THE RAPID CRYSTALLIZATION STRATEGY FOR STRUCTURE-BASED INHIBITOR DESIGN

TERESE BERGFORS[*]
Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, 751 24 Uppsala, Sweden

Abstract. RAPID (Rapid Approaches to Pathogen Inhibitor Discovery) is an integrated center for structural biology, computational chemistry, and medicinal chemistry at Uppsala University, Sweden. The main target of the structural biology section is *Mycobacterium tuberculosis*. Key concepts in the crystallization strategy include minimal screening and buffer optimization. Examples are presented showing how these concepts have been successful in RAPID projects. Three screening methods are used: vapor-diffusion, micro-batch, and microfluidics. Our experiences may be relevant for other small, academic laboratories involved in structure-based inhibitor design.

Keywords: Buffer effects, crystallization strategy, manual screening, *Mycobacterium tuberculosis*, protein crystallization, protein–inhibitor complexes, storage of protein, structure–based inhibitor design

1. Introduction

RAPID stands for Rapid Approaches to Pathogen Inhibitor Discovery; it is an integrated center for structural biology, computational chemistry, and medicinal chemistry at Uppsala University, Sweden. The goal of RAPID is structure-based inhibitor design against proteins from the micro-organisms that cause tuberculosis, malaria, leishmaniasis and trypanosomiasis. The structural biology section focuses on tuberculosis, which is caused by *Mycobacterium tuberculosis*.

---

[*] To whom correspondence should be addressed. Terese Bergfors, Department of Cell and Molecular Biology, Uppsala University, Biomedical Center Box 596, 751 24 Uppsala, Sweden; e-mail: Terese.Bergfors@icm.uu.se

The three sections of RAPID interact closely with each other and with their industrial partners. The structural biology section performs target selection, cloning of the gene, expression and purification of the protein, followed by crystallization screening, data collection and structure determination of the proteins and protein-inhibitor complexes. The medicinal/combinatorial chemistry section synthesizes and optimizes the inhibitors for the structural biology section. This chemistry section also performs enzyme inhibition assays and metabolic stability tests. The third section of RAPID is comprised of the computational chemists who perform homology-based modeling, virtual screening, library design, docking routines, scoring functions, and ADME (adsorption, distribution, metabolism and excretion) prediction.

RAPID has been funded since January 2003 by the Swedish Foundation for Strategic Research. The structural biology section has deposited 22 structures from *Mycobacterium tuberculosis* in the PDB (see Table 1); 10 of these are protein–inhibitor or protein–ligand complexes. The structural biology section employs ten graduate students and four principal investigators (PIs).

TABLE 1. Deposited *M. tuberculosis* structures from RAPID 2003–2007.

Rv	Protein	PDB ID
Rv0009	Peptidyl-prolyl cis-trans isomerase A	1W74
Rv0130	Conserved hypothetical	2C2I
Rv0216	Conserved hypothetical	2BI0
Rv1284	β-carbonic anhydrase related protein	1YLK
Rv1295	Threonine synthase	2D1F
Rv2220	Glutamine synthetase	2BVC
Rv2461c	ClpP1	2C8T
Rv2465c	Ribose-5-phosphate isomerase B	1USL
	Ribose-5-phosphate isomerase B	2BES
	Ribose-5-phosphate isomerase B	2BET
Rv2740	Epoxide hydrolase	2BNG
Rv2870c	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2C82
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JCV
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JCX
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JCY
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JCZ
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JD0
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JD1
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2JD2
Rv3588c	β-carbonic anhydrase (dimer)	1YM3
	β-carbonic anhydrase (tetramer)	2A5V
Rv3778c	Possible oxido-reductase	3CAI

The PIs are responsible for their particular area (protein expression, crystallization, methods development, structure solution) whereas students are trained in the entire process, from cloning to structure refinement. The two chemistry sections together comprise 14 scientists and students. As the PI responsible for crystallization, I will focus below on the crystallization strategy within the structural biology section.

