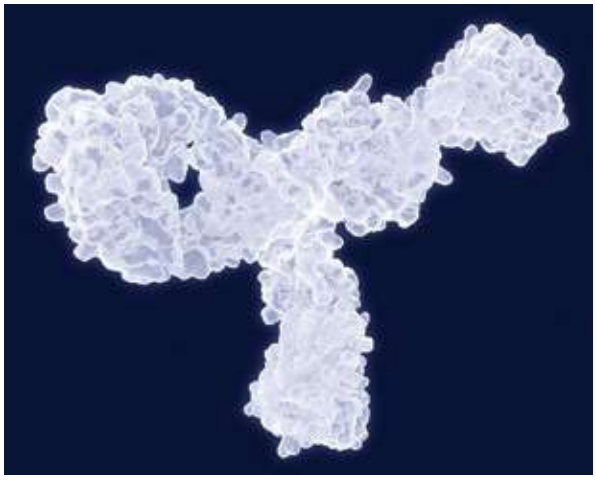
WHITE PAPER

Crystallization of challenging pharmaceutical targets to aid drug discovery

Mary Koszelak-Rosenblum, PhD, Senior Research Scientist, Curia
Guangyu Zhu, PhD, Senior Research Scientist, Curia

X-ray crystallography is the major structural biology method used in drug discovery research. It allows an analysis of the ways a potential drug interacts with its protein target and allows new drugs to be designed. Compared with other structural biology methods, crystallizing the target protein remains challenging and, in most cases, is still the bottleneck step in the entire process of X-ray crystallography.

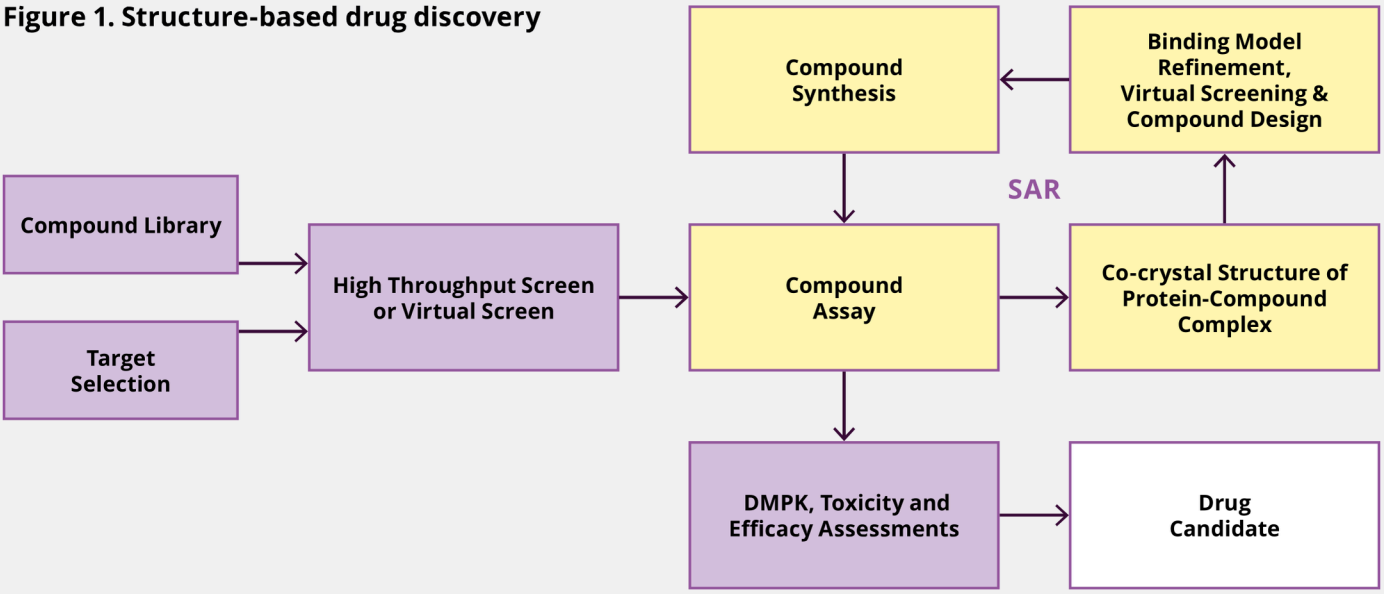
This white paper focuses on a few successful strategies used when growing crystals of challenging pharmaceutical targets to aid drug discovery. Included are two case studies performed by our experienced team of crystallographers at Curia.



