Sample will be eluted from the column with the following profile, which requires optimization. Short gradient from 0% anion-exchange buffer B to X% over 4 mL. This is followed by a longer gradient of X% – X + 10% buffer B over 24 mL. Finally, wash the column with 100% buffer B for 4 mL. The value of X will depend on the length of the oligonucleotide. For DNA oligos of 10–20 bases 25% buffer B

is a good place to start and 35% buffer B for oligos of 20–30 bases. The elution of the oligonucleotide should occur during the second, shallower gradient. If it elutes early, then lower the %B during elution (lower value of X) or raise %B if it elutes after gradient and repeat with another diluted sample. Be sure to re-equilibrate column with buffer A before injection of next sample.

6. Once optimum elution gradients are determined, the remainder of undiluted DNA can be loaded onto the column and purified. During the preparative purification run, fractions should be collected over the entire protocol. Only fractions corresponding to the last two-thirds of desired peak should be kept (**Fig. 1**).
7. Pool peak fractions and neutralize with 10  $\mu\text{L}$  of 1.0 *M* acetic acid solution per milliliter of fraction.
8. Neutralized DNA solution should be desalted by dialysis into 50 mM TEAB buffer or by passage over C18 cartridge (*see Subheading 3.5.*).

### 3.4. Purification of DNA Oligonucleotides by PAGE

Purification of DNA by gel electrophoresis is an alternative to ion-exchange chromatography. This method is more time consuming if multiple samples are to be purified because only one or two samples can be purified at a time. However, it has the advantage that it can be performed without the need for expensive chromatography equipment and may provide better separation for larger oligonucleotides (>25 bases) (**5**).

1. Prepare denaturing gel using 20% denaturing acrylamide stock solution (*see Note 3*). Add 250  $\mu\text{L}$  10% ammonium persulfate and 50  $\mu\text{L}$  TEMED. Quickly pour gel and insert comb containing a single well and allow gel to solidify.
2. Resuspend dried oligonucleotide in 50  $\mu\text{L}$  of  $\text{dH}_2\text{O}$  and mix with equal volume formamide-loading buffer. Load onto gel and run at 300 V (20  $\times$  20-cm gel) or 1000 V (35  $\times$  40-cm gel) until bromophenol blue dye is at least halfway down the gel.
3. Remove gel from plates and place on plastic wrap. Visualize DNA by placing gel on TLC plate with UV indicator and irradiating with a hand-held, long-wavelength UV light. The DNA will appear as dark bands. Using a razor blade cut out gel slice containing the desired band.
4. Chop removed gel slice into small pieces and place in conical tube with 10 mL 50 mM TEAB buffer. Seal the lid of the tube with Parafilm and tumble slowly overnight at room temperature.
5. Remove gel remnants by filtration. Freeze DNA solution on dry ice and lyophilize. Resuspend DNA in 0.5 mL  $\text{dH}_2\text{O}$  and quantitate by UV absorbance (260 nm). Dry sample again and resuspend to the desired concentration.

### 3.5. Desalting DNA

Purified DNA often contains residual salts and/or buffers that are undesirable in the final solution. By exchanging the DNA into a volatile buffer, the solution can be dried down and resuspended in the desired final buffer. This can be achieved by dialyzing DNA oligonucleotides into volatile buffer or alternative-

ly by reverse-phase chromatography using a small C18 column (*see Note 4*). This protocol outlines the steps for buffer exchange of purified DNA using a Sep-Pak C18 cartridge from Waters Corporation.

1. Attach C18 column to a 10-mL syringe and wet column with 3 mL of acetonitrile. Once column is wet it is important to prevent drying by not pushing air through it.
2. Wash acetonitrile out with 5 mL of elution buffer.
3. Equilibrate column with 10 mL wash buffer twice.
4. Load DNA onto column by passing slowly through two times.
5. Wash column with 10 mL wash buffer.
6. Slowly elute with 5 mL elution buffer.
7. Freeze solution on dry ice and lyophilize to dryness.
8. Dissolve DNA in 0.5 mL dH<sub>2</sub>O and quantitate by diluting 5  $\mu$ L into 1 mL and measuring OD<sub>260</sub>.
9. Transfer remaining solution to Eppendorf tube and dry in speed-vac. Resuspend DNA in dH<sub>2</sub>O to desired concentration and store at  $-20^{\circ}\text{C}$ .

### 3.6. Annealing DNA Oligonucleotides

If dsDNA is to be used in crystallization of a protein–DNA complex, the single-strand oligonucleotides will need to be annealed prior to crystallization. Typically, it is easier to anneal concentrated DNA rather than trying to concentrate it after annealing (*see Note 5*) (*see Subheading 3.4.* about suggestions for DNA concentration).

1. Combine equimolar amounts of oligonucleotides to be annealed in screw-top Eppendorf tube.
2. Add 10X DNA-annealing buffer to 1X final concentration.
3. Place tube in a 250-mL beaker containing  $90^{\circ}\text{C}$  water. Let cool to room temperature.
4. Store duplex DNA at  $4^{\circ}\text{C}$ .

### 3.7. Protein–DNA Complex Preparation

In crystallizing a protein–DNA complex, preparing a stable, homogeneous complex is often critical. Determining conditions that promote formation of the complex prior to crystallization trials can often improve the chances of successfully obtaining diffraction-quality crystals. Several considerations come into play when preparing the protein–DNA complex. One is the stoichiometry of the components of the complex. If the complex is a single protein and DNA molecule, a typical starting point is to have a slight excess ( $\sim 10\%$ ) of DNA over protein. If the protein–DNA complex contains multiple proteins or multiple binding sites on the DNA, an excess of one component may not be desirable. Again, the object is to have as homogenous a solution as possible, and not one that could contain mixtures of protein–DNA in multiple binding states that may not be conducive to crystal growth. For this reason it may be worth spending

time optimizing the component ratios. This can be done by examining the complex over a series of protein–DNA ratios using the EMSA (EMSA or gel shift) (5) to identify the ratio for maximal complex formation.

Once the optimal component ratios have been identified, the complex should be formed in the minimal buffer conditions under which the complex is still soluble. Formation of the complex may in some cases affect solubility. Sometimes the solubility of the complex is greater than that of the protein alone, in which case it may be desirable to further concentrate the complex before crystallization trials. This may be particularly useful if the protein has limited solubility (<~5 mg/mL).

Another consideration in obtaining diffraction-quality crystals is the stability of the complex. Several factors come into play here as well. First is the affinity of the protein–DNA interaction. A general rule of thumb is that dissociation constants in the micromolar range or lower are needed for crystallization of a complex. The presence of cofactors, such as metal ions or nucleotides, may greatly affect the interaction of protein with DNA and should be considered when forming the complex. A caveat here is that while cofactors may promote tighter binding they may also facilitate catalytic activity of the protein, which may ultimately compromise the stability of the complex.

If the protein modifies the DNA in some fashion (e.g., polymerases, nucleases, glycosylases, and so on), formation of a stable complex can be trickier still. Addition of chelating agents to remove cations or changing pH may be sufficient to prevent catalytic activity (6), but may also affect DNA binding of the protein. In some cases, incorporation of modified nucleotides or inhibitors in the DNA has been necessary to prevent catalytic activity in order to capture a stable protein–DNA complex (7–10). Alternatively, complexes have been trapped through the formation of covalent protein–DNA interactions (11,12), by either trapping reaction intermediates or by cross-linking agents (13–15).

Another option to form a protein–DNA complex is to soak the DNA into the protein crystals (16,17). This can be accomplished by adding DNA to the crystallization drop after protein crystals have formed or by transferring the crystals to new drops containing DNA (see Note 6). This method presents its own set of difficulties in that it usually only works well for short oligonucleotides, and may disrupt existing crystal contacts, thereby destroying the crystal. The same considerations about protein–DNA affinity and catalytic activity in cocrystallization apply to this method as well.

### 3.8. Crystallization Conditions

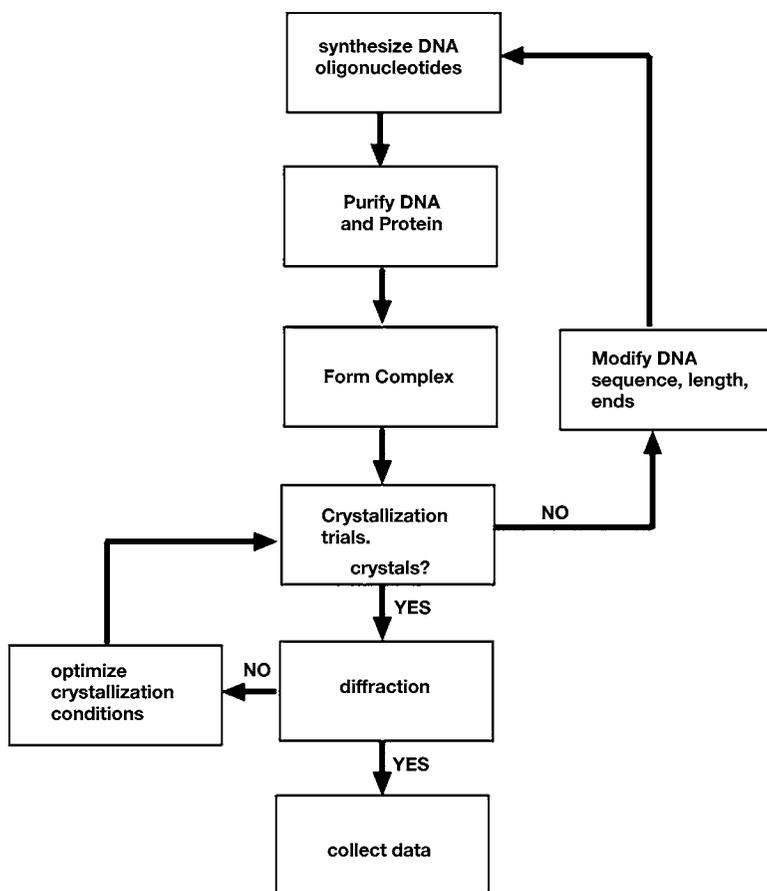
Crystallization techniques for protein–DNA complexes are similar to those of the individual components. The vapor diffusion method (1,18), either hanging or sitting drop, is probably the most practical because it allows for rapid

screening of many conditions. In this method a drop of concentrated protein–DNA solution is placed on a cover slip or on a bridge and mixed with an equal amount of precipitating solution. The cover slip is then inverted and sealed above a well containing the same precipitating solution. Equilibrium of conditions between the well solution and drop are achieved by vapor diffusion, which promotes the slow precipitation (and hopefully crystallization) of the complex. One advantage to using the hanging-drop method is that the cover slips can accommodate several drops at a time. This allows the simultaneous screening of multiple complexes containing different oligonucleotides over a single condition (*see Note 7*).

Crystallization screens are readily available from a variety of companies (Hampton Research, Emerald Biostructures, Jena Bioscience, Molecular Dimensions) and are probably the easiest route to initial crystallization trials. Protein–DNA complexes have been crystallized under a wide variety of conditions, however, a large proportion of the crystals emerged from a relatively limited number of conditions (*1*). A predominant precipitant in crystallization of complexes is polyethylene glycol (PEG) of low-to-medium molecular weight (PEG 400–10,000), either alone or in combination with salts. Frequently, additives are also required such as divalent ions ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ), polyamines (spermine, cobalt hexamine), glycerol, ethylene glycol, or 2-methyl-2,4-pentanediol.

An alternative to using commercial screens having very diverse conditions is to create a more directed PEG-ion screen using an incomplete factorial method (*19,20*). In this method, a subset or “incomplete factorial sampling” of conditions (e.g., 24 or 48 conditions) are derived from a larger group of conditions. The smaller subset is chosen to provide balance or efficient coverage of the entire experimental space. The factorial might sample factors, such as different molecular weight PEGs at varying concentrations, varying values of pH, and the presence or absence of monovalent or divalent cations. For example, if a modest screen of four values for each factor were chosen, a complete factorial screen would be 256 conditions ( $4 \text{ PEGs} \times 4 \text{ concentrations} \times 4 \text{ pH values} \times 4 \text{ cations}$ ). This contains far too many conditions to reasonably screen for each oligonucleotide, but it can be reduced to an easily manageable number by this method. Detailed explanations of sampling, balance, and resolution can be found elsewhere (*1,19,20*). Additionally, resources for generating incomplete factorial screens can be found at several websites, such as CRYSTOOL (<http://porter.llnl.gov/crystool4.1/>), GOSSET (<http://www.research.att.com/~njas/gosset/>), and SAmBA (<http://igs-server.cnrs-mrs.fr/samba/>).

Once an initial crystallization condition is identified, it may be the case that the condition will need optimization in order to improve the quality of the crystals. This is most commonly done by setting up a screen of component concentrations



**Fig. 2.** Flow chart for crystallization of a protein–DNA complex. Broad screening using an array of DNAs under diverse conditions is usually required, followed by more directed screens and optimization once initial crystals are identified.

and pH, bracketing the initial condition. This is a good point to begin testing of additives that can often have dramatic effects on crystal quality (Fig. 2).

Finally, before expending too much effort on optimizing crystallization conditions, it is a good idea to verify that the crystal contains both protein and DNA. This can be done by removing several crystals from the drop, washing them in well solution, and then dissolving them in water to run on denaturing sodium dodecyl sulfate-PAGE gel. It is a good idea to include standards of both protein and DNA on the gel. For visualization, silver staining the gel is probably best. Not only will this verify the presence of protein and DNA in the crystals, but it will also verify that the protein has not undergone degradation. If

degradation has occurred, analysis of the protein by mass spectrometry may be necessary. Additional protein purification or protease inhibitors may be required to prevent degradation, or alternatively the protein fragment could be produced and used in crystallization.

1. When silver staining gels use only clean glassware and wear gloves (*see Note 8*).
2. Soak gels in gel-fixing solution 1 for 20 min, followed by gel fixing solution 2 for 20 min.
3. Rinse gel by soaking in 500-mL water for 1 h. Change water every 15 min during rinse period.
4. Incubate gel for 20 min in 100 mL DTT solution.
5. Pour off DTT solution and without rinsing add 100 mL silver nitrate solution. Incubate for 30–60 min.
6. Rinse gel rapidly with a small amount of water. Soak gel in 100 mL developing solution until the desired level of staining is reached. Typical times are from 2–10 min.
7. Stop the reaction by adding 5 mL of stop solution. Incubate for 30 min before rinsing gel with water.

