Available online at www.sciencedirect.com
SCIENCE DIRECT®
Methods 34 (2004) 373–389
METHODS
www.elsevier.com/locate/ymeth

Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens

Rebecca Page and Raymond C. Stevens*

Department of Molecular Biology, Joint Center for Structural Genomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Received 24 March 2004

Abstract

Recent efforts to collect and mine crystallization data from structural genomics (SG) consortia have led to the identification of minimal screens and novel screening strategies that can be used to streamline the crystallization process. Two groups, the Joint Center for Structural Genomics and the University of Toronto, carried out large-scale crystallization trials on different sets of bacterial targets (539, JCSG and 755, Toronto), using different sample processing and crystallization methods, and then analyzed their results to identify the smallest subset of conditions that would have crystallized the maximum number of protein targets. The JCSG Core Screen contains 67 conditions (from 480) while the Toronto Minimal Screen contains 6 (from 48). While the exact conditions included in the two screens do not overlap, the major precipitants of the conditions are similar and thus both screens can be used to determine if a protein has a natural propensity to crystallize. In addition, studies from other groups including the University of Queensland, the *Mycobacterium tuberculosis* SG group, the Southeast Collaboratory for SG, and the York Structural Biology Laboratory indicate that alternative crystallization strategies may be more successful at identifying initial crystallization conditions than typical sparse matrix screens. These minimal screens and alternative screening strategies are already being used to optimize the crystallization processes within large SG efforts. The differences between these results, however, demonstrate that additional studies which examine the influence of protein biophysical properties and sample preparation methods on crystal formation must also be carried out before more robust screens can be identified.

Keywords: Structural genomics; Protein crystallization; Data mining; Crystallization screens

1. Introduction

By implementing high-throughput and parallel technologies, numerous public and private structural genomics (SG) efforts have the potential to significantly accelerate the structure determination process, and in turn, our understanding of protein function [1–4].[1] These efforts have already resulted in the successful determination of hundreds of novel structures and led to the annotations of proteins with previously unknown functions [5–13]. More importantly, these efforts are also generating considerable amounts of systematically collected positive and negative data for every step in the structure determination process, from cloning to structure determination [14–18]. Since all samples are processed identically, trends in the processing methods and sample characteristics that lead to structure determination success or failure can be identified. These trends will not only greatly facilitate our understanding of protein behavior, but they can also be used to optimize existing processing protocols to make the entire structure determination process more efficient.

Crystallization remains the rate-limiting step in protein structure determination due to the extensive number of variables that must be systematically altered for optimal crystal formation (these variables are collectively

* Corresponding author.
E-mail address: stevens@scripps.edu (R.C. Stevens).

[1] See www.rcsb.org/pdb/strucgen.html for a list of worldwide public and private structural genomics initiatives, with links to corresponding websites.

1046-2023/$ - see front matter © 2004 Elsevier Inc. All rights reserved
doi:10.1016/j.ymeth.2004.03.026

known as 'crystallization space'). Typical variables which influence crystal formation include the concentration and nature of the protein (full length, truncations, and mutations), methods for setting up crystallization trials (vapor diffusion, dialysis, and nanocrystallization), temperature, and the concentration and nature of the chemical components in crystallization solutions (buffer, additives, and precipitants) [19–26]. Over the last 20 years, numerous efforts to make the search for initial crystallization conditions more tractable have led to a number of novel crystallization screens and screening strategies [27–37]. The most widely used screen, the sparse matrix screen, was developed by mining publicly available data for conditions that led to crystallization success [30,32]. Specifically, crystallization conditions that produced diffraction quality crystals were compiled from the literature and the subset which sampled the widest range of buffers and precipitants was identified [32]. Since its initial release, this screen has been used by hundreds of crystallographers to identify initial crystallization leads. More importantly, however, it illustrates the utility of using comprehensive crystallization data mining to streamline the crystallization process.

One of the first large-scale efforts to accumulate and mine crystallization data was initiated over 15 years ago and led to the creation of the Biological Macromolecular Crystallization Database (BMCD; wwwbmcd.nist.gov:8080/bmcd/bmcd.html), an extensive archive of the experimental crystallization details of published protein structures [38,39]. While an incredible resource, the utility of the BMCD for unbiased data mining is limited. First, only successful crystallization results are archived in the BMCD and second, the expression and purification methods used to prepare and screen the protein samples differed significantly among all entries. Since then, other groups have attempted to more systematically survey crystallization conditions in commercially available sparse matrix screens for both crystallization success and failure [27,31,35,40]. While these surveys have also provided useful information, the numbers of conditions and proteins screened were small, limiting their applicability.

SG consortia, which have implemented parallel, pipeline approaches for sample generation, crystallization screening, and tracking, are uniquely positioned to provide these missing data. All samples are processed using identical protocols and the data collected reflect thousands of experiments, making conclusions, both positive and negative, statistically relevant. Two groups, the Joint Center for Structural Genomics (JCSG; www.jcsg.org) and the University of Toronto (Toronto; www.thesgc.com), have mined their crystallization data to identify trends in crystallization success [33,34]. Both studies led to the identification of minimal screens, the smallest sets of conditions that would have produced the maximum number of proteins crystallized. In addition, a third group, the *Mycobacterium tuberculosis* Structural Genomics Consortium (TB; www.doe-mbi.ucla.edu/TB), has used its results to identify crystallization components, especially chemical additives, which maximize crystallization success [41], while a fourth group, the Southeast Collaboratory for Structural Genomics (SEGSG; www.secsg.org), has compared the sparse matrix screening strategy with the incomplete factorial screening strategy to quantify the efficiency of both screening methods [31]. Not only can these minimal screens, additives, and alternative screening strategies be implemented into SG pipelines to make the structure determination process more efficient, but these results can now also be compared to highlight the experimental parameters besides crystallization conditions which are critical for protein crystal formation.

Here, we analyze the results of recent efforts to mine crystallization data for trends in crystallization success. We review the initial efforts to archive all successful crystallization attempts (BMCD), the systematic efforts to compare commercially available sparse matrix screens, and the large-scale efforts to mine SG crystallization data for trends in crystallization success and failure. These studies are then compared, and similarities and differences are discussed.

2. The Biological Macromolecule Crystallization Database: first efforts to archive and mine crystallization data

BMCD was developed out of an effort to archive all successful crystallization experiments [38,39]. Its creation was motivated by the hypothesis that experimental conditions that produced a diffraction quality crystal once might be successful again in future crystallization efforts. First released in 1988, the BMCD now contains crystallization data for 3547 distinct crystal entries from 2526 proteins and it is still growing every week. Each crystal entry, most of which came from published reports in the literature, includes detailed information about the archived crystallization experiment, including protein name, protein concentration, crystallization precipitant, pH, temperature, unit cell, and resolution, among others. In some cases, up to 54 different experimental parameters have been recorded for a single entry. As such, the BMCD is the most comprehensive database for successful crystallization parameters publicly available.

The existence of this database led to some of the first large-scale efforts to mine crystallization data for trends in crystallization success [38,39,42,43]. Within the BMCD itself, the names and number of entries for each macromolecule, prosthetic group, chemical additive, crystallization method, and condition archived have been tabulated. For example, the compiled data show that 3 of the 5 most frequently used methods for

crystallization trials are variations of the vapor diffusion method. Since the tables list all instances of a given parameter, however, the less frequently used methods, such as silica-gel-mediated sitting drops and floating drops, are also listed. The five most frequent entries in the BMCD for prosthetic groups, chemical additives, and crystallization set-up methods are given in Table 1.

More importantly, the public accessibility of this database has enabled other groups to comprehensively analyze these data to identify trends that correlate with crystal formation. For example, the first widely used sparse matrix screens, screens made up of internally unrelated conditions but ones heavily biased towards those which have been successful in the past, were developed by identifying the widest range of crystallization conditions which were very similar to known and/or published conditions, including those contained within the BMCD [30,32]. These data have also been used to carry out detailed statistical analyses to identify the experimental parameters that correlate with the successful crystallization of distinct protein families. In both studies, the analyses led to the identification of conditions which preferentially crystallize distinct classes of macromolecules [42,43]. These results were then incorporated into a computer program which allows users to design screens specific to their molecule of interest [42]. Many of these screens are still in active use today.

Table 1
Data archived in the BMCD

BMCD parameter	No. of different entries/parameter	Top 5 parameter entries
Crystallization methods	37	Vapor diffusion (hanging drop) Vapor diffusion (plates/slides) Batch method Vapor diffusion Microdialysis
Prosthetic groups	91	Heme Zn (II) Ca (II) Glycosylated Mg (II)
Chemical additives	383	Ammonium sulfate Tris–HCl Sodium chloride 2-Methyl-2,4-pentanediol Sodium azide

Crystallization data for 3547 crystals of 2526 proteins are archived in the BMCD and easily accessible via the web at wwwbmcd.nist.gov:8080/bmcd/bmcd.html. Entries can be searched by macromolecule type, species, and crystallization conditions. Tables listing the crystallization methods, prosthetic groups, and crystallization condition chemical additives for all BMCD crystal entries are also available. The five most frequent entries per parameter are listed here.

The primary advantage of the BMCD is that it is a publically available, comprehensive collection of crystallization parameters that have produced crystals suitable for structure determination. However, this is also its primary disadvantage because those conditions and methods that *failed* prior to those used for successful crystal formation are not recorded. If crystallization screens and methods are to be truly optimized, these negative data are essential. While data of this sort are often maintained on a small-scale, they are only just now beginning to be systematically collected for large numbers of protein targets.

3. Efforts to compare and optimize sparse matrix crystallization screens

The data archived in the BMCD led to the development of numerous commercially available sparse matrix screens. While these screens are widely used, the ability of each of these conditions to promote crystal formation for a wide range of targets has only recently begun to be carefully characterized. SG groups are in a unique position to carry out such analyses with large sets of identically processed targets. Other groups have also made medium-scale attempts to systematically evaluate and optimize these conditions for their ability to promote crystal formation which are discussed below.

3.1. Queensland: comparison of three commercially available sparse matrix screens

Wooh et al. (2003; Queensland) compared the crystallization efficiency of 288 conditions of six commercially available crystallization kits (CS: Crystal Screen and Crystal Screen 2, Hampton Research; WS: Wizard Screens 1 and 2, Emerald BioStructures; and PSS: Personal Structure Screens 1 and 2, Molecular Dimensions) [40]. A total of 19 proteins were screened for crystal formation against these conditions. While the protein sample size is small, numerous crystallographers including SG groups use many of these screens for initial crystallization trials, so the results of this study are still relevant. In addition, since a number of conditions overlap (94 conditions of the CS and PSS screens are identical or nearly identical), this study provides data on the ability of these conditions to reproducibly promote crystal formation.

