Published in final edited form as:

Expert Opin Drug Discov. 2011 May 1; 6(5): 465–480. doi:10.1517/17460441.2011.566857.

Lessons from high-throughput protein crystallization screening: 10 years of practical experience

Luft JR, Snell EH, and DeTitta GT

Hauptman-Woodward Medical Research Institute, 700 Ellicott St., Buffalo, NY 14203, USA and
SUNY Buffalo Department of Structural and Computational Biology, 700 Ellicott St., Buffalo, NY
14203, USA

Abstract

Introduction—X-ray crystallography provides the majority of our structural biological knowledge at a molecular level and in terms of pharmaceutical design is a valuable tool to accelerate discovery. It is the premier technique in the field, but its usefulness is significantly limited by the need to grow well-diffracting crystals. It is for this reason that high-throughput crystallization has become a key technology that has matured over the past 10 years through the field of structural genomics.

Areas covered—The authors describe their experiences in high-throughput crystallization screening in the context of structural genomics and the general biomedical community. They focus on the lessons learnt from the operation of a high-throughput crystallization screening laboratory, which to date has screened over 12,500 biological macromolecules. They also describe the approaches taken to maximize the success while minimizing the effort. Through this, the authors hope that the reader will gain an insight into the efficient design of a laboratory and protocols to accomplish high-throughput crystallization on a single-, multiuser-laboratory or industrial scale.

Expert Opinion—High-throughput crystallization screening is readily available but, despite the power of the crystallographic technique, getting crystals is still not a solved problem. High-throughput approaches can help when used skillfully; however, they still require human input in the detailed analysis and interpretation of results to be more successful.

Keywords

crystallization; high-throughput; structural biology; X-ray crystallography

1. Introduction

The language of life is represented by the genome but the alphabet of the primary sequence reveals only part of the story. Structural biology provides the dictionary and grammar, the tertiary and quaternary structure which in turn provide clues to function and mechanism. Three principal techniques are used to gain this structural knowledge, X-ray crystallography, NMR and cryo-EM. X-ray crystallography is the predominant technique accounting for 86% of our structural knowledge, with NMR and cryo-EM providing 14% and 0.4% respectively¹. The importance of structure in the pharmaceutical field comes from the ability to accelerate discovery; there are more than 200 pharmaceuticals produced from structure based studies to date^{2–5}.

Outside of major pharmaceutical industry research, Structural Genomics (SG) efforts have been a major driver for technological advances in high-throughput crystallographic methods. These advances have been substantial, including developments both upstream of crystallization (molecular biology, cloning, protein expression and purification) and downstream (beamline robotics, software developments etc). Protein Structure Initiative (PSI) efforts, the NIH supported SG program, have produced over 36,000 different soluble purified proteins and have led to ~5,000 protein structures, over 7% of our public structural knowledge to date.⁶ While these numbers are impressive, of immediate note is the loss associated with the process. Only 13% of the purified, soluble, proteins result in structural depositions to the PDB; the vast majority of the proteins remain structurally uncharacterized⁶. Crystallization is a significant problem.

SG relies upon well-defined protocols to achieve high structural output. These are codified in a process termed the pipeline. Oftentimes, once standardized, these protocols are applied uniformly to process every protein. High output is accomplished by increasing the pipeline's capacity. Crystallization screening is one aspect of this pipeline that is readily adapted to increase throughput by using repetitive robotic methods and standardized protocols. A 'standardized' set of screening cocktails is applied to many proteins. This is an approach enabled by technology and governed by resources. This approach is very efficient in terms of throughput; it is an effective means to process large numbers of samples. It will by default neglect the more intense efforts required for samples recalcitrant to the 'standard pipeline' that may be of significant biological and pharmaceutical importance. On the other hand, this pipeline will capture structural information for many important, biologically significant samples using standardized methods.

For over ten years we have been operating a high-throughput crystallization screening service at the Hauptman-Woodward Medical Research Institute. This service is open to the SG community, the general biomedical community and industry. In this review we detail specific aspects of crystallization, automation of those aspects and our experiences operating a high-throughput laboratory. We compare our approaches with other high-throughput crystallization efforts and describe anticipated technological developments that could improve the success rate of a structural pipeline. We focus on soluble proteins and plate-based screening to complement a previous review describing structure based pharmaceutical design and alternative technologies and topics⁵.

2. Crystallization in the laboratory setting

Numerous methods of crystallization have been developed and utilized over the years.⁷ These can be classified into the three basic categories of batch, vapor diffusion, and liquid diffusion. Whatever the method used, identifying the crystallization conditions for macromolecules is a formidable, multi-parametric problem⁸. Exhaustive searches through physical and chemical parameter space are simply not an option, even with high-throughput technologies. Two main approaches have been applied to resolve this problem. Chemical sampling where there has been historical success, or statistically designed sampling strategies to cover larger areas of chemical space. The first is exemplified by the work of Jancarik and Kim⁹, and often referred to as the "sparse matrix screen". Commercialization of this screen and others¹⁰⁻¹² has led to the popularity of this approach and it is often the first strategy used in a crystal growth lab. The second, a statistical sampling approach, was pioneered by Carter and co-workers¹³ who applied principally incomplete factorial designs to crystallization. Relative levels of important chemical factors are sampled to achieve good coverage and good balance in the sampling. Both approaches are complementary and frequently combined providing scaffolds upon which further experiments are built. In each case screening is a first step, with the results guiding subsequent rounds of optimization. The

crystal grower must decide upon both the primary variables to test and the range of those variables (including chemical species and concentrations) to sample. The end product of this step is a set of cocktails; solutions that provide coverage of chemical space relevant to crystallization.

Crystallization methods are used to apply these cocktails to a protein. The combination of a cocktail and a method defines a specific location on a physicochemical map. Two experiments with the same starting chemical conditions but using different methods can have very different final chemical concentrations. Besides having a unique, endpoint, path and rate to equilibration, each method will have other distinct advantages and disadvantages. For the majority of crystallization screening experiments there is little foreknowledge of the specific protein and cocktail solution thermodynamics; the process is empirical.

Batch methods take a protein solution and combine it with a cocktail solution. The initial chemical conditions are essentially the final chemical conditions unless there is a phase change in the solution. Batch experiments require high levels of supersaturation from the onset and ideally place the macromolecule at labile supersaturation, required for nucleation. Because batch experiments do not use a dynamic process to increase the level of supersaturation, the initial chemical conditions are well-defined; experimental outcomes can be readily interpreted. Batch experiments are very efficient and have minimal manipulation during their setup. Vapor-diffusion methods codilute a protein and cocktail drop on a surface and then seal the drop in a container with a larger volume of cocktail solution. The reservoir dehydrates the drop until the vapor pressure of water over the drop and reservoir is equivalent. The drop begins at a lower concentration of precipitating agent and during equilibration approaches the endpoint at a higher concentration of precipitating agent. Crystals can form, and the drop will still continue to dehydrate until the vapor pressure of volatile species (typically water) over the drop and reservoir are equivalent. The rate of equilibration is partially dependent upon the solutions' colligative properties¹⁴. One benefit of this approach is that as sampling takes place along the path towards the endpoint, there is a potential to identify chemical conditions that support crystallization. Although the exact chemical conditions at the time of crystallization are generally unknown, reproducing the experimental conditions is still readily accomplished. According to the RCSB PDB¹⁵ (www.pdb.org), the vapor-diffusion technique is the most common crystallization method. Liquid-diffusion methods such as counterdiffusion¹⁶ are less common, but are analogous to the dynamic process of vapor diffusion, sampling a region of chemical space along a path to a set endpoint.

In each case crystallization cocktails are prepared, the crystallization experiments set up, observed over time and the results noted. Automating this process has the benefit of increasing speed (a great number of experiments can be set up in a given amount of time) and decreasing sample volume (the average automated system can deliver smaller solution volumes than the average manual pipette). Potential pharmaceutical targets and complexes of potential pharmaceuticals with those targets can be screened rapidly with minimal sample. In translating this to high-throughput, any automated process has to duplicate cocktail preparation, setup, imaging and notation.