It is obvious from the previous several sections that crystallization of a protein–DNA complex can be substantially more complex than crystallization of the individual components. Factors such as length and sequence of the DNA, as well as conditions like protein–DNA ratios, buffers, cofactors, and ionic strength can contribute to the overall stability of the complex. Identifying and optimizing these parameters before starting crystallization trials can often be the key to successful crystallization of the complex.

#### 4. Notes

1. Protein buffer should be a nonphosphate buffer. Phosphate buffers commonly form insoluble salts that readily crystallize with components of many crystallization screens.
2. The protocol is outlined for use with a POROS HQ10 column because of the high binding capacity and high flow rates attainable with this media (*see also ref. 21* [[http://www.roche-applied-science.com/PROD\\_INF/BIOCHEMI/No.1\\_96/p12-13.pdf](http://www.roche-applied-science.com/PROD_INF/BIOCHEMI/No.1_96/p12-13.pdf)]).

However, other high-resolution quaternary amine columns such as MonoQ<sup>®</sup> and Source 15Q<sup>®</sup> (Amersham Biosciences, Uppsala, Sweden) can be used with comparable purification resolution. If using one of these other columns, the flow rate will have to be adjusted so as not to exceed the pressure limit of the media. Additionally, wash and elution volumes will have to be adjusted proportionally, depending on the bed volume of column. The UV detector is set at 290 nm to avoid an overload of the UV signal.

3. This procedure can be done either using 20 × 20-cm plates or 35 × 40-cm sequencing plates. Forty milliliters of acrylamide solution is sufficient to pour the small

size gel and 80 mL is sufficient for the larger gel. Use appropriate safety precautions when handling acrylamide.

4. Dialysis of oligonucleotide DNA is usually done in low molecular weight cutoff dialysis tubing available from Sigma. Although dialysis takes longer, it is often much more convenient than the C18 column if working with many samples. Dialysis should be done using four changes of buffer with a volume 500-fold larger than that of the oligo.
5. Consider the concentration of protein that will be used in crystallization trials when resuspending the purified DNA. If possible, keep the DNA as concentrated as possible to avoid having to add large volumes to the protein solution when forming the complex. Remember that the DNA concentration will also be diluted in half when annealing strands. For example, a 5 mM oligo concentration will allow for about an equal molar ratio with a 25-kD protein at 7.5 mg/mL after annealing and a 1:10 dilution into protein solution.
6. To minimize crystal degradation, it is probably best to identify a crystal harvest buffer (artificial mother liquor). The DNA can be added to this condition and the crystal transferred to this drop to allow the DNA to soak in.
7. Glass cover slips should be siliconized to prevent drops from spreading. An alternative is to use plastic cover slips rinsed in ethanol and deionized water. The plastic does not require siliconization and is more durable than glass.
8. Silver staining is a very sensitive detection method capable of detecting 1–10 ng per band. Fingerprints will stain and dirty glassware can affect the sensitivity of the reaction. Therefore, it is imperative to wear gloves during the entire procedure and to use only freshly cleaned glassware and Milli-Q® quality water.

## References

1. Ducruix, A. and R. Giegé, R. (1999) *Crystallization of Nucleic Acids and Proteins. A Practical Approach*, ed. Oxford University Press, Oxford, UK, pp. 121–238.
2. Nucleic Acid Database, Rutgers University. (<http://ndbserver.rutgers.edu>). Accessed June, 2006.
3. Scopes, R. K. (1987) *Protein Purification Principles and Practice*, 2nd ed. Springer-Verlag, New York, NY.
4. Deutscher, M. P. (1990) *Guide to Protein Purification. Methods in Enzymology*, vol. 182. Academic Press, New York, NY.
5. Ausubel, F. M. (1992) *Short Protocols in Molecular Biology*, 2nd ed. Greene Publishing Assoc. and John Wiley and Sons, New York, NY.
6. Beese, L. and Steitz, T. (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33.
7. Hollis, T. and Ellenberger, T. (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA. *EMBO J.* **19**, 758–766.
8. Lau, A., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258.

9. Davies, D. R., Interthal, H., Champoux, J. J., and Hol, W. G. (2003) Crystal structure of a transition state mimic for tdp1 assembled from vanadate, DNA, and a topoisomerase I-derived peptide. *Chem. Biol.* **10**, 139–147.
10. Doublé, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258.
11. Fromme, J. and Verdine, G. (2003) Structure of a trapped endonuclease III–DNA covalent intermediate. *EMBO J.* **22**, 3461–3471.
12. Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J., and Hol, W. G. (1998) Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* **279**, 1504–1513.
13. He, C. and Verdine, G. (2002) Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking. *Chem. Biol.* **9**, 1297–1303.
14. Huang, H., S. Harrison, and G. Verdine. (2000) Trapping of a catalytic HIV reverse transcriptase–template:primer complex through a disulfide bond. *Chem. Biol.* **7**, 355–364.
15. Sarafianos, S., Clark, A. D., Jr, Tuske, S., et al. (2003) Trapping HIV-1 reverse transcriptase before and after translocation on DNA. *J. Biol. Chem.* **278**, 16,280–16,288.
16. Brautigam, C., Aschheim, K., and Steitz, T. (1999) Structural elucidation of the binding and inhibitory properties of lanthanide (III) ions at the 3′–5′ exonucleolytic active site of the Klenow fragment. *Chem. Biol.* **6**, 901–908.
17. Brautigam, C. and Steitz, T. (1998) Structural principles for the inhibition of the 3′–5′ exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates. *J. Mol. Biol.* **277**, 363–377.
18. McPherson, A. (1999) *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Carter, C. W. J. and Carter, C. W. (1979) Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12,219–12,223.
20. Betts, L. et al. (1989) Incomplete factorial search for conditions leading to high quality crystals of Escherichia coli cytidine deaminase complexed to a transition state analog inhibitor. *J. Biol. Chem.* **264**, 6737–6740.
21. Ellenberger, T. and Doublé, S. (1996) Use of perfusion chromatography technology for the rapid purification of synthetic oligonucleotides used in DNA/DNA-binding protein structural studies. *Biochemica* **1**, 12–13.



## Preparation and Crystallization of RNA

Barbara L. Golden

### Summary

The field of RNA structure has exploded in recent years, in part owing to advances in crystallography of RNA molecules. This phenomenon can largely be attributed to the development of three modern methods: (1) large-scale *in vitro* RNA synthesis, (2) cryocrystallography, and (3) high-intensity synchrotron beamlines. Milligram quantities of RNA can be routinely synthesized using either chemical or enzymatic syntheses, making it feasible to carry out routine crystallization experiments on RNA. This has allowed crystals of RNA to be readily obtained. Generally, RNA crystals tend to be susceptible to radiation damage and to diffract X-rays more weakly than their protein counterparts. However, cryocrystallography and the high-intensity X-ray sources have overcome many of the difficulties involved in solving crystal structures of RNA. As a result of these advances, we now have a database of RNA structures that span from simple duplexes and hairpins to complex ribozymes and ribosomes. The protocols presented here describe methods to synthesize, purify, crystallize, and derivatize RNA for use in crystallographic studies.

**Key Words:** RNA; ribozyme; RNA–protein interaction; heavy-atom derivatives; T7 RNA transcription; RNA crystallization; RNA purification; RNA synthesis.

### 1. Introduction

Current protocols allow the routine synthesis and purification of RNA molecules for use in crystallization experiments. RNAs can be synthesized in milligram quantities either by use of solid-state chemical methods, or by *in vitro* transcription using T7 RNA polymerase. The purity required for most crystallization experiments can be achieved by preparative acrylamide gel electrophoresis or HPLC. These protocols can be carried out in most laboratories equipped with standard molecular biology equipment. Crystallization of RNA is quite similar to crystallization of protein molecules and requires the same type of equipment and supplies. Crystallization of RNAs is no longer a roadblock to structural studies of these molecules. Unfortunately, there is no technique as simple and powerful as selenomethionine labeling of proteins (*see*

Chapter 5) that can readily be used to synthesize heavy-atom derivatives of RNAs. Few RNAs can be readily metabolically labeled with a handful of heavy atoms for use in calculating crystallographic phases. Alternative strategies for creating heavy-atom derivatives of RNAs must be used. Methods described here will address synthesis, purification, crystallization, and derivatization of RNAs for use in crystallographic studies.

One major concern in laboratories new to working with RNA is degradation of costly material. RNA can be very susceptible to degradation by RNases and exposure to strong bases. The risk of degradation during sample preparation and crystallization may be somewhat exaggerated. In general, the molecules that are used for crystallography are folded into compact structures, like their protein counterparts. They are, therefore, somewhat resistant to degradation. The major causes of RNA degradation are microbial growth and RNase A contamination. For these reasons, it is important to wear clean gloves when performing experiments, to autoclave and/or filter-sterilize solutions, and to use disposable labware when possible. Low pH buffers (6.0–6.5) are also preferred, as they will prevent nicking of the backbone by hydroxide ions. If common sense procedures such as these are followed, the risk of degradation is minimal. It is important, however, to routinely examine the integrity of the RNA samples during synthesis, purification, and crystallization.

## 2. Materials

1. Chemically synthesized RNA or DNA template (linearized plasmid or DNA oligonucleotides).
2. Autoclaved deionized water.
3. Plasmid prep kit or materials for plasmid purification.
4. Heating block or water bath.
5. Appropriate restriction endonuclease and reaction buffer.
6. Horizontal agarose gel electrophoresis apparatus and supplies.
7. Siliconized microcentrifuge tubes.
8. Phenol:chloroform:isoamyl alcohol (25:24:1, [v/v/v]) saturated with 10 mM Tris-HCl, pH 8.0.
9. Chloroform.
10. Ethanol.
11. TE: 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA.
12. 5 M NaCl.
13. T7 RNA polymerase.
14. Nucleoside triphosphates, monosodium salts.
15. 10 M NaOH.
16. 10X Transcription buffer: 400 mM Tris-HCl, pH 8.0, 200 mM MgCl<sub>2</sub>, 100 mM DTT, and 20 mM spermidine.
17. Formamide “stop” buffer: 90% formamide, 50 mM EDTA, 0.1X Tris-boric acid-EDTA (TBE), 0.005% bromphenol blue, and 0.005% xylene cyanol.

18. 10X TBE: 1 M Tris base, 0.83 M boric acid, and 10 mM EDTA.
19. 40% Acrylamide (19:1 crosslinking ratio).
20. *N,N,N',N'*-tetramethylethylenediamine.
21. 10% Ammonium persulfate.
22. Urea.
23. DNA sequencing gel equipment with 3 mM spacers.
24. Stain's All (Sigma, St. Louis, MO, cat. no. E9379).
25. Intensifying screen or TLC plate with fluorescent indicator.
26. Shortwave ultraviolet (UV) lamp.
27. TEN: 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 250 mM NaCl.
28. 5 mM Potassium cacodylate, pH 6.5.
29. 500 mL Disposable 0.2- $\mu$ m filter units (such as Millipore Stericup GP Express).
30. Centricon concentrators (Millipore, Billerica, MA), or equivalent.
31. Linbro 24-well plates (Hampton Research, Aliso Viejo, CA).
32. Siliconized glass cover slips (Hampton Research).
33. Vacuum grease.
34. Crystallization stocks: buffers, precipitants, divalent metal ion solutions (1 M MgCl<sub>2</sub>; 1 M CaCl<sub>2</sub>; 1 M BaCl<sub>2</sub>; 1 M SrCl<sub>2</sub>), monovalent ion solutions (1 M NaCl; 1 M KCl; 1 M LiCl), polyamines (0.5 M spermine; 0.5 M spermidine).

### 3. Methods

#### 3.1. Design of Molecules for Crystallization

The single most important variable in crystallization of RNA molecules is the sequence. Common modes of packing in RNA crystals include stacking of helices end-to-end into a pseudo continuous helix, base pairing, base stacking, minor groove–minor groove packing, and tetraloop–minor groove interactions. Most researchers screen several sequence variants in order to identify those that best pack into regular arrays (1–6). The simplest means of providing variation is to introduce base mutations at the solvent-exposed surface of the molecule. For simple duplex or hairpin molecules, this means changing the length of the helix or introducing base overhangs at the 3' and/or 5' end.

Variations of this approach also work well for RNA molecules that possess complex secondary or tertiary structures. These molecules must be mapped biochemically to identify the surfaces that are exposed to solvent and that are not directly involved in RNA structure or function. This can often be accomplished by phylogenetic analyses or solution structure-probing techniques, including hydroxyl radical protection assays. Solvent-exposed regions of the molecule are likely to be both tolerant of mutations and available to participate in crystal packing. If the wild-type RNA does not crystallize readily, these surfaces can be varied to allow the formation of new or improved crystal contacts.

The simplest variation is to change either the length of exposed helices or the sequence of exposed loops. Replacement of natural sequences with RNA

tetraloops, especially GNRA tetraloops (where N = A,G,C,U and R = A,G), is a common strategy (3,5) because these motifs provide a compact structure and are known to mediate RNA packing (7). Alternately, a protein-binding motif, such as stemloop 2 of U1 RNA, can be introduced at these sites. The RNA can then be cocrystallized with an appropriate RNA-binding protein, such as the RNA-binding domain of U1A protein (8).

