LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-209860

# Automated High Throughput Drug Target Crystallography

B. Rupp

February 18, 2005

American Pharmaceutical Review

# Automated High Throughput Drug Target Crystallography

**Bernhard Rupp**

[1]*Macromolecular Crystallography and TB Structural Genomics Consortium, University of California, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA.*

*br@llnl.gov*

**Abstract** The molecular structures of drug target proteins and receptors form the basis for 'rational' or structure guided drug design. The majority of target structures are experimentally determined by protein X-ray crystallography, which as evolved into a highly automated, high throughput drug discovery and screening tool. Process automation has accelerated tasks from parallel protein expression, fully automated crystallization, and rapid data collection to highly efficient structure determination methods. A thoroughly designed automation technology platform supported by a powerful informatics infrastructure forms the basis for optimal workflow implementation and the data mining and analysis tools to generate new leads from experimental protein drug target structures.

**Keywords: High throughput crystallography, drug discovery, fragment screening.**

## 1. The role of structure based methods in drug discovery

Traditional assay based high throughput compound screening (HTS), well established in pharmaceutical R&D, covers compound diversity largely by combinatorial exploration of the chemical space and/or by screening of natural compound libraries [1]. Typically, millions of

compounds are screened in high density format against a protein target or receptor, at a cost of several dollars per assay and compound. In addition to the relatively high cost, the detailed nature of the target-lead affinity is difficult to identification with biochemical techniques only. Structure-based techniques such as Nuclear Magnetic Resonance (NMR) screening and High Throughput Protein Crystallography [2] (HTPX) enable a more rational and targeted approach to lead discovery and optimization. By avoiding the complexity of target structure determination, modern ligand-based high-field NMR- based screening can de facto achieve throughput of several thousand ligand interactions per day. The additional attractiveness of drug target X-ray crystallography is the fact that the method yields highly detailed structural information of the protein target complexed with its ligand, which has been highly valued for structure guided lead optimization in later stages of preclinical drug development. With recent advances in automated high throughput protein crystallography, the major obstacles to using structure based methods in early stages of drug discovery as a screening tool have been overcome: First, a diffracting protein crystal needs to be produced - a task that can prove challenging and often can only be achieved after substantial protein engineering - and then a full structure refinement for each target-ligand complex structure is necessary.

