Bioinformatics Research and Applications
Volume 9096 of the series Lecture Notes in Computer Science pp 84-95. The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-19048-8_8.

Protein Crystallization Screening using Associative Experimental Design

İmren Dinç[1], Marc L Pusey[2], and Ramazan S. Aygün[1]

[1] Computer Science Department, University of Alabama in Huntsville, Huntsville, Alabama 35899, United States
id0002@uah.edu, marc.pusey@ixpressgenes.com, aygunr@uah.edu,
[2] iXpressGenes, Inc., 601 Genome Way, Huntsville, Alabama 35806, United States

Abstract. Protein crystallization remains a highly empirical process. The purpose of protein crystallization screening is the determination of the main factors of importance leading to protein crystallization. One of the major problems about determining these factors is that screening is often expanded to many hundreds or thousands of conditions to maximize combinatorial chemical space coverage for a successful (crystalline) outcome. In this paper, we propose a new experimental design method called "Associative Experimental Design (*AED*)" that provides a list of screening factors that are likely to lead to higher scoring outcomes or crystals by analyzing preliminary experimental results. We have tested *AED* on Nucleoside diphosphate kinase, HAD superfamily hydrolase, and nucleoside kinase proteins derived from the hyperthermophile Thermococcus thioreducens. After obtaining the candidate novel conditions, we have confirmed that *AED* method yielded high scoring crystals after experimenting in a wet lab.

Keywords: Associative Experimental Design, Protein Crystallization, Screening, Experimental Design

1 Introduction

Protein crystallization is the process of formation of 3-dimensional structure of a protein. One of the significant difficulties in macromolecular crystallization is setting up the parameters that yield a single large crystal for X-ray data collection [2], [3]. The major difficulty in this process is the trial of abundance of parameters with significant number of possible values. Physical, chemical and biochemical factors such as type of precipitants, type of salts, ionic strengths, pH value of the buffer, temperature of the environment, and genetic modifications of the protein affect the crystallization process significantly [4]. Because each protein has a unique primary structure, it is quite challenging to predict the parameters of the experiment that can yield crystal for a specific protein [4]. Theoretically, it is possible to crystallize a protein in a specific solution under certain conditions; however, it may not be possible to crystallize in practice [5].

This means that we can generate hundreds or thousands of conditions to maximize combinatorial chemical space coverage hoping for a crystalline outcome. However, setting up huge number of experiments is not feasible in terms of cost and time.

Basically, there are two main approaches to set parameters for protein crystallization experiments [6], [7]: 1) incomplete factorial design (IFD) [8], [9] or sparse matrix sampling (SMS) [3], [10], and 2) grid screening (GS) of crystallization conditions [4]. The first approach has been widely used by commercial companies such as Hampton Research, Emerald Biostructures, etc. [6]. Carter et al. proposed to set parameters of protein crystallization experiments using incomplete factorial design in their study [8]. The main goal of incomplete factorial design experiments is to identify important factors of the experiments and to produce much less number of experiments compared to full factorial design experiments. The IFD is a very effective method as experts may not afford to set up many experiments or they may not have enough resources to carry out those many experiments [11]. The basic idea of IFD is that after identifying important factors of the experiments; balanced experiments are generated in terms of factors. In the sparse matrix sampling [3], parameters of the experiments are set using fewer major reagents (i.e., pH values, type of precipitants, type of salts, etc.) as in IFD. It can be considered as an optimized version of IFD. In SMS, values of type of salts, pH, and type of precipitants are selected based on past experiences, and these variables are mostly favorable for protein crystallization experiments. The reagents occur based on their frequency in the sparse matrix [10]. This idea was commercialized by Hampton Research [12]. Grid screening of crystallization conditions [4] is an early method that tries possible different solutions exhaustively until the experiments succeed. This takes more time and effort compared to IFD and SMS. However, it could be the only solution for some of the proteins that have never or rarely been crystallized.