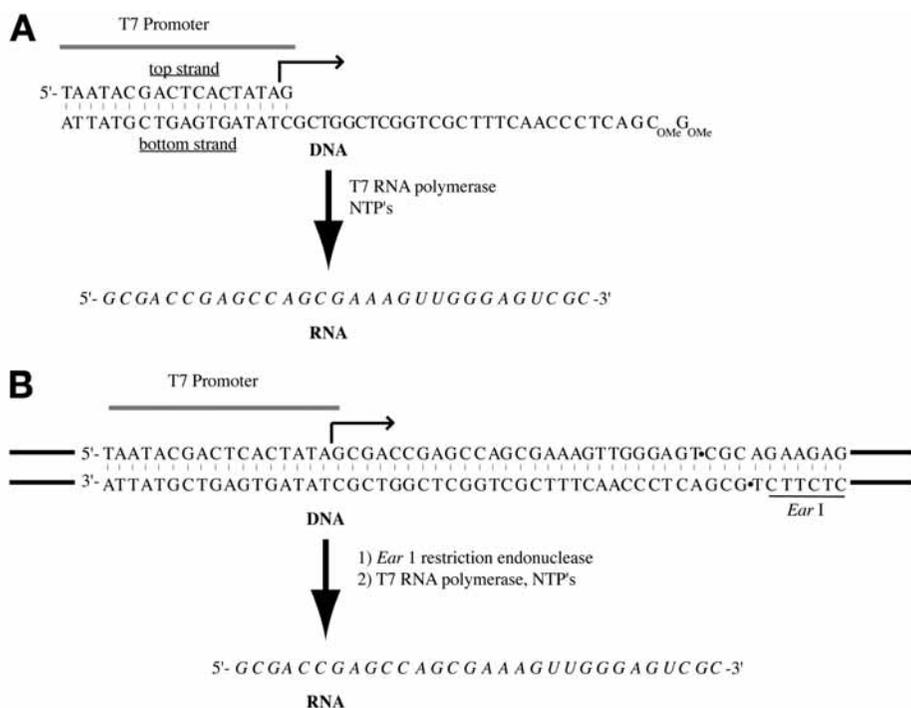
### 3.2. Preparation of Chemically Synthesized RNA

Use of synthetic RNA for crystallography provides several advantages over the use of *in vitro* transcribed molecules: there are little constraints on the sequence and there is less concern with heterogeneity. If the RNA molecules under study are short (30 nucleotides or less), this source of RNA is an option (Dharmacon Research, Lafayette, CO is one source for synthetic RNAs for crystallization). Synthetic RNAs should be deprotected as recommended by the supplier and can be purified by gel electrophoresis (**Subheading 3.5.**) or by HPLC (9) (*see Note 1*). The greatest difference between the synthetic RNA and an *in vitro* transcribed version will be the 5' end. Synthetic RNAs will have an unphosphorylated 5' terminus, whereas the *in vitro* transcribed version will have a triphosphate unless otherwise modified.

### 3.3. Preparation of *In Vitro* Transcribed RNAs

Another major source of RNA for crystallographic studies is *in vitro* transcription from a DNA template using a phage, usually T7, RNA polymerase (**Fig. 1**). In order to obtain significant quantities of RNA by *in vitro* transcription with T7 RNA polymerase, the first nucleotide in the RNA sequence must be a guanosine and the 5' end of the molecule should be purine-rich. Typical start sequences include GAG, GCG, or GGAG. Long runs of guanosines at the 5' end should be avoided to prevent 5'-end heterogeneity owing to stuttering of the polymerase. If a different sequence is required at the 5' end of the molecule, the RNA may be synthesized as a precursor and processed prior to purification to generate an alternate 5' end (*see Note 2*).

The 3' end of the molecule is usually generated by run-off transcription of a linear DNA molecule. The molecule may either be synthesized as a linear DNA (i.e., chemically synthesized DNA or PCR product) or it can be linearized by digestion with a restriction endonuclease. Choice of the restriction endonuclease is critical. The enzyme used should leave either a blunt end (such as *Sma*I) or a 5' overhang (such as *Eco*R1). DNA templates with 3' overhang appear to be inhibitory to the transcription reaction. Of particular use are restriction enzymes such as *Ear*I or *Fok*I, which have recognition sites that are distinct from their cleavage sites (**Fig. 1**). This allows design of templates that are completely independent of the sequence required for restriction enzyme digestion.



**Fig. 1.** Templates for transcription of RNA. **(A)** Transcription from a synthetic DNA template. A DNA with a sequence encoding the RNA of interest fused to the T7 promoter is annealed to a DNA complementary to the promoter sequence to generate a template suitable for large-scale RNA transcription. Methylation of one or two nucleotides at the 5' end of the DNA template has been shown to reduce 3' heterogeneity of the RNA product in some cases. **(B)** Transcription from a double-stranded DNA template. Run-off transcription from a double-stranded template can be an efficient means of obtaining milligram quantities of RNA, especially for larger RNAs. The gene of interest is cloned between the T7 promoter and a suitable restriction site. The DNA is cut with the appropriate restriction endonuclease to generate a DNA template with a homogeneous 3' end prior to transcribing the RNA.

T7 RNA polymerase will add untemplated nucleotides on the 3' end of some fraction of transcribed RNA molecules, leading to heterogeneity in the RNA product that can interfere with crystallization. The amount of heterogeneity appears to be dependent on the construct. These unwanted side products are often the major contaminants of RNA preparations. They can either be rigorously purified from the desired product, or avoided by posttranscriptional processing of the RNA (4,9,10) (see **Note 2**).

There are two common sources of DNA templates for in vitro transcription: single-stranded DNAs made by solid-state synthesis, or linearized plasmids. Double-stranded DNA templates generated by PCR may also be used (*see Note 3*).

### 3.3.1. Synthetic DNA Template

Short RNAs can often be efficiently synthesized from single-stranded DNA templates (*11*). The template strand contains the “bottom” strand of the T7 promoter adjacent to the complement of the desired RNA sequence (**Fig. 1**). It has been reported that incorporation of one or two C2'-methoxy modifications at the 5' end of the DNA template significantly reduces 3'-end heterogeneity of the RNA transcript (*12*). A “top” DNA with the sequence 5'-TAATACGACTCAC-TATAG is annealed to the template to reconstitute the promoter as follows:

1. Dissolve the DNA templates in a low salt buffer, such as TE.
2. Mix equimolar amounts of the “top” and “bottom” strands.
3. Heat the mixture to 95°C for 1 min to melt out any secondary structure in the DNA.
4. Snap cool on ice to anneal the strands.

### 3.3.2. Plasmid DNA as DNA Template

For longer RNAs, it is more efficient to use a high-copy number plasmid as the template for in vitro transcription. Although it is possible to introduce the template into a plasmid that contains a T7 promoter, the most commonly used strategy is to PCR amplify the template of interest fused to the T7 promoter and a restriction site for the run-off transcription reaction. This DNA can then be cloned into a plasmid that does not contain a second promoter region. The commercially available plasmid pUC-19 (New England Biolabs, Beverly, MA) is often used as a cloning vector to make DNA templates for in vitro transcription reactions.

#### 3.3.2.1. PLASMID PURIFICATION

Once the gene is cloned, the plasmid will have to be expressed in a low-nuclease *Escherichia coli* host such as XL-1 Blue (Stratagene, La Jolla, CA) and purified using a plasmid prep protocol. There are many suppliers of kits capable of plasmid purification on the milligram scale. Qiagen kits, especially the QiaFilter kits, are an efficient means of isolating purified plasmid from *E. coli* cultures using ion-exchange chromatography. Follow the protocols enclosed in these kits carefully. Care must be taken using most plasmid preparation kits because RNase A, an essential reagent for purification of the DNA, is very efficient at degrading the desired RNA product. It is a good idea to line benches with a waterproof liner and maintain separate pipetors for use during plasmid preparations. This will avoid cross-contamination of the laboratory with RNase A. Residual nuclease will be removed from the plasmid after restriction.

### 3.3.2.2. LARGE-SCALE RESTRICTION DIGEST

To generate an RNA with a defined 3' end by run-off transcription, the plasmid must be linearized with a restriction endonuclease.

1. A typical large-scale reaction contains 1 mg supercoiled plasmid DNA, 100  $\mu$ L 10X buffer appropriate for the enzyme (usually supplied by the manufacturer), 500 U restriction enzyme, plus sufficient water to bring the reaction to 1 mL total volume.
2. Incubate the restriction reaction at 37°C for 4–6 h.
3. Remove 1  $\mu$ L of the reaction and analyze the extent of cleavage by agarose gel electrophoresis.
4. If the reaction does not appear to be complete, add an additional 50–100 U of enzyme and allow the reaction to proceed overnight. This protocol has been used successfully with many enzymes, including those such as *Eco*R1 that are associated with star activity (a propensity for reduced specificity under certain reaction conditions).

### 3.3.2.3. PHENOL CHLOROFORM EXTRACTION

Once the plasmid has been completely digested, residual protein contamination from the plasmid preparation and restriction digest (RNase A in particular) should be removed by extraction with phenol:chloroform:isoamyl alcohol.

1. Place 500  $\mu$ L of DNA solution in a 1.5-mL microcentrifuge tube.
2. Add 500  $\mu$ L of phenol:chloroform:isoamyl alcohol (25:24:1, [v/v/v]).
3. Mix well and centrifuge briefly to separate the organic phase from the aqueous phase.
4. Carefully pipet the aqueous DNA solution (usually the top layer) into a clean microcentrifuge tube.
5. Add 100  $\mu$ L of TE to the phenolic phase. Mix well, and centrifuge briefly.
6. Carefully pipet the TE away from the phenolic phase and pool with the DNA solution from **step 4**. Discard the phenolic phase in a container dedicated to phenol disposal. Consult your Chemical Safety Department regarding proper disposal of phenol waste.
7. Add 500  $\mu$ L phenol:chloroform:isoamyl alcohol (25:24:1, [v/v/v]) to the pooled DNA solutions and repeat **steps 3–6**.
8. Add 500  $\mu$ L chloroform to the pooled DNA solution.
9. Mix well, centrifuge briefly to separate the phases, and carefully pipet the aqueous DNA solution (upper layer) into a clean microcentrifuge tube.
10. Divide the DNA into 300- $\mu$ L aliquots. Add 30  $\mu$ L 5 M NaCl and 1 mL 100% ethanol to each aliquot to precipitate the DNA. Mix well and place at –20°C for 1 h. Centrifuge at maximum speed in a microcentrifuge for 10 min at 4°C to pellet DNA. Carefully decant off the ethanol. Centrifuge once again briefly to remove excess ethanol from the sides of the centrifuge tube and pipet remaining supernatant from the tube.

11. Dissolve DNA in TE and determine concentration spectrophotometrically (a double-stranded DNA solution of concentration 50  $\mu\text{g}/\text{mL}$  will have an absorbance at 260 nm of  $\sim 1.0$ ).

### 3.3.3. Reagents for In Vitro Transcription Reactions

In vitro transcription reactions can be carried out using enzyme and reagents synthesized in the laboratory. Alternately, Ambion (Austin, TX) currently sells kits for large-scale synthesis of RNA (MegaScript), which may be a useful option for some laboratories.

#### 3.3.3.1. PREPARATION OF NUCLEOSIDE TRIPHOSPHATES

Relatively large quantities of nucleoside triphosphates (NTP) solutions will be required for large-scale synthesis of RNA. It is significantly more economical to prepare these solutions in bulk in the laboratory. The highest available grade of NTP is not necessary for efficient transcription. Nucleotides that are 95% pure are often suitable for in vitro transcription.

1. Dissolve 0.5 mg of ATP in 5 mL water.
2. Check pH by placing 0.5–1.0  $\mu\text{L}$  of this solution onto a pH strip.
3. If pH is below 7.5, add 10  $\mu\text{L}$  of 10 M NaOH, and check pH again.
4. Repeat until pH of the NTP solution is at least 7.0.
5. Make solutions of CTP, UTP, and GTP in a similar manner.
6. Determine the concentration of the NTP stock solutions spectrophotometrically. The  $\epsilon_{260}$  for neutral solutions of ATP, CTP, GTP, and UTP are 15400, 7500, 1170, 9900 M/cm, respectively.
7. Use the concentrated stocks to make up a solution containing 20 mM each ATP, CTP, GTP, and UTP.

#### 3.3.3.2. T7 RNA POLYMERASE

T7 RNA polymerase is available commercially from multiple sources. For large-scale transcriptions, however, most investigators purify their own in the laboratory. Several protocols for large-scale production of T7 RNA polymerase have been published (13,14), and histidine-tagged versions of this enzyme have been described (15).

### 3.3.4. Transcription Trials

Yields of RNAs are maximized by performing transcription trials to determine both the optimal concentration of reagents and duration of the reaction. This procedure is usually followed even when standard conditions for transcription will be used so that plasmid templates can be verified as being free of RNase A. Ideally, the same stocks and exact ratios will be used in the scaled up transcription reaction. A typical 50- $\mu\text{L}$  trial transcription reaction might contain 1  $\mu\text{g}$  restricted plasmid DNA or 250 nM synthetic DNA template and 1 mM

NTPs (1 mM ATP, 1 mM GTP, 1 mM CTP, 1 mM UTP) derived from the 20 mM stock nucleotides prepared as previously described. The reaction buffer is composed of 40 mM Tris-HCl, pH 8.0, 20 mM MgCl<sub>2</sub>, 10 mM DTT, 2 mM spermidine and this buffer is usually prepared as a 10X stock. The T7 RNA polymerase solution usually comprises 10% of volume of the reaction. The amount of enzyme added should correspond to approx 250 U (or ~5 µg) of polymerase (see **Note 4** for variations).

1. Mix NTPs, transcription buffer, DNA, T7 RNA polymerase, and autoclaved water to a final volume of 50 µL in a microcentrifuge tube.
2. Incubate reaction at 37°C.
3. At appropriate time-points (1, 2, 3 h, and so on) remove 10 µL of the reaction and quench with 10 µL formamide stop buffer. Store these aliquots on ice.
4. Analyze fractions by electrophoresis using a 4–20% polyacrylamide (19:1 crosslinking ratio) gel containing 7 M urea and TBE buffer.
5. Remove the gel from the glass plates by coaxing it onto plastic film.
6. Visualize the RNA by UV shadowing or Stains-All staining.
7. Identify the optimal reaction conditions and reaction time.

### 3.3.5. Detection of RNA by UV Shadowing

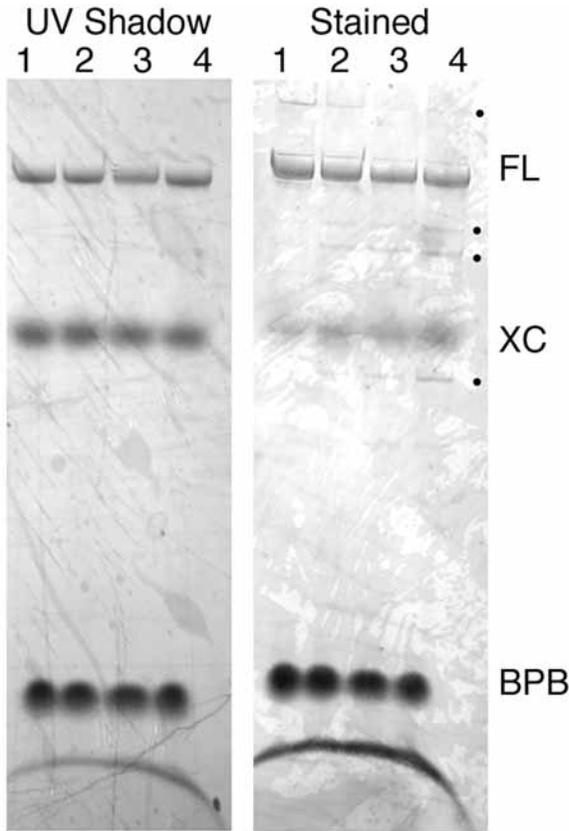
UV shadowing can detect as little as 1 µg or less of nucleic acid. It is the method of choice for identifying bands on a preparative gel. It is also useful for analyzing transcription trials. A UV shadow of an analytical gel is shown in **Fig. 2**.