The experimental design was straightforward. Thirteen proteins were purchased and dissolved in protein crystallization buffer (25 mM Tris (hydroxymethyl)aminomethane hydrochloride [Tris], pH 7.0) at a concentration of 10 mg/ml, while the remaining six were purified by the authors and equilibrated against protein specific buffers at concentrations ranging from 8.6 to 27 mg/ml. The vapor diffusion hanging drop crystallization trials,

using 1 μl protein plus 1 μl crystallization drops equilibrated against 100 μl crystallization buffer, were set up and plates were incubated at 16 °C. Drops were inspected for crystal formation immediately after set-up and 1 week later. All needles, plates and 3D crystals were considered successful crystallizations.

Eighteen of nineteen proteins crystallized, with 9 proteins crystallizing in 10 or more conditions and 2 crystallizing in only one. The most successful screen for promoting crystal formation was the PSS screen, with all 18 proteins crystallizing in at least one of the PSS conditions. The other screens were less successful. The CS and the WS screens both crystallized only 13 proteins. The most successful condition was Hampton Crystal Screen 14, which crystallized 10 (53%) of the 19 proteins.

One of the most interesting results of this study was the observation that most of the 94 identical (or nearly identical) conditions between the CS and PSS screens did not produce crystals for the same proteins. The observed differences could be due to a number of factors, such as normal fluctuations in nucleation from one drop to another, but could also be due to differences in the conditions themselves, even though they are listed as identical by the manufacturers. The authors showed that in spite of having the same formulations, the measured pH values between corresponding conditions in the two screens often differed from one another, with 22 conditions differing by at least 0.5 pH units and one condition differing by 5. These differences could be due to alternative preparation methods and/or screen storage and handling. Premade conditions change over time in storage. A study on the long-term stability of commercially available crystal screens showed that the same conditions from two different lots can differ in pH by 0.5 units and that the pH change observed within a condition over time (6 weeks at 4 °C in this study) can change more than 0.5 pH units [44]. Thus, the differences in crystallization behavior of 'formulation identical' conditions observed in this study, and other similar studies, could also be due to differences in the conditions themselves.

3.2. YSBL: two simple, flexible screens for the rational crystallization of enzymes

While the number of commercially available screening conditions is continually growing, there is still no guarantee that including these new conditions in an initial crystallization experiment will lead to a comparable increase in initial crystallization leads. On the contrary, it is likely that these conditions will just add additional sample and reagent requirements, without necessarily providing additional information about which conditions will ultimately result in crystal formation. To address this issue, Brzozowski and Walton (2001, YSBL) developed two novel crystallization screens for the efficient crystallization of a distinct class of macromolecules, enzymes [27]. The screens are simple and small. The Clear Strategy Screen I (CSS-I) conditions contain various combinations of polyethylene glycols (PEGs) and salts, while the 24 Clear Strategy Screen II (CSS-II) conditions contain salts, organics (such as ethanol, jeffamine, and isopropanol), polyalcohols (such as 2-methyl-2,4-pentanediol, 1,2-butanediol) or additional combinations of PEGs and salts. The advantage of these screens is that the investigator can use information about the protein target to guide the experimental design. The user determines the pH of these screen conditions, and the chemical components, such as salts, can easily be replaced by ones that might be more appropriate for the target under study. More importantly, the results of one experiment can be easily interpreted to rationally guide the design of a next round of crystallization trials.

The authors showed that these screens were very successful for the identification of initial crystallization leads, and moreover, were often successful when commercially available sparse matrix screens were not [27]. Specifically, CSS-I and CSS-II resulted in the successful crystallization of at least 10 targets which could not be crystallized by the conditions in the Hampton Crystal Screen and Crystal Screen 2. While some proteins crystallized in multiple PSS conditions (up to 24), others crystallized in only one. In spite of this variation in crystallization success, they all eventually led to the production of crystals suitable for diffraction studies for each target (final crystals diffracted from 1.3 to 2.6 Å). Although the average time required to obtain the diffraction quality target from the initial lead was not reported, many of the crystals produced by these screens were already of sufficient quality for diffraction studies and no further optimization was required. Clearly, these screens, which provide a more rational approach to crystallization since the results of one screen can be used to efficiently refine the next round crystallization conditions, can be used to effectively identify initial crystallization leads. Their current use is optimal for investigators that have detailed knowledge of the protein of interest. Their applicability to large-scale SG programs is somewhat limited since there is little to no functional information known about many of the SG targets, making tailoring of the screen to the target more difficult. It is expected, however, that the crystallization data being produced by these SG efforts will eventually lead to the production of crystallization screens that are tailored to specific classes of proteins based not only on function, but on sequence as well.

4. Structural genomics efforts to mine crystallization data

Recent SG efforts have implemented high-throughput, parallel technologies for the rapid pipeline production of

protein samples for crystallization trials [15,18], ensuring that the majority of samples are treated identically at every stage of the pipeline. Since the results for each sample are carefully tracked, these experiments are, for the first time providing positive and negative data for thousands of crystallization experiments on hundreds of protein targets. These results can now be mined for trends that can be used to optimize the crystallization screening process. Four SG consortia, the Joint Center for Structural Genomics (JCSG), the University of Toronto (Toronto), the TB Structural Genomics Consortium (TB), and the Southeast Collaboratory for Structural Genomics (SECSG) have made concerted efforts to mine their initial crystallization results to optimize future crystallization screens and in turn streamline the structure determination process [31,33,34,41]. The results of each center are described below. These results are then compared, both with one another and with the results from the smaller-scale studies in the comparison section that follows.

4.1. JCSG: identification of the Core Screen, 67 conditions that crystallize 392 proteins

The JCSG used a crystallization strategy that separated initial crystallization screening from optimization [18]. It is a two-tiered strategy which is founded on the hypothesis that proteins which crystallize readily, even under suboptimal conditions (tier 1, screening), will do so again during focused crystallization attempts (tier 2, optimization). The tier 1 results are reviewed here as this analysis led to the identification of a minimal set of conditions that crystallized more than 84% of the total proteins crystallized in this study (the Core Screen; Table 2).

The first genome selected for processing by the JCSG was that of *Thermotoga maritima*. Every predicted *T. maritima* ORF (1877) was processed through the high-throughput parallel processing pipeline [18]. Of these, 539 (29%) were successfully prepared for crystallization trials. Proteins were expressed and purified using a single chromatography step (metal chelation; all targets contained an N-terminal tag, MGSDKIHHHHHH, to facilitate expression and purification) for a final purity of 90–95%. Each protein was then buffer exchanged into crystallization screening buffer (10 mM Tris, pH 7.9, 150 mM NaCl, and 0.25 mM Tris (2-carboxyethyl) phosphine hydrochloride [TCEP]), concentrated to ~10 mg/ml, and screened for crystallization against 480 commercially available crystallization conditions[2] at 20 °C using the nanocrystallization sitting drop vapor diffusion method [45] with 100 nl drops (50 nl protein plus 50 nl crystallization solution; Fig. 1). This effort resulted in 258,720 distinct crystallization experiments and has provided one of the most extensive, systematic datasets of commonly used crystallization conditions against a wide range of proteins to date [18].

Of the 539 *T. maritima* proteins screened, 465 (86%) crystallized for a total of 5546 crystal hits, indicating that most of the proteins had an inherent propensity to crystallize. Notably, over half of these proteins crystallized in 5 or more conditions, with 19 crystallizing in 50 or more conditions and one, TM0665, crystallizing in over 230. The proteins crystallized sampled a wide range of p*I* and MW values, similar to that observed for the proteome as a whole [18], and there was no observed correlation between protein p*I* and crystallization pH [34].

The conditions used for the crystallization trials were also very successful for promoting crystal formation. Four hundred and seventy-two (98%) of the 480 conditions produced crystals for at least one protein, with 37 conditions producing crystals for 25 or more proteins and one, Core Screen #08 (Table 2), producing crystals for more than 40. Notably, however, the 10 most effective conditions, which account for only 2% of the original conditions used, produced crystals for 196 (42%) different proteins while the best 108 conditions (23%) produced crystals for all 465. Thus, many of the 480 conditions used for these trials significantly oversampled distinct regions of crystallization space and could be eliminated without impacting the final number of distinct proteins crystallized.

To identify the redundant conditions, an iterative selection algorithm, *Min_Cov*, was used to identify the minimum set of conditions that would have produced crystals for the entire set of crystallized proteins (minimal screens) [34]. *Min_Cov* identified 472 minimal screens, one for each condition which successfully crystallized a protein; 415 of these minimal screens were unique. The 67 conditions present in every minimal screen were then identified as those conditions most essential for promoting crystal formation for the most diverse set of proteins. These conditions are collectively referred to as the Core Screen [34][3] and include representatives from all five primary precipitant classes, including high MW PEGs (31 of the original 171 conditions), low MW PEGs (8 of 67), ammonium sulfate/salts (10 of 106), polyalcohols (11 of 83), and remaining organics (7 of 53), although high MW PEG conditions are still the

[2] Conditions compiled from 10 commercially available kits, including Crystal Screen, Crystal Screen 2, Crystal Screen Cryo, PEG/Ion Screen, Grid Screen Ammonium Sulfate, Grid Screen PEG 6000, Grid Screen Peg/LiCl, Grid Screen MPD (Hampton Research, Riverside, CA), Wizard I/II, Cryo I/II (Emerald Biostructures, Bainbridge Island, WA).

[3] Because each condition in the Core Screen was present in all 472 identified minimal screens, they are all *equally* important for the crystallization of the most diverse set of protein targets in this study and thus their numbered position in the Core Screen Table is arbitrary. See Table 2.

