Title: Salvage or Recovery of Failed Targets by \textit{in situ} Proteolysis

Authors: Yufeng Tong*, Aiping Dong\textsuperscript{1}, Xiaohui Xu\textsuperscript{2}, Amy Wernimont\textsuperscript{1}

Affiliations:
\textsuperscript{1} Structural Genomics Consortium, University of Toronto, 101 College Street, Toronto, Ontario M5G 1L7, Canada
\textsuperscript{2} Midwest Center for Structural Genomics, University of Toronto, 112 College Street, Toronto, Ontario, M5G 1L6, Canada

* Corresponding author: yufeng.tong@utoronto.ca

Running head: \textit{In situ} Proteolysis
Summary

In situ proteolysis is the method of proactively adding tiny amounts of non-specific proteases to aid in crystallization of proteins, protein macromolecular complexes. The simplicity of the procedure and high recovery rate make it a method of first choice for recalcitrant targets. Here we describe the improved and updated in situ proteolysis protocol used in our high-throughput structural biology platforms.

Keywords: limited proteolysis; in situ proteolysis; crystallization; non-specific protease; trypsin; chymotrypsin

1. Introduction

Limited proteolysis by prolific proteases has been a classical approach to probe the structure and dynamics of proteins and protein-nucleic acid complexes. It has been widely used to identify stable domains in proteins for crystallization [1-5]. The method normally required identification of the termini of the stable fragment and re-cloning of the corresponding construct where possible, or purification of milligrams of homogenous population of protease-trimmed fragment to set up crystallization trials.

Successful crystallization of protein by the addition of tiny amounts of protease in the crystallization drop without purifying the stable fragments, also known as in situ proteolysis, can be traced back to 1936, when James R. Sumner and Stacey F. Howell at Cornell University crystallized Jack bean canavalin in the presence of trypsin at a 1:6 w/w ratio [6]. They were also able to get similar crystals using proteases ficin, papain, bromelase, and pepsin. Interestingly, the use of protease was inspired by their observation that canavalin contaminated with bacteria were crystallizable; but the crystals were not reproducible unless proteases were added [6]. Serendipitous contamination in the crystallization drop by microbes--bacteria or fungi, or by
unknown proteases present in raw materials, yielding to diffracting protein crystals, has been repeatedly reported in the literature [6-15]. However, proteolysis of protein samples during crystallization was considered something that should be normally avoided [16]. Proactive addition of protease in the crystallization drop was not popular until recently [9,12,14,17-19]. The first modern example of in situ proteolysis may be that of the crystallization of antibody Fv fragment in 1997 [9]. After observation of the degradation of scFv in the initial crystals that took more than 3 weeks to grow, David Rose and coworkers decided to add subtilisin at 1:100 molarity ratio, which resulted in formation of crystals within 3 days and an 8-fold increase in crystal size.

Dong et al. showcased the first large scale trial of in situ proteolysis for protein crystallization using chymotrypsin or trypsin [19]. In a follow-up paper, the list of proteases used was expanded and the method applied to proteins of different species and different families with much success, garnering a 13% rescue rate for proteins recalcitrant to giving diffracting crystals [20]. In situ proteolysis has been proven capable of rescuing targets that failed in previous crystallization trials, and improving crystals with low diffraction resolution. The method is now routinely used for protein crystallization at Structural Genomics Consortium (SGC) and Midwest Center for Structural Genomics (MCSG) and is also gaining wider acceptance in the macromolecular crystallography community. It has been combined with the techniques of mass spectroscopy [21,22] and microseeding [23] for protein crystallization; has been used for successful crystallization of protein-protein complexes [23-26]; and has been commercialized by Hampton Research (Proti-Ace) and Jena Bioscience (JBS Floppy-Choppy).

The advantage of the method lies in its simplicity. Proteases are premixed with the target proteins immediately before setting up crystallization drops, and pilot screening is not necessary.
Proteases are treated like a compound in additive screening. The success of the method may arise from the following reasons when proteases exist in the solution: removal of floppy termini of the target protein to give a stable and crystallizable fragment \textit{in situ}; cleavage in the middle of the polypeptide chain to improve the packing of crystal lattice [20,25]; digestion of denatured or unfolded populations of the target protein or contaminant proteins in the solution.