1. Place the gel, sandwiched between plastic film, onto a TLC plate containing fluorescent dye or an intensifying screen.
2. Hold a shortwave UV lamp ( $\lambda = 254$  nm) over the gel to allow the RNA bands to appear as a dark band on the illuminated screen. UV light will damage the RNA. Thus, work fast to minimize exposure of preparative gels and subsequent damage to the RNA.

### 3.3.6. Detection of RNA by Staining

For analytical gels, staining provides a rapid and more sensitive means of identifying the products of a transcription reaction (**Fig. 2**).

1. Make stock of stain by mixing 1 g of Stains-All in 1 L formamide. Note that this dye is light sensitive and should be stored in amber bottles at 4°C.
2. Make the working solution of Stains-All by mixing 60 mL stock solution with 80 mL of formamide and 160 mL of water. Store this solution at 4°C.
3. Immerse the gel in the working solution of Stains-All, protecting the dye from light using aluminum foil, for approx 10 min.
4. Decant the dye from the gel (the working solution may be reused until it has faded and no longer efficiently stains the gel), and rinse the gel with water.
5. Destain the gel on a light box or under ambient light. The RNA bands should appear purple on a fading magenta background.
6. To record the data, photocopy, photograph, or scan the gel.



**Fig. 2.** Using an analytical gel to analyze a transcription reaction. A transcription reaction was performed as described in **Subheading 3.3.4**. Portions of the reaction were loaded onto a gel containing 6% polyacrylamide, 7 M urea, and Tris-boric acid-EDTA (TBE) buffer. After electrophoresis, the RNA was first visualized by ultraviolet shadowing as described in **Subheading 3.3.5**. (left) and then by staining as described in **Subheading 3.3.6**. Lanes 1, 2, 3, 4 illustrate the progress of the transcription reaction at 1, 2, 3, and 4 h, respectively. Bands corresponding to the full-length RNA, and marker dyes xylene cyanol and bromophenol blue are labeled. Additional bands (•) from degradation of the RNA during long incubations are visible only by staining. BPB, bromophenol blue; FL, full-length; UV, ultraviolet; XC, xylene cyanol.

### 3.4. Large-Scale Synthesis of RNA by *In Vitro* Transcription

Once optimal conditions for transcriptions have been identified, a large-scale transcription is performed to obtain the quantity of RNA required for crystallization experiments.

1. A typical large-scale reaction might contain 20  $\mu\text{g/mL}$  linearized plasmid template or 250 nM synthetic DNA template, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM Tris-HCl, pH 8.0, 25 mM  $\text{MgCl}_2$ , 10 mM DTT, 2 mM spermidine, and T7 RNA polymerase. The polymerase solution usually makes up approx 10% of the reaction mix. This corresponds to approx 0.1 mg/mL of protein or approx 5000 U/mL in the final reaction. It is not unreasonable to perform these reactions on the 10- to 50-mL scale.
2. Place the reaction in a 37°C water bath for the optimal time determined by the trial transcription protocol. Typical times for these reactions range from 1 to 4 h.
3. Transcription reactions often result in precipitation of magnesium pyrophosphate from the reaction. Once the reaction is completed, it is helpful to reverse formation of the precipitate by addition of EDTA to chelate the magnesium. Alternately, formation of the precipitate can be avoided by addition of 1 U/mL *E. coli* inorganic pyrophosphatase (Sigma) to the transcription reaction.
4. For purification by gel electrophoresis or HPLC, at completion of the reaction, the mixture can be concentrated by use of a centricon concentrator, or by ethanol precipitating the reaction and resuspending it in a small volume of autoclaved water.

### 3.5. Purification by Denaturing Gel Electrophoresis

The most common means of purifying RNA for crystallization studies is preparative denaturing gel electrophoresis. The spacers and combs for these gels are often made of 3-mm Teflon, which can usually be purchased locally and cut to provide suitable spacers and combs. The spacers are usually approx 1 cm in width. The combs are cut to provide a single 1.5-cm deep well across the top of the gel. For short RNAs, where it is possible to resolve differences of a single nucleotide, an approx 40-cm tall plate, such as a sequencing gel plate, is usually used. For longer RNAs where single-nucleotide resolution is not attainable, shorter gels (~24 cm) may be used if there are no major side products that migrate near the desired product. In these cases, alternate strategies will be required to eliminate the RNAs with untemplated bases added at the 3' end.

1. The gel matrix contains 6–20% polyacrylamide (19:1 crosslinking ratio), TBE, and 7 M urea.
2. Between 1 and 5 mg of RNA may be loaded on a single gel. To maximize the resolution, the gel should be run such that the RNA reaches the bottom third of the gel matrix.
3. Both glass plates and spacers must be removed from the polyacrylamide and the gel coaxed onto plastic film.
4. The region of the gel that contains the RNA is identified by UV shadowing (**Subheading 3.3.5.**). Because UV light will damage the RNA, exposure should be minimal, long enough only to identify the band, and to outline the region of interest for excision.
5. Cut the band out of the gel using a clean, sterile razor blade.

6. Transfer the gel slices to a sterile, plastic syringe, and force the gel through the syringe tip into a disposable 50-mL centrifuge tube. Each tube should only contain 15–18 mL of crushed gel matrix. (An optional step is to freeze and thaw the crushed gel. This will improve the extraction efficiency in the next step.)
7. Fill the centrifuge tube to the top with TEN buffer to extract the RNA.
8. This mixture should be gently rocked at 4°C overnight to elute the RNA from the polyacrylamide.
9. For maximal recovery of the RNA, the tube is centrifuged to pellet the acrylamide, the TEN buffer is decanted and saved, and the gel matrix is extracted a second time with TEN.
10. To recover the RNA, the TEN extracts are filtered through a 0.2- $\mu$ m sterile filter unit.
11. The RNA is recovered from the TEN by the addition of 3 vol of ethanol, incubation overnight at –20°C, followed by centrifugation (because the RNA solution is concentrated, as little as 3500g are sufficient to pellet the RNA).
12. The pelleted RNA should be resuspended in 1–2 mL of an appropriate buffer (such as 5 mM potassium cacodylate, pH 6.5).
13. The RNA is then transferred to a Centricon concentrator. The RNA is concentrated to 500  $\mu$ L or less by centrifugation, and an appropriate buffer (5 mM potassium cacodylate, pH 6.5) is added to bring the volume up to 2 mL.
14. This step is repeated until small molecules (such as ethanol) that remain as contaminants in the RNA are effectively diluted at least 100-fold by the buffer.
15. The RNA should then be brought to a volume such that the concentration is appropriate for crystallography (10–20 mg/mL).

### **3.6. Renaturing RNA**

Unlike proteins, RNAs are usually purified under denaturing conditions and must be renatured prior to crystallization and the protocol used to anneal the RNA will depend on the folding requirements for the individual molecules (16).

#### *3.6.1. Duplex and Hairpin RNAs*

Symmetrical duplexes have the ability to form hairpin structures and hairpin RNAs may be able to dimerize, forming duplexes. The best means to control this problem is by careful design of the RNA species. Duplex formation is favored under conditions of high-salt and high-RNA concentration. Hairpin formation may be favored by annealing the RNA under dilute conditions and concentrating it to crystallographic concentrations in the native state. Simple RNAs that lack magnesium-dependent structures are often annealed by dissolving the RNAs in dilute buffer and monovalent ion solutions. The mixture is heated to between 50 and 95°C and then either snap-cooled on ice, or slowly allowed to cool to room temperature. Once the mixture is annealed, the RNA should be analyzed by nondenaturing gel electrophoresis with an appropriate size marker to determine if the RNA is monomeric, dimeric, or aggregated.

### 3.6.2. Complex RNAs

RNAs that form complex structures and have magnesium-dependent tertiary folds are annealed using a protocol similar to that for duplexes except magnesium, or other small molecules required for proper folding, are present in the buffer. A sample protocol is described here. **Steps 2 and 3** may not be necessary for some RNAs.

1. Dissolve the RNA in a low-salt buffer in the absence of magnesium.
2. Heat the RNA to 95°C for 1 min to melt out non-native secondary structure.
3. Cool to room temperature.
4. Add magnesium to at least twice the concentration required for folding. The concentration of magnesium needs to be sufficient to fold the RNA even after the RNA solution is diluted with the well solutions during the crystallization experiment. Higher concentrations of magnesium can be detrimental to folding and crystallization of some RNAs.
5. Heat the RNA to 37–50°C for 5–10 min.
6. Cool the RNA to room temperature.

### 3.7. Crystallization

The first variable in crystallization of RNAs is the concentration of the molecules. Typical starting concentrations for small RNAs, including duplexes and hairpins, are 1–5 mM. Larger RNAs are often crystallized at a concentration of 3–10 mg/mL. Sparse matrix approaches and factorial screens are the most commonly used means of screening for initial crystallization conditions (**3,5,17,18**), and some of these are commercially available. A sample sparse matrix screen is shown in **Table 1**.

If these strategies are not appropriate, RNAs may be crystallized by a thorough screening of crystallization conditions, or using a sparse matrix devised in the laboratory. Variations that can be used to obtain or optimize crystals of RNA are discussed in the next section. A surprising number of RNAs will crystallize using a well buffer containing cacodylate buffer, 2-methyl-2,4-pentanediol, monovalent salts, and spermidine or spermine. Other RNAs crystallize readily using PEG/monovalent salt, ammonium sulfate, lithium sulfate, or volatile precipitants such as ethanol. Polyamines, such as spermine or spermidine, can help to stabilize RNA structures and are often included in crystallization.

Crystallization of RNAs is usually accomplished by vapor diffusion using either hanging-drop or sitting-drop plates. Siliconized glass cover slips are recommended for RNA work.

#### 3.7.1. Optimization of Preliminary Crystals or Promising Conditions Observed in a Sparse Matrix Screen

Once initial conditions have been identified that produce an interesting precipitate or small crystals, the experiment will need to be optimized to produce

**Table 1**  
**A 24-Condition Sparse Matrix for Crystallization of RNA**

Well	Buffer	Precipitant	Polyamine
1	Cacodylate, pH 6.5	30% MPD	Spermine
2	Cacodylate, pH 6.5	25% Hexanediol	Spermine
3	Cacodylate, pH 6.5	1.8 M Ammonium sulfate	Spermine
4	Cacodylate, pH 6.5	10% PEG 8K, 50 mM KCl	Spermine
5	Cacodylate, pH 6.5	1.6 M Ammonium sulfate	Spermine
6	Cacodylate, pH 6.5	25% Ethanol	Spermine
7	Cacodylate, pH 6.5	1.8 M Lithium sulfate	Spermine
8	Cacodylate, pH 6.5	30% PEG 400	Spermine
9	MOPS, pH 7.0	1.6 M Lithium sulfate	Spermine
10	MOPS, pH 7.0	25% MPD	Spermine
11	HEPES, pH 7.5	1.8 M Ammonium sulfate	Spermine
12	HEPES, pH 7.5	20% PEG 4K	Spermine
13	HEPES, pH 7.5	25% PEG 4K	Spermine
14	HEPES, pH 7.5	1.8 M Ammonium sulfate	Spermine
15	Tris-HCl, pH 8.0	20% PEG 4K	Spermine
16	Tris-HCl, pH 8.0	30% MPD	Spermine
17	Cacodylate, pH 6.5	25% MPD	Spermidine
18	Cacodylate, pH 6.5	1.8 M Ammonium sulfate	Spermidine
19	Cacodylate, pH 6.5	18% PEG 8K	Spermidine
20	Cacodylate, pH 6.5	20% PEG 1K, 75 mM KCl	Spermidine
21	MOPS, pH 7.0	30% Hexanediol	Spermidine
22	MOPS, pH 7.0	10% PEG 4K, 100 mM NaCl	Spermidine
23	HEPES, pH 7.5	30% MPD	Spermidine
24	Tris-HCl, pH 8.0	1.8 M Lithium sulfate	Spermidine

<sup>a</sup>All buffers are present at a 50 mM concentration and polyamines are present at a 0.5 mM concentration. Cacodylate, MOPS, and HEPES buffers are prepared as the potassium salt. MPD, 2-methyl-2,4-pentanediol; PEG, polyethylene glycol.

diffraction-quality crystals. Most of the strategies employed will be familiar to those working with proteins. The temperature of the experiment, the concentration of the macromolecules, and the concentration of the precipitant can be varied. The identity of the precipitant can be switched in a rational manner ( $\text{Li}_2\text{SO}_4$  instead of  $(\text{NH}_4)_2\text{SO}_4$ , switching to a different molecular weight of PEG, etc.). Some variables, however, are worth particular mention.

1. Buffers. There are no ionizable groups on the surface of RNAs that can be protonated or deprotonated if the pH is kept between 4.5 and 8.5. Crystallization of RNA, in general, tends to be less sensitive to pH than crystallization of proteins. Many RNAs will crystallize over a wide range of pH in a wide variety of buffers.

Potassium cacodylate, pH 6.0–6.5, is a favorite buffer for RNA crystallization because the low pH reduces the amount of backbone nicking and the antimicrobial properties of the arsenic prevent the growth of organisms in the drops during long incubations. There are, of course, exceptions. There are instances where the pKas of functional groups within a RNA are shifted close to neutrality, and these can have a dramatic effect on crystallization. There are also buffer preferences that can be discerned during crystallization of some RNAs. Some commonly used buffers, such as succinate, can chelate divalent metal ions and should be used with great caution in RNA crystallization experiments.