Table 2
Expanded Core Screen identified by the JCSG [34]

	Condition	Screen[a]	Freq[b]
1	50% (w/v) PEG 400, 0.2 M Li_2SO_4, 0.1 M acetate, pH 5.1	W1cryo #47	415
2	20% (w/v) PEG 3000, 0.1 M citrate, pH 5.5	W1 #06	415
3	20% (w/v) PEG 3350, 0.2 M diammonium hydrogen citrate, pH 5.0	PEG/ion #48	415
4	30% (v/v) MPD, 0.02 M $CaCl_2$, 0.1 M NaOAc, pH 4.6	H1 #01	415
5	20% (w/v) PEG 3350, 0.2 M magnesium formate, pH 5.9	PEG/ion #20	415
6	20% (w/v) PEG 1000, 0.2 M Li_2SO_4, phosphate-citrate, pH 4.2	W1 #39	415
7	20% (w/v) PEG 8000, 0.1 M CHES, pH 9.5	W1 #01	415
8	20% (w/v) PEG 3350, 0.2 M ammonium formate, pH 6.6	PEG/ion #23	415
9	20% (w/v) PEG 3350, 0.2 M ammonium chloride, pH 6.3	PEG/ion #09	415
10	20% (w/v) PEG 3350, 0.2 M potassium formate, pH 7.3	PEG/ion #22	415
11	50% MPD, 0.2 M $(NH_4)H_2PO_4$, 0.1 M Tris, pH 8.5	H2 #43	415
12	20% (w/v) PEG 3350, 0.2 M potassium nitrate, pH 6.9	PEG/ion #18	415
13	0.8 M $(NH_4)_2SO_4$, 0.1 M citric acid, pH 4.0	$AmSO_4$ #01	415
14	20% (w/v) PEG 3350, 0.2 M sodium thiocyanate, pH 6.9	PEG/ion #13	415
15	20% (w/v) PEG 6000, 0.1 M bicine, pH 9.0	P6K #18	415
16	10% (w/v) PEG 8000, 8% ethylene glycol, 0.1 M HEPES, pH 7.5	H2 #37	415
17	40% (v/v) MPD, 5% (w/v) PEG 8000, 0.1 M cacodylate, pH 7.0	W2cryo #01	415
18	40% ethanol, 5% (w/v) PEG 1000, 0.1 M phosphate-citrate, pH 5.2	W1cryo #40	415
19	8% (w/v) PEG 4000, 0.1 M NaOAc, pH 4.6	H1 #37	415
20	10% (w/v) PEG 8000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.0	W2 #43	415
21	20% (w/v) PEG 6000, 0.1 M citric acid, pH 5.0	P6K #14	415
22	50% (v/v) PEG 200, 0.2 M $MgCl_2$, 0.1 M cacodylate, pH 6.6	W2cryo #36	415
23	1.6 M sodium citrate, pH 6.5	H2 #28	415
24	20% (w/v) PEG 3350, 0.2 M tripotassium citrate monohydrate, pH 8.3	PEG/ion #47	415
25	30% MPD, 0.02 M $CaCl_2$, 0.1 M NaOAc, pH 4.6	H1cryo #01	415
26	20% (w/v) PEG 8000, 0.2 M NaCl, 0.1 M phosphate-citrate, pH 4.2	W1 #31	415
27	20% (w/v) PEG 6000, 1.0 M LiCl, 0.1 M citric acid, pH 4.0	P6K/LiCl #13	415
28	20% (w/v) PEG 3350, 0.2 M ammonium nitrate, pH 6.3	PEG/ion #19	415
29	10% (w/v) PEG 6000, 0.1 M HEPES, pH 7.0	P6K #10	415
30	0.8 M NaH_2PO_4/0.8 M KH_2PO_4, 0.1 M HEPES, pH 7.5	H1 #35	415
31	40% (v/v) PEG 300, 0.1 M phosphate-citrate, pH 5.2	W2cryo #18	415
32	10% (w/v) PEG 3000, 0.2 M $Zn(OAc)_2$, 0.1 M acetate, pH 4.5	W2 #01	415
33	20% ethanol, 0.1 M Tris, pH 8.5	H2 #44	415
34	25% (v/v) 1,2-propanediol, 0.1 M Na/K phosphate, 10% (v/v) glycerol, pH 6.8	W2cryo #11	415
35	10% (w/v) PEG 20 000, 2% dioxane, 0.1 M bicine, pH 9.0	H2 #48	415
36	2.0 M $(NH_4)_2SO_4$, 0.1 M acetate, pH 4.6	H1 #47	415
37	10% (w/v) PEG 1000, 10% (w/v) PEG 8000	H2 #07	415
38	24% (w/v) PEG 1500, 20% glycerol	H1cryo #43	415
39	30% (v/v) PEG 400, 0.2 M $MgCl_2$, 0.1 M HEPES, pH 7.5	H1cryo #23	415
40	50% (v/v) PEG 200, 0.2 M NaCl, 0.1 M Na/K phosphate, pH 7.2	W2cryo #15	415
41	30% (w/v) PEG 8000, 0.2 M Li_2SO_4, 0.1 M acetate, pH 4.5	W1 #17	415
42	70% (v/v) MPD, 0.2 M $MgCl_2$, 0.1 M HEPES, pH 7.5	H2 #35	415
43	20% (w/v) PEG 8000, 0.1 M Tris, pH 8.5	W2 #03	415
44	40% (v/v) PEG 400, 0.2 M Li_2SO_4, 0.1 M Tris, pH 8.4	W1cryo #38	415
45	40% (v/v) MPD, 0.1 M Tris, pH 8.0	MPD #17	415
46	25.5% (w/v) PEG 4000, 0.17 M $(NH_4)_2SO_4$, 15% glycerol	H1cryo #31	415
47	40% (v/v) PEG 300, 0.2 M $Ca(OAc)_2$, 0.1 M cacodylate, pH 7.0	W1cryo #37	415
48	14% 2-propanol, 0.14 M $CaCl_2$, 0.07 M acetate, pH 4.6, 30% glycerol	H1cryo #24	415
49	16% (w/v) PEG 8000, 0.04 M KH_2PO_4, 20% glycerol	H1cryo #42	415
50	1.0 M sodium citrate, 0.1 M cacodylate, pH 6.5	W1 #14	415
51	2.0 M $(NH_4)_2SO_4$, 0.2 M NaCl, 0.1 M cacodylate, pH 6.5	W2 #04	415
52	10% 2-propanol, 0.2 M NaCl, 0.1 M HEPES, pH 7.5	W1 #02	415
53	1.26 M $(NH_4)_2SO_4$, 0.2 M Li_2SO_4, 0.1 M Tris, pH 8.5	W1 #47	415
54	40% (v/v) MPD, 0.1 M CAPS, pH 10.1	W2cryo #25	415
55	20% (w/v) PEG 3000, 0.2 M $Zn(OAc)_2$, 0.1 M imidazole, pH 8.0	W2 #40	415
56	10% 2-propanol, 0.2 M $Zn(OAc)_2$, 0.1 M cacodylate, pH 6.5	W2 #11	415
57	1.0 M $(NH_4)_2HPO_4$, 0.1 M acetate, pH 4.5	W1 #09	415
58	1.6 M $MgSO_4$, 0.1 M MES, pH 6.5	H2 #20	415
59	10% (w/v) PEG 6000, 0.1 M bicine, pH 9.0	P6K #12	415
60	14.4% (w/v) PEG 8000, 0.16 M $Ca(OAc)_2$, 0.08 M cacodylate, pH 6.5, 20% glycerol	H1cryo #46	415
61	10% (w/v) PEG 8000, 0.1 M imidazole, pH 8.0	W2 #34	415
62	30% Jeffamine M-600, 0.05 M CsCl, 0.1 M MES, pH 6.5	H2 #24	415
63	3.2 M $(NH_4)_2SO_4$, 0.1 M citric acid, pH 5.0	$AmSO_4$ #20	415

Table 2 (*continued*)

	Condition	Screen[a]	Freq[b]
64	20% MPD, 0.1 M Tris, pH 8.0	MPD #11	415
65	20% Jeffamine M-600, 0.1 M HEPES, pH 6.5	H2 #31	415
66	50% (v/v) ethylene glycol, 0.2 M $MgCl_2$, 0.1 M Tris, pH 8.5	W1cryo #43	415
67	10% MPD, 0.1 M bicine, pH 9.0	MPD #06	415
68	2.0 M $(NH_4)H_2PO_4$, 0.1 M Tris, pH 8.5	H1 #41	414
69	3.4 M 1.6 hexanediol; 0.2 M $MgCl_2$; 0.1 M Tris, pH 8.5	H2 #39	414
70	20% (w/v) PEG 6000, 0.1 M citric acid, pH 4.0	P6K #13	414
71	0.2 M potassium chloride, 20% (w/v) PEG 3350, pH 6.9	Peg/Ion #08	414
72	35% (v/v) 2-ethoxyethanol, 0.05 M $Ca(OAc)_2$, 0.1 M imidazole, pH 7.5	W1cryo #18	414
73	35 % (v/v) MPD; 0.2 M Li_2SO_4; 0.1 M MES, pH 6.0	W2 #02	414
74	1.26 M $(NH_4)_2SO_4$; 0.2 M NaCl; 0.1 *M* CHES, pH 9.5	W2 #29	414
75	10 % (w/v) PEG 3000; 0.2 M NaCl; 0.1 M phosphate-citrate, pH 4.2	W2 #36	414
76	40% (v/v) PEG-600, 0.1 M CHES, pH 9.6	W2cryo #31	414
77	40% (v/v) PEG-400, 0.1 M imidazole, pH 7.4	W2cryo #43	414
78	25 % (w/v) PEG 4000, 0.2 M $(NH_4)_2SO_4$, 0.1 M acetate, pH 4.6	H1 #20	413
79	2.0 M $(NH_4)_2SO_4$	H1 #32	413
80	8% (w/v) PEG 8000; 0.1 M Tris, pH 8.5	H1 #36	413
81	35% (v/v) Dioxane	H2 #04	413
82	1.0 M Hexanediol, 0.01 M $CoCl_2$, 0.1 M Na-acetate, pH 4.6	H2 #11	413
83	20% (v/v) PEG 1000, 0.1 M Tris, pH 7.0	W1 #19	413
84	2.5 M NaCl, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.0	W2 #17	413
85	20% (w/v) PEG 8000, 0.2 M $Ca(Ac)_2$, 0.1 M MES, pH 6.0	W2 #28	413
86	20% (w/v) PEG 3350, 0.2 M di-ammonium tartrate, pH 6.6	Peg/ion #38	411
87	40% (v/v) ethylene glycol, 0.1 M acetate, pH 5.0	W1cryo #02	411
88	20% (w/v) PEG 6000, 1.0 M LiCl, 0.1 M Tris, pH 8.0	P6K/LiCl #17	410
89	1.26 M $(NH_4)_2SO_4$, 0.1 M cacodylate, pH 6.5	W1 #13	407
90	12% (w/v) PEG 20000, 0.1 M MES, pH 6.5	H2 #22	403
91	20% (w/v) PEG 3350, 0.2 M lithium acetate dihydrate, pH 7.8	Peg/ion #24	390
92	20% (w/v) PEG 3350, 0.2 M sodium formate, pH 7.2	Peg/ion #21	388
93	20% (v/v) PEG 2000; 0.1 M Tris, pH 7.0	W1 #10	381
94	2.0 M $(NH_4)_2SO_4$; phosphate/citrate, pH 4.2	W2 #09	365
95	1.6 M $(NH_4)H_2PO_4$, 0.08 M Tris, pH 8.5, 20% glycerol	H1cryo #48	364
96	40% (v/v) ethanol, 0.05 M $MgCl_2$, 0.1 M Tris, pH 8.4	W1cryo #07	343

The Core Screens are those conditions that crystallize the maximum number of distinct proteins with the minimal number of conditions for the JCSG study. The original Core Screen (Conditions 1–67) crystallized 383 (84%) of all proteins crystallized. The 29 next most effective conditions were added to make the Expanded Core Screen, so it is compatible with 96-well crystallization formats. The 96 conditions of the Expanded Core Screen crystallized 448 (98%) of proteins crystallized.