\textbf{2. Materials}

\textbf{2.1. Target proteins}

Routine application of \textit{in situ} proteolysis requires a large amount of protein samples--each protease will require an equal amount of protein as that used for regular crystallization screens. Well expressing and soluble proteins are thus better candidates for the method. Typical protein concentration is 10-40 mg/mL in 10-20 mM buffer devoid of chelating reagents like EDTA or EGTA (see \textbf{Note 1}). While highest purity and homogeneity of the protein preparation is always desirable for crystallization, there is no absolute requirement for the level of purity (see \textbf{Note 2}). \textit{In situ} proteolysis can be applied to stock proteins frozen and stored at -80 °C, to macromolecular complexes, and to multi-domain proteins (see \textbf{Note 3}).

\textbf{2.2. Proteases}

We routinely use eight proteases (see \textbf{Note 4, 5}). Stock solutions of proteases at 1 mg/mL were prepared according to the following formulations:

1. Trypsin from bovine pancreas (Sigma-Aldrich, catalog number T8003) is dissolved in 1 mM hydrochloric acid with 2 mM CaCl\textsubscript{2} (see \textbf{Note 6});

2. $\alpha$-Chymotrypsin from bovine pancreas (Sigma-Aldrich, catalog number C3142) is dissolved in 1 mM hydrochloric acid with 2 mM CaCl\textsubscript{2} (see \textbf{Note 6});
3. Subtilisin A (Sigma-Aldrich, catalog number P5380) is dissolved in 10 mM NaAc with 5 mM CaAc₂ (see Note 6);

4. Thermolysin from *Bacillus thermoproteolyticus rokko* (Sigma-Aldrich, catalog number P1512) is dissolved in 10 mM Tris-HCl at pH 7.5 with 150 mM NaCl, 2 mM CaCl₂ (see Note 6);

5. Papain from papaya latex (Sigma-Aldrich, catalog number P5306) is dissolved in milli-Q grade water;

6. Elastase, pancreatic from porcine pancreas (Sigma-Aldrich, catalog number E0127) is dissolved in 10 mM Tris-HCl at pH 8.8;

7. Dispase I from *Bacillus polymyxa* (Sigma-Aldrich, catalog number D4818) is dissolved in 10 mM Tris-HCl at pH 7.5 with 100 mM NaCl;

8. Endoproteinase Glu-C from *Staphylococcus aureus* V8 (Sigma-Aldrich, catalog number P2922) is dissolved in milli-Q grade water.

The stock solutions are aliquoted into 20 µL each, added to 200 µL PCR strip tubes, flash frozen in liquid nitrogen, and stored at -80°C until the time of use.

### 2.3. Crystallization supply and miscellaneous

1. Commercial or home-made crystallization screens (see Note 7);

2. 96-well sitting drop vapor diffusion crystallization plates;

3. CrystalClear Sealing Tape;

4. SDS-PAGE: 4-20% Mini-PROTEAN TGX precast gel from Bio-Rad;

5. SilverQuest silver staining kit from Life Technology;

6. 0.1% trifluoroacetic acid solution for mass spectroscopy
3. Methods

3.1 Crystallization

Handling of target proteins and proteases should be carried out on ice until ready to set up crystallization drops manually using multichannel pipettes or using liquid handling robot. The following procedure assumes use of one protease from the list of eight and one sparse matrix crystallization screen of 96 conditions. If more than one protease and crystallization screen are chosen, change the volumes correspondingly.

1. Prepare 50 µL target protein solution—recombinant or from natural sources—purified to a reasonable purity and concentrated to 10-40 mg/mL depending on the solubility of the protein (see also Note 1 and Note 2). Measure the protein concentration using NanoDrop based on UV absorbance at 280 nm or Bradford method. At the same time, thaw the protease stock solution on ice.

2. Transfer 100 µL for each of the 96 screen buffers to the corresponding reservoir wells on the 96-well plates, and seal with tape temporarily.

3. Mix 5 µL protease stock solution of choice with the 50 µL target protein solution in a 1.5 mL Eppendorf tube, finger tapping a few times to mix the samples thoroughly, and sediment any aggregation by spinning the sample in a benchtop centrifuge pre-chilled to 4 °C at top speed for 5 minutes (see Note 8).

4. Dispense 96×0.5 µL protein solution to the sample drops on the crystallization plate; transfer 0.5 µL screen buffer from reservoir wells to the sample drops and mix.

5. Seal the crystallization plate with CrystalClear sealing tape and check each drop under a microscope to make sure solutions are well mixed and no bubbles exist in the drops.
Make note of unusual drops and store the plate at room temperature on a stable platform or in an imaging system.

