Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The application and use of chemical space mapping to interpret crystallization screening results

Edward H. Snell,[a,b]* Ray M. Nagel,[a] Ann Wojtaszcyk,[a] Hugh O'Neill,[c] Jennifer L. Wolfley[a] and Joseph R. Luft[a,b]

[a]Hauptman–Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, NY 14203, USA, [b]Department of Structural Biology, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA, and [c]Center for Structural Molecular Biology, Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Correspondence e-mail: esnell@hwi.buffalo.edu

Received 17 June 2008
Accepted 7 October 2008

Macromolecular crystallization screening is an empirical process. It often begins by setting up experiments with a number of chemically diverse cocktails designed to sample chemical space known to promote crystallization. Where a potential crystal is seen a refined screen is set up, optimizing around that condition. By using an incomplete factorial sampling of chemical space to formulate the cocktails and presenting the results graphically, it is possible to readily identify trends relevant to crystallization, coarsely sample the phase diagram and help guide the optimization process. In this paper, chemical space mapping is applied to both single macromolecules and to a diverse set of macromolecules in order to illustrate how visual information is more readily understood and assimilated than the same information presented textually.

1. Introduction

Macromolecular crystallization *a priori* often begins by setting up screening experiments. These experiments combine the macromolecule with a diverse set of chemical cocktails designed to supersaturate the macromolecule. A chemical environment that promotes ordered aggregation of the macromolecules will produce a well defined set of outcomes, which will indicate to a trained observer those chemical conditions that may be optimized to produce larger crystals. An early example of chemical cocktails specifically developed to identify crystallization conditions is the incomplete factorial design of Jancarik & Kim (1991), which was originally commercialized by Hampton Research (Aliso Viejo, California, USA). Many different formulations of cocktails designed to identify crystallization conditions have followed, but the original screen provided the first practical standard that was (and still is) used by many laboratories. With the advent of structural genomics efforts and the development of high-throughput technologies, it is routine now to study several hundred possible crystallization conditions for any single macromolecule. At the Hauptman–Woodward Medical Research Institute (HWI), the high-throughput screening (HTS) laboratory sets up crystallization experiments in 1536-well experiment plates (Greiner BioOne, Frickenhausen, Germany) using the microbatch-under-oil method (Chayen *et al.*, 1992). Each experiment plate contains a single common macromolecule solution arrayed with an equal volume of 1536 different crystallization cocktails (400 nl total) under mineral oil. An automated imaging system records the experiments' outcomes 1 d after the addition of the protein solution and weekly thereafter for four weeks. The laboratory has been

described in detail elsewhere (Luft *et al.*, 2003). The HTS laboratory operates as a service to the general scientific community and as a crystallization screening center for several structural genomics groups. Up to 200 samples per month can be in the screening pipeline; the resulting quantity of standardized data is considerable. During a typical month, the HTS laboratory generates 1.8 million images of 300 000 crystallization experiments, plus associated ancillary data describing the macromolecules and cocktails.

Currently, an in-house software package called *MacroScope* is used to examine and classify the images manually. There are seven classification categories: crystal, precipitate, clear, phase separation, skin, unknown and garbage. When images contain combinations of outcomes (*e.g.* crystals and precipitate), they are classified in multiple categories. This approach has been used to establish a training set for automated image analysis (Snell, Lauricella *et al.*, 2008, Snell, Luft *et al.*, 2008) and represents a compromise between a comprehensive detailed description of outcomes, practical utility and the ease of manual classification. Manual sorting of these images is tedious and time-consuming. Even when classification software fully automates the process, it will not alleviate the difficulty of gaining insight and deriving useful information about the crystallization of a macromolecule from 1536 separate results. A method to highlight correlations between the individual experiments is required for interpretation and comprehension of the data.

Crystallization begins when the biochemical conditions promoting supersaturation of the macromolecule produce ordered aggregates. These ordered aggregates (nuclei) form in the labile zone, a higher level of supersaturation than required for crystal growth. As the solution conditions change, protein is sequestered in the nuclei; this can decrease the level of supersaturation and shift the conditions to a metastable region that will support crystal growth but will not support additional nucleation. These conditions are bounded by a precipitation region (highly supersaturated, leading to random disordered aggregation of the macromolecule) and an undersaturated region where the macromolecule remains in solution, viewed experimentally as a clear solution. In Fig. 1, a schematic of a crystallization phase diagram is shown to illustrate these regions. While this form of diagram serves as a sound pictorial description of the process and is common in many discussions regarding crystallization, the representation is greatly simplified. Crystallization is affected by significantly more variables than the simple example of the macromolecule and precipitant concentrations alone and the effect of any single variable is often nonlinear. Additionally, a macromolecule supersaturated in the metastable region will exhibit experimental results that are visually indistinguishable from one that is in the undersaturated region. Until a nucleation event occurs both experiments appear as clear drops, yet the experiment in the metastable region is ideal for seeding and crystal growth (Bergfors, 2003). While it may not be possible to distinguish the location of a clear drop relative to the phase diagram as an isolated data point, it is possible to determine the proximity of a clear drop to the metastable zone empirically by reference to the surrounding experiments when they are presented in chemical space. For example, clear drops located near a precipitation boundary and near or next to crystals along a chemically sensible vector are more likely to be located in a metastable region than those bounded by other clear drops (Fig. 1). Similarly, precipitation indicates that the solution is significantly more supersaturated than required for crystallization; the presence of crystals indicates that the experiment has entered or passed through the labile region. By arranging crystallization screening results for a particular macromolecule in chemical space, we globally portray the complexity and sensitivity of the phase diagram to specific chemical variables. In this manner, we do not observe isolated data but rather observe outcomes in the information-rich context of a phase diagram.

To this end, within the 1536 different chemical cocktails used in the HWI crystallization screen, 984 conditions are designed around an incomplete factorial matrix: they sample chemical space in a matrix that provides good coverage of three primary variables (pH, precipitating agents and inorganic salts). The remaining screens are commercially available and relate the incomplete factorial results to significantly expand the area of chemical space while in other instances providing a finer sampling grid of conditions that have been most successful for macromolecules crystallized previously. Combining this 1536-condition screen with manual image scoring and a graphical representation of crystallization results in chemical space readily identifies regions where crystallization is promoted. Formatting in this manner makes comprehension and correlation of the significant amount of data generated by the high-throughput experiments immediately interpretable. In this paper, we describe these developments, employing them to provide a chemical direction for crystallization optimization and graphically analyzing the crystallization history for several macromolecules.