3. Automating laboratory crystallization processes

Complicated, advanced automation techniques rely upon the same methods, thermodynamics and kinetics as manual crystallization experiments. The seeds of the necessary technologies for automation can be traced back to the 1980's.

In 1987 syringe pumps were used to deliver reservoir and experiment drop solutions to two separate plates, the plates were then adjoined using foil with pinholes to facilitate slow

equilibration¹⁷. The combined preparation of solutions and setup of experiments provided two of the four key ingredients for successful automation. Another early automation system used two syringe pumps coupled with a valve control box to dispense hanging drop grid screen experiments into a 4×4 array¹⁸. This system introduced an information tracking system developed for solution preparation, the third key ingredient in automating the process. Parallel to these studies the APOCALYPSE system was constructed by Jones and co-workers¹⁹ leading to a robot friendly American Crystallography Association (ACA) plate to set up hanging, sitting or sandwich drop crystallization experiments²⁰. This evolved into a system including image analysis software to detect crystals (REVELATIONS)²¹, the fourth and final key ingredient for a fully automated system.

A complete automated crystallization laboratory, combining these key ingredients, was developed by Ward and co-workers²². The system used syringe pumps to deliver solutions to a specially designed plate²⁰ with image analysis of the results. These early developments focused mainly on the most common crystallization method in the laboratory, vapor diffusion. Chayen *et al.*²³ developed a system for automated microbatch-under-oil crystallization. Douglass Instruments, Ltd (Berkshire, UK) sells updated versions of one of the first commercially available robotic systems for the home laboratory.

Enhancing the process by reducing human intervention was one of the goals set for a system developed by Rubin *et al.*²⁴ This system used a single glass plate to hold the experiment drops and had the capacity, without human intervention, to set up 51 standard 24-well hanging drop experiment plates in 40 hours. On a smaller scale a partially automated crystallization system, the PIPEX, was developed around a computer controlled, motorized pipette²⁵. The system prepared crystallization solutions interactively with the investigator, using information entered into a spreadsheet; an example of a semi-automated method that could aid in the laboratory's overall throughput, without a major financial investment. At the other extreme a system capable of setting up as many as ten 24-well crystallization plates without operator intervention was developed using a five-axis robot²⁶.

There are important liquid-handling considerations for implementing any crystallization strategy in an automated manner. Most if not all of the automated formulation of cocktail solutions is achieved through volumetric co-dilution of a set of concentrated stock solutions. These solutions typically consist of a range of chemicals e.g. buffers, PEGs, salts, cryoprotectants and water, covering a wide range of physicochemical properties. Particularly relevant with respect to liquid-handling are viscosity and surface tension. Concentrated PEG solutions, especially those of high M_r , have a high viscosity. The solutions can have a consistency similar to that of room-temperature honey. A system has to have exceptional precision and accuracy across the whole range of stock solutions that are used in the typical protein crystallization laboratory to have reproducible results. Commercial liquid-handling robots, capable of dispensing these solutions for crystallization screens, made their way into the laboratory. Two contemporaneous efforts used the Robbins Scientific Hydra, a 96-syringe bank liquid-handling system, but took very different approaches. The Protein Structure Factory²⁷ used the Hydra robot to deliver reservoir solutions to 96 well vapor diffusion experiment plates, while an approach at the Hauptman-Woodward Institute used the same robot to set up microbatch-under-oil crystallization experiments in 1536-well microassay plates²⁸. Both approaches are valid and offer advantages that fit the particular goals of the groups using them. There have been many developments since then including a proliferation of robots available commercially. We will focus on our own efforts and results, and on where we are heading to use them, as practical advice in implementing high-throughput crystallization screening.

4. The HWI High-throughput Crystallization Screening Experience

Our experiences in automated crystallization screening, from the perspective of a High-throughput Screening (HTS) laboratory that started operations in early 2000, were built upon the application of commercial liquid-handling robots to the microbatch-under-oil technique. To date we have conducted crystallization trials on over 12,500 macromolecules, ~20 million experiments, which have generated over 120 million imaged outcomes at regularized time points. While vapor-diffusion is the most popular technique for structural crystallography, we have adopted the batch method due to the small volume requirements for both macromolecule and cocktails, and relative thermal and vibrational stability. The batch method provides easily traceable chemical parameters; the solution concentration of chemicals in the crystallization experiment remains constant until there is a phase change. Batch under oil offers practical advantages for automation. The experiments do not require additional sealing processes to prevent experiments from evaporating to dryness, there are no coverslips to manipulate and there are no additional reservoir solutions required for controlled-dehydration of the experiment drop. Our history, operation and developments are summarized in Figure 1. Two things are immediately obvious in this figure, the first being the steady progression in samples and the second being the development and implementation of improved technologies and techniques. We describe the process and outcomes, and then touch on the developments that have helped us to continuously improve our operations.

4.1 The process

Screening plates are prepared with paraffin oil and cocktails in advance of the crystallization experiments. The plates are centrifuged and stored at 4°C for up to 2 weeks. The 1536 different cocktails cover a range of chemical space that spans properties thought to be important determinants of crystallogenesis including: pH; Hofmeister series; surface tension of water; viscosity; activity of water; excluded volume; density; osmolality; and ionic strength. Rather than focusing strictly on minimizing the sample volume, we took the approach of maximizing the information obtained from initial screening while keeping reasonable sample requirements. The cocktails are formulated on an annual basis with formulations changing significantly during the first three generations (learning with experience), then less dramatic changes during subsequent years. The refractive index of the stock solutions and final cocktails is measured to fingerprint the solutions. This technique requires small amounts of sample and is a sensitive, rapid and reliable technique to identify changes in chemical concentrations e.g. as little as 1%(v/v) in the case of PEG.²⁹ This quality control step ensures cocktails are accurately prepared to maintain generational consistency. Periodic checks of the cocktails' refractive indices ensure consistency during the course of the screen's lifetime.

The current generation of cocktails contains three groups; Group 1, Salt/buffers and Group 2, PEG/Salt(at low concentrations)/buffers are both constructed using an incomplete factorial design²⁷ and Group 3 commercial screens. The Groups 1 and 2 sample 8 buffers at 100 mM concentrations; CAPS (10.0), TAPS (9.0), Tris (8.0), HEPES (7.5), Bis-Tris Propane (7.0), MES (6.0), sodium acetate (5.0), and sodium citrate (4.0). The 229 salt cocktails of Group 1 contain 33 different, highly soluble salts at ~ 30, 60, and 90% of room temperature saturation in water. The 687 PEG cocktails of Group 2 include 5 different molecular weights of PEG, 20 KDa, 8 KDa, 4 KDa, 1 KDa and 400 Da combined with 36 salts at 100mM concentration. The 620 commercial conditions of Group 3 are from Hampton Research (Aliso Viejo, CA) and include Silver Bullets™, Silver Bullets Bio™, PEG/Ion HT™, Crystal Screen HT™, Index™, Crystal Screen Cryo™ and Grid Screen Salt HT™. The combination of the incomplete factorial sampling of chemical space coupled with commercial screens allows the results to be related to the wider universe of proteins

crystallized from these commercial screens and those results to be related to data from our high-throughput laboratory.