6. Inspect the drops on the second day, the third day, one week, and two weeks from date of setup (see Note 9, 10).

3.2 Verification of protein fragments in the crystals

Once crystals grow to a mountable size, they should be harvested and diffraction tested on an X-ray diffractometer. The size of the protein fragments in the crystals can be roughly estimated using SDS-PAGE or more accurately measured using mass spectroscopy (see Note 3, 11, 12). To harvest crystals for SDS-PAGE verification:

1. Add 8 µL milli-Q grade water to the cap of a 1.5 mL Eppendorf tube;
2. Open the drop with crystals and pipette two 2µL reservoir solution onto a glass cover;
3. Harvest a crystal using harvesting loop, wash the crystal twice in the two drops of reservoir solution and dissolve the crystal in the water;
4. For crystals of 0.05 mm×0.05 mm×0.05 mm dimensions, harvest about 1 or 2 crystals and dissolve in the water (Note 13);
5. Add 3 uL 4×SDS-PAGE loading dye to the tube, close the cap, spin down the liquid for one minute in a benchtop centrifuge at top speed, heat the tube at 95 °C for 5 minutes and load all the sample onto a 4-20% precast polyacrylamide gel and run the electrophoresis;
6. Silver stain the gel using the silver staining kit.

To harvest crystals for mass spectroscopy verification:

1. Find a non-PEG and non-detergent based stabilizing solution in which crystal is not cracked or dissolved by soaking the crystal in the solution and observing under a microscope (Note 12);
2. Add 8 µL milli-Q grade water to the cap of a 1.5 mL Eppendorf tube;

3. Pipette 3-5 drops of 2 µL stabilizing solution found in step 1 onto a glass cover;

4. Harvest a crystal using harvesting loop, and wash 3-5 times sequentially in the drops of the stabilizing solution, then dissolve in the water;

5. For crystals of 0.1 mm×0.1 mm×0.1 mm dimensions, harvest 3 to 5 crystals and dissolve in the water (Note 13);

6. Add 50 µL 0.1% trifluoroacetic acid solution and run mass spectroscopy according to protocol.

4. Notes

1. Sample concentrations higher than that used for normal crystallization--at least 20-30 mg/mL--were recommended in Dong’s original protocol [19]. This gave overall higher success rate. The reason could be that higher protein concentrations lead to higher concentrations of crystallizable fragments after protease treatment. However, successful cases have also been observed for much lower protein concentrations. Chelating reagents in the buffer will deplete metal ions required for the activation or stabilization of several proteases and should thus be avoided.

2. Typical purification procedures in high-throughput structural proteomics settings use immobilized metal affinity chromatography (e.g. Ni-NTA for His6-tagged recombinant proteins) followed by size exclusion chromatography. This can produce proteins of around 95% purity judging by SDS-PAGE for highly expressing constructs. However, in certain cases, contaminant proteins are very difficult to remove and the purity of the target protein may be as low as 50% or less coming out of the standard purification procedure; in situ proteolysis may service the purpose of in situ cleansing of
contaminants (Fig. 1).

3. Due to the nature of proteolysis, the size of the protein fragments in the crystals should be estimated using SDS-PAGE and silver-stained, or the molecular weight measured using ESI or MALDI mass spectroscopy. But the latter method is not always feasible and requires dedicated instrument (See also Note 11, 12, 13). Multi-domain proteins impose a particular challenge for in situ proteolysis and often end up with one stable domain out of the multiple domains in the initial polypeptide[27] (Fig. 2)

4. These eight proteases have all proven successful in our hands. While trypsin and chymotrypsin yielded more structures than other proteases did, the numbers may be quite biased because the sample size tried with these two proteases are much bigger than that with other proteases. Elastase has the lowest success rate so far from our experience. Proteases 3-8 are listed in an increasing order of price per unit weight. When the quantity of the target protein for crystallization is limited, in situ proteolysis can be tried for the proteases from the top of the list first. Exploitation of other proteases not on the list is also encouraged.

5. A previous protocol [20] suggested pilot screening to identify a “promising” protease that gives the largest stable domain. Given the fact that protein fragments from a compact but nicked domain can still interact with each other, and that there is no direct relationship between SDS-PAGE band pattern and the success of in situ proteolysis, pilot screening is not necessary if protein sample is abundant. When the amount of proteins is limited, a “promising” protease from pilot screening is the one that will digest the target protein to smaller pieces rather than the one that does not cleave the target protein at all.

6. While the pH optimum for trypsin and \(\alpha\)-chymotrypsin is around 8.0, the buffer of 1 mM
HCl with 2 mM CaCl$_2$ is to retard autolysis of the proteases. Calcium is also required for the activation and stabilization of subtilisin A, dispase I, and thermolysin.

7. While certain conditions in the commercial screens contain heavy atoms that may inhibit the activity of proteases, this is not generally a concern because: first, proteases are pre-mixed with proteins; second, as it is a sparse matrix screen, missing a few conditions should not make a statistical difference.