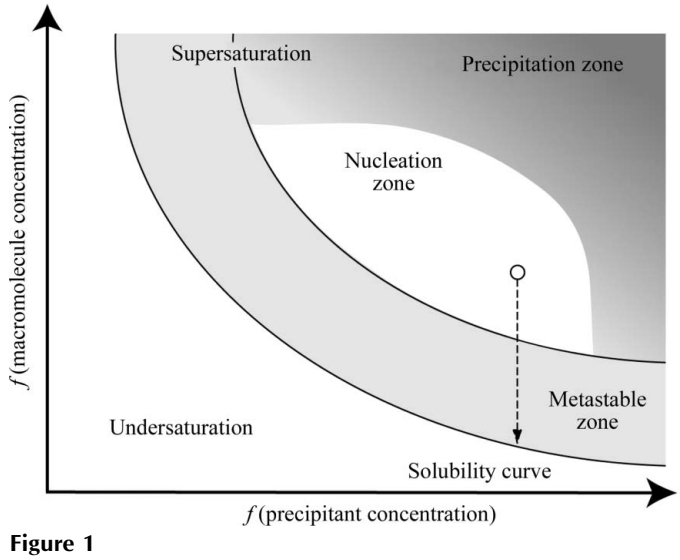

Figure 1
Simplified phase diagram for crystallization, showing the path that a successful batch experiment takes. The horizontal and vertical axes in this case are undetermined functions of the precipitant and macromolecule concentration, respectively.

2. Experimental

2.1. The Hauptman–Woodward 1536-cocktail screen

The cocktails used to screen macromolecules within the HWI screening laboratory have been developed over 8 y and are now in their ninth generation. Each generation has been optimized using lessons from the past. The current cocktails can be divided into three groups: salts, PEGs and a commercial screen group. The salts and PEGs, groups 1 and 2, were constructed using an incomplete factorial design (Audic *et al.*, 1997) and are buffered to the pH values indicated in parentheses with 100 m*M* concentrations of CAPS (10.0), TAPS (9.0), Tris (8.0), HEPES (7.5), MOPS (7.0), MES (6.0), sodium acetate (5.0) and sodium citrate (4.0). Group 1 cocktails are highly soluble salts (262 cocktails), including 36 different salts (11 cations and 14 anions) at ∼30%, ∼60% and ∼90% saturation, buffered as described. Group 2, PEG/salt (722 cocktails), includes five different molecular-weight PEGs, 20 kDa, 8 kDa, 4 kDa, 1 kDa and 400 Da, combined with 35 salts at 100 m*M* concentration, also buffered as described. Group 3 contains the commercial screens (552 cocktails). At the time of writing, this group is comprised of Hampton Research Natrix, Quick, PEG/Ion, Grid Screens (PEG, Ammonium Sulfate, Sodium Chloride), Crystal Screen HT, Index and SaltRx screens. For historical reasons, the first 22 cocktails from Hampton Research Crystal Screen Cryo are distributed within groups 1 and 2. These and other occurrences of Hampton Research cryocondition cocktails serve as controls during the experimental process.

2.2. Crystallization and analysis

Two studies were conducted, one an investigation of the use of chemical space mapping for individual crystallization and the other for the analysis of multiple crystallization experiments. For the individual cases, apoferritin and ubiquitin were used as examples. For the analysis of multiple crystallization experiments, 106 samples from the North East Structural Genomics consortium (NESG) and 163 samples from the Structural Genomics of Pathogenic Protozoa (SGPP) structural genomics centers have been included. Each sample was set up at approximately 10 mg ml^{-1} concentration with 200 nl of the macromolecule solution and 200 nl of each of the 1536 crystallization cocktails. These are set up under 5 µl mineral oil using the microbatch method of crystallization as described elsewhere (Luft *et al.*, 2003). Images were recorded over time and the four-week images were analyzed and classified manually (Snell, Lauricella *et al.*, 2008; Snell, Luft *et al.*, 2008). In the case of apoferritin and ubiquitin, all categories of outcomes were classified. Where a region of interest was indicated by this analysis, a crystallization optimization protocol called drop volume ratio/temperature (Luft *et al.*, 2007) was applied. For the macromolecules from the structural genomics community, only the crystal category was classified. Chemical space mapping for single macromolecules was achieved using the program *AutoSherlock* (Nagel *et al.*, 2008). For the study of multiple macromolecules, the chemical space maps were added and again displayed as a color-coded spreadsheet to identify trends in crystallization behavior averaged over the entire group of 1536 cocktails. The input for *AutoSherlock* is the classified data, image names and the chemical makeup of the experiment. The output is a Sherlock plot, a color-coded spreadsheet linked to the crystallization images, sorted in a chemically sensible manner.

3. Results

The program *MacroScope* displays the 1536 crystallization screening results for a single macromolecule in 16 arrays of 96 thumbnail images. Selecting a thumbnail image enlarges it to the full-size image and provides the option of displaying