Our sample flow starts with a monthly call for applications sent to all the laboratories that have previously used our service – there is no general solicitation to groups that have not used the service but groups can be placed on the mailing list by request. Samples are not judged on merit; we do not have the resources to accomplish this, but are accepted on a first come first serve basis. Basic information is collected on samples to be sent to the laboratory to ensure that they are non-hazardous and help us improve the crystallization screening process. Samples are shipped to the laboratory by commercial shipping companies, usually for overnight delivery on dry-ice, wet-ice, or at room temperature. If the sample was received frozen, it is thawed as directed by the investigator who supplied the sample. The sample (in a 1.5 ml microcentrifuge tube) is centrifuged to pellet any precipitate that may have formed during transit. Notes are taken on the condition of the sample, and then the sample is manually loaded into 12 wells (one row) of a 96-well source plate. A liquid-handling system (equipped with a 12-syringe bank head to minimize dead volume loss) is used to aspirate protein from the source plate and deliver 200 nanoliters to each well of the 1536-well experiment plate (already containing oil and cocktail). The experiment plate is centrifuged at low speed to ensure that the aqueous cocktail and protein solutions merge under the oil. The plate is then ready for imaging. The majority of samples are kept at 22°C although special cases that require other temperatures can be accommodated. Temperature is an important variable affecting solubility³⁰ and therefore crystallization, but adds to the number of experiments. In the majority of cases we explore temperature during the optimization, rather than the initial screening stage.

Imaging takes place on one of three automated digital photomicrography systems designed in-house. Each accommodates up to 28 experimental plates (43,008 experimental conditions). Each individual experiment is imaged by translating the well of interest under a digital camera with a medium magnification zoom lens and illumination from below. Each image is saved as a grayscale image in compressed tiff format (lossless compression). The full complement of 1536 images are packaged (WinRAR compressed) with an XML file that describes the chemical cocktail used to formulate the experiment drop (*i.e.* image) and converted to a JPEG format for distribution to collaborators through a secure FTP server. Multiple backup systems are incorporated. When a plate reaches the end of its imaging cycle Tiff images are written to DVD-R and JPEG images are written to CD-R. Optical discs and tapes are archived. Our current imaging schedule is one image prior to the addition of protein (cocktail control), then additional images 1 day after adding protein, then at 1,2,3,4 and 6 weeks. This has been revised from the original schedule based on the performance of the cocktails used in the screen. Plates are sent to the investigators on request, but typically discarded due to the difficulty of harvesting crystals.

4.2 Results

For the general biomedical community we do not track results after images have been recorded, with the notable exception of a set of manually classified images from a study of 96 proteins that underwent high-throughput screening.^{31 32} We ask that investigators reference a paper describing the early high-throughput screening laboratory and track citations to this. For our SG collaborators, the status of each target is tracked in detail. A summary of the number of samples we see and the corresponding number of structures resulting in the SG case is shown in Table 1. Not shown are those from industry which are dealt with separately.

Selected structures from the general biomedical community are shown in Table 2. These demonstrate the diverse range of samples where initial conditions for crystallization have

been successfully determined. It is important to note that although initial screening was performed using the HTS-laboratory the vast majority of the work required to produce these structures occurred outside of the HTS-laboratory, with our crystallization lead providing only the initial starting point for optimization.

We have been analyzing the SG data in detail order to develop automated image classification techniques (described later). As part of this analysis for 269 macromolecules out of a selection of 823 targets where crystals occur (32.7%), the average percentage of conditions that show hits in the screen is 3.58%, or ~55 hits per sample - these can be crystals that visually appear to require little optimization, or those that may require significant effort to optimize. Figure 2 shows an example of results that we currently classify as a hit compared to ones that are ignored.

4.3. Practical implementation of high-throughput screening

In implementing the screening-laboratory practically perhaps the most significant development effort has been made in the cocktail development. These are formulated on a yearly basis with the first generation not designed for crystallization screening but as a means to fingerprint macromolecules based upon their precipitation behavior. The cocktails were initially divided into five groups; simple inorganic salts of high solubility, PEG / salt combinations, PEG / buffer combinations, fine screens for three highly soluble salts, and Hampton Research Crystal Screens I and II to relate the results to standard commercially available screens. The concentrations were chosen based on crystallization results available in the Biological Macromolecular Crystallization Database⁶⁸⁻⁷⁰. We used the software SAmBA⁷¹ to formulate an optimal set of cocktails that tested all variables and pairwise interactions between them symmetrically. In the first group are 46 distinct, highly soluble (>1M), simple salts, each at ten concentrations surveying 11 distinct cations and 14 distinct anions. In the second group are PEG / salt combinations. Three distinct PEGs (1 kDa, 8 kDa, and 20 kDa) at five concentrations were combined with each of the 46 salts from the first group. The salts were uniformly at 200 mM concentration. In the third group were PEG / buffer combinations. Eight different PEGs, ranging in molecular weight from 200 Da to 35 kDa, at five concentrations, were combined with 100 mM buffers ranging in pH from 4.8 to 10.4. In the fourth group were 20-step (0.2M interval) fine screens for Ammonium sulfate (0.2 – 3.8M), Lithium chloride (0.2 – 10.0M) and Potassium thiocyanate (0.2 – 9.7M).

The first generation of cocktails was successful in producing precipitation profiles for biological samples. The second generation of cocktails was specifically designed for crystallization screening. The overall screen was significantly re-structured and the majority of the cocktails buffered. Ten salts were eliminated from those used in the first generation due to poor solubility. The fine screens of Ammonium sulfate, Lithium chloride and Potassium thiocyanate were replaced with a buffered salt screen sampling three concentrations of the 36 salts based on 90%, 60% and 30% of their maximum solubility in water at room temperature. SAmBA was used to reformulate the cocktails. The second generation, and all subsequent generations of cocktail including the latest are grouped into three major categories; buffered salt, buffered PEG/salt and commercial screens. Buffering of the cocktails for Generation 2 led to precipitation in a number of conditions with or without the macromolecular sample. These cocktails were removed in formulating the Generation 3 cocktails. We did not re-implement the incomplete factorial for this generation as this would have simply re-populated regions of chemical space that produced precipitated cocktails. The selected cocktails were simply eliminated from the screen, and replaced with further commercial cocktails from Hampton Research, adding the Grid Screen MPD, PEG-6000, Ammonium Sulfate, Sodium Malonate and PEG 6000/Lithium Chloride. Minor changes have been implemented in cocktails from generation 3 onward. Of the largest changes, for Generation 5 HEPES buffer was used at pH 7.0 rather than 7.5 to compare

buffer influence on the crystallization result. In Generation 8 the Hampton Research Quik Screen, and PEG Ion 2 cocktails were duplicated. This allowed us to monitor reproducibility in preparation for replacement of these duplicates by a set of 96 novel cocktails (in generation 8A) that use small molecules to promote macromolecular contacts⁷².

We maintain a deliberate balance between the incomplete-factorial derived portion of the screen and the sampling of commercial screens. This provides a means to derive chemical space information while the commercial screens allow us to relate our screening results to a much larger population of samples that makes use of those screens. The commercial screens themselves operate on several different principles. The conditions from Hampton Research (Aliso Viejo, CA) include Silver Bullets™ and Silver Bullets Bio™ (use libraries of small molecules to establish stabilizing, intermolecular, hydrogen bonding, hydrophobic and electrostatic interactions to promote lattice formation and crystallization), PEG/Ion HT™ (PEG produced 60% of published crystals, uses purified PEG 3350 as a precipitating agent with salts, pH, neutralized organic acids and multivalent ions to promote crystallization) Crystal Screen HT™, (a sparse matrix sampling approach, 48 of the conditions coming from Jancarik and Kim⁷³ and 48 from Cudney et al.⁷⁴), Index™ (targets zones of different chemical space and pH to identify what specific chemical space is effective for crystallization or reducing solubility), Crystal Screen Cryo™ (a selection of Crystal Screen HT™ conditions with the proper amount of glycerol added for cryoprotection) and Grid Screen Salt HT™ (a 96 condition grid of salt concentrations and pH for Ammonium Sulfate, Sodium Malonate, Sodium Potassium Phosphate, and Sodium Chloride). The commercial screens provide a complementary set of cocktails to the incomplete factorial sampling of chemical space.