Structural biology for drug discovery process

The drug discovery pipeline usually involves multiple steps, including early discovery, preclinical studies, clinical trials, FDA approval, and post-market monitoring. The early discovery phase can be further separated into various subcategories, including target selection and validation, assay development, high-throughput screening (HTS) to identify lead candidates, hit validation (hit-to-lead), and lead optimization (**Figure 1**, next page).

Figure 1. Structure-based drug discovery



Structural biology contributes to the lead optimization stage when the complex structures of the target and the lead compound are available. This allows a direct analysis of the interaction between the compound and its protein target. From there, the information is used to understand the structure-activity relationship (SAR) and guide the compound's further development and design by computational chemists (**Figure 1**). The structural information accelerates the iteration cycles of lead optimization. In many cases, the IC50 of the lead compounds can be improved from micromolar to nanomolar concentrations within a few months.

Structural biology has also been used as an initial lead compound screening strategy. A high-throughput assay (e.g., biochemistry assay) is usually preferred as a primary screen against a large compound library (100,000–millions of compounds). In some cases, a lower throughput assay has to be used when a high-throughput assay is impractical. Both X-ray crystallography and nuclear magnetic resonance (NMR) have been used in this manner. Although the throughput of these structural biology methods is very low, they allow the identification of low binding affinity compounds, which makes fragment-based library screening possible.

X-ray crystallography for drug discovery

Atomic-resolution biomolecular structures can be obtained by NMR, cryo-electron microscopy (cryo-EM), and X-ray crystallography. Traditionally, NMR is restricted to small proteins while cryo-EM is used for larger proteins. With new technology being developed to provide enhanced detectors, the resolution of cryo-EM has been improved to better than 3Å for some proteins with a molecular weight less than 100 kDa. This makes cryo-EM popular for the structural determination of large biomolecular assemblies at high resolution.

When the target can be crystallized, X-ray crystallography generally provides higher resolution structures compared to cryo-EM.

Even with these advancements, X-ray crystallography is still the dominant technique for drug discovery and will be for the foreseeable future. When the target can be crystallized, X-ray crystallography generally provides higher resolution structures compared to cryo-EM. It is also much faster and far less expensive, especially during experimental data collection. For most current synchrotron beamlines, data collection from a single crystal takes only a few minutes. For the beamlines with full automation, processing of one crystal can take less than three minutes, including mounting, centering, and collecting data.

Computational power and newer versions of software have made data processing faster and more efficient. When a target cannot be crystallized, the decision of which experimental technique to use to get the necessary structure information is sometimes unclear. In most instances, such a decision is based on the availability of funding for the project.

CASE STUDY: CRYSTALLIZING PERK-INHIBITOR COMPLEXES

Although some complex structures of the catalytic domain of protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK) with a variety of molecular inhibitors had been solved, we found that not all PERK inhibitors could be crystallized with the protein in the same way. The kinase inhibitors can be classified in different types, such as type I, I ½, II, III, and IV. All these types of inhibitors bind between two lobes of the kinase, but the protein adapts to different conformations upon the inhibitor interaction, such as DFG motif flip and αC-helix moving. Therefore, these conformational changes may affect crystal packing. When we worked on PERK inhibitors with different scaffolds, PERK crystallized to different crystal forms. All of these were caused by subtle conformation changes of the protein. However, PERK could not be crystallized with one inhibitor scaffold (Calvo *et al.* 2021) around existing crystallization conditions and in initial screening.

To obtain a crystallization hit for this inhibitor scaffold, we re-screened with more than 1,200 cocktail solutions at different conditions (e.g., temperature, protein concentration, inhibitor-PERK ratio). At the same time, other PERK-inhibitor crystals were used as seed stocks for cross-seeding to promote crystal formation. PERK was crystallized in a new crystal form with a condition using the precipitate 2-Methyl-2,4-pentanediol (MPD.)
The crystals were extremely twinned even after optimizations (**Figure 2**). Fortunately, we were able to isolate a small piece from the crystal cluster for synchrotron data collection. The crystal diffracted to 2.8Å, and the inhibitor was clearly visible in the electron density map in the solved structure.