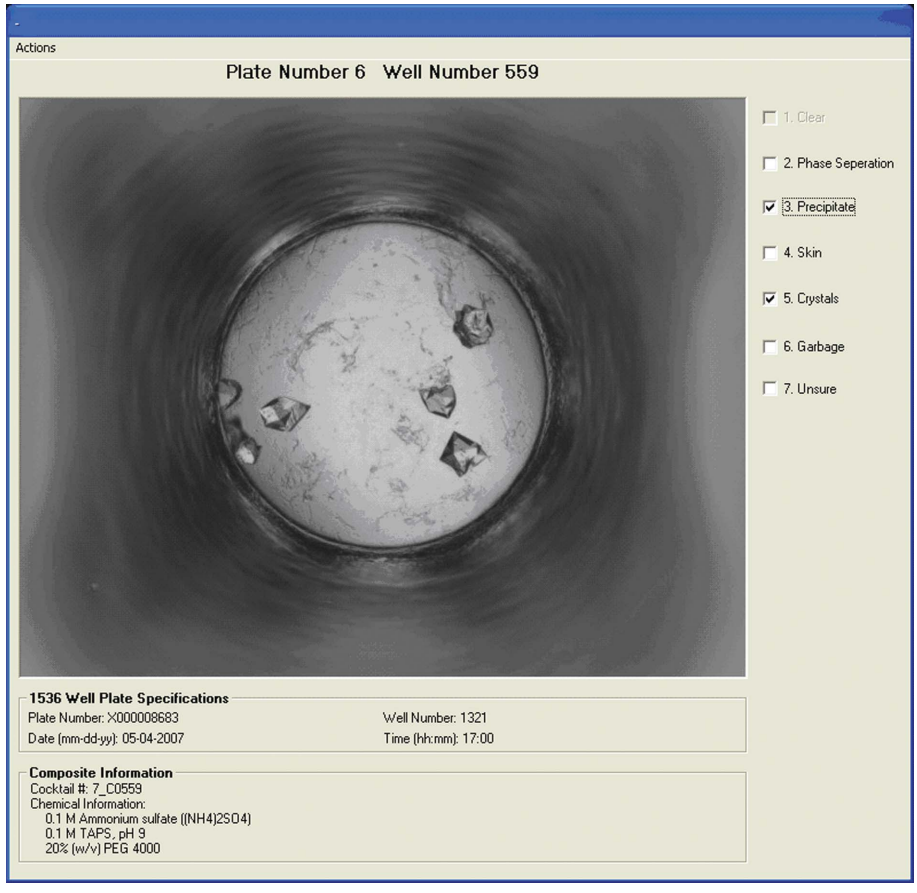

Figure 2
An example showing the image and chemical data presented for scoring with *MacroScope* along with the options available for scoring. Images can be classified as a function of multiple categories, with the exception of clear which can only be selected as a single category.

chemical information and/or a scoring box (Fig. 2). The display of the 96 blocks is strictly sequential, with diverse chemical conditions often adjacent to each other, providing no indication of the larger chemical space that remains unsampled

	PEG 1000	Na Citrate	Na Acetate	MES	MOPS	HEPES	Tris	TAPS	CAPS	PEG 4000	Na Citrate	Na Acetate	MES	MOPS	HEPES	Tris	TAPS	CAPS	PEG 8000	Na Citrate	Na Acetate	MES	MOPS	HEPES	Tris	TAPS	CAPS
pH		4	5	6	7	7.5	8	9	10		4	5	6	7	7.5	8	9	10		4	5	6	7	7.5	8	9	10
	Ammonium																											
phosphate-monobasic	20%			0704	0703					20%	0553			0554			0552		20%		0399		0396		0397	0398	
	40%	0773			0775	0774			0772	40%					0629			0630	40%		0477	0478					
phosphate-dibasic	20%	0705						0706		20%	0555			0557			0556	0631	20%		0479	0401		0480		0400	
	40%	0776																									
sulfate	20%			0707	0708					20%					0558		0559		20%					0402			0403
	40%				0777		0779		0778	40%	0633		0634					0632	40%		0481	0482				0483	
chloride	20%	0696	0697				0699	0698		20%		0550						0549	20%				0393				
	40%						0770			40%				0624				0623	40%	0472	0471				0473		

Figure 3
A chemical space map with the cells colored according to the crystallization result. The red color indicates a crystal hit, light green a precipitate, dark green a skin and turquoise a clear result. Multiple category results are shown by cross-hatched colors, with each color representing one of the categories. In this illustration, the blocks which appear orange are cross-hatched with alternating red and light green to indicate a combination of crystals and precipitate.