Practically we ensure that the effort involved in formulation is as minimal as possible. We have several quality control steps in the process that have maintained consistency over the 10 years of operation including measuring the refractive index of the cocktail solutions and a set of standard proteins to test reproducibility of screening experiments. Preparation uses stock solutions either made up and tested in-house or purchased commercially from Hampton Research (Aliso Viejo, CA).

Contamination and the use of washing or disposable items was a practical consideration. Disposable tips are a viable option for many laboratories. They can save a significant amount of time, as they do not require wash cycles, but increase supply costs and typically require additional manual intervention. Systems using syringes with non-disposable tips require thorough washing to prevent cross-contamination. Chemicals such as PEGs require significant numbers of wash cycles to thoroughly clean the capillary-like needles of a syringe. We chose to use non-disposable syringes. They are well-suited for dispensing into the narrow diameter, high-density 1536 well plates, estimations on throughput were within our target range, and it made financial sense as our protocols would have required thousands of disposable tips. This cycle does not affect productivity as the most time consuming task, waiting on crystallization, is not affected by this.

We use volumes of 200 nl of cocktail and 200 nl of sample solution. This is an important consideration. There is some debate in the literature over the best sample volume to use for crystallization screening. Sample volumes as low as 20 nL are used for automated vapor-diffusion experiments; up to a two-order of magnitude reduction in sample volume and a tenfold reduction in the time required for crystal formation was reported⁷⁵. Low volume drops have to be accommodated by not only the liquid-handling systems, but also the imaging/crystal detection and if desired crystal retrieval steps. The 200 nanoL + 200 nanoL drops are readily dispensed by our liquid-handling systems, and are well-accommodated by our imaging systems.

The data acquisition was an important design consideration from the start for data-mining. Information is captured in a plate-centric manner. Each plate is associated with an investigator, protein sample(s), cocktails and a set up date. Currently we have 1,815 investigator accounts in our database; these accounts capture who should have access and be notified of any changes in the experiment's status, the preferred method of delivering information to the user along and the confidentiality status of the sample. There are three categories of confidentiality, fully open access for those samples associated with the PSI efforts or where an investigator tells us details can be released, those samples from individual investigators where we protect the sample details (the majority of our cases) and those from industrial collaborators where often we do not know the identity of the sample. Each sample receives a consecutive P-code for identification (to maintain confidentiality of the sample) and information about the sample and the crystallization screening conditions is recorded.

Dozens of scripts run in the background to generate imaging schedules, data archiving and distribution, and notifications to users and staff. A plate imaging schedule is scripted, along with software to control archiving. This scripting that runs behind the database is invisible, but critical for successful operation of the laboratory.

The key aspects of our operation are instrumentation, scale, staff and expertise. We have invested significant resources in dedicated robotics for many steps of the process. These are in constant use due to the number of samples we receive from the external community. A dedicated staff supports these robotic systems interfacing between the different stages where manual steps are more efficient than a completely automated process. For example, our current maximum throughput is some 200 individual proteins per month. Cocktail plates are manually positioned and removed after filling. This is not an onerous task and for the cost of an automated system to help in this we have a full-time technical staff member who can help with other aspects of the process when the cocktail filling is not taking place. Automation has its place but to be fully cost effective it should be applied in all stages that need it but not every stage that could use it. Robots aren't best at everything; we don't assume robots will always out-perform humans. Talented, conscientious, intelligent and observant people are the most important ingredient for a successful crystallization laboratory, not the robotics.

5. Optimization

We have a successful system to perform crystallization screening and record results. Others have similar capabilities^{76–78} and commercial instruments are available to replicate this process on a slightly smaller scale in the home laboratory. The high-throughput screening process has reached a stage of maturity in that we can replicate the setup and imaging of crystallization trials set up by hand but in a massive way. Where we have data on failure as well as success, we advance from a purified, soluble protein to a crystal for 21% of samples and advance from a purified soluble protein to a structure for only 8% of samples⁶. Optimization of initial crystallization hits to produce crystals suitable for structural determination is a significant bottleneck.

What is currently lacking is the ability to go from the initial screening results to a well-diffracting crystal in an automated manner. A good deal of manual effort is needed after the screening stage to interpret the results and proceed to data collection. Individually, many of the processes are there but they are not linked into a pipeline. For example, automated image classification shows promise in identifying defined regions of the crystallization phase diagram (even if crystals are not seen in the screening⁷⁹). Chemical space mapping can use that information together with a knowledge of theoretical phase diagrams to identify the key trends promoting crystallization and thereby rationally design the cocktails needed to sample

the most likely regions of chemical space for crystallization)^{80, 81}. By X-raying lead conditions *in-situ*^{82–85} or harvesting them directly from the screening plates (currently in development) we can use diffraction quality rather than visual quality as a quantitative metric for optimization. By the use of concentration variation and temperature in a technique we have termed Drop Volume Ratio and Temperature (DVR/T)⁸⁶ we can rapidly automate the optimization process.

Despite significant improvements, optimization remains a significant technical challenge for high-throughput methods. The optimization steps tend to be less standardized and often rely upon protein-specific strategies to improve the crystals' diffraction properties.

6. Conclusions

There are no fundamental limits to producing a completely automated pipeline where a purified protein sample enters and, if it can exist, a well diffracting crystal results. High-throughput screening techniques work well. We have described those based on plate technologies but others exist, in particular exploiting the use of microfluidic devices^{82, 87–89}. Practically the number of conditions that can be sampled increases and/or the volume of sample required decreases. We have focused on soluble samples but similar approaches can be applied to membrane proteins⁹⁰. In terms of pharmaceutical discovery, efficient crystallization screening strategies exist but the results require interpretation to utilized most effectively.

High-throughput optimization will be the key development that shifts the paradigm of structural biology. In screening for the SG program a standard set of screens is used and lead conditions are optimized manually. The manual step imposes a finite number on the leads that can be followed up - frequently only a fraction of the most promising cocktails are chosen. If crystals are not observed with different constructs or species, we can make use of solubility information provided by the clear and precipitated experiments to refine our search.⁸⁰ If the same automation technologies can be applied to optimization, coupled with computational classification and analysis, we may be able to considerably increase the success rate of going from a new protein to a structure. While we do not yet know what this increase would be, our own Small angle X-ray scattering studies on several hundred SG targets (complementary to X-ray crystallographic and NMR studies)⁹¹ characterized the majority of the samples as globular and well-folded. More excitingly from the pharmaceutical view is the increase in number of lead compounds we could incorporate at the screening stage. High-throughput crystallization screening has an important role in structure based pharmaceutical design.

7. Expert opinion

High-throughput crystallization screening is a mature technology. Thousands of proteins per year can be screened at large facilities such as ours. Commercially available systems allow smaller, but still significant, efforts to take place in individual laboratories. There are some practical aspects that when taken into account will make it possible to maximize the research advantages that a high-throughput approach has to offer. High-throughput means high-input and produces high-output. To clarify this important statement, unless you can produce sufficient samples or experiments to feed robotic systems and have the capacity to collect and then interpret the results produced, you should give serious consideration and ask yourself why automation will be beneficial. High-throughput crystallization is well-suited for drug-discovery; large numbers of identical crystals can be prepared for ligand-soaking, or co-crystallization experiments.⁹² Automation is best suited for building core facilities, linking several laboratories or operating as a large facility serving many investigators. This

achieves the high sample input needed, provides expertise in the operation and maintenance of the systems, and allows the analysis of results to be carried out in more detail by each individual. Idle automated systems, especially those designed for liquid-handling, often require the operator to relearn the software and operation. The instrument itself will likely require significant preparation if not in use continuously.