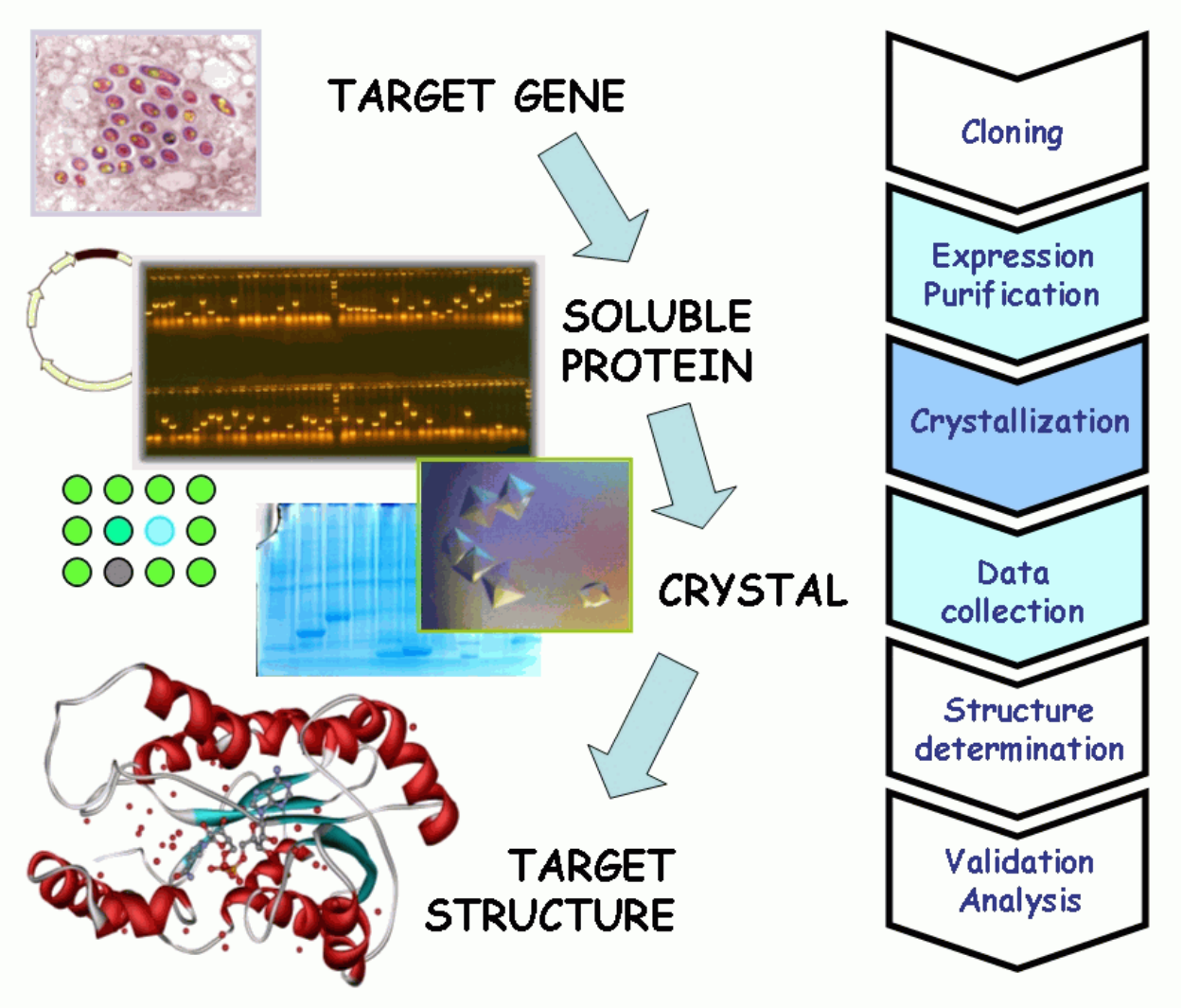


**Figure 1 Elements of crystallographic structure determination**. In the right hand flow diagram, highlights indicate process steps with high dependence on advanced robotic automation. Structure determination, validation and analysis are conducted *in silico* and already highly automated (image reproduced with permission from GIT publications, Germany).

## 2. High throughput X-ray crystallography in drug lead discovery

Historically, protein structures have been determined by an expert in laborious cycles of repeated manual rebuilding of the structure models into electron density and subsequent refinement. With the advance of automated protein purification and crystallization techniques and powerful phasing and model building methods [3], determination of a target-ligand complex can now be accomplished in a nearly automatic mode. The resulting rapid availability of ligand-complex crystal structures via HTPX enables novel strategies of crystallographic fragment screening and fragment evolution [4]. Protein targets are co-crystallized with smaller molecules of a fragment library, where high concentrations of ligands allow discovery of weaker binders, starting with affinities in the low mM range. In contrast, classical, bioassay based HT compound screening focuses on initial high affinity hits of already drug-like molecules in the nano-molar range,

Fragment libraries are developed to either allow classical combinatorial chemistry, or to support novel methods such as *in-situ* click chemistry between fragments [4]. From crystal structures of fragments located in the binding pocket of the target, larger molecules with higher binding affinities can be assembled, and the process of fragment-evolution repeated. Starting from a collection of relatively few, low affinity, low-molecule weight fragments, a targeted, structure based evolution of molecules with increased drug like properties and low nM affinity can be achieved in few steps. Crystallographic library fragments generally follow a 'rule of 3' [5], similar to Lipinski's 'rule of 5' for drug-like properties. During the design process, ADME/T properties, cross-reactivity with other targets, metabolites and drugs can be predicted already *in silico*, thus increasing the chance to arrive at a successful drug candidate.