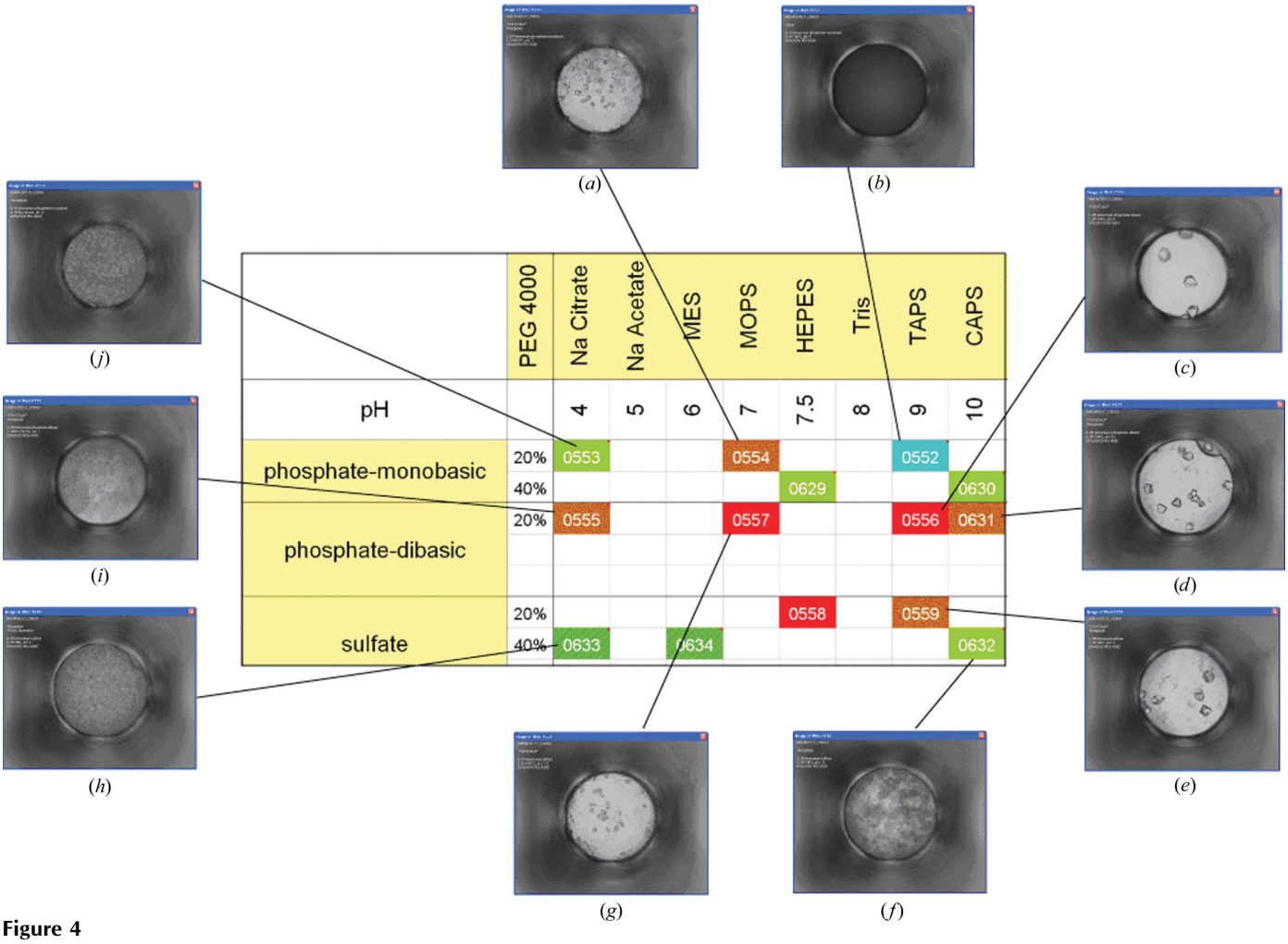

Figure 4
Magnified view of Fig. 3 with selected crystallization results shown. Images (*a*), (*d*), (*e*) and (*i*) have been classified as a mixture of crystals and precipitate. Image (*b*) has been classified as clear with a dark color but no apparent texture. In the context of its chemically related neighbors, this classification is probably incorrect and the image should be reclassified as precipitate. Images (*c*) and (*g*) are crystal only and finally image (*h*) shows evidence of a skin and precipitate. The uncolored blocks indicate areas of the incomplete factorial screen that are unsampled.

between images. Scoring is currently manual and this method of display has proved to be most effective for this purpose.

The images are classified and then represented by *Auto-Sherlock* on a two-dimensional matrix in a chemical space

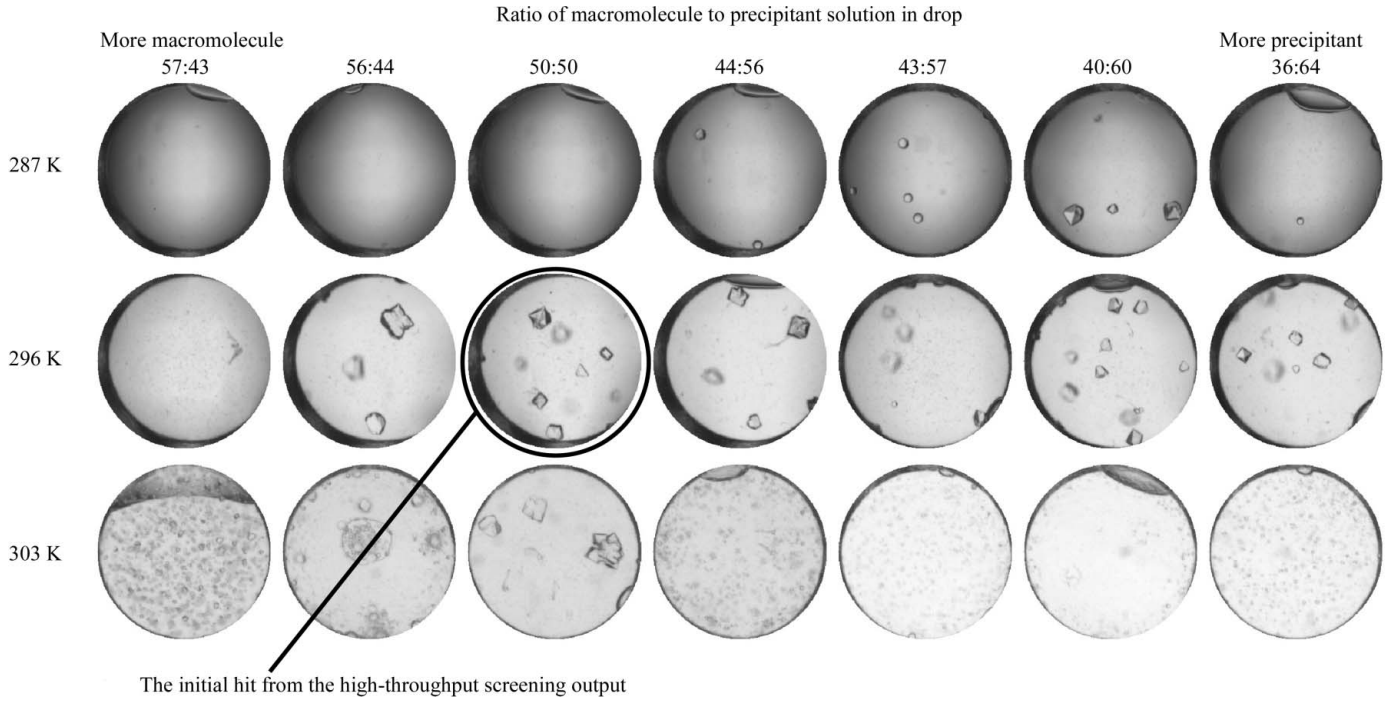

Figure 5
Partial results from drop volume ratio and temperature screening (DVR/T; Luft *et al.*, 2007) applied to the hit seen in Fig. 4(*c*). The full optimization technique uses 16 ratios at five different temperatures and provides a phase diagram for optimization.