While high-throughput crystallization screening is mature, the vast majority of biological macromolecules are recalcitrant to crystallization. Consider that for the 'easy' cases of soluble proteins the odds of structural success are roughly 1:5. This ignores the truly challenging classes of membrane proteins and protein complexes. The *rate of structural success* for samples recalcitrant to crystallization won't increase by simply increasing the number or the rate of experimental set ups. In the initial SG studies, a fold library was created where it did not matter if 4 out of 5 samples failed; there were multiple targets to build this library. As the technology turns to biological pathways the high failure rate of crystallization will leave important gaps in our knowledge. Studies have identified physical properties of the protein that affect crystallization.⁹³ We can approach this problem from the perspective of modifying the protein targets to improve crystallization success. Methods such as truncation,⁹⁴ reductive methylation,⁹⁵ surface entropy reduction,⁹⁶ limited and *in situ* proteolysis⁹⁷ and small molecule additives that can act as freezers to reduce conformational flexibility⁷² have all proven effective means to promote crystallization. These are just a few of the salvage pathways for recalcitrant samples. We can push the envelopes of other techniques such as NMR and Cryo-EM. We can also take aim at the optimization process, developing more suitable protocols for automation rather than trying to adapt manual methods.

Optimization is a skilled aspect of the crystallization process. The initial information from the crystallization screening is multi-parametric and has to be digested and used to guide the next experiments. Initial analysis predominantly relies on a trained observer characterizing crystals visually. Robotic crystal mounting which is being actively developed^{98, 99}, promises to apply reproducible mounting protocols such that initial leads can be examined with an X-ray beam. Similarly *in situ* diffraction can enhance the process considerably with less sample manipulation. Both provide a quantitative metric than can be interpreted algorithmically and used in automated optimization of the sample¹⁰⁰. Developments outside of crystallography in detector, beamline and X-ray sources are pushing the definition of what we consider a suitable crystal for structural determination to the extremes. Smaller and smaller crystals are being used for structural determination. It is technically feasible to collect structural data from a crystal 1–2 microns in size¹⁰¹. Intense 4th generation X-ray sources such as the Linac Coherent Light Source (LCLS) collected femtosecond microdiffraction patterns from nanocrystals of photosystem 1 that were used for structural determination.¹⁰² These advances could in many cases actually eliminate the need for optimization. One path for the future of structural crystallography seems clear; we will use smaller crystals to produce structures of larger and more complex biological macromolecules. It is unclear how much these developments will improve the overall success rate, going from a soluble protein to a crystallographic structure. However, it is clear that scientific and technological developments in the steps that proceed and follow crystallization will enable X-ray crystallographic techniques to be applied to a broader range of targets for a broader sector of the biological community. These advances will provide structural information to those who are experts in the biological process rather than specialists in structure. This will lead to an improved understanding of biology and will specifically benefit research into novel drug targets.

While setting up the required crystallization experiments, it is possible to simultaneously use the high-throughput technologies to conduct basic crystallization research. From their very

nature high-throughput automated technologies can produce statistically meaningful data on crystallization with precisely controlled conditions, reproducible volumes and outcomes recorded in detail. The crystal growth literature is full of relevant research on the fundamentals of crystallization. However most of these studies used a very narrow subset of soluble proteins to establish these fundamentals. With many parameters to be studied, high-throughput approaches also offer the most efficient means to truly gain a better understanding of the multi-parametric crystallization problem. These approaches have the capacity, to rapidly duplicate all the previous studies and extend them to multiple proteins and conditions, building up a more complete picture of the entire range of outcomes seen in crystallization trials. The demands placed upon most high-throughput labs for required crystallization efforts often leave little time to initiate such fundamental studies. With an increase in both availability and the number automated systems for crystallization, and appropriate support for these efforts, the resulting knowledge could significantly increase the success rate in going from a purified protein to structural knowledge of that protein.

In summary, a dedicated facility can perform high-throughput crystallization screening in an efficient manner. The tools used, staff to keep them running and the computational infrastructure are all key aspects of success. High-throughput optimization of crystals is not yet a mature, fully automated method. Humans remain the cornerstone of any laboratory efforts. Technology does not remove people from the pipeline; it will focus their strengths, which cannot be duplicated by robotics, to tasks where their talents are better applied. High-throughput approaches achieve results most effectively when they are used in a high-input manner. They are not the panacea for every crystallization problem. Automation can be used to guide crystallization efforts towards samples that are more likely to provide structures. This provides a means to obtain crystallographic structures faster and can help to understand and address fundamental crystallization challenges. High-throughput techniques can provide both applied and fundamental crystallization knowledge in an efficient, cost-effective manner. Ultimately, humans define the experiments and interpret the outcomes. Crystallization robotics are tools to improve human efficiency.

Acknowledgments

The members of the crystallization screening laboratory are acknowledged for their tireless efforts to maintain an efficient laboratory. We thank Bob Cudney for his informative discussions and insight. The many users of the screening laboratory are also thanked, in particular Guy Montelione from the NESG for access to detailed information about all of the samples run through the screening laboratory.

Declaration of Interest JR Luft, EH Snell and GT DeTitta are all supported by NIH grants GM074899 through Center for High-Throughput Structural Biology (CHTSB) and GM094597 through the NorthEast Genomics Consortium. EH Snell and JR Luft are also supported by NIH grant GM088396 as well as the Mae Stone Goode, James H Cummings and Richard W Foundations.

References

1. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28(1): 235–42. [PubMed: 10592235]
2. Leach AR, Gillet VJ, Lewis RA, et al. Three-dimensional pharmacophore methods in drug discovery. *J Med Chem.* Jan 28; 2010 53(2):539–58. [PubMed: 19831387]
3. Scapin G. Structural biology and drug discovery. *Curr Pharm Des.* 2006; 12(17):2087–97. [PubMed: 16796557]
4. Arinaminpathy Y, Khurana E, Engelman DM, et al. Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov Today.* Dec; 2009 14(23–24):1130–5. [PubMed: 19733256]