2. Divalent metal ions. Many RNAs require the presence of divalent metal ions, especially  $Mg^{2+}$ , to stabilize their biologically relevant conformation.  $Mg^{2+}$  can also participate in RNA–RNA crystal contacts. Magnesium concentration can be a variable in the initial sparse matrix screen. Altering the magnesium concentration can also significantly improve preliminary crystals, but high concentrations of magnesium ( $>50$  mM) can increase precipitation of RNA or even stabilize alternate RNA conformation. If appropriate, additional divalent cations ( $Ca^{2+}$ ,  $Ba^{2+}$ ,  $Sr^{2+}$ ) or the  $Mg(H_2O)^{2+}$  analog  $Co(NH_3)^{3+}$  can be used as an additive screen or used to replace the magnesium. Beware the appearance of salt crystals when using these ions, especially when phosphate or sulfate ions are present.
3. Monovalent metal ions.  $K^+$ ,  $Na^+$ ,  $Li^+$  (especially potassium ions) also appear to stabilize RNA structures. Variation of the amount and type of monovalent metal ions is an excellent means of perturbing the solubility properties of RNA and can have a profound effect on the quality of crystals obtained. Potassium ions have also been observed to mediate crystal contacts (2).
4. Additives. Additive screening can also be used to improve the quality of RNA crystals. Many RNAs are crystallized in the presence of polyamines, usually spermine or spermidine. A typical concentration range for these compounds is 0.25–2.0 mM. Crystals can be optimized by varying the identity and the concentration of these compounds. Very high concentrations of polyamines can aggregate the RNA, and thereby inhibit crystallization.

### 3.8. Derivatives

#### 3.8.1. Heavy-Atom Soaks

RNA crystals, like protein crystals, can be derivatized by soaking in heavy-atom compounds. The most commonly used strategy is to employ a metal that can bind at one of the high-affinity magnesium-binding sites that are often present within an RNA. Lanthanide ions such as  $Sm^{3+}$  can bind at some sites where atoms from the RNA serve as inner sphere ligands to magnesium ions (19). Many magnesium ions are not coordinated directly by the RNA, but are bound as the hydrated  $Mg(H_2O)_6^{2+}$  species and are recognized by hydrogen bonding between RNA atoms and the water ligands. Cobalt hexammine, rhodium hexammine, osmium hexammine, and iridium hexammine seem to be able to mimic hydrated magnesium ions at some level, binding mostly in the major groove of

RNA, and can be used to produce heavy-atom derivatives (20). Useful concentrations of both lanthanide salts and the hexammines start at about 300  $\mu\text{M}$ .

RNA molecules also have specific binding sites for monovalent ions such as  $\text{K}^+$ . These are thought to be visualized by soaking crystals in thallium salts (20–30 mM thallium acetate) and calculating anomalous difference maps (2,21). Clear sites for one to three atoms per molecule can often be visualized. It is theoretically possible to harness thallium binding to RNAs as a derivatization strategy, although this technique has not yet been used.

### 3.8.2. Site-Specific Covalent Modification of Nucleic Acid

Derivatives of RNA crystals can be obtained by site-specific incorporation of heavy atoms. Halogens, such as bromine and iodine, can be readily incorporated into RNA at the 5 position of pyrimidines. Halogenated RNAs are light sensitive and exposure to light during purification and crystallization should be minimized. There has also been a report that synchrotron radiation can readily debrominate RNA (22). This should be taken into consideration in the data collection strategy. There have been recent reports of phosphoramidites that can be used to synthesize selenium derivatives of RNAs (23). These may prove useful for those studying smaller molecules. In protein crystallography, sulfur substitution to yield a mercury-binding site has been a means to create heavy-atom derivatives. This may also be accomplished with RNA molecules by introduction of a phosphorothioate modification into the RNA backbone. Phosphorothioate modification has been harnessed to solve structures that contain DNA (24,25). Use of this strategy in RNA crystallography has not yet come to fruition, although it is theoretically possible. If a phosphorothioate modification is introduced into a RNA, the base 5' to the modification should be substituted with a deoxyribonucleotide to prevent loss of the sulfur atom during mercury treatment. Additionally, to maximize occupancy of the heavy atom, the  $R_P$  and  $S_P$  diastereomers should be resolved by reverse-phase HPLC (26).

There are several strategies for site-specific modification of RNAs to produce heavy-atom derivatives. If the RNA has been, or can be made by chemical synthesis, it is very easy to target one or two positions within the sequence for modification. Several variants may need to be screened in order to find sites that do not interfere with proper folding or crystallization of the RNA. Phosphorylation of the 5' end of the synthetic RNA may be required for crystallization if the native crystals had been obtained using transcribed RNA.

If the molecules were previously made by transcription and are too large to be made by chemical synthesis, site-specific modification is more challenging. It is relatively simple to modify all of the residues of a particular identity, for example, bromination of all of the uracil bases in the molecule, by transcription with the appropriate modified NTP. However, each site of modification runs the

risk of interfering with either the RNA structure or the crystal contacts. Modification of every position amplifies this risk. The molecule can be synthesized in two fragments, one of which is small enough to be synthesized chemically. The two pieces are then reassembled either by annealing (**Subheading 3.7.**) to yield a nicked RNA species, or ligating the two RNAs with T4 DNA ligase and a DNA splint (27). Modification of the chemically synthesized species using one of the strategies previously described results in derivatization of the RNA (28).

#### 4. Notes

1. An alternative purification strategy for RNAs, is HPLC using an ion-exchange column such as Dionex DNAPac PA-100 semi-preparative (9 × 250 mm) ion-exchange column (9) (Dionex, Sunnyvale, CA). This technique, combined with posttranscriptional 3'-end processing, can allow rapid purification of RNA for structural studies. The column is often heated to 85–90°C to denature residual secondary structure in the RNA that may affect the resolution of the column. RNAs tend to elute between 400 and 800 mM NH<sub>4</sub>Cl. Up to 2 mg of RNA can be injected.
2. There is a strong sequence constraint at the 5' end of any RNA produced by transcription with T7 RNA polymerase. The molecule absolutely must begin with a guanosine residue. If this is unacceptable for the molecules under investigation, the molecule can be synthesized as a precursor and processed posttranscriptionally by *cis*- or *trans*-acting ribozymes (3,4). Heterogeneity at the 3' end of the molecule can also be eliminated using *cis*- or *trans*-acting ribozymes (4,9,10). Processing of nascent transcripts prior to purification can result in an RNA sample in which there are no significant contaminants with molecular weights near the desired product. This, in turn, makes purification significantly simpler (9).
3. It is feasible to use a PCR product to generate DNA templates for the transcription reaction. This approach is especially seductive if many variants will be screened for their ability to crystallize. A library of different PCR products can be synthesized rapidly at a cost of little more than that of the primers. Incorporation of 2'-methoxy modifications on the reverse primer results in a modification of the 3' end of the template that can reduce 3'-end heterogeneity of the RNA. It is very important to use a proofreading DNA polymerase when using PCR to generate DNA templates. Point mutations introduced by PCR are not likely to be detected by sequencing of the PCR product and they will not be removed prior to transcription. The heterogeneous mixtures of RNAs that result from the transcription reaction may not crystallize readily. If this strategy is used, it may be wise to sequence a few clones and get a feel for the heterogeneity of the material that will be produced.
4. To optimize the amount of RNA produced by T7 transcription, it may be necessary to deviate from the "standard" reaction. Variables in the reaction include the concentrations of DNA, NTPs, T7 RNA polymerase, and magnesium. Additionally, MOPS buffer may be substituted for Tris. The concentration of NTPs in the reaction may be increased to 4 mM each. As the NTP concentration is varied, however, the free magnesium is decreased because of chelation by the triphosphate moiety. In

general, the magnesium concentration should be at least 4 mM greater than the total concentration of NTPs. The transcription reaction can also be carried out at temperatures as low as 4°C, although the reaction may need to proceed for a considerable length of time in order for sufficient product to accumulate.

## References

1. Anderson, A. C., Earp, B. E., and Frederick, C. A. (1996) Sequence variation as a strategy for crystallizing RNA motifs. *J. Mol. Biol.* **259**, 696–703.
2. Correll, C. C., Wool, I. G., and Munishkin, A. (1999) The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J. Mol. Biol.* **292**, 275–287.
3. Golden, B. L., Podell, E. R., Gooding, A. R., and Cech, T. R. (1997) Crystals by design: a strategy for crystallization of a ribozyme derived from the Tetrahymena group I intron. *J. Mol. Biol.* **270**, 711–723.
4. Price, S. R., Ito, N., Oubridge, C., Avis, J. M., and Nagai, K. (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398–408.
5. Scott, W. G., Finch, J. T., Grenfell, R., et al. (1995) Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.* **250**, 327–332.
6. Hoggan, D. B., Chao, J. A., Prasad, G. S., Stout, C. D., and Williamson, J. R. (2003) Combinatorial crystallization of an RNA-protein complex. *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 466–473.
7. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111–113.
8. Ferré-D'Amaré, A. R. and Doudna, J. A. (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556.
9. Shields, T. P., Mollova, E., Ste Marie, L., Hansen, M. R., and Pardi, A. (1999) High-performance liquid chromatography purification of homogenous-length RNA produced by trans cleavage with a hammerhead ribozyme. *RNA* **5**, 1259–1267.
10. Grosshans, C. A. and Cech, T. R. (1991) A hammerhead ribozyme allows synthesis of a new form of the tetrahymena ribozyme homogeneous in length with a 3' end blocked for transesterification. *Nucleic Acids Res.* **19**, 3875–3880.
11. Milligan, J. F. and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Meth. Enzymol.* **180**, 51–62.
12. Kao, C., Zheng, M., and Rudisser, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272.
13. Davanloo, P., Rosenberg, A. H., Dunn, J. J., and Studier, F. W. (1984) Cloning and expression of the gene for bacteriophage-T7 RNA-polymerase. *Proc. Natl. Acad. Sci. USA* **81**, 2035–2039.

14. Li, Y., Wang, E., and Wang, Y. (1999) A modified procedure for fast purification of T7 RNA polymerase. *Protein Expr. Purif.* **16**, 355–358.
15. Chen, C. J., Liu, Z. J., Rose, J. P., and Wang, B. C. (1999) Low-salt crystallization of T7 RNA polymerase: a first step towards the transcription bubble complex. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 1188–1192.
16. Uhlenbeck, O. C. (1995) Keeping RNA happy. *RNA* **1**, 4–6.
17. Doudna, J. A., Grosshans, C., Gooding, A., and Kundrot, C. E. (1993) Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl. Acad. Sci. USA* **90**, 7829–7833.
18. Berger, I., Kang, C. H., Sinha, N., Wolters, M., and Rich, A. (1996) A highly efficient 24-condition matrix for the crystallization of nucleic acid fragments. *Acta Crystallogr. D. Biol. Crystallogr.* **52**, 465–468.
19. Cate, J. H. and Doudna, J. A. (1996) Metal-binding sites in the major groove of a large ribozyme domain. *Structure* **4**, 1221–1229.
20. Cruse, W., Saludjian, P., Neuman, A., and Prange, T. (2001) Destabilizing effect of a fluorouracil extra base in a hybrid RNA duplex compared with bromo and chloro analogues. *Acta Crystallogr. D. Biol. Crystallogr.* **57**, 1609–1613.
21. Basu, S., Rambo, R. P., Strauss-Soukup, J., et al. (1998) A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **5**, 986–992.
22. Ennifar, E., Carpentier, P., Ferrer, J. L., Walter, P., and Dumas, P. (2002) X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1262–1268.
23. Carrasco, N., Ginsburg, D., Du, Q., and Huang, Z. (2001) Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography. *Nucleosides Nucleotides Nucleic Acids* **20**, 1723–1734.
24. Yang, W. and Steitz, T. A. (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell* **82**, 193–207.
25. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A., and Schultz, S. C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95**, 963–974.
26. Slim, G. and Gait, M. J. (1991) Configurationally defined phosphorothioate-containing oligoribonucleotides in the study of the mechanism of cleavage of hammerhead ribozymes. *Nucleic Acids Res.* **19**, 1183–1188.
27. Moore, M. J. and Query, C. C. (2000) Joining of RNAs by splinted ligation. *Meth. Enzymol.* **317**, 109–123.
28. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1996) X-ray crystallography of large RNAs: heavy-atom derivatives by RNA engineering. *RNA* **2**, 1295–1305.



## Crystallization of RNA–Protein Complexes

Eiji Obayashi, Chris Oubridge, Daniel Pomeranz Krummel,  
and Kiyoshi Nagai

### Summary

RNA-binding proteins play crucial roles in many biological processes, such as transcription, pre-mRNA splicing, nuclear-cytoplasmic transport of RNA, and translation of mRNA. Specific RNA–protein interactions are key to the correct assembly of ribonucleoprotein complexes and their biological functions. To date, more than 100 unique RNA–protein crystals have been prepared and there are more than 300 entries of RNA–protein complex structures in the Protein Data Bank. This chapter focuses on methods of RNA–protein complex crystallization discussed in six sections: determination of protein-binding sites in RNA, preparation of RNA, preparation of protein, annealing of RNA, reconstitution of RNA–protein complex, and searching crystallization conditions.

**Key Words:** RNA–protein complex; crystals; RNA-binding proteins.

### 1. Introduction

RNA is a fascinating molecule, which can carry catalytic activity as well as genetic information (1). DNA is found predominantly in the double-stranded form, whereas RNA can fold into complex three-dimensional structures. In eukaryotic cells mRNA precursors undergo capping, splicing, and polyadenylation (2). Precursors of tRNAs are processed by RNase P (3) and undergo base modifications (4), and in eukaryotes and archaea CCA is enzymatically added to the 3'-end of tRNA (5). Many RNA-binding proteins play crucial roles in these processes. Aminoacyl-tRNA synthetases specifically recognize and aminoacylate their cognate tRNAs (6), whereas EF-Tu recognizes a common feature of all elongator tRNAs (7). This represents a fascinating problem of tRNA recognition. RNA is also an integral component of the ribosome (8), spliceosome (9), snoRNP (10), RNase P (3), signal recognition particle (SRP) (11), and telomerase (12). Specific RNA–protein interactions are key to the correct assembly of these complexes and their biological functions.