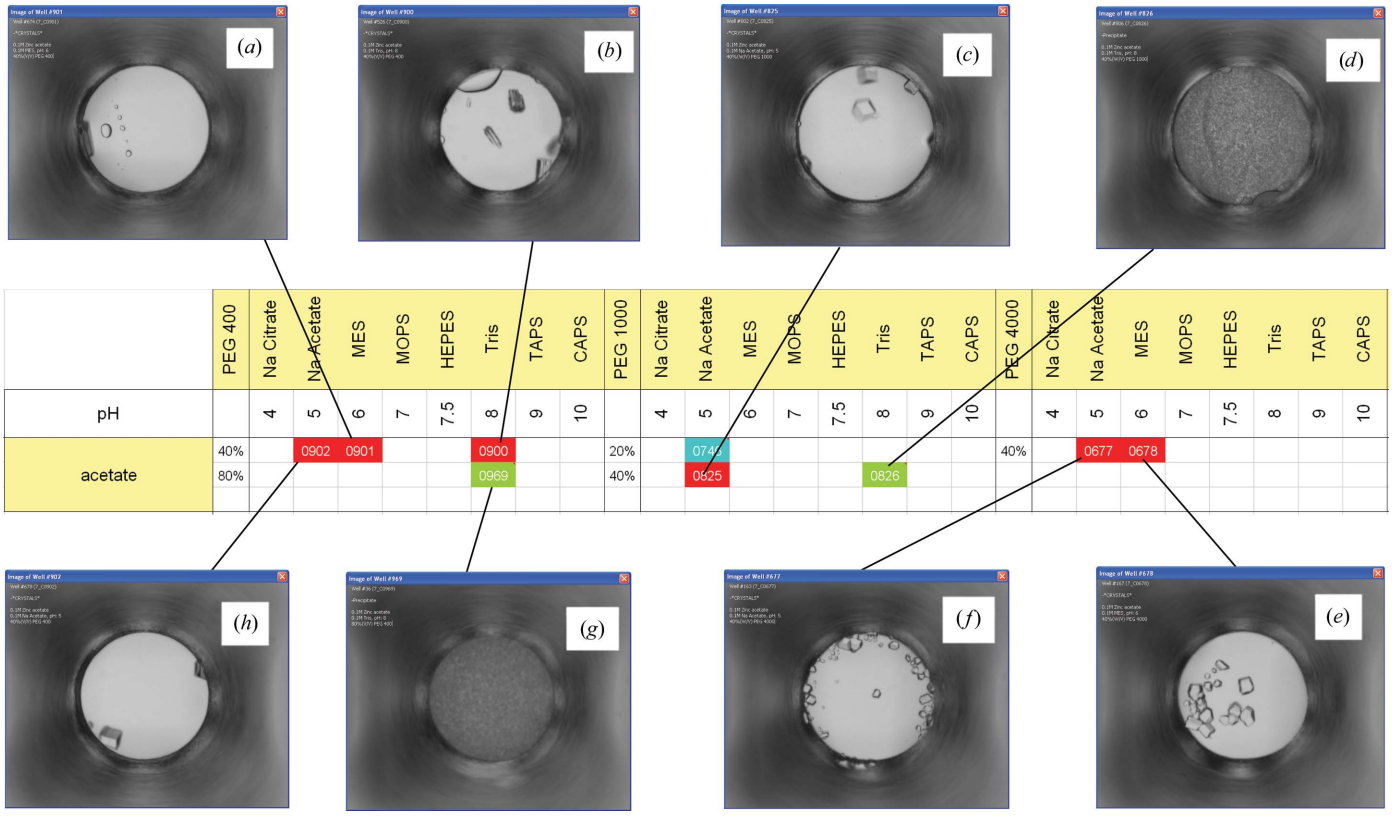

Figure 6
Enlargement of a chemical space map showing the influence of zinc acetate on the crystallization results.

determined by the user. For an individual macromolecule, the matrix is colored according to the classifications (if any) assigned to the image. Where no classification is present (but an experiment was conducted) the matrix is colored gray and where no experiment took place the matrix is white. For a series of macromolecules and multiple sets of experiments, the same matrix is used but the shade of the color represents the number of crystallization hits observed within the samples (lighter, fewer observations; darker, more observations).

3.1. Chemical space mapping for optimization of crystallization conditions

In Fig. 3, part of a chemical space map for apoferritin is illustrated. Images classified as crystal hits are colored red, those that were clear are shown in turquoise and those showing precipitate are colored green. Other colors represent different outcomes, with crosshatched squares representing multiple classifications (Nagel *et al.*, 2008). The results are shown as a function of PEG molecular weight and pH on the horizontal axis and the cation and anion on the vertical axis. On this scale it is difficult to see individual details, but one result that does stand out is the island of crystallization hits associated with ammonium phosphate and PEG 4000 at 20%(*v*/*v*). Enlarging this area further for the ammonium phosphate and sulfate portion (Fig. 4) shows that low pH encourages the formation of precipitate and microcrystals. Similarly, precipitate is favored at higher PEG concentration. Increasing the pH and reducing the PEG concentration results in well developed crystals of equidimensional morphology. By plotting and analyzing the results in the form of chemical space, the influence of PEG concentration and pH becomes readily apparent. In the same figure, we can also see that there are areas of space than have not been sampled, *e.g.* 20% PEG 4000, ammonium phosphate dibasic and pH 7.5 and 8.0. Based upon the surrounding data, these look to be good areas for further optimization. The drop-volume ratio variation and temperature (DVR/T; Luft *et al.*, 2007) technique was applied to the pH 9, 20% phosphate dibasic condition (Fig. 4*c*). By varying the drop-volume ratio and temperature, we sample different solute concentrations and, in the case of TAPS, the buffer pH by approximately $\pm$0.5 pH units. The results from applying this optimization technique to the initial crystallization conditions identified from the chemical space mapping are shown in Fig. 5. As indicated by the chemical space map, this is indeed a fertile area for crystallization optimization. An interesting aspect of the chemical space mapping is illustrated by Fig. 4(*b*). This image was originally classified as clear, the image was dark and there was no texture visible to indicate precipitate. However, rather than looking at this image as a single data point, if we compare the image to others surrounding it in a chemical context it becomes apparent that the image should be reclassified as precipitate. This has implications in improving automated image analysis by using chemically adjacent results to weight the interpretation of ambiguously classified images.

A second example with ubiquitin is shown in Fig. 6. In this case there were a succession of hits associated with zinc acetate and 40% of various molecular-weight PEGs as the precipitant. Outside of the displayed areas there were few hits. On examining the chemical space, the optimum PEG concentration lies below 80% and above 20%. Trends indicate that crystallization favors lower pH values, although few conditions at high pH are sampled. The protein seems to be equally crystallizable in PEG 400, 1000 and 4000. Microcrystals were seen in PEG 8000 and no crystals were found in PEG 8000 or 20 000 or in those conditions that did not contain PEG or zinc acetate. In this case, the zinc acetate and low-molecular-weight PEG combination seems to be the most favorable for crystallization. If we choose another representation of chemical space where zinc acetate is held constant and plot the data as a function of the pH and PEG molecular weight (Fig. 7), it becomes clear that these are the two most significant principal variables for crystallization. As the molecular weight of the PEG and the pH value decrease, the outcomes transition from precipitate through crystals to clear, *i.e.* the biochemical conditions sample through the precipitation, nucleation, metastable and possibly the undersaturated zones. One can quickly determine that in the unsampled regions PEG 4000 at pH 7 and PEG 1000 at both pH 7 and pH 6 follow a trend that makes them appear suitable for further investigation. The DVR/T optimization results (Fig. 8) would tend to support this. Although higher molecular weights are not sampled, the concentration of PEG 400 is varied across one axis. As the ratio of PEG decreases precipitate is seen and as it increases well formed single crystals appear. Temperature significantly changes the pH of the Tris buffer used in these experiments. In this example with Tris buffer, DVR/T optimization finely samples a range of pH values that are relevant to crystallization. In Fig. 9, we have scaled up the experiment and grown crystals under the previously unsampled conditions PEG 4000 pH 7 and PEG 1000 pH 6 and 7. At PEG 4000 pH 7