In the literature, there are also other optimization methods based on IFD and GS [2], [11]. We do not intend to give detail about these methods in this paper. These studies in macromolecular crystallization generally try to optimize available conditions, which are supposed to yield crystals. Most of these conditions are commercially available. In Mimer [7], it may be possible to change the pH value of the buffer and the weight of the precipitant rather than just changing one value at a time by visualizing the result. The traditional optimization techniques do not consider new combinations of reagents. In this paper, we propose a new experimental design method called Associative Experimental Design (AED) that generates candidate conditions by analyzing preliminary experimental data. This existing data is analyzed to determine which screening factors are most likely to lead to higher scoring outcomes, crystals. Unlike IFD, AED generates unbalanced experiments for protein crystallization that may include novel conditions. This means AED is not a typical optimization method for crystallization conditions. In the literature, optimization steps usually include changing the pH value, concentration, weight of precipitants and salts. The AED method finds small but effective number of conditions that may lead

to crystallization. The main idea of the *AED* method is to generate novel conditions for crystallization by keeping at least two reagents from promising conditions. Basically, the *AED* analyzes other possible interactions between reagents to determine new crystallization conditions. In this study, we have generated candidate conditions for Nucleoside diphosphate kinase, HAD superfamily hydrolase, and nucleoside kinase proteins using preliminary experimental results. After obtaining the candidate novel conditions, we have confirmed that *AED* method yielded high scoring crystals after experimenting in a wet lab.

The rest of the paper is structured as follows. Background information is provided in Section 2. The proposed method, "Associative Experimental Design (*AED*)," is explained in Section 3. Experimental results are provided in Section 4. Finally, our paper is concluded with the last section.

2 Background

In this section, we provide some information about the phase diagram, which is a useful diagram for setting up protein crystallization experiments. We develop *AED* based on the phase diagram, and we believe that a brief explanation of it would help reader to understand the problem domain and our method. In addition, we are going to provide some brief explanation of Hampton scores in this section, since we are going to refer those scores throughout the paper.

2.1 Phase Diagram

Normally, a protein is going to dissolve in a liquid up to a point such that the solvent will not be able to solve any more protein. At that point, the solution is said to be in equilibrium, and that concentration of the solution is called solubility limit [4]. If the concentration of a solution is below the solubility limit, then that solution is said to be undersaturated; if it is exactly on the solubility limit, then it is called saturated. When the solution reaches the solubility limit, it is possible to increase its solubility by changing some physical factors such as pH, temperature, etc. If the concentration of the solution is above the solubility limit, then the solution will be supersaturated. This is the only region that a protein crystal can be grown [4]. However, supersaturated solution is not enough for crystallization process by itself. A specific amount of activation energy and some rare sequence of chemical reactions require initiating nucleation of protein crystallization that eventually yields a protein crystal [5], [4].

In chemistry, a phase diagram is a graphical representation of different phases (solid, liquid and gas) of a substance with respect to temperature and pressure. In structural biochemistry, a phase diagram mostly represents solubility curve of a protein with respect to some parameters such as precipitant, pH, etc. Since the proteins can grow only in supersaturated solutions, it is important to locate solubility curve based on these parameters [5], [13]. Thus, the phase diagram is useful to set parameters for the experiments properly for X-ray diffraction studies [14]. Figure 1 shows a visual representation of a phase diagram.

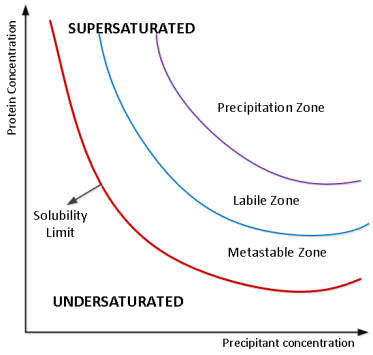

Fig. 1. Phase Diagram.

The phase diagram mainly has two main zones: undersaturated region and supersaturated region. The supersaturated region consists of three subdivisions as can be seen in Figure 1. The first region is labile zone, where nuclei of protein crystals can form and continue growing its structure if the certain conditions are provided. Once the nucleation starts, protein crystals start using the nutrients of the solution, which will reduce the concentration of the solution. While the concentration of the solution reduces, the solution will be in metastable region. In this region, protein crystal may continue to grow up to its concentration equal to the solubility limit, if there are nuclei that have formed before. In other words, new nuclei cannot form in that region [4]. If the supersaturation is too high, amorphous precipitates can also appear in precipitation zone instead of crystals, which is not a desirable outcome for crystallization process [5]. Furthermore, the amorphous precipitates do not yield crystals, when they complete their formations. Since nucleation can only occur in labile zone, *AED* focuses the conditions that fall into that region. Detailed information will be provided in Section 3.