*From: Methods in Molecular Biology, vol. 363: Macromolecular Crystallography Protocols: Volume 1: Preparation and Crystallization of Macromolecules Edited by: S. Doublé © Humana Press Inc., Totowa, NJ*

The first RNA–protein complex structure to be determined was of *Escherichia coli* glutamyl-tRNA synthetase in complex with its cognate tRNA (13). The crystal structures of yeast aspartyl-tRNA synthetase (14) and *Thermus thermophilus* seryl-tRNA synthetase (15) in complex with their cognate tRNAs soon followed. For these complexes tRNAs were purified from over-producing strains. Crystals of the MS2 bacteriophage capsid protein with a fragment of viral mRNA was prepared by soaking a small chemically synthesized RNA oligonucleotide into preformed crystals of spherical viral capsid (16). The complex between the spliceosomal U1A protein and a fragment of U1 snRNA was the first to be cocrystallized with a chemically synthesized RNA (17,18). To date (up to April 2005), there are greater than 300 PDB entries of RNA–protein complex structures including MS2 capsid–RNA variant complex structures and crystal structures of the large and small ribosomal subunits with various antibiotics and fragments of various tRNAs. Over 100 unique RNA–protein crystals have been prepared. These structures have provided answers to many fascinating biological problems involving RNA. This chapter focuses on the generation of RNA–protein complex crystals for such studies.

## 2. Materials

1. Standard buffer: 20 mM HEPES-Na, pH 7.5, 200 mM NaCl, and 5 mM MgCl<sub>2</sub>.
2. Phenol equilibrated with 0.1 M Tris-HCl, pH 8.0.
3. Chloroform.
4. Isoamylalcohol.
5. 3.0 M Na-acetate, pH 5.2.
6. Ethanol.
7. Equipment for denaturing/native polyacrylamide electrophoresis.
8. TBE buffer: 10X stock solution (275 g boric acid, 37.8 g EDTA Na<sub>2</sub>, 540 g Tris base in 5 L).
9. Formamide dye: 100 mL formamide (deionized with 5 g of amberlite MB-3), 0.1 g bromophenol blue, 0.1 g xylene cyanol, and 2 mL 0.5 M EDTA.
10. Toluidine blue stain solution: 0.1% (w/v) in water.
11. 8 M urea.

## 3. Methods

The methods described next outline: (1) prediction of secondary structure of RNA; (2) large-scale preparation of RNA; (3) preparation of protein for RNA–protein complex crystallization; (4) method for annealing of RNA; (5) assembly of proteins with RNA; and (6) crystallization of RNA–protein complexes.

### 3.1. Determining the Three-Dimensional Fold of RNA and Mapping of the Protein-Binding Site

Proteins bind RNAs with diverse structures, from the single-stranded form to complex folded domains. Biochemical characterization of RNA fold and the

protein-binding site is, therefore, a first key step for crystallization of a RNA–protein complex. Secondary structures of RNA can be predicted with reasonable accuracy using programs such as MFOLD (19). The secondary structure of a RNA is well conserved during evolution such that a mutation of a base paired nucleotide is usually compensated by a mutation of its base pairing partner. Hence, a predicted secondary structure can be verified by covariation analysis of evolutionarily related RNA sequences (20). Structural information of RNA can also be obtained experimentally by chemical probing. Functional groups of nucleotide bases react with chemicals such as dimethylsulfate (DMS), kethoxal, diethylpyrocarbonate (DEPC), and so on, and their reactivity is modulated by hydrogen bonding and their chemical environment (21). Hydroxyl-radical and ethyl-nitrosourea foot-printing is also used to probe the structure of the sugar-phosphate backbone (22). Chemical and enzymatic probing are also used to map protein-binding sites within RNA, but care must be taken in interpreting the results as a protein-induced structural change of RNA can also alter the reactivity of functional groups (*see Note 1*). Once a minimal protein-binding domain is identified a next important step is to produce a corresponding fragment of RNA in isolation to see if it indeed binds protein in the expected manner (*see Note 1*).

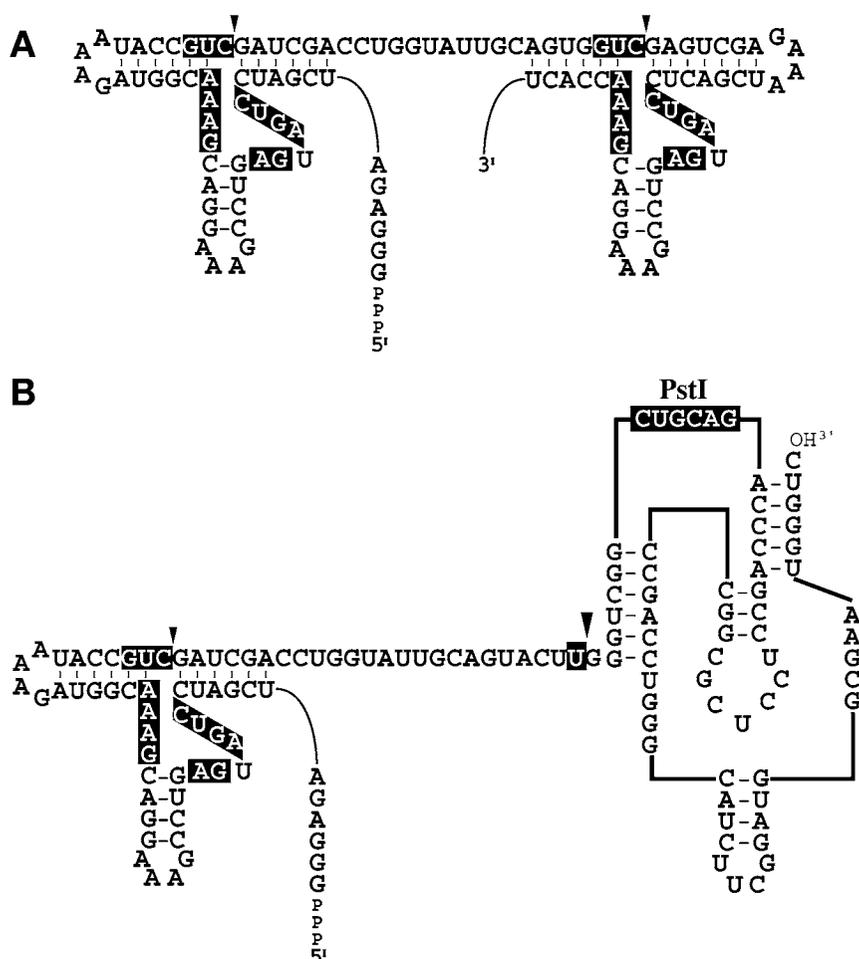
### 3.2. Large-Scale Preparations of RNA for Crystallization

Large amounts of RNA suitable for crystallization can be synthesized either chemically or enzymatically (*see Chapter 12*). Short RNA oligonucleotides can be synthesized on a standard DNA synthesizer using phosphoramidite chemistry. For the crystallization of the U1A–hairpin RNA complex (17,18) and the U2B''–U2A'–RNA complex (23) we used phosphoramidites with the 2'-*o*-triisopropylsilyloxymethyl-protecting group (24) from Glen Research. Dharmacon provides custom synthesis of RNA including a variety of modified nucleotides using phosphoramidites with the 2'ACE (2'-*O*-bis[2-acetoxyethoxy]methyl) protecting group (25).

RNA longer than 30 nucleotides can be best produced by *in vitro* transcription using bacteriophage RNA polymerases such as T7 RNA polymerase (26). A large amount of hexahistidine-tagged T7 RNA polymerase can be purified free of RNases by metal chelate chromatography (Shrader, T., personal communication). *In vitro* transcription with bacteriophage RNA polymerases works less well for short RNAs as T7 RNA polymerase tends to fall off from the template before it undergoes a large conformational switch from the initiation to the elongation form (27–29). For RNA of about 20–30 nucleotides a deoxy-oligonucleotide template can be used together with a short oligonucleotide of the complementary strand in the promoter region (30,31). For a longer RNA the purity (homogeneity) of the oligonucleotide template could become an issue.

RNA longer than 30 nucleotides can be best synthesized by run-off transcription using a linearized plasmid template. A desired template DNA sequence is assembled using overlapping oligonucleotides and is cloned into a high copy-number plasmid such as a member of the pUC family (32). Correctly assembled template sequence should be verified by DNA sequencing. Restriction enzymes that give 3' overhanging nucleotides should be avoided for linearization of plasmid because in this case RNA polymerase often fails to fall off from the template and continues to synthesize RNA in the opposite direction (33). We use type II restriction enzymes such as *Bsm*AI and *Bbs*I for linearization of plasmid template because they cleave outside their recognition sites so that plasmids can be cleaved at any sequence (34). The yield of RNA largely depends on the sequence downstream of the transcription initiation site and a purine-rich sequence (particularly a G-rich sequence) is much preferred at this site (25). However, a run of Gs sometimes results in a heterogeneous 5'-end (35,36). Run off transcription usually results in heterogeneous 3'-end as polymerase tends to add a few nucleotides in a template-independent manner before it falls off (30). It is almost impossible to separate the desired product from such contaminants at preparative scale for a long RNA (see Note 2).

Heterogeneous 5'- and 3'-ends may be less of an issue for a large piece of RNA to be crystallized as the probability of the 5'- and 3'-ends to be involved in crystal contacts is low. However, for shorter RNA the ends are likely to be involved in crystal contacts and it is essential to avoid heterogeneity. In order to get around this problem, RNA can be cotranscribed with flanking self-cleaving ribozymes (34). We use a hammerhead ribozyme at the 5'-end as there is no specific sequence requirement following the scissile bond. The 5'-end of the transcript is cleaved off so that a strong transcription initiation sequence such as that of the  $\phi$ 10 promoter (GGGAGA) can be used. We use either a hammerhead or hepatitis  $\delta$  virus ribozyme on the 3'-end (Fig. 1A,B). In order for hammerhead ribozyme to cleave RNA efficiently the scissile bond must be preceded by a GUC sequence. Hence, the product must end with GUC when a hammerhead ribozyme is used on the 3'-end (37–39). In the case of hepatitis  $\delta$  virus ribozyme the scissile bond can be preceded by any nucleotide except C, which apparently base pairs with the ribozyme sequence and prevents cleavage (40). Transcription reaction should be carried out at high magnesium concentration (30 mM) to allow cotranscriptional ribozyme cleavage. However, the hepatitis  $\delta$  virus ribozyme folds less efficiently and, hence, a complete cleavage can only be achieved by extra annealing steps. For example,  $MgCl_2$  concentration is raised to 50 mM and the reaction is heated to 90°C, snap cooled to 55°C, and held at 55°C, for at least 5 h (41). For details of RNA preparation and protocols for crystallization readers should refer to Price et al. (32) (see also Note 2).



**Fig. 1.** Self-cleaving RNA constructs. **(A)** Double hammerhead construct. Nucleotides on the black background are essential for cleavage. Scissile bonds are shown with arrows. The 3'-end of RNA must be preceded by GUC to allow cleavage of the 3' hammerhead ribozyme. **(B)** RNA flanked by 5' hammerhead and 3' hepatitis  $\delta$  virus ribozymes. The *PstI* site can be used for the cloning of new sequences. The nucleotide preceding the 3' ribozyme scissile bond can be anything except C.

### 3.3. Protein Preparation for RNA-Protein Complex Crystallization

Many RNA-binding proteins consist of multiple domains that are connected by flexible linker peptides. For example U1A (42), U2B'' (42), *Drosophila* sex-lethal protein (43) and poly-A binding protein (44) consists of two to four copies of the RNA-recognition motifs (RRMs), whereas TFIIIA consists of

nine Zn-fingers (45). To crystallize these proteins in complex with their target RNA-binding sites it was crucial to identify a minimal RNA-binding fragment that binds the target site strongly (46). RNA-binding studies of protein fragments have been very informative (41,46–49). For mapping of a minimal RNA-binding fragment a limited protease digestion of free and RNA-bound protein and a subsequent analysis of proteolytic fragments by mass spectrometry can provide vital information in identifying an RNA-binding fragment (50). Fine mapping of a minimal binding fragment should also be carried out by expressing PCR-generated fragments in *E. coli*, for example.

For cocrystallization with RNA extra care is needed to prepare proteins free of even trace amounts of RNases. Ambion® RNA alert system provides a quick and sensitive RNase contamination assay of chromatographic fractions. The RNA substrate contains a fluorescent dye on one end and a quencher on the other so that its fluorescence increases when it is cleaved. Fractions rich in RNase can be identified quickly and excluded from subsequent purification steps. However, we have experienced significant RNA degradation when RNA was mixed with a protein sample that appeared virtually negative by this assay (Obayashi, E., unpublished results). The protocol for checking degradation of RNA is shown next. The most pertinent assay for RNase contamination is to incubate together protein and RNA to be crystallized at room temperature for a few weeks. For RNA components of a large ribonucleoprotein assembly, the incubation of RNA with individual protein components is a stringent test as RNA can be well protected from RNases within the assembled RNP particle (Obayashi, E. et al., unpublished results). Ideally RNA should remain intact for 2 wk or longer but this cannot easily be achieved. For the RNA–protein complexes, which we have crystallized so far, crystals grew to a maximum size within 1 wk (18,23,51,52). RNA in the crystal was intact but RNA in the mother liquor was often degraded (18). It appears that degraded RNA was excluded from the crystal lattice even though RNA was degraded during the course of crystallization. Slight RNA degradation may, therefore, not always be a problem but monitoring RNA degradation during the course of crystallization is absolutely essential.

### 3.3.1. Checking RNA Degradation

1. Mix 5–10 µg RNA with equivalent molar protein(s) in the standard buffer or crystallization buffer in an Eppendorf tube.
2. Incubate at 37°C or at room temperature overnight or for a few days.
3. Add a half volume of phenol equilibrated with Tris-HCl, pH 8.0, and of chloroform:isoamyl-alcohol (24:1) solution, then mix well.
4. Separate phases by brief centrifugation (16,000g) and carefully transfer the upper phase into a fresh tube.
5. Repeat process three to four times.

6. Add one-tenth volume of 3 M Na-acetate and 2.5 vol of ethanol, then mix well.
7. Place tube at  $-20^{\circ}\text{C}$  for 10 min.
8. Centrifuge at 16,000g for 30 min at  $4^{\circ}\text{C}$ .
9. Wash pellet with 80% ethanol.
10. Resuspend pellet with 10  $\mu\text{L}$  formamide dye or 9 M urea/dye.
11. Load it on a  $20 \times 20$ -cm denaturing polyacrylamide gel; run gel at 35 W for 30 min.
12. Stain gel with 0.1% toluidine (or methylene) blue solution for 2 min, then destain it with water.