2.   Materials and methods

2.1. PROTEIN PRODUCTION

After target selection, the gene is cloned into a pCR®T7/CT-TOPO® or pEXP5-CT/TOPO® vector (Invitrogen). Each construct carries an N-terminal 6-His tag without a linker. The His-tag is not removed for the crystallization trials. The plasmid is transformed into *Escherichia coli* TOP10 cells (Invitrogen). Positive clones are sequenced to confirm correctness, after which they are used to transform *E. coli* strain BL21/AI. Cultures are grown in 2.8 L Buchner flasks containing 1 L LB-medium supplemented with ampicillin, grown to log phase, then induced with 0.02% arabinose. Growth at 37°C continues a further 2–4 h before harvesting by centrifugation. Cell pellets not processed immediately are stored at –20°C.

The standard lysis buffer is 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.1% Triton X-100. After cell debris is removed by centrifugation, the supernatant is applied to a Ni-IMAC column (Qiagen) and the His-tagged protein is eluted with an imidazole gradient. The second purification step is size-exclusion chromatography (SEC), usually a Superdex 75 column, (GE Healthcare), equilibrated in 10 mM Tris-HCl, pH 8.0 and 150 mM NaCl. The protein is always assayed by SDS-PAGE and sometimes with native PAGE as well.

The pure protein fractions are pooled and concentrated at 15°C in centrifugal concentration devices (VivaSciences). The choice of buffer can be critical to the outcome of the concentration step. The protein is eluted in the SEC buffer, which serves as the default buffer in the concentration step. Centrifugation is paused every 5 min to monitor the behavior of the protein. Should the protein show signs of precipitation, the centrifugation step is discontinued, the protein solution (supernatant) is cleared of precipitate, and the supernatant is tested in a buffer screen. The buffer screen is performed as a vapor-diffusion setup where the experimental droplet consists of a 1:1 mixture of protein and reservoir solutions. The reservoir solutions in this case do not contain precipitants but only buffers, from pH 3.5 to 10.5 at

concentrations of 100 mM. The droplet is equilibrated over the reservoir for 1 day or longer and observed for signs of precipitation. To be able to see the precipitation, the protein concentration needs to be high enough. Therefore I recommend using a concentration from 3 to 10 mg/mL, but it can be lower if the protein is already precipitating. Here the goal of the experiment is to find buffer conditions where the droplet remains clear, i.e., the protein remains soluble. The SEC buffer is then exchanged, by diafiltration or dialysis, for one of those found in the screen. The concentration by centrifugation step can be resumed after this buffer exchange – with the aim of achieving a concentration from 3 to 25 mg/mL for the crystallization screen. This method, under the name Optimum Solubility Screening, as well as variations of it, have been recently described in the literature.[1–3] There are now commercial buffer screens available for this purpose (Jena BioSciences, Molecular Dimensions, etc.).

After the protein is concentrated, it is immediately submitted to crystallization screening. Any surplus protein is flash-frozen according to the protocol developed in the laboratory of Prof. Wim Hol.[4]

2.2. CRYSTALLIZATION SCREENING

Crystallization screening is performed on an Oryx 6 robot (Douglas Instruments, UK) as sitting-drop vapor-diffusion trials with 100 µL precipitant solution in the reservoirs and drop volumes of 150 nL protein and 150 nL precipitant. Two different screens kits are used: JCSG+ (available from Qiagen, Molecular Dimensions, etc.), containing 96 conditions, and Mini (Molecular Dimensions), with 24 conditions.

In parallel with the vapor-diffusion trials, the same screen is set up in two additional geometries: as microbatch experiments and in microfluidic chips, (Microlytic, Denmark, www.microlytic.com). The microfluidic setup is shown in Fig. 1. In the microbatch trials, the volumes are identical to those in the vapor-diffusion droplet. A 1:1 mixture of parafin:silicone oil is used to cover the microbatch droplets.