2.2 Hampton Scoring

Hampton scoring is used to evaluate the growth of the protein during the crystallization experiments. In Hampton scoring, there are 9 scores from 1 to 9. In most of the experiments, a score that is greater than 7 is desired by the crystallographers, although scores between 5-7 are also classified as crystals. It should be noted that mostly crystals that have either score 8 or 9 are able to provide sufficient information about their 3D structures. Table 1 shows the list of Hampton scores. The brief explanations of some scores [1] are provided below.

[1] http://hamptonresearch.com/tips.aspx

Table 1. List of Hampton scores.

Score	Outcome	Score	Outcome
1	Clear solution	6	1D needles
2	Phase change (oiling out)	7	2D plates
3	Regular granular precipitate:	8	3D crystals small, < 200m
4	Birefringent precipitate or bright spots	9	3D crystals large, > 200m
5	Spheroids, dendrites, urchins		

Clear Drop: Clear drop means that the condition falls into either undersaturated region or metastable region in the phase diagram. This region also indicates that either protein or protein concentration (or both) is not enough to initiate nucleation.

Phase Separation: This phase occurs when the protein concentration is too high such that the protein is separated from the entire solution. Altering pH value or temperature could be one way to initiate nucleation.

Regular Granular Precipitate: This phase falls into precipitation zone in the phase diagram, so the condition yields aggregates, which is not a favorable for crystallization. Altering the precipitant concentration may be one solution to have a crystal.

Birefringent Precipitate or Microcrystal: This score is a good starting point for optimization, since it falls around line between liable and precipitation region in the phase diagram. This score is a good candidate to have successful crystallization.

Posettes or Spherulites: This outcome indicates that our condition fell into the liable region in the phase diagram, where nucleation can be initiated, and it continues to grow until the solution reaches the solubility limit. Even if this score is categorized as a crystal, it is not appropriate for X-ray diffraction. Therefore, this score is a good candidate for optimizations.

1D Needles: This score yields one dimensional needle like crystals, which is not enough for X-ray diffraction as well. Similar to posettes, these conditions can also be optimized to have higher scores. This crystal also starts growing in liable region in the phase diagram and completes its formation.

2D Plates: This category yields two dimensional quadrangular crystals. Similar to other crystals, this type of crystal starts growing in liable region. Since plates have 2 dimensions, this category may be used for further optimizations as well.

3D Small($< 200\mu m$) & Large($> 200\mu m$) Crystals: 3D crystals category yields three dimensional crystals. These crystals are generally appropriate to study protein structures. Thus, optimization is not necessary for this category.

In this study, we focused on scores from 4 to 7 to generate novel conditions using *AED* method. The details about *AED* are provided in Section 3.

3 Proposed Method: Associative Experimental Design(AED)

3.1 Motivation

In this research, we have generated some crystal screens for a few specific proteins using preliminary crystal screen data with their Hampton scores. We use 4 different proteins to test our approach. There are 86 different crystal screens in our dataset for the protein $Tt189$ without considering the conditions having multiple types of salts or precipitants. This data set contains 9 different salt concentration values, 23 different type of salts, 7 different pH values, 45 different precipitant concentration values, 85 different precipitants, 3 different protein concentration values, where the concentrations and pH values are continuous data and the other features are categorical data. (Note that type of buffer is not considered, since it is generally correlated with pH value.) If we consider full factorial design, it means that we need to set up approximately 16,627,275 different experiments for a single protein based on this dataset without considering the continuity of some of the variables and this is not feasible. In this research, our goal is to generate less number of conditions rather than 16.6M that is more likely to form a crystal. To achieve this goal, we proposed a method called "Associative Experimental Design (AED)."

3.2 Method

Associative experimental design generates a new set of experiments by analyzing the scores of experiments already evaluated in the lab. We use almost the same scores from 1 to 9 provided in Table 1. Since we are using trace fluorescent labeling (TFL) [15], a score of 4 is assigned to outcomes giving "bright spot" lead conditions as an exceptional case.