### 3.4. Annealing of RNA

It may be important to anneal RNA before mixing with protein to ensure correct secondary and ternary structure. RNAs behave in different ways and, therefore, the annealing condition should be optimized for each RNA. RNA should be annealed under different conditions and then analyzed on a native polyacrylamide gel run slowly in cold room to avoid excessive heating, shown in **Subheading 3.4.1**. Golden and Kundrot (53) recommend annealing by slow cooling in  $\text{Mg}^{2+}$  but this method is not applicable to all RNA. Slow cooling in the presence of  $\text{Mg}^{2+}$  often result in dimerization and multimerization of RNA (54). For many RNAs we have worked on it is preferable to anneal RNA in water by snap cooling on ice and add potassium or other monovalent ions first and then divalent ions. Readers should also be reminded that structural conversion of RNA can take place much faster than one might anticipate. When we set up crystallization of a fragment of *E. coli* 4.5S RNA (55) we ensured that the RNA was completely in the hairpin form. When we solved the structure the RNA was found to be in the dimeric form. Similar observations have been made by others (56,57). All hairpin RNAs can also form dimers with the same base pairing scheme and can slowly convert to the dimeric form even at fairly low temperatures ( $10$ – $30^{\circ}\text{C}$ ) until it reaches equilibrium. The form that can be readily packed into a crystal lattice is depleted from the solution and this drives equilibrium toward the dimeric form. This conversion may require a large activation energy and Dumas et al. (57) noted that crystal of the dimeric form appeared more quickly when crystallization was carried out at  $37^{\circ}\text{C}$ . It is always desirable to set up crystallization with a structurally homogeneous RNA preparation but this may not be absolutely crucial as we probably end up with crystals of the form that can be packed most readily into a crystal lattice, rather than what we started with. The hairpin RNA, which was crystallized with the spliceosomal U2B''–U2A' protein complex, could form stable dimers with 11 consecutive base pairs (23). We denatured the RNA at  $80^{\circ}\text{C}$  in the presence of urea and quickly diluted the RNA solution into a solution of the U2B''–U2A' on ice to trap the hairpin form, as shown next (23). It may be necessary to fold some RNAs in the presence of its binding protein.

### 3.4.1. Methods for Annealing of RNA

#### 3.4.1.1. SNAP COOLING

1. Dissolve RNA in water to less than 10  $\mu$ M.
2. Incubate at 90°C for 3 min and quick cool on ice for more than 10 min.
3. Load RNA sample on a 20  $\times$  20-cm native polyacrylamide gel; run gel at 35 W for 30 min.
4. Stain gel with 0.1% toluidine blue solution for 2 min, then destain with water.

#### 3.4.1.2. DILUTION OF UREA

1. Dissolve RNA in 8 M urea to less than 10  $\mu$ M.
2. Incubate at 80°C for 3 min.
3. Dilute the RNA solution with continuous, gentle stirring more than 10-fold into a solution containing an RNA-binding protein on ice.

## 3.5. RNA-Protein Complex Formation

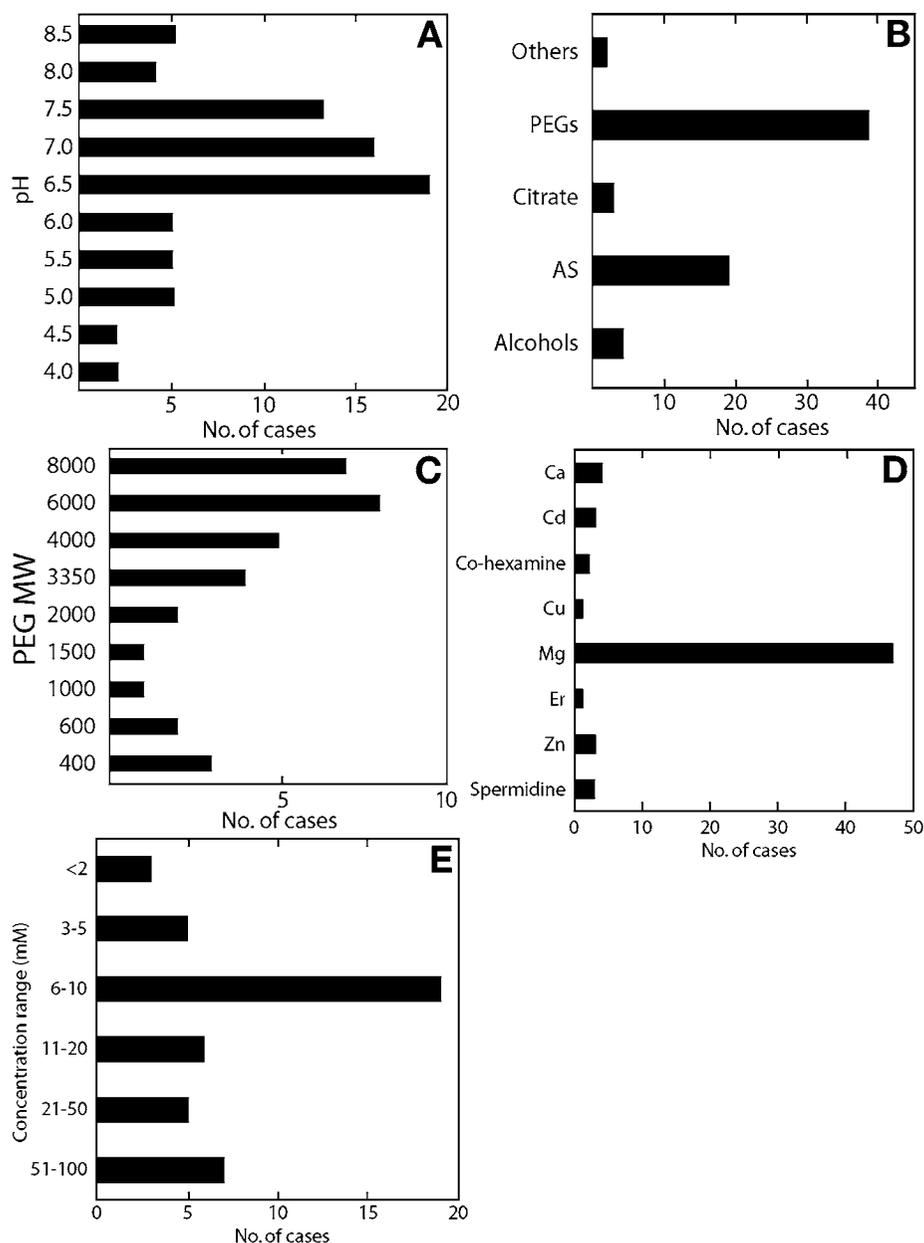
RNA-protein complexes are normally prepared by simply mixing RNA and protein together. Some of the proteins we have worked on are not soluble in low-salt buffer. We normally mix protein in a buffer with sufficient salt for its solubility and then add RNA to the protein solution. This procedure avoids the exposure of protein to low-salt conditions. When U1A and its cognate RNA hairpin were first mixed, they aggregated and white precipitate immediately formed. After standing on ice for a while, the solution became slowly clear (18). This shows that RNA and protein first bind nonspecifically and aggregate but a specific complex gradually forms and the solution becomes clear.

For in vitro assembly of multiprotein ribonucleoprotein complexes such as the U1 snRNP we found that protein components have to be added in a certain order to achieve efficient assembly of the particle (58). In the 1960s Nomura et al. (59) reported a complete in vitro assembly of the 30S ribosomal subunit. It was necessary to add protein subunits to 16S rRNA in a certain order and incubate their assembly intermediates at high temperature. This procedure avoids the formation of unfavorable assembly intermediates. Nature avoids dead-end assembly intermediates by shifting RNA into different cellular compartments and using RNA-modifying enzymes (60) or other RNA-binding proteins. Some proteins may bind RNA cotranscriptionally. The spliceosomal snRNPs are assembled by an elaborate mechanism (61). For example U1 snRNA is transcribed by RNA polymerase II and transported into the cytoplasm in a cap-dependent manner. A large protein assembly called the SMN (survival of motor neuron) complex traps (61) both RNA and some of the protein components (core proteins) and facilitates the assembly of the U1 snRNP. The 5'-end of U1 snRNA in the intermediate state is then hyper-methylated to form the 2,2,7-trimethyl guanosine cap (62). This cap is then recognized by snurportin and

transported back to the nucleus where the remaining subunits (U1A, U1-70K, U1C) are believed to assemble. Consistent with nuclear assembly, the U1A can be imported into the nucleus without interacting with U1 snRNA (63). We have been able to achieve efficient *in vitro* assembly of the U1 snRNP by mimicking this process. It will be necessary to optimize the *in vitro* assembly procedure for each RNP complex (see Note 3). Robinson et al. (64) have shown that a modified mass spectrometry machine can be used to analyze noncovalent protein and protein–nucleic acid complexes as large as GroEL–GroES complex, ribosome, and viruses. The assembly process can be monitored by native gel electrophoresis and mass spectrometry (see Note 3).

### 3.6. Crystallization Conditions

We have compiled crystallization conditions of all published RNA–protein complexes on our webpage: <http://www2.mrc-lmb.cam.ac.uk/personal/kn/NewFiles/crystal.html>. General trends in crystallization of RNA–protein complexes are summarized in Fig. 2 (see also Note 4). Most crystals of RNA–protein complexes grow at a neutral pH range (6.5–7.5) (Fig. 2A), whereas a slightly acidic pH range (pH 6.0–6.5) is preferred for RNA crystallization (65). As shown in Fig. 2B, polyethylene glycol (PEG) is the most commonly used precipitant for crystallization of a RNA–protein complexes (66,67). Figure 2C shows the molecular weight of PEG (apart from PEG-monomethylether) preferred for crystallization of a RNA–protein complex. On the other hand, most complexes of tRNA and tRNA binding proteins (aminoacyl-tRNA synthetases, EF-Tu [7], and methionyl-tRNA formylase [68]) were crystallized from ammonium sulfate. The complex of the spliceosomal U1A protein and its cognate hairpin was also crystallized from ammonium sulfate (17,18). Divalent ions especially magnesium, calcium, or manganese ions are important components for RNA–protein crystallization. Figure 2D,E shows additives and their concentration used for RNA–protein crystallizations. Magnesium ions or hydrated magnesium ions are often seen to interact with the phosphate backbone of RNA in high-resolution structures of nucleic acid or nucleic acid complex crystals (69). Hexamine cobalt or osmium hexamine are used in crystallizations because they mimic hydrated magnesium ions. Polyamines, such as spermine and spermidine, can interact with the phosphate backbone of RNA and are also often used for crystallization. Sauter et al. (70) reported that cyclic polyamines are particularly effective in producing ordered crystals. Many of the important aspects of nucleic acid–protein complexes crystallization have been discussed (66,67). In the majority of DNA–protein complex crystals, DNA forms pseudocontinuous helices. Hence, the lengths of DNA should be close to multiples of half a turn of helix and the crystal quality is optimized by fine-tuning the end-to-end interaction of DNA. Except for crystallization of



**Fig. 2.** General trends in crystallization conditions of RNA-protein complexes. **(A)** pH range used for RNA-protein crystallizations; **(B)** precipitants used for RNA-protein crystallizations; **(C)** molecular weights of polyethylene glycol; **(D)** additives; **(E)** concentrations of additives. These are extracted from our original protein-RNA crystallization table: (<http://www2.mrc-lmb.cam.ac.uk/personal/kn/NewFiles/crystal.html>).

double-stranded RNA-binding protein (71) this strategy is non-applicable. However, the ends of RNA helices are often involved in crystal contacts. **Figure 3** shows the effect of the helix end on the crystallization of the spliceosomal U2B''/U2A' protein with RNA (23). The end of the helix is packed against an RNA base and both the length and the choice of even the penultimate base pair affect crystallization. A similar effect has been observed for the crystallization of the U1A–RNA complex (18). Altering nonessential parts of RNA is an important variable in crystallization.

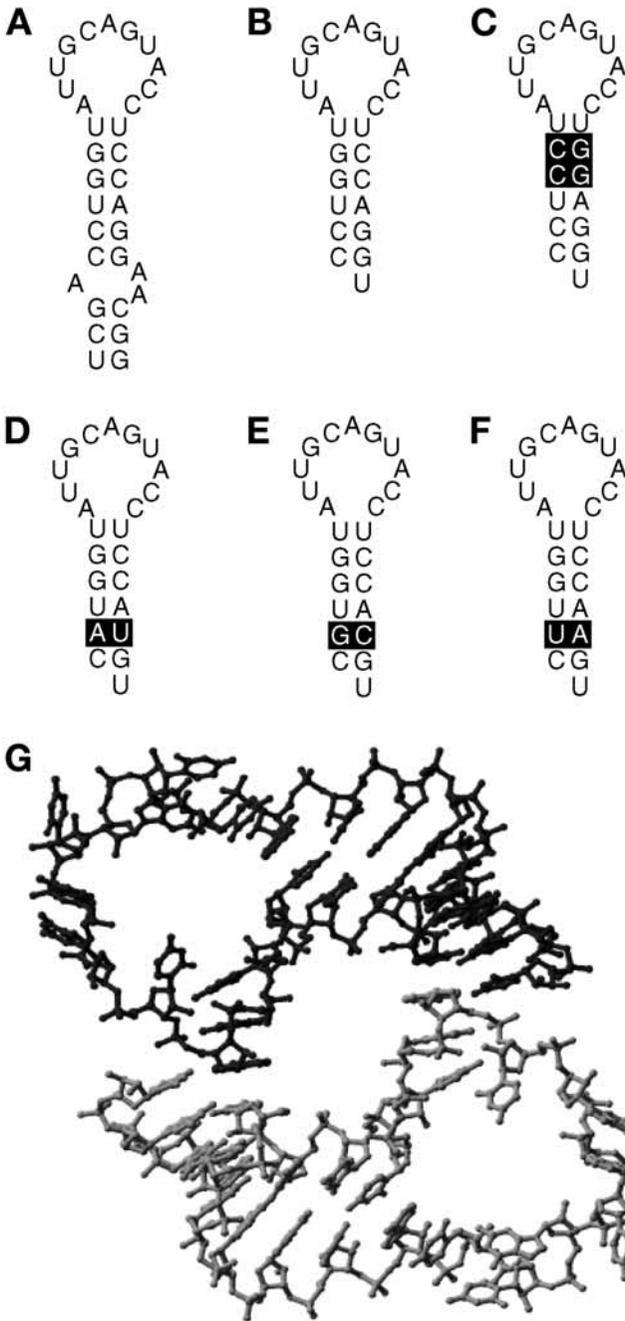
It is usually not essential to set up crystallization with purified RNA–protein complexes or stoichiometric amounts of RNA and protein. Even when RNA and protein bind with high affinity, if one of the components can readily be packed into a crystal lattice, crystals of that component rather than of the complex could form. Aggarwal (66) reported that for crystallization of DNA–protein complexes having excess DNA often results in better crystals. For crystallization of a ternary SRP complex it was important to have 1.5-fold molar excess of proteins over RNA (52). It may not be necessary to change the RNA-to-protein ratio for initial screening but this could be one of the variables at the optimization stage. Crystals of TRAP (tryptophan RNA attenuation protein) in complex with a 33mer RNA was obtained only with a RNA:protein ratio of 1:2 (72). In the crystal both a free TRAP 11mer ring and an RNA-bound ring are seen to be packed tightly in the crystal lattice. In this case the ratio between the two components was essential.