The setups are incubated at 20°C. Other temperatures (4°C, 27°C) might also be tested, but not until the second tier of experiments. The crystallization experiments are observed and the results are recorded in Xtrack, a laboratory information management system developed in our laboratory.[5] The setups are monitored immediately upon setup, then daily for a week, and thereafter on a weekly basis for about 3 months. Visual assessment and recording of the results are performed manually.
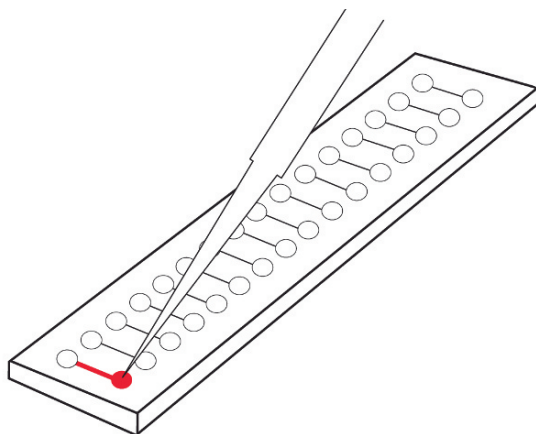
Figure 1. Sketch of the Crystal Former^TM from Microlytic. Here a single channel pipette is used to fill the inlets; a multichannel pipette can be used for simultaneous filling of the inlets. The chip is SBS-compatible for robot loading. The 16 protein inlets are loaded with 150–400 nL each; the channels fill by capillarity. The precipitant is then added to the inlet at the opposite end of the channel. Both rows of inlets are sealed with tape or foil. The figure is reprinted with permission from Microlytic.

2.3. SECOND- AND THIRD-TIER SCREENING

There is no shortage of commercially available screening kits to try, should the first two fail to produce any promising leads. The second tier of experiments varies the temperature and protein concentration and may be expanded to include three other screens: Pact (Qiagen, Molecular Dimensions, etc.), Quik (a phosphate/pH screen, Hampton Research) and Silver Bullet (Hampton Research). The His-tag is still retained at this level of the screening.

Microseeding with any promising solid phase produced in the first round of screening is always done in the second-tier experiments. Promising solid phases include microcrystals, but even crystalline precipitates, spherulites, or seemingly amorphous precipitate. Many amorphous precipitates harbor some crystallinity which is not obvious in visual inspection through the microscope. A seed slurry is generated from the precipitate or other solid phase and a small fraction of it is included as an additive to the new drops. The procedure whereby seeds originating in one mother liquor are used to "innoculate" drops with unrelated mother liquors can be done robotically.[6] It has been dubbed "matrix seeding".[7]

If a third tier of experiments should be necessary, a new construct is made, sometimes without the His-tag. Our construct in the first and second tiers does not have a cleavage site for the His-tag. As a result, removal of

the tag requires a second cloning step. However, all the structures in Table 1 were solved with N-terminal 6-His tags and we have not yet encountered any examples in the RAPID project where removal of the His tag was critical to obtaining the crystals. The most comprehensive analysis to date of His tags in the PDB concludes that they are generally benign.[8] For our project other deletions, usually from the N- and C-termini, have proved to be more effective than His-tag removal for making the proteins more "crystallizable". Certainly by this stage, if not already in the initial cloning step, the amino acid sequence of the protein is analyzed with the bioinformatics programs available at www.disprot.org for evidence of disordered regions that could interfere with crystallization. These are removed in the new constructs.