We start with the notation for screening conditions and scores. Let

$$D = \{(C_i, H_i) \mid (C_1, H_1), (C_2, H_2), ..., (C_n, H_n)\} \tag{1}$$

be our dataset consisting of the pairs that include features of the conditions C_i and their scores H_i for the i^{th} solution in the dataset. For simplicity we discarded conditions that have more than one type of salt or precipitant. We only focused on three main components of the remaining conditions: type of precipitant, type of salt and pH value of the solution, while separating their concentrations. Let

$$C_i = \{S_i\,[sc_i], pH_i, P_i\,[pc_i]\} \tag{2}$$

be the set of all the features of i^{th} crystal screen where i is $1 \leq i \leq n$, n is the number of samples in our dataset, $S_i\,[sc_i]$ represents type of salt with the concentration of sc_i, pH_i value represents the pH of i^{th} solution, and $P_i\,[pc_i]$ represents type of precipitant with the concentration of pc_i. Let R be a subset

of D that contains the crystal screen pairs having a score greater than or equal to low_H and less than or equal to $high_H$:

$$R = \{(C_i, H_i) \mid (C_i, H_i) \in D, low_H \leq H_i \leq high_H, 1 \leq i \leq n\} \tag{3}$$

In our preliminary experiments, we set $low_H = 4$ and $high_H = 7$. Thus, the samples that have a score of 8 or 9 are excluded to generate unbiased conditions for the proteins. Similarly, for simplicity the samples that have score from 1 to 3 have not been included in the result set.

The AED analysis process consists of two major stages. In the first stage, we process the data to reduce its size as we stated before. Let

$$R_c = \{C_i \mid (C_i, H_i) \in R\} \tag{4}$$

be the set of conditions of R, where $SC_i = \{sc_1, sc_2, ..., sc_k\}$ represents the all unique concentration values of the i^{th} salt, and $PC_i = \{pc_1, pc_2, ..., pc_k\}$ represents the all unique concentration values of i^{th} precipitant. Then, we compare each C_i and C_j condition pairs where $i \neq j$ in R_C. If there is a common component between C_i and C_j, then we generate the candidate conditions set Z based on these two sets. For example, let $C_i = \{S_i\,[SC_i]\,, pH_i, P_i\,[PC_i]\}$ and $C_j = \{S_j\,[SC_j]\,, pH_j, P_j\,[PC_j]\}$ where $S_i = S_j$ (i.e., the type of salt is common in C_i and C_j). We generate two new conditions Z by switching the other components among each other. Thus,

$$Z = \{\{S_i\,[SC_i]\,, \mathbf{pH_j}, \mathbf{P_i}\,[\mathbf{PC_i}]\}\,, \{S_i\,[SC_i]\,, \mathbf{pH_i}, \mathbf{P_j}\,[\mathbf{PC_j}]\}\} \tag{5}$$

is the set of candidate crystal screens for the pair C_i and C_j. Similarly, candidate screens can be generated where pH value or precipitant is common between the pairs as well. After we generate candidate combinations using these components, we remove conditions that are replicated or are already in the training data. In the second stage of our method, we assign unique values of concentrations, generate SC_i and PC_i, and unique type of buffers that were used in the preliminary data to generate finalized crystal screens. At the end, we merge generated results from two stages of the method. The identified significant factors are output and used to generate condition screens with factor concentrations varied over the indicated ranges from the analysis. These screens are then used to prepare a new plate. Figure 2 shows the flow diagram of AED.

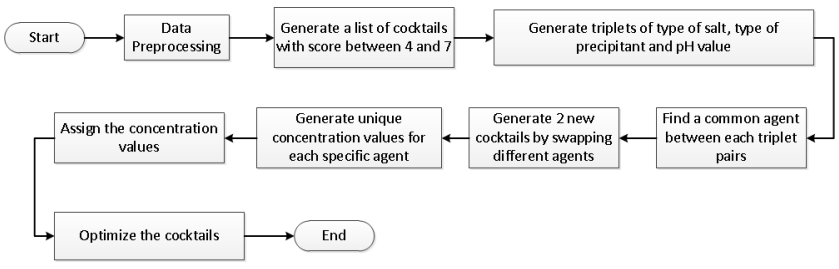

Fig. 2. Flow diagram of AED.