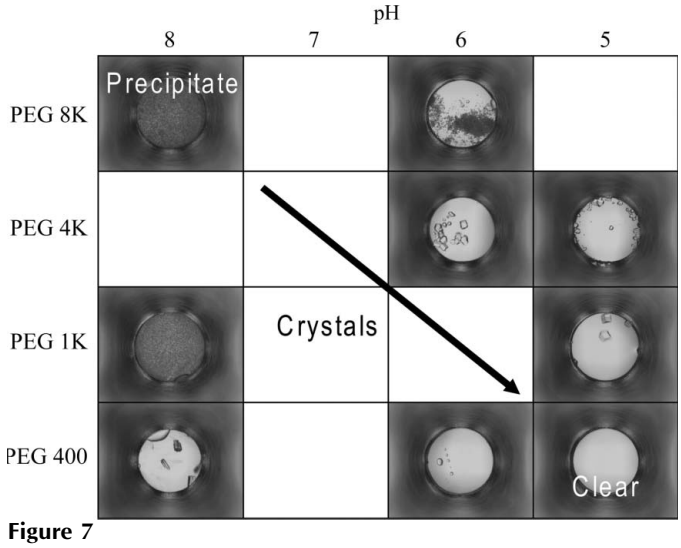

Figure 7
Re-plotting the chemical space data as a function of pH and PEG molecular weight to show the dependence of the results on these two variables.

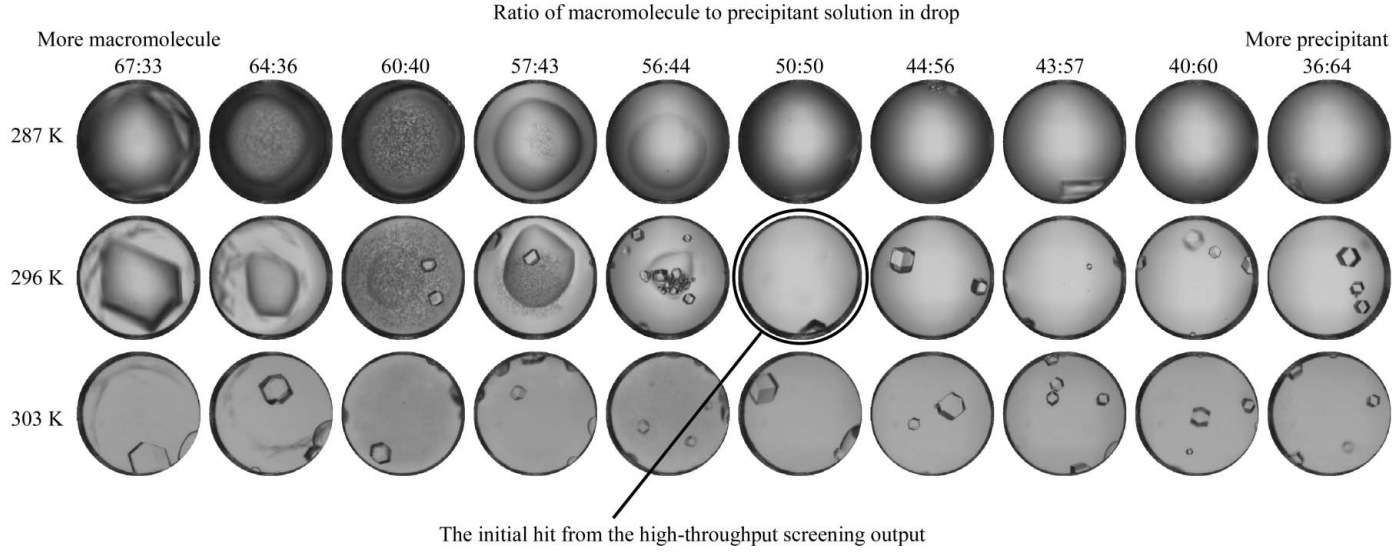

Figure 8
Partial results from drop volume ratio and temperature screening (DVR/T; Luft *et al.*, 2007) applied to the hit seen in Fig. 6(*b*). The full optimization technique uses 16 ratios at five different temperatures and provides a phase diagram for optimization.

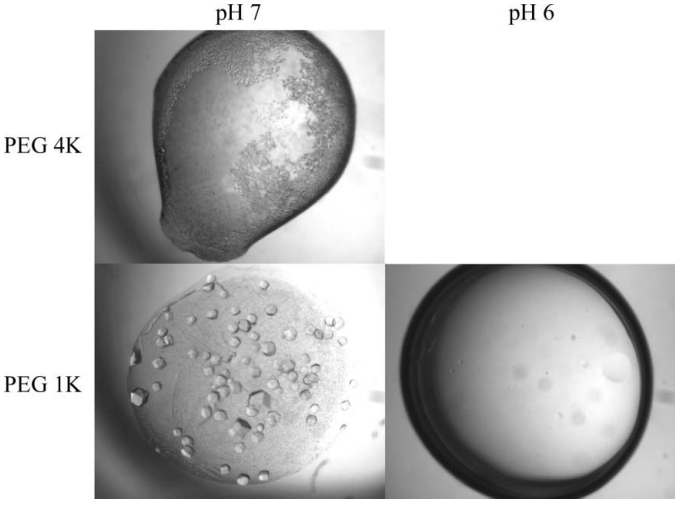

Figure 9
Scaled-up crystallization trials for promising conditions not sampled by the incomplete factorial matrix in Fig. 7.

we see a crystalline precipitate. As we decrease the PEG molecular weight to 1000, crystals appear on a faint background of precipitate. Shifting the pH one unit to pH 6 results in a clear drop. The trends revealed by chemical space mapping of the incomplete factorial sampling are confirmed with these experiments that fill in the missing conditions.