### 3.7. Conclusions

We have described our limited knowledge and experience previously, which we hope is of some use to the readers. Some may think that crystallization is not really a science and only requires “mix and pray.” We believe that like any biological science good knowledge of chemistry and physics, careful observation, profound thinking, and persistent effort are the foundation of successful crystallization. Unfortunately our efforts will not pay off until the structure is solved. However, if we choose an important biological problem, then the structure will reveal something important that is well worth the effort.

### 4. Notes

1. Caution should be exercised in the interpretation of experiments using either enzymes or chemicals to probe RNA in the absence and presence of protein(s). It is important to verify, before doing such experiments, that a complex between RNA and protein forms efficiently and exists in a single conformational state. This can be done on a native acrylamide or agarose gel. If the complex does not form a stable unit then the probing pattern will arise from a combination of free RNA as well as the protein-bound form(s). On the quest to find a minimal RNA-binding site it should be remembered that although small RNA constructs may be favor-



---

**Fig. 3.** The effect of RNA sequence on the crystallization of the U2B''/U2A' protein–RNA ternary complex. **(A)** The binding site for the U2B''/U2A' protein complex within U2 snRNA stem-loop IV. **(B)** RNA used for the final crystallization. **(C)** No crystals are obtained when two GC base pairs in the stem are replaced by CG base pairs. **(D)** Crystals were obtained when the penultimate CG base pair is replaced by AU base pair. **(E)** Crystals were obtained when the penultimate CG base pair is replaced by GC base pair. **(F)** No crystal grew when the penultimate CG base pair is replaced by UA base pair. **(G)** Packing of RNA in the crystal of the U2B''/U2A'/RNA complex (23).

able for crystallization they may reduce the solubility of the RNA–protein complex.

2. We routinely purify RNA by urea-denaturing polyacrylamide gel electrophoresis, either using slab gels or a Bio-Rad prep cell. By using slab gels we can achieve excellent separation of RNA from degradation products, as well as from the ribozyme(s). It has also been observed that acrylamide does carry over with RNA purified by polyacrylamide gel electrophoresis. Gel-filtration has been used to separate RNA products from NTPs and small abortive transcripts (Lukavsky and Puglisi, personal communication).
3. Native-gel mobility shift assay is commonly utilized to demonstrate formation of a stable RNA–protein complex. Temperature can contribute significantly to the efficiency of reconstitution of RNA–protein complexes. Protein may not be able to access RNA that has a rigid form at low temperature, so in cases where efficiency of reconstitution is low it may be useful to incubate RNA with protein at an elevated temperature. This was the case for the reconstitution of the 30S ribosomal subunit (60).

Sodium chloride is often the salt of choice, but it is not generally prevalent under physiological conditions and may not favor stability of macromolecular complexes. One should consider the use of an anion such as potassium and a counterion such as glutamate or acetate. Ordered potassium ions have been observed in the RNA–protein complex structures.

4. Although commercial screens allow one to sample a diverse range of conditions, many of these may not be suitable for RNA–protein complex crystallization. It is particularly inefficient to use such screens when sample is limiting. In such cases it is important to understand the character of your protein–RNA complex and determine the precipitation point and trends in its behavior. Avoid phosphate buffer because it leads to formation of salt crystals in the presence of magnesium.

## Acknowledgments

The authors would like to dedicate this article to the late Max Perutz for his encouragement. The authors also would like to thank all their past and present colleagues, particularly Stephen Price for allowing us to use a figure from his thesis,

Adelaine Leung, Elena Menichelli, Luca Jovine, and Andreas Kuglstatter, for their contributions to the knowledge and experiences accumulated within the group.

## References

1. Cech, T. R. and Bass, B. L. (1986) Biological catalysis by RNA. *Annu. Rev. Biochem.* **55**, 599–629.
2. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M. (1987) *Molecular Biology of the Gene*, 4th ed. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.
3. Gopalan, V., Talbot, S. J., and Altman, S. (1994) RNA–protein interactions in RNase P, in *RNA–Protein Interactions*, (Nagai, K. and Mattaj, I. W., eds.), Oxford University Press, New York, pp. 103–126.
4. McCloskey, J. A. and Crain, P. (1998) The RNA modification database-1998. *Nucleic Acids Res.* **26**, 192–197.
5. Li, F., Xiong, Y., Wang, J., et al. (2002) Crystal structures of the *Bacillus stearothermophilus* CCA-adding enzyme and its complexes with ATP or CTP. *Cell* **111**, 816–824.
6. Cusack, S. (1997) Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* **7**, 881–889.
7. Nissen, P., Kjeldgaard, M., Thirup, S., et al. (1995) Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472.
8. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572.
9. Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) Splicing of Precursors to mRNA by the spliceosome, in *The RNA World*, 2nd ed., (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.
10. Yu, Y. -T., Scharl, E. C., Smith, C. M., and Steitz, J. A. (1999) The growing world of small nuclear ribonucleoproteins, in *The RNA World*, 2nd ed., (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 487–524.
11. Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., and Jovine, L. (2003) Structure, function and evolution of the signal recognition particle. *EMBO J.* **22**, 3479–3485.
12. Blackburn, E. H. (1999) Telomerases, in *The RNA World*, 2nd ed., (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 609–635.
13. Rould, M. A., Perona, J. J., Söll, D., and Steitz, T. A. (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science* **246**, 1135–1142.
14. Ruff, M., Krishnaswamy, S., Boeglin, M., et al. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252**, 1682–1689.
15. Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser). *Science* **263**, 1404–1410.

16. Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J., and Liljas, L. (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**, 623–626.
17. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.
18. Oubridge, C., Ito, N., Teo, C. H., Fearnley, I., and Nagai, K. (1995) Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. *J. Mol. Biol.* **249**, 409–423.
19. Jaeger, J. A., Turner, D. H., and Zuker, M. (1990) Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol.* **183**, 281–306.
20. Westhof, E. and Michel, F. (1994) Prediction and experimental investigation of RNA secondary and tertiary foldings. In: *RNA-Protein Interactions*, (Nagai, K. and Mattaj, I. W., eds.), Oxford University Press, New York, pp. 25–51.
21. Brunel, C. and Romby, P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol.* **318**, 3–21.
22. Merryman, C. and Noller, H. F. (1998) Footprinting and modification-interference analysis of binding sites on RNA. In: *RNA:Protein Interactions; A Practical Approach*, (Smith, C. W. J., ed.), Oxford University Press, New York, pp. 237–253.
23. Price, S. R., Evans, P. R., and Nagai, K. (1998) Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650.
24. Wu, X. and Pitsch, S. (1998) Synthesis and pairing properties of oligoribonucleotide analogues containing a metal-binding site attached to β-D-allofuranosyl cytosine. *Nucleic Acids Res.* **26**, 4315–4323.
25. Scaringe, S. A., Wincott, F. E., and Caruthers, M. H. (1998) Novel RNA synthesis method using 5'-O-silyl-2'-O-orthoester protecting groups. *J. Am. Chem. Soc.* **120**, 11,820–11,821.
26. Davanloo, P., Rosenberg, A. H., Dunn, J. J., and Studier, F. W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **81**, 2035–2039.
27. Cheetham, G. M. and Steitz, T. A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309.
28. Yin, Y. W. and Steitz, T. A. (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **298**, 1387–1395.
29. Tahirou, T. H., Temiakov, D., Anikin, M., et al. (2002) Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* **420**, 43–50.
30. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucl. Acids Res.* **15**, 8783–8798.
31. Milligan, J. F. and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62.
32. Price, S. R., Oubridge, C., Varani, G., and Nagai, K. (1998) Preparation of RNA: protein complexes for x-ray crystallography and NMR. In: *RNA:Protein*

- Interactions; A Practical Approach*, (Smith, C. W. J., ed.), Oxford University Press, New York, pp. 37–74.
33. Schenborn, E. T. and Mierendorf, R. C., Jr. (1985) A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucl. Acids Res.* **13**, 6223–6236.
  34. Price, S. R., Ito, N., Oubridge, C., Avis, J. M., and Nagai, K. (1995) Crystallisation of RNA–protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398–408.
  35. Pleiss, J. A., Derrick, M. L., and Uhlenbeck, O. C. (1998) T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* **4**, 1313–1317.
  36. Helm, M., Brule, H., Giege, R., and Florentz, C. (1999) More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **5**, 618–621.
  37. Sheldon, C. C. and Symons, R. H. (1989) Mutagenesis analysis of a self-cleaving RNA. *Nucl. Acids Res.* **17**, 5679–5685.
  38. Perriman, R., Delves, A., and Gerlach, W. L. (1992) Extended target site specificity for a hammerhead ribozyme. *Gene* **113**, 157–163.
  39. Koizumi, M., Hayase, Y., Iwai, S., Kamiya, H., Inoue, H., and Ohtsuka, E. (1989) Design of RNA enzymes distinguishing a single base mutation in RNA. *Nucl. Acids Res.* **17**, 7059–7071.
  40. Ferré-D'Amaré, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574.
  41. Searles, M. A., Lu, D., and Klug, A. (2000) The role of the central zinc fingers of transcription factor IIIA in binding to 5S RNA. *J. Mol. Biol.* **301**, 47–60.
  42. Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., and Mattaj, I. W. (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. *Nature* **345**, 502–506.
  43. Cline, T. W. and Meyer, B. J. (1996) Vive la difference: males vs females in flies vs worms. *Annu. Rev. Genet.* **30**, 637–702.
  44. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell Biol.* **6**, 2932–2943.
  45. Miller, J., McLachlan, A. D., and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609–1614.
  46. Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., and Nagai, K. (1991) Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447–3456.
  47. Sachs, A. B., Bond, M. W., and Kornberg, R. D. (1986) A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. *Cell* **45**, 827–835.
  48. Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., and Mattaj, I. W. (1989) Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163–4170.

49. Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., and Pieler, T. (1992) RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. *Cell* **71**, 679–690.
50. Cohen, S. L., Ferré-D'Amaré, A. R., Burley, S. K., and Chait, B. T. (1995) Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Sci.* **4**, 1088–1099.
51. Oubridge, C., Kuglstatter, A., Jovine, L., and Nagai, K. (2002) Crystal structure of SRP19 in complex with the S domain of SRP RNA and its implication for the assembly of the signal recognition particle. *Mol. Cell* **9**, 1251–1261.
52. Kuglstatter, A., Oubridge, C., and Nagai, K. (2002) Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nat. Struct. Biol.* **9**, 740–744.
53. Golden, B. L. and Kundrot, C. E. (2003) RNA crystallisation. *J. Struct. Biol.* **142**, 98–107.
54. Jovine, L., Hainzl, T., Oubridge, C., and Nagai, K. (2000) Crystallisation and preliminary X-ray analysis of the conserved domain IV of *Escherichia coli* 4.5S RNA. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1512.
55. Jovine, L., Hainzl, T., Oubridge, C., et al. (2000) Crystal structure of the ffh and EF-G binding sites in the conserved domain IV of *Escherichia coli* 4.5S RNA. *Structure Fold. Des.* **8**, 527–540.
56. Holbrook, S. R., Cheong, C., Tinoco, I. Jr., and Kim, S. H. (1991) Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature* **353**, 579–581.
57. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* **8**, 1064–1068.
58. Muto, Y., Pomeranz Krummel, D., Oubridge, C., et al. (2005) The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol.* **341**, 185–198.
59. Nomura, M., Traub, P., and Bechmann, H. (1968) Hybrid 30S ribosomal particles reconstituted from components of different bacterial origins. *Nature* **219**, 793–799.
60. Will, C. L. and Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301.
61. Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002) Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775–1779.
62. Mattaj, I. W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**, 905–911.
63. Hetzer, M. and Mattaj, I. W. (2000) An ATP-dependent, Ran-independent mechanism for nuclear import of the U1A and U2B' spliceosome proteins. *J. Cell. Biol.* **148**, 293–303.
64. Sobott, F. and Robinson, C. V. (2002) Protein complexes gain momentum. *Curr. Opin. Struct. Biol.* **12**, 729–734.
65. Golden, B. L. and Kundrot, C. E. (2003) RNA crystallisation. *J. Struct. Biol.* **142**, 98–107.
66. Aggarwal, A. K. (1990) Crystallisation of DNA binding proteins with oligodeoxynucleotides. *Methods* **1**, 83–90.

67. Perona, J. J. (1990) Crystallisation of RNA–protein complexes. *Methods* **1**, 75–82.
68. Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1998) Crystal structure of methionyl-tRNA<sup>fMet</sup> transformylase complexed with the initiator formyl-methionyl-tRNA<sup>fMet</sup>. *EMBO J.* **17**, 6819–6826.
69. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B., and Doudna, J. A. (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**, 1232–1239.
70. Sauter, C., Ng, J. D., Lorber, B., et al. (1999) Additives for the crystallisation of proteins and nucleic acids. *J. Cryst. Growth* **196**, 365–376.
71. Ryter, J. M. and Schultz, S. C. (1998) Molecular basis of double-stranded RNA–protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* **17**, 7505–7513.
72. Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X., and Gollnick, P. (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**, 235–242.