[a] H1, H2, H1cryo, PEG/Ion, $AmSO_4$, P6K, P6K/LiCl, MPD: Crystal Screen, Crystal Screen 2, Crystal Screen Cryo, PEG/Ion Screen, Grid Screen Ammonium Sulfate, Grid Screen PEG 6000, Grid Screen PEG/LiCl, Grid Screen MPD, respectively (Hampton Research). W1, W2, W1cryo, W2cryo: Wizard I and II and Cryo I and II, respectively (Emerald Biostructures).

[b] Frequency the condition was identified in a minimal screen using the Min_Cov algorithm. 473 minimal screens identified, 415 of which were unique.

most prevalent (47%). Using only the Core Screen conditions, greater than 84% (392/465) of the proteins would have been crystallized. This screen has since been expanded to include the 29 next most effective conditions for promoting crystal formation to make it compatible with 96-well crystallization plates (Table 2). This Expanded Core Screen would have crystallized 448 of the total 465 proteins crystallized in this study, for a success rate of 96%. This screen is now regularly used for initial crystallization trials.

4.2. Toronto: identification of a 24 condition minimal screen that crystallizes 308 proteins

In parallel to the JCSG efforts, the Toronto SG group also processed a significant number of samples using parallel processing methods to identify minimal crystallization screens [33]. In this effort, 755 distinct protein samples from 6 bacterial genomes (*Staphylococcus aureaus*, *Helicobacter pylori*, *Escherichia coli*, *Methanobacterium thermoautotrophicum*, *T. maritima*, and *Pseudomonas aeruginosa*) were screened for crystal formation. This contrasts to the JCSG effort, in which only 539 proteins from a single organism (*T. maritima*) were screened.

The sample preparation methods used in the Toronto study also differed from those used by the JCSG (sample preparation methods are compared in Table 3). Similar to the JCSG, proteins were expressed and purified using a single chromatography step (metal chelation; all Toronto targets contained a different N-terminal tag, HHHHHHSSGLVPRGSH, to facilitate purification, and allow for the removal of the tag using thrombin). Unlike the JCSG, the thermophilic samples from *M. ther-*

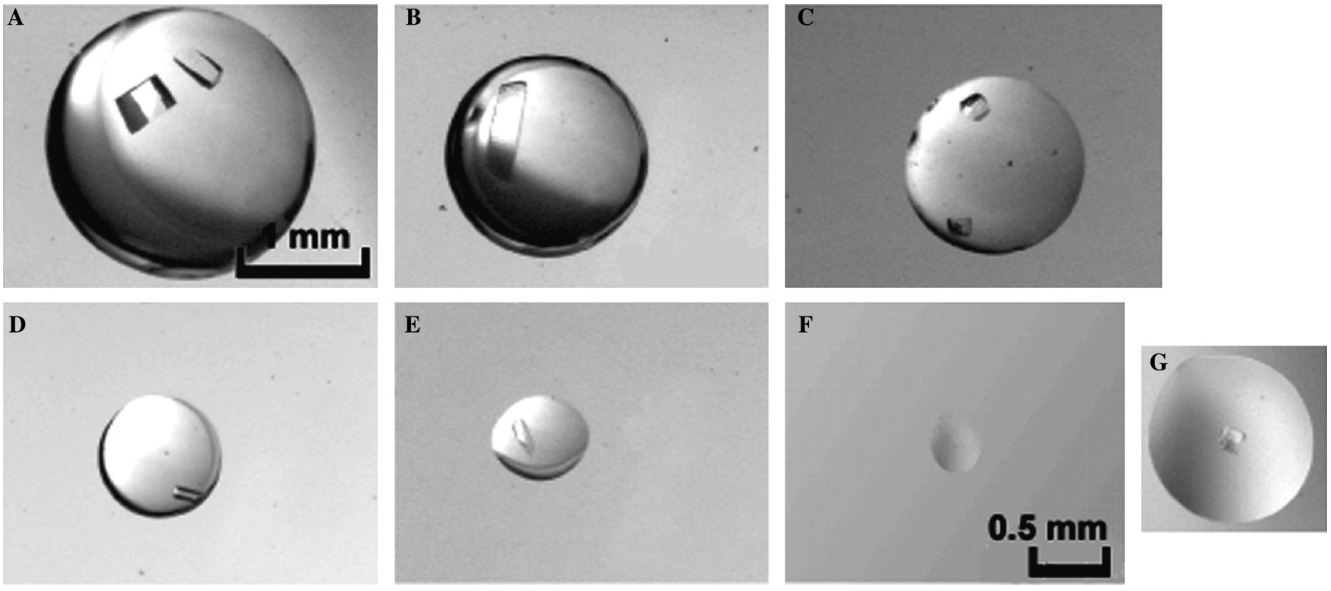


Fig. 1. Nanodrop crystallization: the rates of protein drop equilibration and crystal formation differ with drop size. Representative images of lysozyme crystallization trials, obtained at 277 K, with various total drop volumes. Total drop volumes for crystallization contain equal volumes of protein and reservoir solutions (e.g., 4 ml total drop volume equals 2 ml protein plus 2 ml reservoir solution). The time listed after total drop size is the amount of time to see a crystal of ~50 μm size: (A) 4 ml drop volume, 48 ± 72 h for initial crystallization; (B) 2 ml, 36 ± 48 h; (C) 1 ml, 24 ± 36 h; (D) 400 nl, 18 h; (E) 100 nl, 10 ± 12 h; and (F) 40 nl, 2 h (G) is a magnified view of (F) showing crystal formation, roughly 50 μm on the longest edge (previously published, [45]).

Table 3
Protocols used by the JCSG and Toronto groups for protein crystallization screening

Protocol	JCSG	Toronto
Purification	1 Chromatographic step (MC affinity)	Heating (55 °C)[a] 1 Chromatographic step (MC affinity)
Purification tag	Not cleaved	Cleaved in many cases[b]
Estimated purify	⩾90–95%	⩾95–99%
Protein crystallization buffer	20 mM Tris, pH 7.9 150 mM NaCl 0.25 mM TCEP	20 mM HEPES, pH 7.5 500 mM NaCl
Protein concentration	~10 mg/ml 1 protein concentration screened per protein	5–40 mg/ml 2–4 protein concentrations screened per protein and results pooled
Set-up method	Vapor diffusion sitting drop 100 nl drop; 75 μl well solution; 96-well plate format	Vapor diffusion hanging drop 2 μl drop size; 700 μl well solution; 24-well plate format
Temperature	20	20
Conditions screened	480	48

There are key differences in nearly every protocol used by the JCSG and Toronto groups for protein crystallization trials. Each can have a significant effect on the crystallization process. The differences observed between the two groups with respect to those conditions most likely to promote crystal formation reflect the differences in the experimental protocols listed here.

[a] For only some samples.

[b] Cleaved/uncleaved versions of the same protein were treated as distinct protein samples.

moautotrophicum were usually heated prior to purification, causing an enrichment of many of these proteins in the lysate. In addition, in cases where the N-terminal 6× His tags were cleaved, the proteins underwent a second purification step for tag removal. Tagged and untagged versions of the same protein were treated as separate samples for this study. The final purity of most samples was judged to be 99% by Coomassie staining. The samples were exchanged into a different crystallization buffer (20 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), pH 7.5; 500 mM NaCl) and screened for crystallization trials against only 48 sparse matrix coarse screen conditions at a single temperature (293 K) using the vapor diffusion hanging drop method with 4 μl drops (2 μl protein plus 2 μl crystallization solution). Finally, each protein was screened for crystallization at 2–4 different protein concentrations from 5 to 40 mg/ml (versus only 1 protein concentration at 10 mg/ml in the

JCSG study), and the results were pooled for each target.

Three hundred and thirty-eight (45%) of the proteins in the Toronto study crystallized in at least one condition, averaging nearly 5 crystal hits per protein target, with 99 proteins crystallizing in only a single condition. In addition, protein crystallization behavior varied with protein species. Proteins from *T. maritima* produced the greatest number of crystallized proteins, with 68% (23/34) proteins crystallized, while those of *H. pylori* produced the least, with 37% (47/128). This percentage of crystallized *T. maritima* proteins (68%) is smaller than that obtained in the JCSG study (84%), but is due in part to the smaller number of targets (34, Toronto and 539, JCSG) and conditions screened (48 Toronto and 480, JCSG). In fact, when the JCSG crystallization results were limited to only the 48 conditions used in this study (instead of the entire set of 480), only 222 (49%) of the *T. maritima* proteins successfully crystallized. This observation is discussed further in the section on screen similarities and differences.

Each of the 48 conditions used in the Toronto study produced crystals for at least four of the 338 crystallized proteins. Some conditions were extremely successful at promoting crystal formation, with 20 conditions (42%) producing crystals for 35 or more proteins and 3 (6%) producing crystals for 70 or more. Notably, some conditions were significantly less successful, with 5 (10%) of the conditions each producing crystals for no more than 10 distinct proteins. As stated above, 99 of the proteins screened crystallized in only a single condition and 38 (78%) of the conditions screened crystallized at least one of these 99 proteins. Interestingly, the total number of proteins crystallized per single condition was not correlated with the total number of proteins that had only crystallized in that condition, indicating that the likelihood of a condition to crystallize a more intractable target (one of the 99 'single crystal hit' proteins) was not necessarily due to an inherent ability of the condition to crystallize more proteins.

Remarkably, the six most productive conditions would have produced crystals for 180 (53%) of the crystallized proteins, indicating that the conditions in this screen, like those of the JCSG screen, are significantly oversampled. By systematically, generating all possible combinations of conditions, the set most effective for promoting crystal formation of the widest range of protein targets was identified (Table 4). A minimal screen consisting of just 6 conditions would have produced crystal hits for 205 (61%) of the 338 proteins crystallized, while a minimal screen containing 24 of the conditions would have produced crystal hits for 318 (94%). Reducing the number of initial screening conditions by half doubles the number of experiments that can be tested for a given target, with almost no impact on the final number of targets crystallized.