Sample Scnerio. Figure 3 shows the scores from four experiments from a commercial screen. The figure shows a partial graph of scores for pH value of 6.5. These conditions led to four scores: 1, 1, 4, and 4. As it can be seen, none of the conditions lead to a good crystallization condition. Our AED method finds the common reagent between solutions that could lead crystallization conditions. In this scenario, there are only two promising conditions (with score 4): $[Zn(O_2CCH_3)_2, PEG\ 8K, pH = 6.5]$ and $[(NH_4)_2SO_4, PEG\ MME\ 5K, pH = 6.5]$. The AED draws a rectangle where these conditions (with score 4) are the corners of this rectangle (Figure 4). The other corners represent the candidate conditions. There are two possible conditions for this scenario. One of them ($[(NH_4)_2SO_4, PEG\ 8K, pH = 6.5]$) already appeared in the commercial screen with a low score. When we generate the experiment for the other condition ($[Zn(O_2CCH_3)_2, PEG\ MME\ 5K, pH = 6.5]$), we were able to get a score of 7 after optimizations. The experiments have not been conducted for others since they were not on the corners of conditions with promising scores.

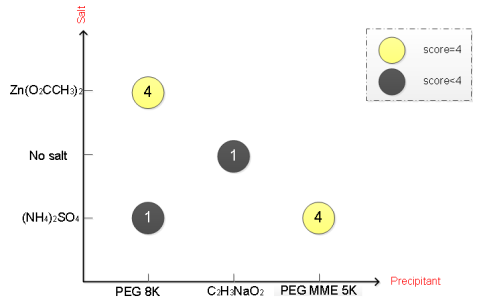

Fig. 3. Preliminary screen results.

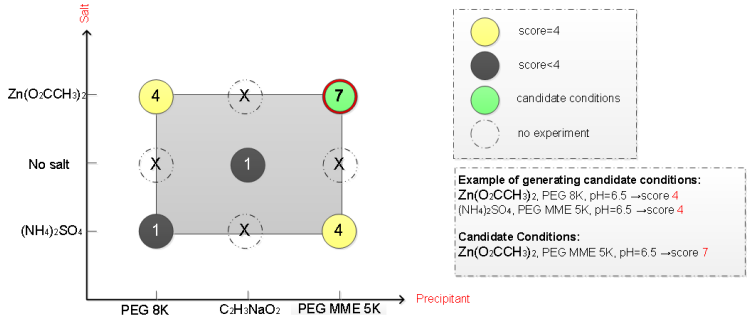

Fig. 4. Candidate (green node) conditions that AED generated based on preliminary data.

We wanted to check that *AED* is able to generate novel conditions (leading to crystallization) that do not appear in any commercial screen. A question was where to draw the distinction between identical, similar, or different screen conditions in comparison to those present in the original or all commercially available screens. Using the C6 webtool [16], an exact match to an existing commercial screen condition gives a score of 0. Variations on that condition (change in one or both precipitant concentrations, or the buffer and/or pH), give scores > 0, ranging to 1 for completely different conditions. The C6 web tool gives the top 10 matches to the input conditions. Our usual first pass optimization approach to a lead condition, having precipitants A and B, is to use four solutions; one at 100% A and B, one at 50% A and 100% B, one at 100% A and 50% B, and one at 50% each A and B. The buffer is unchanged for all four conditions. Using the C6 webtool the greatest difference between the starting and optimization conditions is for the 50% A and B, with a score of 0.269, using a reference condition of 0.5M ammonium sulfate, 30% PEG 4K, 0.1M Tris-HCl pH 8.5. This is rounded to 0.3 for our threshold score for a different screen condition. Scores > 0 but ≤ 0.3 are taken to be similar to an existing screen condition, with a score of 0 indicating identity.

4 Experiments

4.1 Dataset

The Associated Experimental Design (*AED*) approach was evaluated using proteins derived from the hyperthermophile Thermococcus thioreducens [1]. Six crystallization screening plates, three using *TFL*'d and three unlabeled protein, all with the Hampton Research High Throughput screen (HR2-130) had been set up for each of these proteins as part of a separate experimental program. For this preliminary test the scores of the results from the second (of the three) plates for the *TFL*'d protein were used, as this also includes scores of potential cryptic leads indicated by *TFL*. One was a difficult crystallizer (*Tt*106, annotated as a nucleotide kinase) with no conditions giving needles, plates or 3D crystals; one a moderate crystallizer (*Tt*82, annotated as a HAD superfamily hydrolase), with one condition giving 2D plates but none giving needles or 3D crystals; one an easy crystallizer (*Tt*189, annotated as a nucleotide diphosphate kinase) having five conditions that gave 3D crystals).