5. Grey J, Thompson D. Challenges and Opportunities for New Protein Crystallization Strategies in Structure-Based Drug Design. *Expert Opin Drug Discov.* Nov 1; 2010 5(11):1039–45. [PubMed: 21116481]
6. Chen L, Oughtred R, Berman HM, et al. TargetDB: a target registration database for structural genomics projects. *Bioinformatics.* Nov 1; 2004 20(16):2860–2. [PubMed: 15130928]
7. Luft, JR.; DeTitta, GT. Protein Crystallization. In: Bergfors, TM., editor. *Protein Crystallization*. second ed.. International University Line; La Jolla, California: 2009. p. 11–45.
8. Giege, R.; Ducruix, A. *Crystallization of Nucleic Acids and Proteins - A Practical Approach*. Oxford University Press; Oxford: 1992.
9. Jancarik J, Kim S-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Cryst.* 1991; 24:409–11.
10. Cudney B, Patel S, Weisgraber K, et al. Screening and Optimization Strategies for Macromolecular Crystal-Growth. *Acta Crystallogr D.* Jul 1.1994 50:414–4123. [PubMed: 15299395]
11. Berger I, Kang CH, Sinha N, et al. A highly efficient 24-condition matrix for the crystallization of nucleic acid fragments. *Acta Crystallogr D.* May 1.1996 52:465–8. [PubMed: 11539196]
12. Garavito, M. *Crystallization of Membrane Proteins*. Michel, H., editor. CRC Press; 1991. p. 89–105.
13. Carter CW. Response surface methods for optimizing and improving reproducibility of crystal growth. *Macromolecular Crystallography, Pt A.* 1997; 276:74–99.
14. Arakali SV, Luft JR, Detitta GT. Nonideality of Aqueous-Solutions of Polyethylene-Glycol - Consequences for Its Use as a Macromolecular Crystallizing Agent in Vapor-Diffusion Experiments. *Acta Crystallographica Section D-Biological Crystallography.* Sep 1.1995 51:772–9.
15. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Research.* Jan 1; 2000 28(1):235–42. [PubMed: 10592235]
16. Garcia-Ruiz JM, Otalora F, Novella ML, et al. A supersaturation wave of protein crystallization. *Journal of Crystal Growth.* Nov; 2001 232(1–4):149–55.
17. Kelders HA, H.Kalk K, Cros P, et al. Automated protein crystallization and a new crystal form of a subtilisin:eglin complex. *Protein Engineering.* 1987; 1(4):301–3. [PubMed: 3334088]
18. Cox MJ, Weber PC. Experiments with Automated Protein Crystallization. *Journal of Applied Crystallography.* Oct 1.1987 20:366–73.
19. Jones, N.; Deeter, J.; Swartzendruber, J., et al. APOCALYPSE: An Automated Protein Crystallization System. American Crystallographic Association Annual Meeting; Austin Texas. 1987; University of Texas; 1987. p. 27
20. Jones, N.; Ward, K.; Perozzo, M. CrystalPlate: The new ACA protein crystallization plate. American Crystallographic Association Annual Meeting; Philadelphia, PA. 1988; 1988. p. 109
21. Swartzendruber, J.; Jones, N. APOCALYPSE: An automated Protein Crystallization System; III. The Beginning: The Genesis software; American Crystallographic Association Annual Meeting; Philadelphia, PA. 1988; 1988.
22. Ward KB, Perozzo MA, Zuk WM. Automatic Preparation Of Protein Crystals Using Laboratory Robotics and Automated Visual Inspection. *J Cryst Growth.* 1988; 90:325–39.
23. Chayen NE, Stewart PDS, Maeder DL, et al. An Automated-System for Microbatch Protein Crystallization and Screening. *Journal of Applied Crystallography.* Aug 1.1990 23:297–302.
24. Rubin B, Talafous J, Larson D. Minimal Intervention Robotic Protein Crystallization. *Journal of Crystal Growth.* Mar; 1991 110(1–2):156–63.
25. Eisele JL. Preparation of Protein Crystallization Buffers with a Computer-Controlled Motorized Pipette - Pipex. *Journal of Applied Crystallography.* Feb 1.1993 26:92–6.
26. Sadaoui N, Janin J, Lewitbentley A. Taos - an Automatic System for Protein Crystallization. *Journal of Applied Crystallography.* Aug 1.1994 27:622–6.
27. Mueller U, Nyarsik L, Horn M, et al. Development of a technology for automation and miniaturization of protein crystallization. *Journal of Biotechnology.* 2001; 85:7–14. [PubMed: 11164957]
28. Luft JR, Wolfley J, Jurisica I, et al. Macromolecular Crystallization in a High Throughput Laboratory - The Search Phase. *Journal of Crystal Growth.* 2001; 232:591–5.

29. Mohsen-Nia M, Modarress H, Rasa H. Measurement and modeling of density, kinematic viscosity, and refractive index for poly(ethylene glycol) aqueous solution at different temperatures. *Journal of Chemical and Engineering Data*. Sep-Oct;2005 50(5):1662–6.
30. Christopher GK, Phipps AG, Gray RJ. Temperature-dependent solubility of selected proteins. *J Cryst Growth*. Aug; 1998 191(4):820–6.
31. Snell EH, Luft JR, Potter SA, et al. Establishing a training set through the visual analysis of crystallization trials. Part I: similar to 150 000 images. *Acta Crystallogr D*. Nov.2008 64:1123–30. [PubMed: 19020350]
32. Snell EH, Lauricella AM, Potter SA, et al. Establishing a training set through the visual analysis of crystallization trials. Part II: crystal examples. *Acta Crystallogr D*. Nov.2008 64:1131–7. [PubMed: 19020351]
33. Larson ET, Parussini F, Huynh MH, et al. Toxoplasma gondii cathepsin L is the primary target of the invasion-inhibitory compound morpholinurea-leucyl-homophenylvinyl sulfone phenyl. *J Biol Chem*. Sep 25; 2009 284(39):26839–50. [PubMed: 19596863]
34. Johnson TA, Qiu J, Plaut AG, et al. Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of haemophilus influenzae immunoglobulin A1 protease. *J Mol Biol*. Jun 12; 2009 389(3):559–74. [PubMed: 19393662]
35. Liu L, O'Grady C, Dalrymple SA, et al. Crystallization and preliminary X-ray studies of the N-domain of the Wilson disease associated protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. Jun 1; 2009 65(Pt 6):621–4.
36. Shah MB, Ingram-Smith C, Cooper LL, et al. The 2.1 Å crystal structure of an acyl-CoA synthetase from Methanosarcina acetivorans reveals an alternate acyl-binding pocket for small branched acyl substrates. *Proteins*. May 18.2009
37. Guhaniyogi J, Sohar I, Das K, et al. Crystal structure and autoactivation pathway of the precursor form of human tripeptidyl-peptidase 1, the enzyme deficient in late infantile ceroid lipofuscinosis. *J Biol Chem*. Feb 6; 2009 284(6):3985–97. [PubMed: 19038967]
38. Ho JD, Yeh R, Sandstrom A, et al. Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proc Natl Acad Sci U S A*. May 5; 2009 106(18):7437–42. [PubMed: 19383790]
39. Larson ET, Parussini F, Huynh MH, et al. Toxoplasma gondii cathepsin L is the primary target of the invasion inhibitory compound LHVS. *J Biol Chem*. Jul 17.2009
40. Lee MM, Isaza CE, White JD, et al. Insight into the substrate length restriction of M32 carboxypeptidases: Characterization of two distinct subfamilies. *Proteins*. May 11.2009
41. Immormino RM, Metzger LEt, Reardon PN, et al. Different poses for ligand and chaperone in inhibitor-bound Hsp90 and GRP94: implications for paralog-specific drug design. *J Mol Biol*. May 22; 2009 388(5):1033–42. [PubMed: 19361515]
42. Prudden J, Perry JJ, Arvai AS, et al. Molecular mimicry of SUMO promotes DNA repair. *Nat Struct Mol Biol*. May; 2009 16(5):509–16. [PubMed: 19363481]
43. Fallon JL, Baker MR, Xiong L, et al. Crystal structure of dimeric cardiac L-type calcium channel regulatory domains bridged by Ca²⁺ calmodulins. *Proc Natl Acad Sci U S A*. Mar 31; 2009 106(13):5135–40. [PubMed: 19279214]
44. Lima S, Sundararaju B, Huang C, et al. The crystal structure of the Pseudomonas dacunhae aspartate-beta-decarboxylase dodecamer reveals an unknown oligomeric assembly for a pyridoxal-5'-phosphate-dependent enzyme. *J Mol Biol*. Apr 24; 2009 388(1):98–108. [PubMed: 19265705]
45. Koksall AC, Nardozi JD, Cingolani G. Dimeric quaternary structure of the prototypical dual specificity phosphatase VH1. *J Biol Chem*. Apr 10; 2009 284(15):10129–37. [PubMed: 19211553]
46. Alontaga AY, Rodriguez JC, Schonbrunn E, et al. Structural characterization of the hemophore HasAp from Pseudomonas aeruginosa: NMR spectroscopy reveals protein-protein interactions between Holo-HasAp and hemoglobin. *Biochemistry*. Jan 13; 2009 48(1):96–109. [PubMed: 19072037]
47. Xia S, Monzingo AF, Robertus JD. Structure of NS1A effector domain from the influenza A/Udm/72 virus. *Acta Crystallogr D Biol Crystallogr*. Jan; 2009 65(Pt 1):11–7. [PubMed: 19153461]