Table 4
Minimal screens 6, 12, and 24 identified by the Toronto Group

J–K/Hampton CS 1	Condition	Total crystals	Minimal screen
6	30% PEG 4000 Tris–HCl, pH 8.5, 0.2 M $MgCl_2$	65	6, 12, 24
10	30% PEG 4000 Na acetate, pH 4.6, 0.2 M NH_4 acetate	49	6, 12, 24
18	20% PEG 8000 Na cacodylate, pH 6.5, 0.2 M Mg Acetate	72	6, 12, 24
38	1.4 M Na citrate Na HEPES, pH 7.5	62	6, 12, 24
39	2.0 M NH_4 sulfate Na HEPES, pH 7.5, 2% PEG 400	63	6, 12, 24
43	30% PEG 1500	54	6, 12, 24
4	2.0 M NH_4 sulfate Tris–HCl, pH 8.5	50	12, 24
17	30% PEG 4000 Tris–HCl, pH 8.5, 0.2 M Li sulfate	70	12, 24
30	30% PEG 8000 0.2 M NH_4 sulfate	38	12, 24
36	8% PEG 8000 Tris–HCl, pH 8.5	22	12, 24
41	10% Isopropanol + 20% PEG 4000 Na HEPES, pH 7.5	58	12, 24
45	18% PEG 8000 Na cacodylate, pH 6.5, 0.2 M Zn acetate	19	12, 24
1	30% MPD Na acetate, pH 4.6, 0.02 M $CaCl_2$	17	24
11	1.0 M NH_4 phosphate Na citrate, pH 5.6	15	24
13	30% PEG 400 Tris–HCl, pH 8.5, 0.2 M Na citrate	10	24
14	28% PEG 400 Na HEPES, pH 7.5, 0.2 M $CaCl_2$	25	24
16	1.5 M Li sulfate Na HEPES, pH 7.5	30	24
20	25% PEG 4000 Na acetate, pH 4.6, 0.2 M NH_4 sulfate	37	24
21	30% MPD Na cacodylate, pH 6.5, 0.2 M Mg acetate	19	24
28	30% PEG 8000 Na cacodylate, pH 6.5, 0.2 M Na acetate	65	24
33	4.0 M Na formate	36	24
34	2.0 M Na formate Na acetate, pH 4.6	34	24
35	1.6 M K,Na phosphate, Na HEPES, pH 7.5	15	24
42	20% PEG 8000 0.05 M K phosphate	52	24

Minimal screens composed of those conditions that crystallized the maximum number of distinct proteins for the Toronto study. Minimal Screen 6 crystallized 205 (61%) of the total proteins crystallized, minimal screen 12 crystallized 268 (79%), and minimal screen 24 crystallized 318 (94%).

4.3. TB consortium: additives that facilitate crystal formation

Not all SG groups use the sparse matrix screening method. The findings of Segelke [35] and Rupp [41] (TB) suggest that random sampling of crystallization space is the most efficient method for screening novel protein targets, particularly when success rates are low or clustered. The basic idea behind this approach is that a comprehensive coverage of crystallization space is achieved by using each condition only once. The 90 stock reagents used are classified into the following four classes: precipitants, buffers, additives, and detergents, with a chemical basis set of 50 distinct reagents.[4] Based on frequency and success of previous crystallization experiments, the TB group determined that 288 crystallization trials are normally sufficient to identify successful crystallization conditions for targets with natural propensities to crystallize.

Interestingly, the results of 203 successful random sampling experiments indicate that components besides those present in most commercially available sparse matrix screens are also very effective for promoting crystallization. Reagent frequency results show that additives and detergents such as lauryldimethylamine-*N*-oxide (LDAO), glycerol, dimethyl sulfoxide (DMSO), and β-octyl glucoside (BOG) effectively promote crystallization. The most successful components were LDAO and glycerol, both with frequencies of 28, with the remaining components, including common reagents such as PEGs and buffers, present at frequencies of 21 or less. Although the concentrations of LDAO and glycerol used in these experiments are not listed, earlier descriptions of the random sampling protocol indicate these reagents are normally present in small amounts, such as 0.5% (v/v) [35]. These reagents are not present at additive concentration amounts in most sparse matrix conditions. LDAO, in particular, was not present in any of the conditions used by either the JCSG or the Toronto groups for their initial crystallization screening trials. The incorporation of these additives to any screening strategy, especially with the use of minimal screens, may enable previously uncrystallizable proteins to form crystals suitable for structure determination.

4.4. SECSG: incomplete factorial screens identify initial crystallization leads more successfully than traditional sparse matrix screens

In parallel to the TB group, DeLucas et al. (2003; SECSG group) compared an alternative screen strategy, incomplete factorial screens, with sparse matrix screens to determine the relative crystallization efficiency of the two methods [31]. Unlike random sampling, or sparse matrix screens, the incomplete factorial screen is designed to enable the most important factors and interactions that facilitate crystal formation to be identified from an initial set of screening experiments [28]. The incomplete factorial screen used in this study was designed to sample six anionic precipitants 60 times each, with the binary combinations balanced and the higher order combinations randomly distributed. Five proteins were then subjected to crystallization trials using both this incomplete factorial screen (360 experiments) and a sparse matrix screen made up of commercially available conditions (290 experiments).

The results showed that the incomplete factorial screen was consistently more successful at identifying initial crystallization leads than those of the sparse matrix screens, with the factorial screens providing between 1 and 7 crystal hits for each of the 5 proteins screened, and the sparse matrix screen producing crystals for only 2 of them. A detailed analysis of these results and the identification of those components which were most successful for promoting crystal formation, are eagerly anticipated.

5. Similarities and differences in crystallization data mining results

The availability of these recent efforts to optimize crystallization screens and screening strategies make detailed comparisons between the results possible (Table 5). The similarities and differences observed between the studies will highlight those experimental parameters, besides the crystallization conditions, which are also critical for crystal formation. These comparisons will also indicate which conclusions are generally applicable, and which apply only to targets of a specific type or prepared using specific protocols. There are similarities between the studies. For example, all results indicate that many of the sparse matrix conditions oversample distinct regions of chemical space, especially PEG/ion conditions. In addition, these studies also show that the current screens are still insufficient for crystallizing all targets; additional protein characteristics must also be evaluated and altered in a high-throughput manner if the number of targets processed for crystal formation by SG consortia is to become more efficient. There are also a number of differences between the studies. Two groups, the JCSG and TB groups, found that glycerol has different effects on crystallization behavior. In addition, the exact identity of the conditions in the Core and Minimal Screens of the JCSG and Toronto groups, respectively, differs. Thus, while these data mining studies are still extremely useful for guiding future crystallization efforts, they clearly indicate that more data must be collected if optimal, reproducible strategies for the crystallization of a wide range of targets are to become a reality.

[4] A computer program, CRYSTOOL, can be accessed at porter.llnl.gov/crystool4.1 for the generation of random screens based on input parameters and frequencies defined by the user.

Table 5
Strengths and weaknesses of recent efforts to archive and mine crystallization data

Dataset	Database characteristics	Strengths	Weaknesses
BMCD	• 2526 proteins • 3547 conditions which produced diffraction quality crystals	• Most comprehensive database of the conditions which have produced diffraction quality protein crystals • Detailed information available for each target	• No negative data; those conditions which failed to produce crystals are not reported • Protein samples prepared very differently from one another
JCSG	• 539 proteins screened (465 crystallized) • 480 conditions at 20 °C used for screening (472 produced crystals for at least 1 protein)	• Positive + negative crystallization data • All proteins (from a single organism) purified and screened for crystallization using identical protocols • Largest number of distinct crystallization conditions screened	• Only commercially available conditions tested
Toronto	• 755 proteins screened (338 crystallized) • 48 conditions used for screening (all 48 produced crystals for at least 1 protein)	• Positive + negative crystallization data • Most proteins (from six organisms) prepared and screened for crystallization using identical protocols • Largest number of distinct proteins screened	• Only commercially available conditions tested
Queensland	• 19 proteins (18 crystallized) • 288 conditions tested	• Positive + negative crystallization data • Comparison of two sets of commercially available screens which are listed as having identical crystallization solutions	• Most protein samples prepared very differently from one another • Small sample size
TB	• 288 conditions screened for most proteins • 203 successful random screening trials used for initial frequency results	• Positive + negative crystallization data • Random Matrix crystallization solutions designed to include additive/detergents in addition to traditional buffer/precipitants	• Awaiting detailed crystallization data mining results
SEGSG	• 5 proteins screened • 290 sparse matrix conditions compared to 360 incomplete factorial conditions	• Positive + negative crystallization data • Comparison of two crystallization strategies (sparse matrix screens and an incomplete factorial screen)	• Awaiting detailed crystallization data mining results
YSBL	• 10 proteins crystallized • 48 conditions screened	• Comparison of two crystallization strategies (sparse matrix and novel crystallization screens developed by the authors)	• Only successful results reported • Detailed information about sample preparation of all proteins not reported

5.1. Similarities: many sparse matrix crystallization conditions, especially those containing PEGs, are redundant

All data mining efforts reviewed here indicate many of the conditions commonly used in sparse matrix screens are redundant and can be eliminated from initial crystallization screens. The JCSG results showed that the greatest number of distinct proteins (358/456) crystallized in conditions containing high MW PEGs, while the fewest (210/456) crystallized in conditions containing organics [34]. When the number of distinct proteins was normalized to the number of conditions, organic precipitants crystallized the largest number of distinct proteins per condition tested, whereas high MW PEG precipitants produced the fewest, indicating many of the high MW PEG conditions were redundant. Similar to the JCSG, the Toronto group found that the greatest number of distinct proteins (229/338) crystallized in conditions containing high MW PEGs [33]. In fact, the six most productive conditions for promoting crystal formation contained high MW PEGs. However, many of the proteins which crystallized in one of these six conditions often crystallized in another, indicating these high MW PEG conditions were redundant and could also be eliminated with only a minimal impact on the final number of proteins crystallized.

The number and types of conditions that were eliminated during the identification of minimal screens further illustrate the oversampling of distinct regions of crystallization space. Conditions containing high MW PEGs made up the highest percentage of conditions in both the JCSG Core Screen (31 of 67, 46.3%) and the Toronto minimal screen (4 of 6, 66.7%), illustrating their efficacy for promoting crystallization of a wide range of targets (Fig. 2). However, the total numbers of high MW PEG conditions in the original screens were highly redundant, as 140 (of 171; 81.9%) and 15 (of 19; 79.0%) were eliminated from the original JCSG and Toronto screens, respectively, once the minimal screens were identified. Salts were the second most redundant component, with 96 (of 106; 90.6%) and 15 (of 17; 88.2%) eliminated from the JCSG and Toronto screens, respectively.