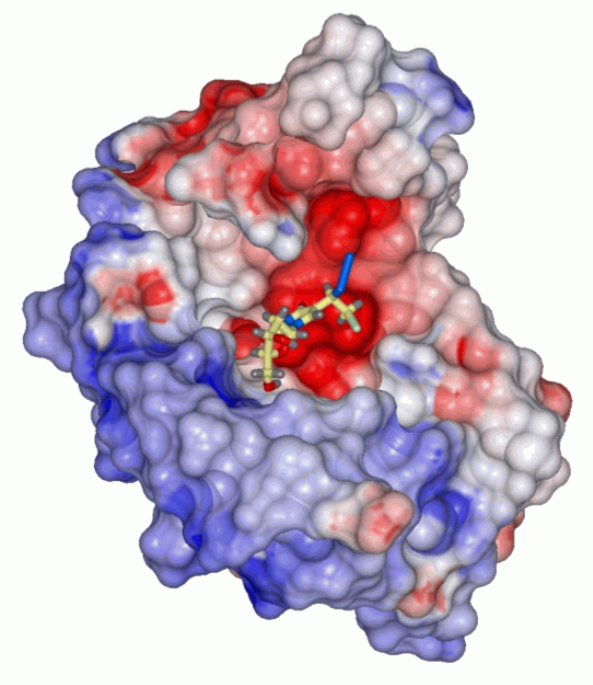

**Figure 2 Example for exploration of X-ray structures in drug discovery**. A surface charge representation of a *Mycobacterium tuberculosis* target molecule (PDB entry 1W30), with a lead compound fragment docked *in silico* by HT virtual ligand screening.

### 3. Drug target structures as a basis for virtual ligand screening

The value of fully detailed structural information extends beyond the obvious goal of ligand screening and lead optimization. Knowledge of the structure of a potential drug target *per se*, even without a ligand, informs about the drugability of the target. An array of structural bioinformatics tools can be deployed early on, to screen *in silico* for properties of a protein that indicate its suitability as a drug target [6]. Similar to experimental fragment screening, *in silico* virtual ligand and fragment screening (VLS) can be used on the native target structure to identify a smaller, target-specific subset of effective compounds [7]. Although in early stages virtual screening struggled to meet its initial promise and drug discovery remained dominated by

empirical screening, recent successes in predicting new ligands and their receptor-bound structures [1] and better rates of ligand discovery compared to empirical screening, have led to widespread use of VLS in drug discovery to compliment empirical screening [8]. Additional attractiveness is the early inclusion of drug-likeness and ADME/T data in the ranking of the compounds, thus providing early indications that a lead may fail in later drug development stages [9].

## 4. Automation in high throughput protein crystallography

Synergistic developments in technology, in particular laboratory automation, cryo-, synchrotron-, and computational techniques, combined with influx of substantial public and venture capital funding, have made HTPX possible. Given the potentially enormous rewards of structure guided drug development [2], it is not surprising that commercial biotech ventures have been able to attract funds to develop and implement advanced robotic HTPX pipelines. Public funding of structural genomics pilot projects in the USA and similar efforts in Europe, Japan and Israel (http://www.isgo.org/ lists all SG initiatives world wide), provide the means for development of non-proprietary high throughput structure determination methods, which has benefited not only drug discovery, but practically every structural biology effort.

### 4.1. Task-centric and data-centric process analysis.

The process of crystallographic structure determination can be broken up into a number of successive task blocks (Figure 3). A large number of these tasks are successive screening experiments, each creating a substantial and rapidly multiplying amount of data. As these data form the basis for knowledge discovery, predictive modelling, and process optimization - thus

translating directly into intellectual property - the importance of a data-centric view complementing the robotic automation of the various tasks can hardly be overstated [10].

The highest demand for task automation is generally found in screening steps, where repetitive operations of modest complexity can be conveniently handled by robotics. Protein crystallization is currently the most prominent candidate for full automation, while at the same time the front-end of protein production begins to undergo a similar transformation with increasing use of small scale, high throughput parallel expression and purification screening techniques [11]. Novel plasmid vectors, expression autoinduction, affinity-tagged proteases of high specificity, and modular parallel chromatography equipment have contributed major advances towards automation in protein production. Structure solution, model building and refinement on the other hand are conducted entirely *in silico*, and given the rapid development of automated computational crystallography are generally not considered throughput limiting factors.