48. Craven SH, Ezezika OC, Haddad S, et al. Inducer responses of BenM, a LysR-type transcriptional regulator from *Acinetobacter baylyi* ADP1. *Mol Microbiol.* May; 2009 72(4):881–94. [PubMed: 19400783]
49. Drake EJ, Gulick AM. Three-dimensional structures of *Pseudomonas aeruginosa* PvcA and PvcB, two proteins involved in the synthesis of 2-isocyano-6,7-dihydroxycoumarin. *J Mol Biol.* Dec 5; 2008 384(1):193–205. [PubMed: 18824174]
50. Persson M, Tars K, Liljas L. The capsid of the small RNA phage PRR1 is stabilized by metal ions. *J Mol Biol.* Nov 21; 2008 383(4):914–22. [PubMed: 18786545]
51. Larson ET, Deng W, Krumm BE, et al. Structures of substrate- and inhibitor-bound adenosine deaminase from a human malaria parasite show a dramatic conformational change and shed light on drug selectivity. *J Mol Biol.* Sep 12; 2008 381(4):975–88. [PubMed: 18602399]
52. Reger AS, Wu R, Dunaway-Mariano D, et al. Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry.* Aug 5; 2008 47(31):8016–25. [PubMed: 18620418]
53. Merritt EA, Holmes M, Buckner FS, et al. Structure of a *Trypanosoma brucei* alpha/beta-hydrolase fold protein with unknown function. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* Jun 1; 2008 64(Pt 6):474–8.
54. Slade DJ, Lovelace LL, Chruszcz M, et al. Crystal structure of the MACPF domain of human complement protein C8 alpha in complex with the C8 gamma subunit. *J Mol Biol.* May 30; 2008 379(2):331–42. [PubMed: 18440555]
55. Biswas T, Tsodikov OV. Hexameric ring structure of the N-terminal domain of *Mycobacterium tuberculosis* DnaB helicase. *FEBS J.* Jun; 2008 275(12):3064–71. [PubMed: 18479467]
56. Korotkov KV, Hol WG. Structure of the GspK-GspI-GspJ complex from the enterotoxigenic *Escherichia coli* type 2 secretion system. *Nat Struct Mol Biol.* May; 2008 15(5):462–8. [PubMed: 18438417]
57. Greenhagen BT, Shi K, Robinson H, et al. Crystal structure of the pyocyanin biosynthetic protein PhzS. *Biochemistry.* May 13; 2008 47(19):5281–9. [PubMed: 18416536]
58. Stack CM, Caffrey CR, Donnelly SM, et al. Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. *J Biol Chem.* Apr 11; 2008 283(15):9896–908. [PubMed: 18160404]
59. Arakaki TL, Buckner FS, Gillespie JR, et al. Characterization of *Trypanosoma brucei* dihydroorotate dehydrogenase as a possible drug target; structural, kinetic and RNAi studies. *Mol Microbiol.* Apr; 2008 68(1):37–50. [PubMed: 18312275]
60. Wang A, Zeng Y, Han H, et al. Biochemical and structural characterization of *Pseudomonas aeruginosa* Bfd and FPR: ferredoxin NADP+ reductase and not ferredoxin is the redox partner of heme oxygenase under iron-starvation conditions. *Biochemistry.* Oct 30; 2007 46(43):12198–211. [PubMed: 17915950]
61. Lee J, Page R, Garcia-Contreras R, et al. Structure and function of the *Escherichia coli* protein YmgB: a protein critical for biofilm formation and acid-resistance. *J Mol Biol.* Oct 12; 2007 373(1):11–26. [PubMed: 17765265]
62. Sullivan SM, Holyoak T. Structures of rat cytosolic PEPCK: insight into the mechanism of phosphorylation and decarboxylation of oxaloacetic acid. *Biochemistry.* Sep 4; 2007 46(35):10078–88. [PubMed: 17685635]
63. Lima S, Khristoforov R, Momany C, et al. Crystal structure of *Homo sapiens* kynureninase. *Biochemistry.* Mar 13; 2007 46(10):2735–44. [PubMed: 17300176]
64. Ross B, Kristensen O, Favre D, et al. High resolution crystal structures of the p120 RasGAP SH3 domain. *Biochem Biophys Res Commun.* Feb 9; 2007 353(2):463–8. [PubMed: 17188236]
65. Kajander T, Cortajarena AL, Mochrie S, et al. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr D Biol Crystallogr.* Jul; 2007 63(Pt 7):800–11. [PubMed: 17582171]
66. Ezezika OC, Haddad S, Clark TJ, et al. Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator. *J Mol Biol.* Mar 30; 2007 367(3):616–29. [PubMed: 17291527]

67. Paddock ML, Wiley SE, Axelrod HL, et al. MitoNEET is a uniquely folded 2Fe 2S outer mitochondrial membrane protein stabilized by pioglitazone. *Proc Natl Acad Sci U S A*. Sep 4; 2007 104(36):14342–7. [PubMed: 17766440]
68. Gilliland GL. A Biological Macromolecule Crystallization Database - a Basis for a Crystallization Strategy. *J Cryst Growth*. Jul; 1988 90(1–3):51–9.
69. Gilliland GL, Tung M, Ladner J. The biological macromolecule crystallization database and NASA protein crystal growth archive. *J Res Natl Inst Stan*. May-Jun;1996 101(3):309–20.
70. Gilliland GL, Tung M, Ladner JE. The biological macromolecule crystallization database: crystallization procedures and strategies. *Acta Crystallogr D*. Jun.2002 58:916–20. [PubMed: 12037329]
71. Audic S, Lopez F, Claverie JM, et al. SAMBA: An interactive software for optimizing the design of biological macromolecules crystallization experiments. *Proteins-Structure Function and Genetics*. Oct; 1997 29(2):252–7.
72. McPherson A, Cudney B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol*. Dec; 2006 156(3):387–406. [PubMed: 17101277]
73. Jancarik J, Kim SH. Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins. *Journal of Applied Crystallography*. Aug 1.1991 24:409–11.
74. Cudney R, Patel S, Weisgraber K, et al. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr*. Jul 1; 1994 50(Pt 4):414–23. [PubMed: 15299395]
75. Santarsiero BD, Yegian DT, Lee CC, et al. An approach to rapid protein crystallization using nanodroplets. *Journal of Applied Crystallography*. Apr.2002 35:278–81.
76. Stock D, Perisic O, Lowe J. Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology. *Progress in Biophysics & Molecular Biology*. Jul; 2005 88(3):311–27. [PubMed: 15652247]
77. Newman J, Pham TM, Peat TS. Phoenix experiments: combining the strengths of commercial crystallization automation. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. Nov 1; 2008 64(Pt 11):991–6.
78. Mueller-Dieckmann J. The open-access high-throughput crystallization facility at EMBL Hamburg. *Acta Crystallogr D*. Dec.2006 62:1446–52. [PubMed: 17139079]
79. Cumbaa CA, Jurisica I. Protein crystallization analysis on the World Community Grid. *J Struct Funct Genomics*. Mar; 2010 11(1):61–9. [PubMed: 20072819]
80. Snell EH, Nagel RM, Wojtaszyk A, et al. The application and use of Chemical Space Mapping to interpret crystallization screening results. *Acta Cryst D*. 2008; 64:1240–9. [PubMed: 19018100]
81. Nagel RM, Luft JR, Snell EH. AutoSherlock: a program for effective crystallization data analysis. *J Appl Crystallogr*. Dec 1; 2008 41(Pt 6):1173–6. [PubMed: 19529834]
82. Zheng B, Tice JD, Roach LS, et al. A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction. *Angewandte Chemie-International Edition*. 2004; 43(19): 2508–11.
83. Yadav MK, Gerdt CJ, Sanishvili R, et al. In situ data collection and structure refinement from microcapillary protein crystallization. *J Appl Crystallogr*. Dec; 2005 38(6):900–5. [PubMed: 17468785]
84. Jacquamet L, Ohana J, Joly J, et al. Automated analysis of vapor diffusion crystallization drops with an X-ray beam. *Structure*. Jul; 2004 12(7):1219–25. [PubMed: 15242598]
85. Ng JD, Clark PJ, Stevens RC, et al. In situ X-ray analysis of protein crystals in low-birefringent and X-ray transmissive plastic microchannels. *Acta Crystallographica Section D-Biological Crystallography*. Feb.2008 64:189–97.
86. Luft JR, Wolfley JR, Said MI, et al. Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. *Protein Sci*. Apr; 2007 16(4):715–22. [PubMed: 17327388]
87. van der Woerd M, Ferree D, Pusey M. The promise of macromolecular crystallization in microfluidic chips. *Journal of Structural Biology*. Apr; 2003 142(1):180–7. [PubMed: 12718930]