4.2 Results & Discussion

The crystallization screen components that were determined to have the greatest positive effect were determined by the *AED* software, and a 96 condition optimization screen generated using those components for each protein. Optimization was in 96 well sitting drop plates, with the protein being *TFL*'d to facilitate results analysis. The successful conditions were identified and scored. Those conditions giving 2D and 3D crystals were then used to search the C6

database [16] for similar conditions across all commercially available screens as a determination of their uniqueness. As the optimization screens had different concentration ratios for the same precipitant pairs, each ratio where a hit was obtained was searched and the lowest C6 score was used.

The moderate and difficult proteins, $Tt82$ and $Tt106$ respectively, were subjected to a second round of optimization based on the results from the first. In the case of $Tt82$ the second round was a grid screen around a condition that gave an aggregated mass of plates. Many of the second round optimization wells also showed clusters of plates. However, in one case a single plate was observed. Although not pursued, the plate clusters could be excellent starting material for seeded crystallizations, both with the original and first stage optimization screening conditions. The second optimization round for protein $Tt106$ used ionic liquids as an additive [17], with the lead conditions selected from those outcomes giving "bright spots" in the first round. Within one week one family of conditions had 3D crystals, Figure 5. Novelty of the second round conditions was determined from the grid screen condition for $Tt82$, while it was based on the parent condition for $Tt106$. Additional lead conditions were apparent in the optimization screens for $Tt82$ and $Tt106$.

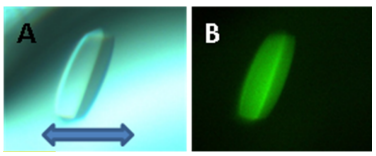

Fig. 5. White light (A) and fluorescent images (B) of second round optimization crystals of $Tt106$. Crystallization conditions: 0.2M Na/K Tartrate, 0.75M Ammonium Sulfate, 0.1 M NaCitrate, pH 5.6, 0.1M 1-hexyl, 3-methyl imidazolium chloride. Scale bar is $300\mu m$. All images are to the same scale.

The results are summarized in Table 2. The numerical values in the first two columns after the protein name refer to the number of conditions with that score in the original screening experiment (numerator) vs. those with that score in the optimization screen (denominator). For example, (0/2) indicates two novel crystallization conditions with the score of 7 (for Tt189), which did not have score 7 in the original screening experiment. The third column lists the number of optimization conditions that are novel compared to the original screen, while the last column lists those that are novel compared to all available screens. All found conditions were judged to be novel compared to the original screen on the basis of our cutoff score criteria. For $Tt189$, one optimization condition was identical to an existing commercial screen condition.

Table 2. Summary of Experiments

Protein Annotated Function	*HSHT* **Screen** [b] **Score = 7**	**Optimize Screen Score = 8, 9**	**Novel Cond. vs HSHT Screen***	**Novel Cond. vs All Screens***
*Tt*189 (Nucleoside diphosphate kinase)	0 / 2	5 / 3	5	4
*Tt*82 (HAD superfamily hydrolase)	1 / 1	0 / 1	2	2
*Tt*106 (Nucleoside kinase)	0 / 0	0 / 1	1	1
*** Using C6 tool for scores of 7, 8, & 9 threshold value of 0.3**				

[b] *HSHT*:Hampton Screen High-Throughput

5 Conclusion & Future Work

According to Table 2, *AED* generated 7 novel conditions compared to commercially available conditions for 3 different proteins derived from the hyperthermophile Thermococcus thioreducens [1]. The results obtained indicate that scored results from commercially available screens can be analyzed, and that components that may contribute to the crystallization of the macromolecule can be derived. Not surprisingly, a number of novel conditions were found for the facile crystallizer (*Tt*189). However, conditions were also found for both the moderate and difficult crystallizers, one of which had not shown any results of needles or better in the original screens (*Tt*106). For all three proteins crystallization conditions were obtained that were novel combinations of the identified factors.