5.2. Similarities: current screens are still insufficient for the successful crystallization of all protein targets

While many of the proteins screened in these studies readily crystallized in one or more conditions, many did not. Specifically, 76 (14.1%) of the JCSG proteins and

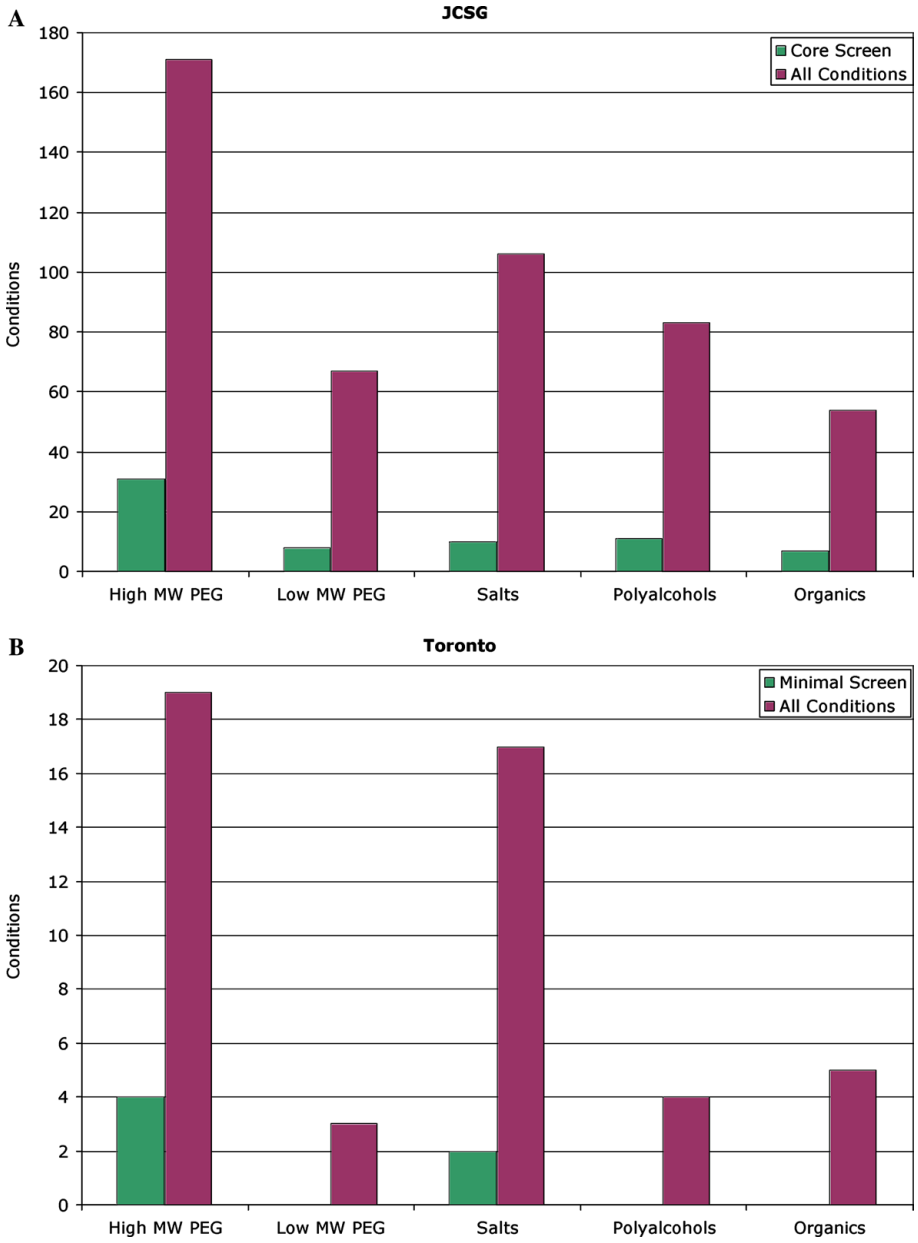


Fig. 2. Conditions of the original and minimal screens identified by the JCSG and Toronto groups categorized by precipitant type. Conditions classified by their major precipitant: high MW PEG, low MW PEG, salts, polyalcohols, and organics. Conditions used for screening in magenta, those included in minimal screens in green. (A) JCSG: 480 conditions screened, 67 make up the Core Screen. (B) Toronto: 48 conditions screened, 6 make up Minimal Screen 6. The *y*-axis scale reflects the number of conditions used by the JCSG and Toronto groups for their crystallization studies.

417 (55.2%) of the Toronto proteins failed to form crystals, in spite of the extensive efforts made by both groups to crystallize these targets. The inability of these proteins to form diffraction quality crystals may have been due to any number of causes. First, they may not have been screened against their optimal crystallization conditions. Since most sparse matrix screen conditions are heavily biased towards those that have previously produced protein crystals, it is expected that some protein families that have never crystallized before may require totally novel conditions for crystal formation. For some of these targets, a random sampling screen [35,41], factorial screen [28,31] or novel screen [27] might certainly be more appropriate. In addition, there are a number of other parameters besides the nature of the crystallization conditions that can be manipulated to maximize the likelihood a target will crystallize. For example, the proteins that failed to crystallize in these studies may have required binding partners or small molecule cofactors for folding and stability, or the protein sequences may have included purification tags or unstructured loops and termini that prohibited crystal formation. Experimentally introduced point mutations have been used to crystallize or improve existing

Table 6
Conditions from the Hampton Crystal Screen (same as Jancarik–Kim Sparse Matrix Screen [32]) that produced the most crystals for the Toronto, JCSG, and Queensland studies

J–K/Hampton CS 1	Conditions which produced the maximum number of crystals	Crystal hits	Study
6	30% PEG 4000; 0.2 M $MgCl_2$; 0.1 M Tris–HCl, pH 8.5	65	Toronto
9	30% PEG 4000; 0.2 M NH_4 acetate; 0.1 M Na citrate, pH 5.6	76	Toronto
17	30% PEG 4000; 0.2 M Li sulfate; 0.1 M Tris–HCl, pH 8.5	70	Toronto
18	20% PEG 8000; 0.2 M Mg acetate; 0.1 M Na cacodylate, pH 6.5	72	Toronto
22	30% PEG 4000; 0.2 M Na acetate; 0.1 M Tris–HCl, pH 8.5	65	Toronto
28	30% PEG 8000; 0.2 M Na acetate; 0.1 M Na cacodylate, pH 6.5	65	Toronto
1	30% MPD; 0.02 M $CaCl_2$; 0.1 M Na acetate, pH 4.6	37	JCSG
20	25% PEG 4000; 0.2 M NH_4 sulfate; 0.1 M Na acetate, pH 4.6	23	JCSG
37	8% PEG 4000; 0.1 M Na acetate, pH 4.6	22	JCSG
40	20% Isopropanol/20% PEG 4000; 0.1 M citrate, pH 5.6	19	JCSG
41	10% Isopropanol + 20% PEG 4000; 0.1 M Na HEPES,pH 7.5	23	JCSG
42	20% PEG 8000; 0.05 M K phosphate	19	JCSG
1	30% MPD; 0.02 M $CaCl_2$; 0.1 M Na acetate, pH 4.6	6	Queensland
14	28% PEG 400; 0.2 M $CaCl_2$; 0.1 M Na HEPES, pH 7.5	10	Queensland
17	30% PEG 4000; 0.2 M Li sulfate; 0.1 M Tris–HCl, pH 8.5	7	Queensland
24	20% Isopropanol; 0.2 M $CaCl_2$; 0.1 M acetate, pH 4.6	5	Queensland
30	30% PEG 8000; 0.2 M NH_4 sulfate	6	Queensland

crystals of a number of proteins [46–48] as have truncations and deletions to remove disordered domains [49,50]. A new technique to identify unstructured regions in proteins, deuterium exchange mass spectrometry (DXMS), has also recently been used to experimentally define to the structural domain boundaries of a *T. maritima* protein, TM0160, which failed to crystallize as a full-length protein. When a deletion mutant of this protein was expressed with only the ordered domain, it crystallized readily and the structure was determined to 1.9 Å resolution [34].

These examples illustrate that efforts to improve the efficiency of the crystallization process must focus not only on the crystallization conditions themselves, but also on strategies to efficiently alter protein characteristics to enhance their propensity to crystallize. In particular, studies that parallel those described here must be carried out to determine which protein characteristics correlate with crystallization success and failure [51]. These studies can then be used to develop new, high-throughput methods to experimentally alter these characteristics prior to initial screening in order to maximize their crystallization potential.