3.2. Multiple experiments

The chemical space representation produced by *AutoSherlock* can also be used to enable the analysis of 1536 screening experiments for multiple macromolecules using an identical formatting of the chemical space. This provides insight into the behavior of particular chemical cocktails averaged over different macromolecules. In Fig. 10, we have used this approach to analyze crystal hits from 269 macromolecules supplied by two structural genomics groups. While it is difficult to read the text in this figure, the trends are immediately apparent. The shading of the cells denotes the number of crystal hits observed for a particular cocktail. In this case, dark blue indicates five or more hits, medium blue 3–4 hits and light blue 1–2 hits. Unsampled areas of chemical space are white and those cocktails that showed no hits are colored grey. Of the 984 cocktails sampled, only four showed no hits with the 269 macromolecules. These were 0.1 *M* ammonium chloride and 40%(*v*/*v*) PEG 4000 at pH 6, 1.69 *M* potassium carbonate pH 8 and 0.1 *M* zinc acetate and 40%(*v*/*v*) PEG 8000 at both pH 8 and pH 5. These have been successful in other crystallization studies; it appears to be chance that they were not successful here rather than that they lack any crystallization potential. What is quite remarkable is that so many of the cocktails provided crystallization hits. The diagram is organized in this representation so that there are six columns. The first column is group 1 of the HWI screen: the highly soluble salts. The second to sixth columns go from PEG 20K to PEG 400, representing group 2 of the HWI screen. The columns are divided into cations, indicated by yellow stripes, and associated anions. While now well known, it is immediately apparent from a single glance that the PEG groups strongly favor crystallization. The graphical representation of crystallization data for large sets of macromolecules demonstrates the potential power of this method when applied to analyze data from high-throughput crystallization studies.

4. Discussion

Visual information is more rapidly understood and assimilated than the same information presented numerically. This is especially true for large quantities of interrelated data. From the *AutoSherlock* plot of chemical space, we can rapidly

identify common chemical conditions that promote crystallization and see trends in crystallization caused by variables such as pH or precipitant type or concentration. This is made possible owing to the incomplete factorial design of the crystallization screen. Crystallization cocktails are chemically related in space and therefore the results can be mapped and interpreted in that same space. It is relatively intuitive, with knowledge of a typical crystallization phase diagram and the

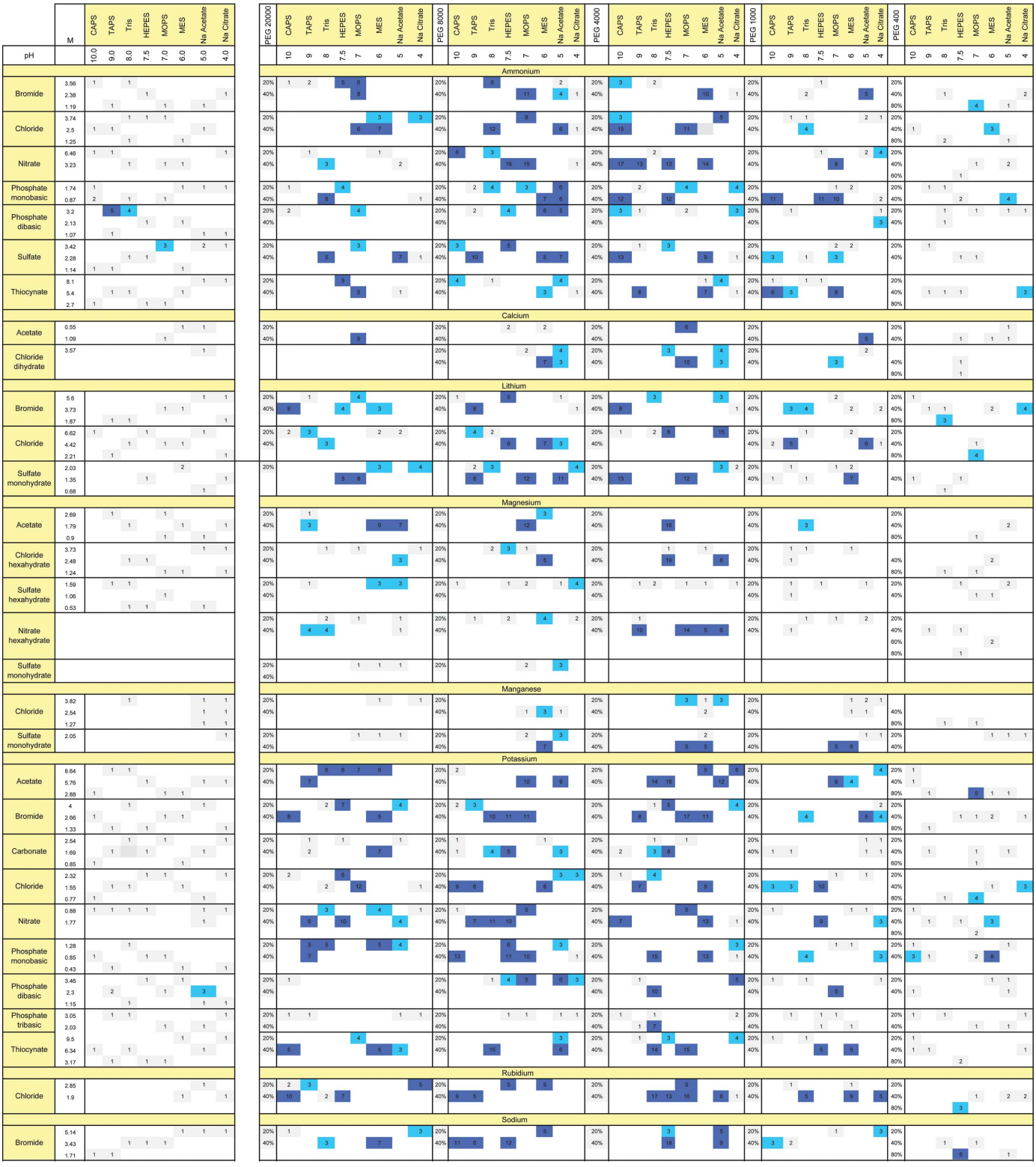

Figure 10
Plot of crystal hits as a function of color for 269 macromolecules from the structural genomics community. Dark blue indicates five or more crystal hits for that cocktail, medium blue 3–4 and light blue 1–2. White areas are unsampled areas of chemical space. The columns are arranged as described in the text.

correct chemical space mapping, to identify directions for crystallization optimization. Another important aspect of mapping the results into chemical space is the ready identification of conditions that have not been sampled and the results surrounding those conditions. For example, if precipitation and clear results were seen along a chemical vector that included a large unsampled area of space, this would be a strong indication that this region is a good candidate for further exploration.