2.4. SCREENING OF PROTEIN–INHIBITOR COMPLEXES

Two methods for introducing an inhibitor into the protein are co-crystallization and soaking. Others are discussed in an excellent review.[9] In co-crystallization, the protein is incubated together with the inhibitor for a defined time and then the protein–inhibitor complex is set up in crystallization droplets. In the other method, soaking, the protein is crystallized without the inhibitor, and then the inhibitor is soaked into the protein crystal. There are advantages and disadvantages to both methods, but soaking is usually the easier of the two. However, the inhibitor may cause such a conformational change in the protein that the crystal contacts are disrupted. At RAPID, co-crystallization experiments on the protein–inhibitor complex are screened as described above. If crystals of the apo protein are available, they are used in soaking experiments with the ligands and for microseeding the co-crystallization experiments. Soaking experiments are performed in parallel with the co-crystallization ones (when apo crystals are available) to increase the chances of obtaining a crystal of the complex.

The limited solubility in aqueous buffers of the majority of the inhibitors in this project is a major complication, regardless of whether the binding attempts are made as co-crystallization or soaking experiments. The inhibitors are usually dissolved in neat (100%) DMSO (dimethyl sulfoxide). When the inhibitor is added to the protein solution (co-crystallization) or the mother liquor containing the crystal (soaking), the resulting dilution of the solvent (in this case, DMSO) leads to precipitation of the inhibitor. Enough inhibitor might still remain in solution to bind to the protein, but this is not known until the crystal structure is solved. Apart from solubility issues, the inhibitor binding is also dependent in varying degrees upon the buffer, pH,

and other mother liquor components. Therefore, second-tier experiments may include exchanging the mother liquor before soaking or co-crystallization with the inhibitor.

3. Discussion

Three concepts in our crystallization strategy at RAPID are discussed more in depth below. These deal with the questions of how many conditions are "enough" to test in the crystallization screening; the protocol for storing and freezing the protein to improve reproducibility in the crystallization trials; and the role of the protein buffer.

3.1. THE CONCEPT OF MINIMAL SCREENING

Our initial screening strategy uses only 120 precipitant/mother liquor combinations (the 96 in the JCSG+ screening kit and the 24 in the Mini). These are applied to the protein in up to three different geometries: vapor-diffusion, microbatch, and microfluidics. The three geometries affect the equilibration kinetics so differently that each format can generate "hits" that are unique to it. We are currently compiling the success rates of the three geometries for our proteins, but so far our results show that each of the three geometries produce overlaps with each other as well as hits that are geometry-specific. Thus, with only two screening kits and three geometries, 360 conditions can be tested per protein concentration and temperature. The number of screening kits commercially available nowadays is enormous and maintaining an entire stock of them, reformatting them to Deep Well blocks, etc., are expensive and laborious tasks. For simplicity and cost-effectiveness, we therefore use only two screens in the first tier of experiments. At this stage the goal is not to obtain well-diffracting crystals, although that is a welcome side-effect, should it happen. Instead, the goal of the initial screen is to answer the question: "Is this protein likely to crystallize or not?" Extensive screening with hundreds and hundreds of conditions has a limited return on the investment it requires. The efficiency study by Segelke[10] showed that a screen consisting of 300 conditions is a reasonable enough size to determine if "a protein is likely to crystallize or not". Another study by Newman et al.[11] found similar results. The advantage of minimal screening in a first tier of experiments is that it may produce results with little investment of time and effort, but it does not preclude further screening in a second tier. One must also consider the time and effort involved in visual examination of hundreds of drops.

3.2. IMPORTANCE OF THE FREEZING PROTOCOL

Given that the initial screen does not usually produce X-ray ready crystals, optimization of the promising hits is the second step. Even if the initial crystals do exhibit excellent diffraction quality, a drug-discovery program needs to produce more of them for further experiments with the inhibitors. This requires a reproducible and steady supply of the crystals. Batch to batch differences in protein production can lead to irreproducibility in the crystallization, which is why it is clearly an advantage to repeat the crystallization with one and the same batch. At the same time, storage of the batch introduces variations because the protein ages with time. To improve reproducible outcomes from stored protein batches, we use a method which involves flash-freezing the protein solution in aliquots of less than 100 μL in thin-walled Eppendorf tubes for storage at –70°C and then rapid thawing at 37°C.[4]