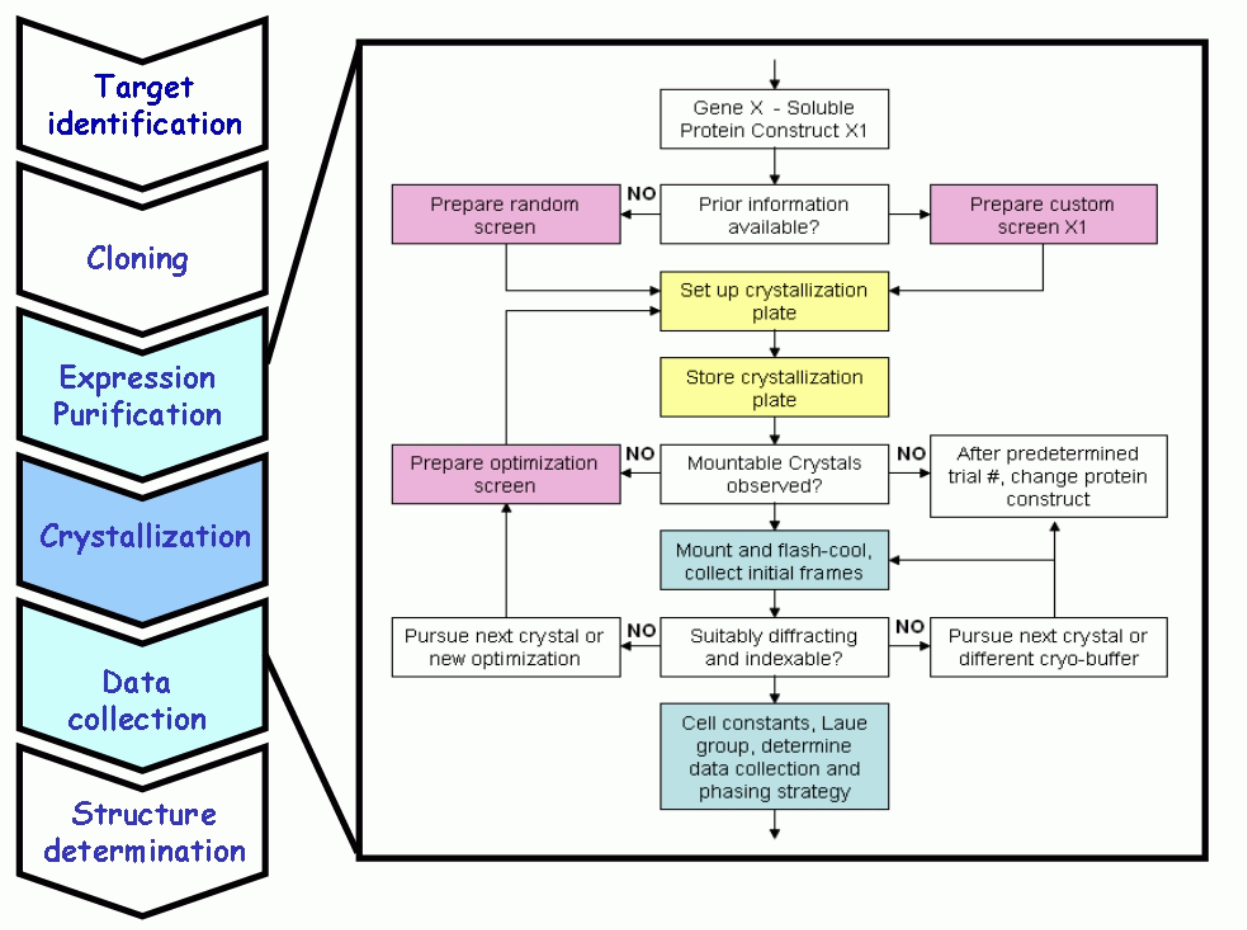


**Figure 3** **Basic tasks in automated protein crystallization**. Shading indicates: basic liquid handling (magenta); crystallization plate setup and handling (yellow); and mounting and data collation tasks (cyan). Omitted for clarity are the numerous feedback pathways from structure analysis to target modification and in iterative ligand screening and optimization.

One of the most substantial advantages of automation is miniaturization, allowing comprehensive parallel screening of large sample sets (multiple constructs, orthologs) with very little material. As a consequence, statistically sound go/no-go decisions can be made early in each successive screening step, and the pursuit of a target already showing warning signs of limited likelihood of success can be avoided. Early go/no-go decisions are common practice in

pharmaceutical industry, as failures in later stages of the process (exemplified by losing harvestable crystals or the failure of new therapeutic drug candidates in late phases of clinical trials) become increasingly costly.