Crystallization is a multidimensional problem; for ease of human interpretation we derived a two-dimensional representation. It is readily apparent that at the expense of additional experiments the analysis could become more powerful. We accomplish this refinement using the DVR/T method, in which both temperature and the concentration of the solutes are experimental variables (Luft *et al.*, 2007), as seen in Figs. 5 and 8. This allows us to optimize the chemical concentrations and pH in fine gradients. The complexity of interpreting DVR/T experiments increases with the realisation that the temperature causes secondary chemical effects that alter solubility. We can use the chemical space mapping to guide and improve optimization by identifying key variables. Where a dramatic pH effect is seen, the DVR/T optimization could be geared toward a constant temperature with a buffer-driven pH sampling (similar to the experiments conducted for Fig. 9 but with a finer degree of pH sampling). Thus, specific optimization approaches can be identified at the crystallization screening stage. There remains a huge potential for powerful computational analysis incorporating and capable of interpreting the interplay of crystallization variables in multi-dimensional space in order to isolate and identify the most significant parameters affecting crystallization, *e.g.* in the case of pH and PEG molecular weight in Fig. 7.

A time-consuming aspect of the process is that images are manually classified using *MacroScope*. We have found that presenting the images globally in chemical space facilitates a more accurate classification of results than those obtained from the review of individual experiments, as discussed for Fig. 4(*b*). Current image-analysis software for crystallization considers images independently, but the accuracy of the system could be improved by weighting the images as a function of chemically related results. Rather than develop an automated image-analysis system, an expert system could be developed that makes use not only of the image results but also of knowledge of crystallization and the analysis of related images. Automated image-analysis techniques have had some success, but have not perfected the identification of crystals within samples. However, they are reasonably successful at identifying precipitated or clear conditions with accuracy. Populating the chemical space diagram with only these results rapidly points out areas of chemical space to explore further for crystallization and minimizes the manual scoring effort by eliminating clear and precipitated experiments. This approach offers a way to make use of reliable clear and precipitate classifications using automated image analysis and couples these simple classifications with chemical space plots to enhance and expedite scoring of outcomes pending the development of more sophisticated systems.

Adapting the *AutoSherlock* plot to look at global trends proves informative in revealing cocktails or groups of cocktails that perform well or badly. While we have only sampled a small proportion of the 10 000 macromolecules that have passed through the HWI crystallization screening service, we intend to utilize the same analysis coupled with automated image analysis to look at a much greater proportion of these experiments, examining the cocktails both as a function of success and as a function of time. Other structural genomics groups have noted that only a small number of conditions are necessary for successful crystallization of targets (Page *et al.*, 2003); however, while requiring slightly more sample, using a set of 1536 crystallization cocktails based largely on an incomplete factorial design provides individual cocktail targets for optimization and, perhaps more importantly, a vector for that optimization.

Our crystallization screening cocktails are formulated from a relatively small number of stock solutions. The chemical space map can identify chemical conditions falling between those sampled that might be appropriate. For example, where potential crystallization results are seen using similar conditions with PEG 4000 and PEG 8000, it may be worthwhile to consider using PEG 6000 for optimization. In some instances, chemicals which are not used for the incomplete factorial portion of the screen are present in the commercial screens. Again, mapping the results in chemical space emphasizes this, helping to correlate seemingly distant relationships between the incomplete factorial-based design and commercial screens contained within the 1536 cocktails.

In our optimization process, we judge our results visually. This is appropriate for the initial stages of crystallization screening, but offers no information on the X-ray scattering of the crystal. A future development will be to link rapid X-ray diffraction feedback into the chemical space diagram to determine chemical shifts that help to promote short-range order.

5. Conclusions

Representing crystallization results in chemical space quickly identifies optimization pathways and meaningful gaps in the sampled chemistry that could be exploited. It allows the rapid visualization and interpretation of a complex experimental space. At this point we have not incorporated any chemical knowledge beyond the breakdown of the space into the basic variables of anion, cation, pH and PEG. The pH is taken as the pH of the starting buffer solution and is assumed to be constant over time; no attempt is made to account for the kinetics of dehydration which, even with the microbatch approach, we know takes place over time. The microbatch crystallization method is perhaps the most simple to analyze as the starting conditions are well known. Vapor diffusion slowly comes to equilibrium and this may pose potential complications when presenting the results in chemical space. There is much potential to further refine and improve upon the

chemical space mapping and the resulting representation of crystallization data. The current version of *AutoSherlock* (Nagel *et al.*, 2008) that produces the chemical space map is a powerful but admittedly primitive software tool which simplifies the interpretation and identification of complex crystallization trends and provides significant insight and information for further optimization.

This work was supported by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the US Department of Energy, by NIH U54 GM074899 and the John R. Oishei Foundation. We would like to thank NESG and SGPP for allowing us to use their images in this analysis. At Oak Ridge National Laboratory, Anna Gardberg is thanked for useful discussions and testing the *AutoSherlock* software. Dean Myles and Flora Meilleur are thanked for useful discussions. Finally, Dr George DeTitta at HWI is acknowledged for insightful comments and support with this analysis. *AutoSherlock* is available for users of the HTS screening service. Further information is available at http://www.hwi.buffalo.edu/Services/Services.html.

References

Audic, S., Lopez, F., Claverie, J. M., Poirot, O. & Abergel, C. (1997). *Proteins*, **29**, 252–257.
Bergfors, T. (2003). *J. Struct. Biol.* **142**, 66–76.
Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. (1992). *J. Cryst. Growth*, **122**, 176–180.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K. & DeTitta, G. T. (2003). *J. Struct. Biol.* **142**, 170–179.
Luft, J. R., Wolfley, J. R., Said, M. I., Nagel, R. M., Lauricella, A. M., Smith, J. L., Thayer, M. H., Veatch, C. K., Snell, E. H., Malkowski, M. G. & DeTitta, G. T. (2007). *Protein Sci.* **16**, 715–722.
Nagel, R. M., Luft, J. R. & Snell, E. H. (2008). *J. Appl. Cryst.* **41**, 1173–1176.
Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003). *Acta Cryst.* D**59**, 1028–1037.
Snell, E. H., Lauricella, A. M., Potter, S. A., Luft, J. R., Gulde, S. M., Collins, R. J., Cumbaa, C., Jurisica, I., Malkowski, M. G. & DeTitta, G. T. (2008). *Acta Cryst.* D**64**, 1131–1137.
Snell, E. H., Luft, J. R. *et al.* (2008). *Acta Cryst.* D**64**, 1123–1130.