8. Concentration of the protease stock solutions are normalized to 1 mg/mL, so that mixing of 5 µL protease solution with 50 µL protein solution at 10 mg/mL gives a protease:protein ratio of 1:100 (w/w). A protease:protein ratio in the range of 1:100 to 1:500 (w/w) has been a good starting point from our experience.

9. In a few cases, we noticed that crystals from \textit{in situ} proteolysis drops formed overnight but started to deteriorate visibly from the second day, and completely disappeared within 3 days. It could probably be that the protease continues to digest target protein even when they were packed in the crystal lattice. Thus the crystallization drops should be inspected promptly and crystals harvested when the size is mountable.

10. We also noticed in cases where a target protein crystallized with or without the presence of proteases, the proteolysis method usually gives crystals much faster, which is consistent of the observation by other groups, e.g. [9].

11. Different proteases can give crystals containing different protein fragments. One same protease can lead to crystals grown in different conditions with different protein fragment contents (Fig. 3). Thus it is important to verify the content of crystals from different crystallization drops.

12. Triton, Tween, SDS, or PEG molecules will swamp the polypeptide signal in an ESI or
MALDI mass spectroscopy, and thus should be carefully removed from the crystals by washing in a solution that does not contain PEG. If the crystals grow from non-PEG, non-detergent based condition, the reservoir solution may well be a stabilizing solution for the crystal. If the crystal grows from a PEG based condition, sodium chloride solution of 1.5 M to 2.0 M is a good starting point to test if it can stabilize the crystal. A stabilizing solution may not always be possible to find.

13. Silver-staining has a practical lower detection limit of about 50 ng protein while a sample for mass spectroscopy normally requires about 5 μg protein. For a 3D crystal of 0.05 mm×0.05 mm×0.05 mm size, it contains about 1.35 g/cm³ × 125×10⁻⁹ cm³ ≅ 168 ng proteins [28], such that one crystal of this size should be detectable using silver staining. Mass spectroscopy requires much more protein samples and it may not be easily obtainable. One 0.1 mm×0.1 mm×0.1 mm crystal contains about 1.35 μg protein. But usually the sizes of protein crystals are much smaller or not of 3D, in which case many crystals are required to run a decent mass spectroscopic analysis.

**Acknowledgment**

We are grateful for members of SGC and MCSG for the domestication of the *in situ* proteolysis method on the high-throughput structural biology platforms.
**Figure Legend**

**Figure 1. Crystallization of the ArfGAP domain of ARFGAP1:** (A) Constructs A2 (a.a. 1-128), A3 (a.a. 1-143) were purified using Ni-NTA (Lane Ni), followed by gel filtration (Lane GF). There were high molecular weight contaminants difficult to remove, which were digested by α-chymotrypsin treatment for 30 minutes at room temperature (Lane Chymo); (B) Both constructs A2 and A3 gave similar plate shaped crystals among heavy precipitates in the drop. Deposited structure (PDB: 3DWD) was solved from a crystal of construct A2. Residues after 120 are not visible in the electron density map.

**Figure 2. In situ proteolysis applied to a multiple domain protein:** (A) All 4 constructs gave crystals of similar shape in the presence of papain 1:100(w/w); (B) Domain architectures of the C-terminal part of ITSN2 protein corresponding to the 4 constructs; (C) Silver staining result of the crystals suggests they are of same size. Solved structure (PDB:3GF9) contains the RhoGEF domain only.
Figure 3. Crystallization of IQGAP2 GRD domain under different conditions: (A) SDS-PAGE of subtilisin generated crystals in condition SGC-A05 (1.6 M (NH₄)₂SO₄, 0.2 M NaAc, 0.1 M HEPES at pH 7.5); (B) SDS-PAGE of subtilisin generated crystals from the Red Wings Screen-H12 (Lane Subst, 20% PEG 4K, 20% isopropanol, 0.1 M NaCitrate at pH 6.5), Chymotrypsin generated crystals from RW-C01 (Lane Chymo, 1.5 M (NH₄)₂SO₄, 0.1 M Tris at pH 8.5); (C) Crystals from different proteases treatment: α-chymotrypsin (RW-C01), subtilisin (RW-H12), dispase (SGC-F3), thermolysin (SGC-G4). Crystals were not seen from drops without in situ proteolysis in the time frame that proteases generated crystals. See http://www.sgc.utoronto.ca/SGC-WebPages/toronto-technology-crystallization.php for the details of the formulations of the SGC Screen and Red Wings Screen.

References