3.3. THE ROLE OF THE PROTEIN BUFER BEFORE CRYSTALLIZATION SCREENING

Nucleation occurs at high levels of supersaturation. The more protein molecules that are in solution, the more likely it is that a critical mass is reached which can lead to a stable nucleus upon which further growth can occur. If the protein is poorly soluble in a particular buffer or pH, it may never reach a high enough level of supersaturation to support a nucleation event. Thus a higher, rather than lower, protein concentration in the crystallization screening is advantageous. The buffer choice can be critical, but it is often not optimized after the last purification step. Instead the buffer used in the elution of the last chromatographic column becomes the buffer by default in which the protein is concentrated for the crystallization trials. For example, we had one protein that would not concentrate to more than 0.1 mg/mL in the SEC buffer of 10 mM Tris-HCl, 150 mM NaCl, pH 8.0. The protein could be concentrated to 10 mg/mL after exchanging the SEC buffer for a phosphate buffer at the same pH of 8.0. In another case, a protein that precipitated heavily after a few hours in the SEC buffer, crystallized in one of one of the screen buffers without any precipitant. The protein solubility as a function of buffer/pH is easy to test and can therefore be done at an early level of the screening. It is especially useful to examine when the protein does not concentrate to more than 1–2 mg/mL in the SEC buffer.

4. Summary

This chapter covers some of the tenets of the crystallization approach used by our academic laboratory. It is a small laboratory with no automation except a crystallization robot. The suggestions here are not used to the exclusion of the many other options available, such as Thermofluor stability studies, dynamic light scattering, modification of the surface entropy, domain refinement, etc. We use these and other methods when the first screens fail. I have chosen to focus on the ones that I have in this chapter because they are simple to implement. For that reason, they should be considered as a first recourse. For example, changing the buffer of the protein is certainly easier and quicker than cloning a new construct of it.

The size and type of laboratory dictates what approaches are practical, cost-effective, and efficient. The approaches presented here have met these three criteria in our laboratory and they have proved successful. Our experiences may be relevant for other academic laboratories or drug-discovery programs.

References

1. J. Jancarik and S.H. Kim, *Acta Crystallographica* D60, 1670–1673 (2004).
2. A. Izaac, C. Schall and T. Mueser, *Acta Crystallographica* D62, 833–842 (2006).
3. B.K. Collins, R. Stevens and R. Page, *Acta Crystallographica* F61, 1035–1038 (2005).
4. J. Deng, D.R. Davies, G. Wisedchaisri, M. Wu, W. Hol and C. Mehlin, *Acta Crystallographica* D60, 203–204 (2004).
5. M. Harris and T.A. Jones, *Acta Crystallographica* D58, 1889–1891 (2002).
6. A. D'Arcy, F. Villard and M. March, *Acta Crystallographica* D63, 550–554 (2007).
7. G.C. Ireton and B.L. Stoddard, *Acta Crystallographica* D60, 601–605 (2003).
8. M. Carson, D.H. Johnson, H. McDonald, C. Brouillette and L.J. DeLucas, *Acta Crystallographica* D63, 295–301.
9. A. Hassell, G. An, R.K. Bledsoe, J.M. Bynum, H.L. Carter III, S-J. J. Deng, R.T. Gampe, G.T.E. Grisard, K.P. Madauss, R.T. Nolte, W.J. Rocque, L. Wang, K.L. Weaver, S.P. Williams, G.B. Wisely, R. Xu and L.M. Shewchuk, *Acta Crystallographica* D63, 72–79 (2007).
10. B.W. Segelke, *Journal of Crystal Growth*, 232, 553–562 (2001).
11. J. Newman, D. Egan, T.S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J.L. Sussman, D.I. Stuart and A. Perrakis, *Acta Crystallographica* D61, 1426–1431 (2005).