88. Hansen CL, Skordalakes E, Berger JM, et al. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proceedings of the National Academy of Sciences* 2002 December 24. 2002; 99(26):16531–6.
89. Segelke B. Macromolecular crystallization with microfluidic free-interface diffusion. *Expert Review of Proteomics*. Apr; 2005 2(2):165–72. [PubMed: 15892562]
90. Koszelak-Rosenblum M, Krol A, Mozumdar N, et al. Determination and application of empirically derived detergent phase boundaries to effectively crystallize membrane proteins. *Protein Science*. Sep; 2009 18(9):1828–39. [PubMed: 19554626]
91. Grant T, Luft JR, Montelione GT, et al. Small Angle X-ray Scattering as a complementary tool in a High-throughput Crystallization Laboratory. *Biopolymers*. 2011
92. Kuhn P, Wilson K, Patch MG, et al. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr Opin Chem Biol*. Oct; 2002 6(5):704–10. [PubMed: 12413557]
93. Price WN 2nd, Chen Y, Handelman SK, et al. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol*. Jan; 2009 27(1):51–7. [PubMed: 19079241]
94. Mooij WT, Mitsiki E, Perrakis A. ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. *Nucleic Acids Res*. Jul 1; 2009 37(Web Server issue):W402–5. [PubMed: 19395596]
95. Sledz P, Zheng HP, Murzyn K, et al. New surface contacts formed upon reductive lysine methylation: Improving the probability of protein crystallization. *Protein Science*. Jul; 2010 19(7): 1395–404. [PubMed: 20506323]
96. Cooper DR, Boczek T, Grelewska K, et al. Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr D Biol Crystallogr*. May; 2007 63(Pt 5):636–45. [PubMed: 17452789]
97. Dong A, Xu X, Edwards AM, et al. In situ proteolysis for protein crystallization and structure determination. *Nat Methods*. Dec; 2007 4(12):1019–21. [PubMed: 17982461]
98. Viola R, Carman P, Walsh J, et al. Automated robotic harvesting of protein crystals-addressing a critical bottleneck or instrumentation overkill? *J Struct Funct Genomics*. Dec; 2007 8(4):145–52. [PubMed: 17965947]
99. Viola R, Carman P, Walsh J, et al. Operator-assisted harvesting of protein crystals using a universal micromanipulation robot. *J Appl Crystallogr*. Jun 1; 2007 40(Pt 3):539–45. [PubMed: 19461845]
100. Zhang Z, Sauter NK, van den Bedem H, et al. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *Journal of Applied Crystallography*. Feb.2006 39:112–9.
101. Holton JM, Frankel KA. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallogr D Biol Crystallogr*. Apr; 66(Pt 4):393–408. [PubMed: 20382993]
102. Chapman HN, Fromme P, Barty A, et al. Femtosecond X-ray protein nanocrystallography. *Nature*. Feb 3; 2011 470(7332):73–7. [PubMed: 21293373]

Figure 1.

A historical perspective of the Hauptman-Woodward high-throughput crystallization screening laboratory illustrating the cumulative and yearly number of macromolecules screened against time and notable events in the laboratory. Detail is given in the text.

Figure 2.

High-throughput microbatch-under-oil screening outcomes from our laboratory that show examples of crystal leads for several different proteins. Each image displays a well of 0.9 mm diameter. Some hits are obvious crystals, others need closer examination and are more difficult to discern on the scale of the figure.

Table 1

Samples submitted to the HTS-laboratory and the number of structural depositions in the PDB by year divided into those from the general biomedical community and those from our SG collaborators. Note that structures are deposited some time after the sample has been through the laboratory and trends in that column will lag trends in the sample column.

Year	Total samples	General biomedical community samples	Structural Genomics		
			Samples	Structures	%
2010	910	406	504	75	14.8
2009	748	345	403	96	23.8
2008	1,185	617	568	86	15.1
2007	1,505	868	637	82	12.8
2006	1,886	1,468	418	68	16.2
2005	1,784	1,008	776	86	11.1
2004	1,833	1,130	703	59	8.4
2003	1,268	1,043	225	20	8.9
2002	749	644	105	12	11.4
2001	426	398	28	1	3.6
2000	227	214	13	0	0
Total	11,581	7735	3876	485	12.5

Table 2
Examples of structures from individual laboratories where the initial crystallization lead was obtained in the HTS laboratory.

Name	PDB	Ref		Name	PDB	Ref	
<i>Toxoplasma gondii</i> Cathepsin L	3F75	33		<i>Haemophilus influenzae</i> Immunoglobulin A1 Protease	3H09	34
N-domain of Wilson disease associated protein	N/A	35	N/A	acyl-CoA synthetase from <i>Methanosarcina acetivorans</i>	3ETC	36
Precursor Form of Human Tripeptidyl-peptidase 1	3EDY	37		human aquaporin 4*	3GD8	38
<i>T. gondii</i> cysteine protease cathepsin L (TgCPL)	3F75	39		<i>Thermus thermophilus</i> and <i>Bacillus subtilis</i> M32 carboxypeptidases	3HQ2 3HOA	40

Name	PDB	Ref		Name	PDB	Ref	
Inhibitor-Bound Hsp90 and GRP94	2FXS 2GFD 2EXL	41		Rad60 SLD2	3GOE	42
Dimeric cardiac L-type calcium channel regulatory domains	3G43	43		<i>Pseudomonas dacunhae</i> Aspartate- β -Decarboxylase Dodecamer	3FDD	44
VH1, a dual specificity phosphatase	3CM3	45		Hemophore HasAp from <i>Pseudomonas aeruginosa</i>	1B2V	46
NS1A effector domain from the influenza A/Udm/72 virus	3EE9	47		BenM, a LysR-type transcriptional regulator from <i>Acinetobacter baylyi</i> ADP1	2H99	48

Name	PDB	Ref		Name	PDB	Ref	
<i>Pseudomonas aeruginosa</i> PvcA and PvcB	3E59 3EAT	49		Small RNA Phage PRR1 capsid	2VF9	50
Adenosine Deaminase from a Human Malaria Parasite	2PGF 2PGR 2QVN	51		CoA ligase characterization	3CW8 3CW9	52
<i>Trypanosoma brucei</i> alpha/beta-hydrolase fold	2Q0X	53		MACPF domain of human complement protein C8 complex	2RD7	54
N-terminal domain of <i>Mycobacterium tuberculosis</i> DnaB helicase	2R5U	55		GspK-GspJ-GspJ complex from the enterotoxigenic <i>Escherichia coli</i> type 2 secretion system	3CI0	56

Name	PDB	Ref		Name	PDB	Ref	
Pyocyanin biosynthetic protein PhzS	2RGJ	57		Virulence-associated cathepsin L proteases of <i>Fasciola hepatica</i>	2O6X	58
<i>Trypanosoma brucei</i> dihydroorotate dehydrogenase	2B4G	59		<i>Pseudomonas aeruginosa</i> bId and FPR	2QDX	60
<i>Escherichia coli</i> protein YmgB	2OXL	61		Rat cytosolic PEPCK	2QEW 2QEY 2QE1 2QF2	62
<i>Homo sapiens</i> kynureninase	2HZP	63		p120 RasGAP SH3 domain	2J05 2J06	64

Name	PDB	Ref		Name	PDB	Ref	
TPR protein superhelice	2HYZ	65		Effector Binding Domain of BenM	2F97 2F8D	66
P120 Rasgap SH3 Domain	2IO5	64		2Fe 2S outer mitochondrial membrane protein	2QH7	67