Figure 2. PERK protein kinase inhibitor complex crystals. **Left:** PERK crystals with an easy-to-crystallize inhibitor. These were used as seeds to produce for the next round of optimization. **Right:** The crystal drop of PERK protein with a challenging inhibitor appeared after cross-seeding using crystal seeds from the crystals on the left. This cluster was broken into individual crystals for X-ray analysis.

CRYSTALLIZATION STRATEGY

For all the advances in automation and equipment, forming a high-quality crystal still remains the bottleneck in the crystallography process. Although X-ray crystallography is still the gold standard of structural biology techniques, crystallizing the macromolecule to a well-diffracting crystal can be challenging and unpredictable. What is needed is the know-how to form a crystal with enough clarity to lead to X-ray diffraction patterns that provide sufficient data to solve the structure; knowing how to form these crystals is a mixture of both art and science.

Given the limitless potential screening iterations, it is only with experience that researchers can reduce the number of conditions to ensure efficiency and success. In-depth experience helps guide the decisions when faced with numerous experimental strategies to choose from when setting up crystallization experiments. These include:

- Designing different construct
- Modifying the protein (e.g., with limited proteolysis or lysine methylation)
- Selecting a similar homolog
- Mutating surface residues
- Applying alternate crystallization screens
- Changing the incubation temperature
- Using different ligands for co-crystallization
- Incubating with a chaperone or partner (e.g., antibody, lysozyme, aptamer)
- Altering the protein concentration

As for the number of crystallization conditions used for the high-throughput crystallization screen, 400 conditions might be a good starting point for some targets, though this might not be enough. In the past, a 1,536-condition screen was designed for crystallizing membrane proteins (Koszelak-Rosenblum *et al.* 2009), which are more difficult to crystallize than soluble proteins. In some cases, only one hit was found in the 1,536-condition screen for crystallizing membrane proteins, suggesting that more conditions might be needed for certain targets.

By combining all the strategies mentioned above and many other crystallization techniques we have not introduced here, crystallization efforts can be endless if the macromolecule is difficult to crystallize. In the real world, this multidimensional problem has to fit within the limited resources available. Therefore, a strategic plan must be designed based on the properties of the targets, the experience of the scientists, and the availability of the resources.

Given the limitless potential screening iterations, it is only with experience that researchers can reduce the number of conditions to ensure efficiency and success.

Initial crystal hit screening can be performed by various liquid handling systems, such as SPT Labtech's mosquito® or Labcyte's Echo® liquid handling system, which are used at Curia. Here we describe the process of setting up sitting drop vapor diffusion crystallization in a 96-well format using the mosquito®. The crystallization drop consists of 100 nL of protein and 100 nL of the crystallization cocktail pipetted together using a mosquito® liquid handling system equilibrated with 40 µL of cocktail. If a crystal seed solution is available, 20 nL seed solution can be pipetted to the crystallization drop as well. Two crystallization drops with a variety of conditions (e.g., ligands, protein concentration, and protein-to-cocktail ratio) are set up in each well. Robotics are

CASE STUDY: CRYSTALLIZING THE ARTEMIS ENDONUCLEASE

The Artemis endonuclease is required for V(D)J recombination in immune cells and DNA repair. It forms an active complex with the DNA-dependent protein kinase subunit. Inhibition of the Artemis endonuclease makes it attractive as a potential therapeutic for certain cancers, but discovering inhibitors of this enzyme had been limited by difficulty in determining the crystal structure of the protein. As far as we knew, the catalytic domain of Artemis had not been crystallized successfully, despite attempts by a number of academic and industrial organizations. Some related nucleases (SNM1A and SNM1B) in the β-CASP family have been solved.

Human Artemis has 692 amino acid residues and its catalytic domain has fewer than 400 residues. Different truncations were used for the structural determination of the catalytic domain, and those proteins were produced in insect cells. By analyzing the alignment, purification results, and activities of available constructs, only construct 1–368 was selected for all crystallization experiments. Since this construct could not be crystallized directly, we focused on proteolysis (pretreatment or *in situ*) and crystallizing with ligands. Artemis is a nuclease without stringent selectivity. Therefore, we expected Artemis to have an extended positively charged surface for recruiting negatively charged DNA. By varying DNA double-stranded constructs — 9–15 nucleotides per chain, with different overhangs — we hoped that crystal packing might be generated for Artemis.