Full walk-away automation up to but not including harvesting is conceivable given the current equipment on the market, and has been demonstrated (albeit at substantial cost) in several custom-made industrial designs [12]. Cocktail preparation, plate setup, automated crystal recognition, and subsequent optimization can be integrated with plate handling robotics and provide no principal (nonetheless financial) challenges. Process automation currently stops at the harvesting stage, largely due to the expense of micromanipulation, and the need for rather advanced machine vision tools to allow real-time processing of the events during crystal harvesting. However, reproducibly manufactured mounting loop designs and micro-manipulation actuators for robots are being developed [13], and will eventually address this remaining manual bottleneck. Once the crystals are safely cryo-protected, however, robotic mounting of sample pins has become standard on HTPX synchrotron beam lines [14] and in larger biotech companies and laboratories.

#### 4.2. Data processing and handling

An often underestimated consequence of task automation is the rapid generation of data during multiple levels of successively branched screening steps such as expression, solubilization, crystallization, cryo-protection (figure 4). Despite considerable sample attrition, the amount of data generated at each screening step can rapidly outstrip the capability to analyse them. Consequently, even for a small effort, capturing of primary data at the source, directly into a relational data base via automated scripts or a laboratory information management system

(LIMS), is important. In addition, with increasing automation, tightly integrated process control, multiple feedback steps, real time data processing and decision making, and machine learning for predictive purposes [15] are becoming major components of any automation intensive laboratory. Direct capture of all experiments also assures the collection of negative experimental results, a valuable and necessary component in the analysis of complex and multivariate crystallization data [16, 17].

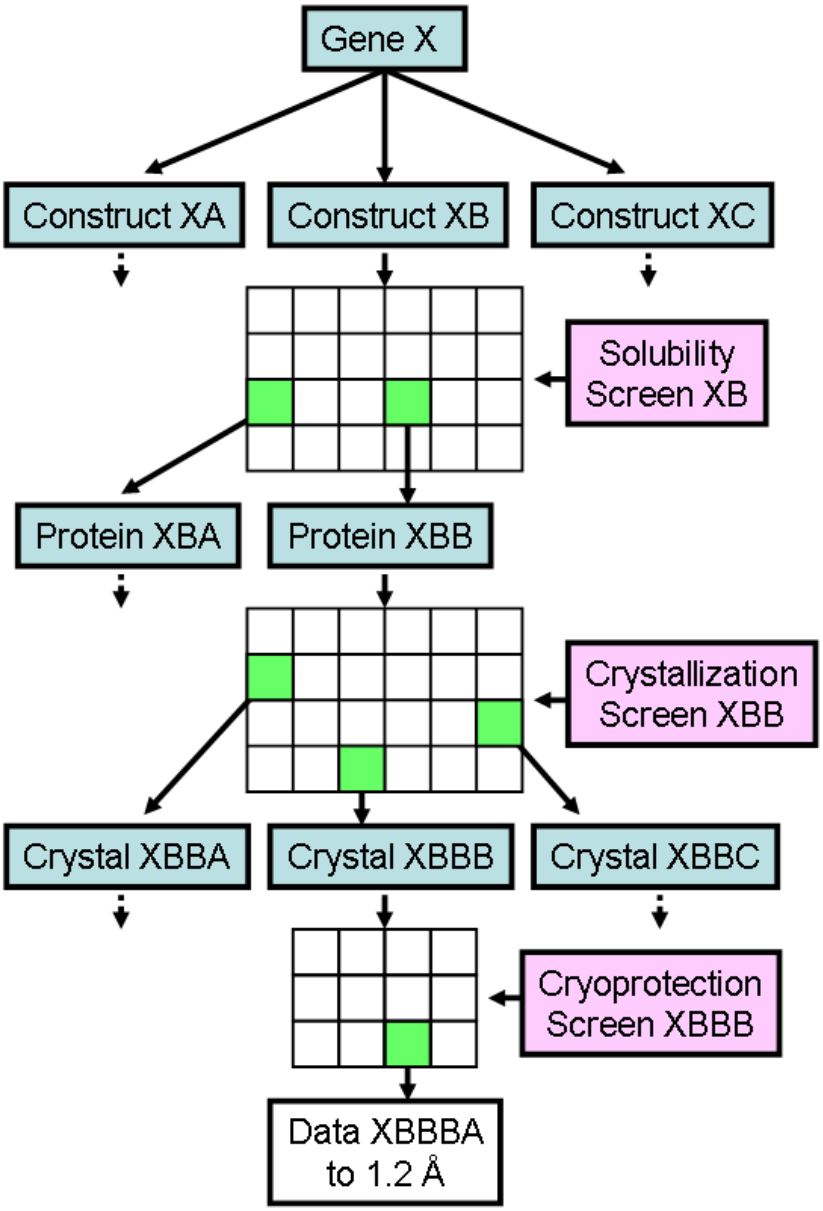


**Figure 4** Path through a basic 3-step screening tree in a protein production and crystallization laboratory. Despite substantial attrition, the amount of data multiplies rapidly [18] with each screening step, and an adequate IT infrastructure must be provided for data capture, data analysis and machine learning, with tight integration to process control and process refinement. Figure reproduced from [15].