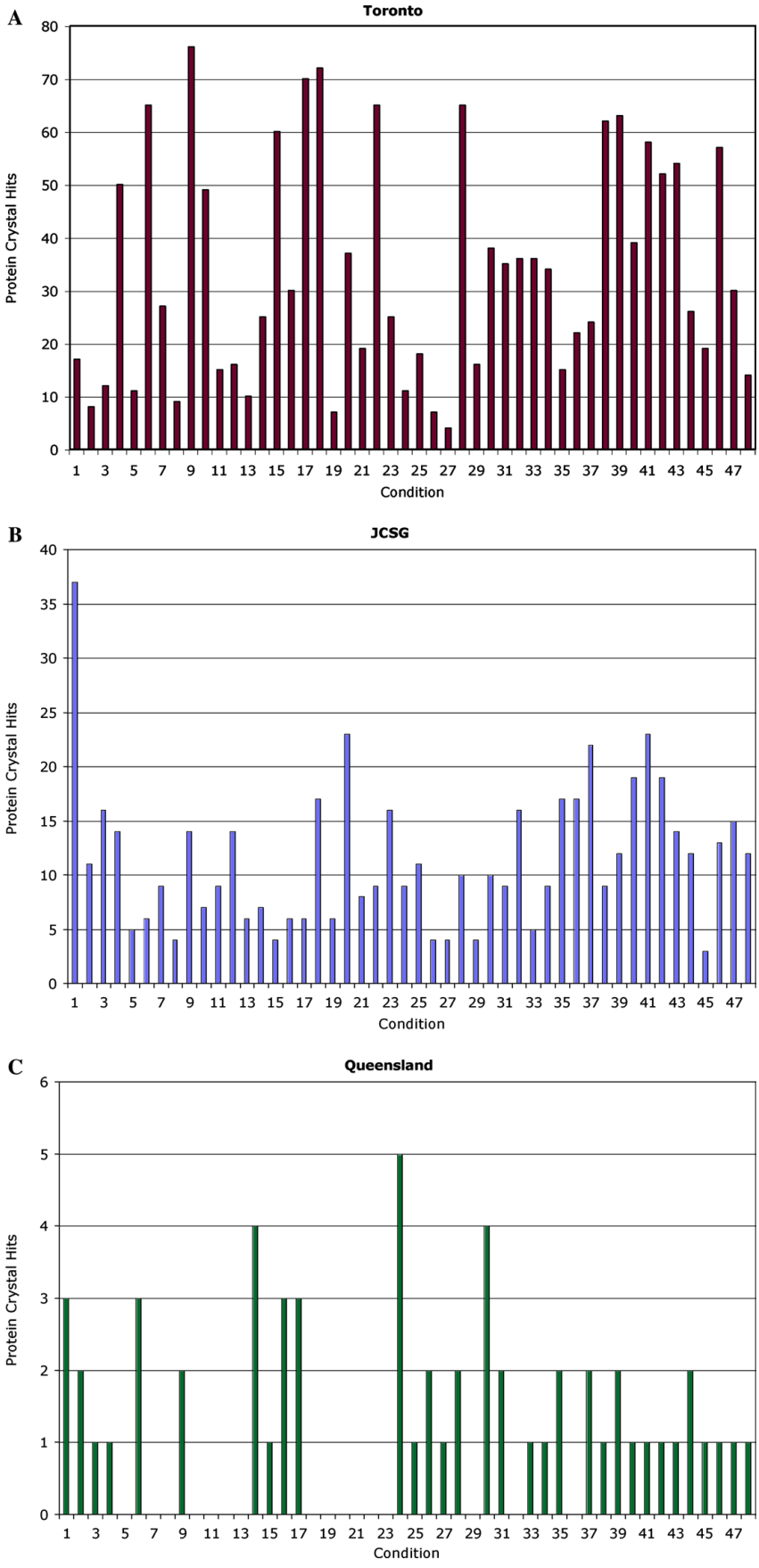
5.3. Differences: is glycerol a facilitator or inhibitor of crystal formation?

It is interesting to note that glycerol was the most frequent reagent of the 203 successful random crystallization trials reported by TB group [52]. This contrasts with the results of the JCSG study in which glycerol was found to inhibit, not facilitate, crystallization [34]. These observations, however, are not mutually exclusive. In the JCSG study, conditions with glycerol normally contained it in concentrations sufficient for cryoprotection (up to 35% (v/v)). The concentration of glycerol in the 203 random sampling experiments of the TB group was not reported, but earlier descriptions of the random screening protocol indicate the concentrations of glycerol were probably significantly lower, closer to 0.5% (v/v). This volume is similar to that used in traditional Additive Screens (Hampton Additive Screen 1, 3% (v/v)). Thus, lower concentrations of glycerol may in fact promote rather than inhibit crystal formation, and its use as an additive in commonly used crystallization conditions, especially those in minimal screens, may enable the targets which were unable to crystallize in these screens to crystallize.

5.4. Differences: different minimal screens are identified depending on the SG data analyzed

The most significant difference observed between the studies reviewed here is the minimal correlation between the results of the JCSG group and those of the Toronto group, especially with respect to the exact identity of the conditions in the Minimal Screen (Toronto) and the Core Screen (JCSG). The majority of conditions, however, are still quite similar in the nature of the precipitants (for example, many are PEG/ion conditions) and thus both screens can be used to identify targets with a natural propensity to crystallize. Of the 539 *T. maritima* proteins screened using the JCSG two-tiered strategy, only 48% (222/465) crystallized in the 48 conditions used by the Toronto group for their screening experiments. This percentage is similar to the overall success rate observed by the Toronto group of 45% (338/755).

What is more interesting, however, is that the identity of the conditions determined to be the most productive for crystal formation differed between the two groups when the JCSG crystallization results were limited to only those conditions used for the Toronto study (Hampton Crystal Screen, conditions 1–48). The conditions that


A
Toronto
Protein Crystal Hits
Condition
B
JCSG
Protein Crystal Hits
Condition
C
Queensland
Protein Crystal Hits
Condition

crystallized the greatest number of proteins for the Toronto group were 6, 9, 17, 18, 22, and 28 (Table 6) while those that crystallized the greatest number of total proteins for the JCSG group were 1, 20, 37, 40, 41, and 42. There is no overlap between the two sets. This lack of overlap extends to the entire screen (Fig. 3); the correlation coefficient of the number of crystal hits per condition is only 0.11 between all 48 conditions. In addition, since the Queensland group also used this screen for their crystallization studies, their results, although representative of a much smaller protein sample set, can also be compared. The Queensland study identified a third set of most productive conditions including 1, 14, 17, 24, and 30. Thus, only two conditions, Hampton Crystal Screen conditions 1 and 17, were identified by more than one group; the rest differed.

While it may be surprising that, among all of these studies, there was no single condition that was consistently better than another for promoting crystal formation, it likely reflects the differences between these studies in the methods used for the crystallization experiments. Like the JCSG, the Toronto group used a single purification step (MC affinity chromatography) for protein purification. For at least the *M. thermoautotrophicum* samples, however, an additional heating step was also used prior the MC purification, contributing to an increase in the final purity of these samples (99%, compared to an average 90–95% purity observed for JCSG samples). In addition, in some cases, the N-terminal tags of the Toronto proteins were cleaved prior to crystallization trials, while those of the JCSG proteins were not. Equally important, the sample buffers of the proteins differed between the two groups. The protein buffer used in the Toronto study had a higher concentration of salt (500 mM NaCl, Toronto vs. 150 mM NaCl, JCSG) and different buffer (20 mM HEPES, pH 7.5, Toronto vs. 20 mM Tris, pH 7.9, JCSG) than that used in the JCSG study. Finally, the crystallization experiments were setup differently between the two groups. The Toronto group used 4 μl hanging drops (2 μl protein plus 2 μl mother liquor) in 24-well Linbro plates, while the JCSG proteins were screened using 100 nl sitting drops (50 nl protein plus 50 nl crystallization solution) in 96-well crystallization plates. Both types of crystallization experiments have different kinetic properties for crystal drop equilibration and crystal formation (Fig. 1) [45,53]. Finally, the Toronto proteins were screened at 2–4 distinct protein concentrations, while the JCSG proteins were screened at only a single protein concentration. The samples in the Queensland experiment were prepared differently as well, even among the 19 samples screened within the study. Some were purchased and solubilized in crystallization buffer (25 mM Tris, pH 7.0; 10 mg/ml protein concentration) while the rest were purified and equilibrated in protein specific buffers and concentrated between 8.6 and 27 mg/ml. The crystallization trials were also setup differently from both the JCSG and Toronto studies, using a 1 μl protein plus 1 μl crystallization solution hanging drop equilibrated against 100 μl of well solution.

Each of these factors—purity, the presence or absence of purification tags, protein buffer, crystallization setup methods, and protein concentration—is known to contribute significantly to the protein crystallization process [23,54] and is likely to have contributed to many of the differences in the crystallization results observed between these groups. As more studies are carried out, the effects of these factors on crystallization success will be able to be evaluated. For now, however, it seems that identified minimal screens are still somewhat correlated to the methods used for sample preparation and screening and thus one must take these additional preparation methods into consideration when deciding which screen might be appropriate for their own experiments. In addition, however, one must also consider that these screens were developed as initial screens with the purpose of identifying proteins with a natural propensity to crystallize and not necessarily to result in diffraction quality crystals. Thus, although the exact nature of the condition differs between the two minimal screens, both sets of conditions are expected to successfully identify proteins which are amenable to further crystallization studies.

6. Conclusions

It is clear from the results presented here that the ability of certain conditions to promote crystallization more effectively than others depends significantly on experimental parameters other than just the conditions themselves. In particular, the JCSG, Toronto, and Queensland groups each identified a unique set of conditions that were the most effective for promoting crystal formation from the same original set, reflecting the very different crystallization sample preparation methods used by each for their crystallization trials. As more data are made available from other SG groups, and as further

Fig. 3. Differences in SG crystallization data mining results. Number of proteins crystallized per condition of the Jancarik and Kim [32] (1991) sparse matrix screen by both the (A) Toronto and (B) JCSG and (C) Queensland groups. (A) Toronto: 755 proteins were screened for crystallization against 48 sparse matrix screen conditions, of which 338 crystallized. (B) JCSG: 539 proteins were screened for crystallization against the same set of conditions, of which 222 crystallized. (C) Queensland: 19 proteins were screened for crystallization, of which 18 crystallized. The conditions which were most successful for promoting crystallization differed substantially between the three groups, reflecting the differences in the protein preparation protocols, targets, and crystallization setup methods used by each to carry out these experiments. The *y*-axis scale reflects the number of proteins screened and crystallized by the different groups.

experiments are carried out, these screens will be able to be further optimized, and result in the identification of more robust screens.

One of the most important advantages of using minimal screens, however, is that additional crystallization parameters can be rapidly evaluated without increasing the total number of experiments. For example, one can begin to explore how the addition of additives such as LDAO and glycerol influences protein crystallization behavior. Multiple protein concentrations can also be easily screened, as can multiple crystallization set-up methods. Most importantly, the effect of altering the nature of the protein can also be efficiently and rapidly evaluated using these minimal screens. Using minimal screens in this way will ensure that a high resolution structure will be determined in the minimal amount of time using the fewest resources.

Finally, there are still a considerable number of studies that need to be carried out in order to optimally streamline the crystallization process. While the minimal screens can be used to identify those proteins with a natural propensity to crystallize, they have not yet been shown to provide an accurate indication of which condition will ultimately result in a diffraction quality crystal. It will be interesting to see how well these minimal screens correlate with the conditions that ultimately produce diffraction quality crystals. In addition, the characteristics of the proteins themselves, such as sequence, predicted secondary structure, predicted disordered regions, among others, also need to be analyzed in detail to determine whether or not there are protein sequence characteristics which correlate with crystallization efficiency. Ultimately, it is hoped that the crystallization screening process can become more predictive prior to crystallization trials so that a protein, based on its sequence alone, can more accurately be identified as one with a potential for crystal formation or not. The SG efforts being carried out today are hoped to enable the structure determination process to be streamlined to such an extent that the majority of future efforts will be spent not on identifying targets amenable to crystal formation, but instead on moving a protein of interest into a region of chemical space that will produce diffraction quality crystals. This, in turn, will enable more time to be spent on understanding biological protein function in atomic detail.