The advantage of optimizing the DNA constructs rather than the protein constructs was that it was faster and cheaper, since DNA can be synthesized easily while the insect cell expression system takes more than a month for each new construct. Since we planned to screen multiple DNA constructs, only four 96-condition screens were used for each DNA construct.

Eventually, *in situ* proteolysis of Artemis did not provide any crystal hits, and pretreatment with protease (e.g., trypsin) did not generate a homogeneous digestion product, which showed multiple species over size-exclusion chromatography and SDS-PAGE. On the other hand, we obtained two different crystal forms when we screened DNA constructs (**Figure 3**). In both cases, the cy5 dye linked DNAs had to be used for promoting crystallization. Data sets were collected for these crystals and the structures were determined to 2Å resolution (Karim *et al.* 2020). DNA in these two structures was intrinsically disordered, likely because of low specificity of the DNA with Artemis. This is actually an optimum outcome for future work, since the active site of Artemis is not blocked by the DNA and the crystallization methods can be used for future Artemis inhibitor development.

Figure 3. Artemis crystals in blue color due to cy5 dye linked to DNA. Left: Crystal form 1 from co-crystallization with an overhang DNA. **Right:** Crystal form 2 from co-crystallization with a hairpin DNA.

used to screen multiple 96-well plates at a time to find the ideal conditions to get proteins to crystallize. The crystallization drops are imaged weekly using a Formulatrix® Rock Imager®. UV images are also taken to distinguish salt crystals from protein crystals.

Once the conditions to successfully form a crystal have been identified, it is necessary to scale up the process to get larger, loopable crystals in a 2–4-µL drop. Oftentimes, optimization includes varying precipitant concentration, drop volume ratios, temperature, and protein concentration. Experience plays a vital role here since knowing what you are observing in a drop leads you to understand what the next step to optimization might be.

Summary of case studies

Crystallizing the target protein is the bottleneck step of the process of protein crystallography. As numerous experiments can be designed, varying a multitude of parameters, a well-thought-out strategy needs to be developed based on the property of the target. The strategy for using additional protein constructs and mutations was not necessary in the two case studies presented here because the two target proteins had been confirmed with purity, oligomerization state, crystallization ability (PERK), and *in vivo* activity (Artemis). Designing additional protein variants might or might not overcome the crystallization difficulty, but it will expand a project's workload and timeline by requiring tremendous cloning and purification. For the PERK project, the challenging inhibitor was crystallized with PERK using cross-seeding; for the Artemis project, DNA was designed as a crystallization chaperone to facilitate crystal packing. In both cases, we focused on crystallization screens with only one protein construct and we were able to accomplish structure determination within a few months.

About the Curia protein crystallography team

Curia's crystallography team includes a dozen PhDs working on protein expression, purification, and crystallography projects. The team members have published more than 50 structural biology papers in prestigious journals including *Science*, *Nature Structural and Molecular Biology*, *PNAS*, and *The EMBO Journal* with various targets, such as enzymes, kinases, nucleases, membrane proteins, protein complexes, and protein-DNA complexes.

In addition to its crystallography work, Curia also offers services for integrated discovery described in **Figure 1**:

- Medicinal chemistry
- *In vitro* assay development and testing
- Cell line development and cell-based assay
- Compound libraries with more than 600,000 compounds and high-throughput screening
- Drug metabolism and pharmacokinetic studies (DMPK)
- Protein purification, modification, and analysis
- Computer-aided drug discovery (CADD)
- Antibody discovery, maturation and humanization
- Compound management

To learn more about crystallization of challenging pharmaceutical targets to aid drug discovery, visit curiaglobal.com/research-development/discovery/biology/

References

1. Calvo V *et al.* Discovery of 2-amino-3-amido-5-aryl-pyridines as highly potent, orally bioavailable, and efficacious PERK kinase inhibitors. *Bioorg Med Chem Lett* 2021;43:128058.
2. Karim *et al.* Structural analysis of the catalytic domain of Artemis endonuclease/SNM1C reveals distinct structural features. *J Biol Chem* 2020;295(35):12368–12377.
3. Koszelak-Rosenblum M, Krol A, Mozumdar N, *et al.* Determination and application of empirically derived detergent phase boundaries to effectively crystallize membrane proteins. *Protein Sci* 2009;18(9):1828–1839.