These results show that *AED* is an efficient tool to generate novel conditions based on existing experimental results, which helps to save time and resources, as well as facilitating more rapid progress. In the future, we plan to include the conditions that have scores from 1 to 3 into *AED* analysis. Thus, we may generate novel conditions that may yield a successful outcome. We are also going to work on the correlation between original pair of conditions and candidate conditions by analyzing their scores. By using the scores of the original pairs, we plan to rank the candidate conditions to determine the conditions for a 96-well plate.

References

1. E. V. Pikuta, D. Marsic, T. Itoh, A. K. Bej, J. Tang, W. B. Whitman, J. D. Ng, O. K. Garriott, R. B. Hoover, Thermococcus thioreducens sp. nov., a novel hyperthermophilic, obligately sulfur-reducing archaeon from a deep-sea hydrothermal vent, International journal of systematic and evolutionary microbiology 57 (7) (2007) 1612–1618.
2. A. McPherson, B. Cudney, Optimization of crystallization conditions for biological macromolecules, Structural Biology and Crystallization Communications 70 (11) (2014) 1445–1467.

3. J. Jancarik, S.-H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, Journal of applied crystallography 24 (4) (1991) 409–411.
4. A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, 1999.
URL http://books.google.com/books?id=EDNRAAAAMAAJ
5. N. Asherie, Protein crystallization and phase diagrams, Methods 34 (3) (2004) 266–272.
6. R. C. Stevens, High-throughput protein crystallization, Current opinion in structural biology 10 (5) (2000) 558–563.
7. D. E. Brodersen, G. R. Andersen, C. B. F. Andersen, *Mimer*: an automated spreadsheet-based crystallization screening system, Acta Crystallographica Section F 69 (7) (2013) 815–820. doi:10.1107/S1744309113014425.
URL http://dx.doi.org/10.1107/S1744309113014425
8. C. W. Carter Jr, C. W. Carter, Protein crystallization using incomplete factorial experiments, J. Biol. Chem 254 (23) (1979) 12219–12223.
9. C. Abergel, M. Moulard, H. Moreau, E. Loret, C. Cambillau, J. C. Fontecilla-Camps, Systematic use of the incomplete factorial approach in the design of protein crystallization experiments., Journal of Biological Chemistry 266 (30) (1991) 20131–20138.
10. J. A. Doudna, C. Grosshans, A. Gooding, C. E. Kundrot, Crystallization of ribozymes and small rna motifs by a sparse matrix approach, Proceedings of the National Academy of Sciences 90 (16) (1993) 7829–7833.
11. J. R. Luft, J. Newman, E. H. Snell, Crystallization screening: the influence of history on current practice, Structural Biology and Crystallization Communications 70 (7) (2014) 835–853.
12. E. H. Snell, R. M. Nagel, A. Wojtaszcyk, H. O'Neill, J. L. Wolfley, J. R. Luft, The application and use of chemical space mapping to interpret crystallization screening results, Acta Crystallographica Section D: Biological Crystallography 64 (12) (2008) 1240–1249.
13. H. Yang, Å. C. Rasmuson, Phase equilibrium and mechanisms of crystallization in liquid–liquid phase separating system, Fluid Phase Equilibria 385 (2015) 120–128.
14. K. Baumgartner, L. Galm, J. Nötzold, H. Sigloch, J. Morgenstern, K. Schleining, S. Suhm, S. A. Oelmeier, J. Hubbuch, Determination of protein phase diagrams by microbatch experiments: Exploring the influence of precipitants and ph, International journal of pharmaceutics 479 (1) (2015) 28–40.
15. E. Forsythe, A. Achari, M. L. Pusey, Trace fluorescent labeling for high-throughput crystallography, Acta Crystallographica Section D: Biological Crystallography 62 (3) (2006) 339–346.
16. J. Newman, V. J. Fazio, B. Lawson, T. S. Peat, The c6 web tool: a resource for the rational selection of crystallization conditions, Crystal Growth & Design 10 (6) (2010) 2785–2792.
17. M. L. Pusey, M. S. Paley, M. B. Turner, R. D. Rogers, Protein crystallization using room temperature ionic liquids, Crystal growth & design 7 (4) (2007) 787–793.