## 5. Summary

High throughput crystallization of even a few proteins per day is not sustainable without careful planning of the entire automation setup and thorough operations review. A singular deployment of high speed robotics, in particular if intended to significantly increase throughput, will likely create new bottlenecks downstream. Data capture, warehousing and curating is of substantial importance not just for process control, but also for successful data mining of highly dimensional and complex proteomics data. Realistic planning and full consideration of high throughput process and design principles will go a long way to accomplish a successful and financially sound transition into robotic high throughput screening.

Given a well conceived and implemented design, protein drug target crystallography can be implemented efficiently and has become an increasingly attractive drug discovery and screening tool, allowing in unprecedented ways novel strategies that identify lead candidates early. Structural information combined with *in silico* approaches provides a powerful basis for corrective action during design and optimization of drug candidates - before more resources are expended on low probability targets likely to fail in later steps.

## 6. Acknowledgements

LLNL is operated by University of California for the US DOE under contract W-7405-ENG-48. Work funded under NIH P50 GM62410 grant.

## 7. References

1. Hillisch, A. and R. Hilgenfeld, *Modern Methods of Drug Discovery*. 2002, Basel: Birkhäuser.

2. Blundell, T.L., H. Jhoti, and C. Abell, *High-throughput crystallography for lead discovery in drug design.* Nat Rev Drug Discov, 2002. **1**(1): p. 45-54.

3. Bourne, P.E. and H. Weissig, *Structural Bioinformatics*. 2003, Hoboken, NY: Wiley-Liss.

4. Rees, D.C., et al., *Fragment-based lead discovery.* Nat Rev Drug Discov, 2004. **3**(8): p. 660-72.

5. Congreve, M., et al., *A 'rule of three' for fragment-based lead discovery?* Drug Discov Today, 2003. **8**(19): p. 876-7.

6. Ekins, S., *Perdicting undesirable drug interactions with promiscuous proteins in silico.* Drug Disc Today, 2004. **9**(6): p. 276-285.

7. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications.* Nat Rev Drug Discov, 2004. **3**(11): p. 935-949.

8. Stoichet, B.K., *Virtual screening of chemical libraries.* Nature, 2004. **432**: p. 862-865.

9. Greene, N., *Computer systems for the prediction of toxicity: an update.* Advanced Drug Delivery Reviews, 2002. **54**: p. 417–431.

10. Rupp, B. and J. Wang, *Predictive models for protein crystallization.* Methods, 2004. **34**(3): p. 390-407.

11. Kim, Y., et al., *Automation of protein purification for structural genomics.* Journal of Structural and Functional Genomics, 2004. **5**(1-2): p. 111-118.

12. Hosfield, D., et al., *A fully integrated protein crystallization platform for small-molecule drug discovery.* J Struct Biol, 2003. **142**(1): p. 207-217.

13. Thorne, R.E., et al., *Microfabricated mounts for high-throughput macromolecular cryocrystallography.* J Appl Crystallogr, 2003. **36**(6): p. 1455-1460.

14. Snell, G., et al., *Automatic Sample Mounting and Alignment System for Biological Crystallography at a Synchrotron Source.* Structure, 2004. **12**: p. 1-12.

15. Rupp, B. and J.-W. Wang, *Predictive models for protein crystallization.* Methods, 2004. **34**: p. 390-407.

16. Rupp, B., *Maximum-likelihood crystallization.* J Struct Biol, 2003. **142**(1): p. 162-169.

17. Page, R. and R.C. Stevens, *Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens.* Methods, 2004. **34**(3): p. 373-389.

18. Patterson, S.D., *Data analysis - the Archilles heel of proteomics.* Nature Biotechnol, 2003. **21**: p. 221-222.