Acknowledgments

The authors thank Scott Lesley, Slawomir Grzechnik, and the rest of the JCSG for their contributions. This work was supported in part by National Institutes of Health Protein Structure Initiative Grant GM62411 from the National Institute of General Medical Sciences (www.nigms.nih.gov). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of US Department of Energy, Office of Basic Energy Sciences. The Stanford Synchrotron Radiation Laboratory Structure Molecular Biology Program is supported by Department of Energy, Office of Biological and Environmental Research, the National Institutes of Health, National Center for Research Resources, the Biomedical Technology Program, and the National Institute of General Medical Sciences.

References

[1] S.E. Brenner, Nat. Rev. Genet. 2 (2001) 801–809.
[2] S.K. Burley, Nat. Struct. Biol. 7 (Suppl.) (2000) 932–934.
[3] W.G. Hol, Nat. Struct. Biol. 7 (Suppl.) (2000) 964–966.
[4] R.C. Stevens, S. Yokoyama, I.A. Wilson, Science 294 (2001) 89–92.
[5] K. Das, R. Xiao, E. Wahlberg, F. Hsu, C.H. Arrowsmith, G.T. Montelione, E. Arnold, Proteins 45 (2001) 486–488.
[6] G.L. Gilliland, A. Teplyakov, G. Obmolova, M. Tordova, N. Thanki, J. Ladner, O. Herzberg, K. Lim, H. Zhang, K. Huang, Z. Li, A. Tempczyk, W. Krajewski, L. Parsons, D.C. Yeh, J. Orban, A.J. Howard, E.J.F.P. Eisenstein, N. Bonander, K.E. Fisher, J. Toedt, P. Reddy, C.V. Rao, E. Melamud, J. Moult, Curr. Drug Targets Infect. Disord. 2 (2002) 339–353.
[7] S.H. Kim, D.H. Shin, I.G. Choi, U. Schulze-Gahmen, S. Chen, R. Kim, J. Struct. Funct. Genomics 4 (2003) 129–135.
[8] V. Oganesyan, D. Busso, J. Brandsen, S. Chen, J. Jancarik, R. Kim, S.H. Kim, Acta Crystallogr. D 59 (2003) 1219–1223.
[9] M.J. Romanowski, J.B. Bonanno, S.K. Burley, Proteins 47 (2002) 563–567.
[10] R. Schwarzenbacher, J.M. Canaves, L.S. Brinen, X. Dai, A.M. Deacon, M.A. Elsliger, S. Eshaghi, R. Floyd, A. Godzik, C. Grittini, S.K. Grzechnik, C. Guda, L. Jaroszewski, C. Karlak, H.E. Klock, E. Koesema, J.S. Kovarik, A. Kreusch, P. Kuhn, S.A. Lesley, D. McMullan, T.M. McPhillips, M.A. Miller, M.D. Miller, A. Morse, K. Moy, J. Ouyang, A. Robb, K. Rodrigues, T.L. Selby, G. Spraggon, R.C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, X. Wang, B. West, G. Wolf, K.O. Hodgson, J. Wooley, I.A. Wilson, Proteins 53 (2003) 142–145.
[11] J. Symersky, G. Lin, S. Li, S. Qiu, Carson, M.N. Schormann, M. Luo, C.W. Goulding, A. Parseghian, M.R. Sawaya, D. Cascio, M.I. Apostol, M.L. Gennaro, D. Eisenberg, D. Christendat, V. Saridakis, Y. Kim, P.A. Kumar, X. Xu, A. Semesi, A. Joachimiak, C.H. Arrowsmith, A.M. Edwards, Proteins 53 (2003) 947–949.
[12] C. Zhang, S.H. Kim, Curr. Opin. Chem. Biol. 7 (2003) 28–32.
[13] Q. Zhao, Y.K. Chae, J.L. Markley, Biochemistry 41 (2002) 12284–12296.
[14] M.W. Adams, H.A. Dailey, L.J. DeLucas, M. Luo, J.H. Prestegard, J.P. Rose, B.C. Wang, Acc. Chem. Res. 36 (2003) 191–198.
[15] D. Christendat, A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J.R. Cort, V. Booth, C.D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K.L. Maxwell, N. Wu, L.P. McIntosh, K. Gehring, M.A. Kennedy, A.R. Davidson, E.F. Pai, M. Gerstein, A.M. Edwards, C.H. Arrowsmith, Nat. Struct. Biol. 7 (2000) 903–909.
[16] M.R. Chance, A.R. Bresnick, S.K. Burley, J.S. Jiang, C.D. Lima, A. Sali, S.C. Almo, J.B. Bonanno, J.A. Buglino, S. Boulton, H. Chen, N. Eswar, G. He, R. Huang, V. Ilyin, L. McMahan, U. Pieper, S. Ray, M. Vidal, L.K. Wang, Protein Sci. 11 (2002) 723–738.
[17] A.M. Edwards, C.H. Arrowsmith, D. Christendat, A. Dharamsi, J.D. Friesen, J.F. Greenblatt, M. Vedadi, Nat. Struct. Biol. 7 (Suppl) (2000) 970–972.

[18] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Proc. Natl. Acad. Sci. USA 99 (2002) 11664–11669.
[19] N.E. Chayen, E. Saridakis, Acta Crystallogr. D 58 (2002) 921–927.
[20] G.E. Dale, C. Oefner, A. D'Arcy, J. Struct. Biol. 142 (2003) 88–97.
[21] D. Hennessy, V. Gopalakrishnan, B.G. Buchanan, J.M. Rosenberg, D. Subramanian, Proc. Int. Conf. Intell. Syst. Mol. Biol. 2 (1994) 179–187.
[22] A. McPherson, Methods Enzymol. 114 (1985) 120–125.
[23] A. McPherson, Eur. J. Biochem. 189 (1990) 1–23.
[24] A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1994.
[25] A. McPherson, Protein Sci. 10 (2001) 418–422.
[26] R. Giegi, B. Lorber, A. Theobald-Dietrich, Acta Crystallogr. D D50 (1994) 339–350.
[27] A.M. Brzozowski, J. Walton, J. Appl. Crystallogr. 34 (2001) 97–101.
[28] C.W. Carter Jr., C.W. Carter, J. Biol. Chem. 254 (1979) 12219–12223.
[29] C.W. Carter Jr., Y. Yin, Acta Crystallogr. D 50 (1994) 572–590.
[30] B. Cudney, S. Patel, K. Weisgraber, Y. Newhouse, A. McPherson, Acta Crystallogr. D 50 (1994) 414–423.
[31] L.J. DeLucas, T.L. Bray, L. Nagy, D. McCombs, N. Chernov, D. Hamrick, L. Cosenza, A. Belgovskiy, B. Stoops, A. Chait, J. Struct. Biol. 142 (2003) 188–206.
[32] J. Jancarik, S. Kim, J. Appl. Crystallogr. 24 (1991) 409–411.
[33] M.S. Kimber, F. Vallee, S. Houston, A. Necakov, T. Skarina, E. Evdokimova, S. Beasley, D. Christendat, A. Savchenko, C.H. Arrowsmith, M. Vedadi, M. Gerstein, A.M. Edwards, Proteins 51 (2003) 562–568.
[34] R. Page, S.K. Grzechnik, J.M. Canaves, G. Spraggon, A. Kreusch, P. Kuhn, R.C. Stevens, S.A. Lesley, Acta Crystallogr. D 59 (2003) 1028–1037.
[35] B.W. Segelke, J. Cryst. Growth 232 (2001) 553–562.
[36] S. Radaev, P.D. Sun, J. Appl. Crystallogr. 35 (2002) 674–676.
[37] Q.-q. Huang, M.-k. Teng, L.-w. Niu, Acta Crystallogr. D 55 (1999) 1444–1448.
[38] G.L. Gilliland, M. Tung, J. Ladner, J. Res. Natl. Inst. Stand. Technol. 101 (1996) 309–320.
[39] G.L. Gilliland, M. Tung, J.E. Ladner, Acta Crystallogr. D 58 (2002) 916–920.
[40] J.W. Wooh, R.D. Kidd, J.L. Martin, B. Kobe, Acta Crystallogr. D 59 (2003) 769–772.
[41] B. Rupp, J. Struct. Biol. 142 (2003) 162–169.
[42] D. Hennessy, B. Buchanan, D. Subramanian, P.A. Wilkosz, J.M. Rosenberg, Acta Crystallogr. D 56 (Pt. 7) (2000) 817–827.
[43] C.T. Samudzi, M.J. Fivash, J.M. Rosenberg, J. Cryst. Growth 123 (1992) 47–58.
[44] J.T. Bukrinsky, J.-C.N. Poulsen, J. Appl. Crystallogr. 34 (2001) 533–534.
[45] B.D. Santarsiero, D.T. Yegian, C.C. Lee, G. Spraggon, J. Gu, D. Scheibe, D.C. Uber, E.W. Cornell, R.A. Nordmeyer, W.F. Kolbe, J. Jin, A.L. Jones, J.M. Jaklevic, P.G. Schultz, R.C. Stevens, J. Appl. Crystallogr. 35 (2002) 278–281.
[46] K. Braig, Z. Otwinowski, R. Hegde, D.C. Boisvert, A. Joachimiak, A.L. Horwich, P.B. Sigler, Nature 371 (1994) 578–586.
[47] A. D'Arcy, M. Stihle, D. Kostrewa, G. Dale, Acta Crystallogr. D 55 (Pt 9) (1999) 1623–1625.
[48] D.M. Lawson, P.J. Artymiuk, S.J. Yewdall, J.M. Smith, J.C. Livingstone, A. Treffry, A. Luzzago, S. Levi, P. Arosio, G. Cesareni, et al., Nature 349 (1991) 541–544.
[49] G.E. Dale, D. Kostrewa, B. Gsell, M. Stieger, A. D'Arcy, Acta Crystallogr. D 55 (Pt 9) (1999) 1626–1629.
[50] G.Q. Chen, Y. Sun, R. Jin, E. Gouaux, Protein Sci. 7 (1998) 2623–2630.
[51] C.S. Goh, N. Lan, S.M. Douglas, B. Wu, N. Echols, A. Smith, D. Milburn, G.T. Montelione, H. Zhao, M. Gerstein, J. Mol. Biol. 336 (2004) 115–130.
[52] V. Reddy, S.M. Swanson, B. Segelke, K.A. Kantardjieff, J.C. Sacchettini, B. Rupp, Acta Crystallogr. D 59 (2003) 2200–2210.
[53] D. Hosfield, J. Palan, M. Hilgers, D. Scheibe, D.E. McRee, R.C. Stevens, J. Struct. Biol. 142 (2003) 207–217.
[54] A. McPherson, Methods Enzymol. 114 (1985) 112–120.