Cinder: keeping crystallographers app-y

Nicholas Rosa, Marko Ristic, Bevan Marshall and Janet Newman*

Manufacturing (Biomedical), CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia. *Correspondence e-mail: janet.newman@csiro.au

Received 22 March 2018

Accepted 31 May 2018

Edited by M. W. Bowler, European Molecular Biology Laboratory, France

Keywords: crystallization; images; scoring; machine learning; *Cinder*; mobile apps.

Supporting information: this article has supporting information at journals.iucr.org/f

The process of producing suitable crystals for X-ray diffraction analysis most often involves the setting up of hundreds (or thousands) of individual crystallization trials, each of which must be repeatedly examined for crystals or hints of crystallinity. Currently, the only real way to address this bottleneck is to use an automated imager to capture images of the trials. However, the images still need to be assessed for crystals or other outcomes. Ideally, there would exist some rapid and reliable machine-analysis tool to translate the images into a quantitative result. However, as yet no such tool exists in wide usage, despite this being a well recognized problem. One of the issues in creating robust automatic image-analysis software is the lack of reliable data for training machine-learning algorithms. Here, a mobile application, *Cinder*, has been developed which allows crystallization images to be scored quickly on a smartphone or tablet. The *Cinder* scores are inserted into the appropriate table in a crystallization database and are immediately available to the user through a more sophisticated web interface, allowing more detailed analyses. A sharp increase in the number of scored images was observed after *Cinder* was released, which in turn provides more data for training machine-learning tools.

1. Introduction

Crystals are a prerequisite for structural studies using diffraction techniques, but are notoriously difficult to produce. The general process is to set up a purified protein sample (or any other macromolecule/macromolecular complex) against a number of standard (commercially available) crystallization screens and observe the results over time. Ideally, this will result in suitable single crystals of the protein for X-ray analysis, but more commonly any crystalline hits need to be optimized before they evolve into appropriate crystals. Unfortunately, the initial screening often gives no obvious crystalline starting point at all. To determine whether a bank of trials have been successful, each trial needs to be examined to find crystals or near-crystals. As crystallization is time-dependent, the trials often have to be examined multiple times. The visual identification of positive outcomes was traditionally performed by looking at each experiment under a microscope. This can be enormously time-consuming and, for trials set up in the cold room, very uncomfortable.

Automatic imagers, particularly systems with plate storage and robotics to move plates, have eased this problem. These machines, which developed out of the structural genomics push of the last two decades, automatically capture images of crystallization trials and can be set up to collect a time course of images from the crystallization droplet. Along with automatic imagers, the last two decades have led to the commercial

availability of a number of low-volume dispensing technologies specifically designed for crystallization trials. Combining these means that it has become quite trivial to set up many crystallization experiments, with the corresponding automatic collection of large numbers of images tracking the time course

of each experiment. The easy acquisition of images has not been matched by facile interpretation. Each image still needs to be looked at by a human to determine the outcome of the crystallization trial. This is currently performed by looking at each image using software that presents both the image and

Figure 1

Screenshot of the viewing application *See3* used to inspect images in C3. This is a web-based application which presents images collectively or individually. The top figure shows a section of a crystallization plate set up with lysozyme (here dyed with blue food colouring) and shows that the top row of images has been assigned scores by a human. The scores are shown as a coloured border around each thumbnail. The bottom figure shows a single droplet and shows the scoring history for the droplet. Other views present other information, for example information about the crystallization conditions or the time since the experiment was initially created.

information about the experiment shown in the image. There are a number of image-presenting applications: some of these are thick clients running on individual workstations (for example *ROCK MAKER* from Formulatrix or *CrystalTrak* from Rigaku) and some are web-based applications (for example *ROCK MAKER WEB* from Formulatrix and *CT Web* from Rigaku). In our laboratory, which is a specialist crystallization laboratory (Collaborative Crystallisation Centre, 'C3'; <http://crystal.csiro.au>), we have developed our own web-based viewing software, *See3*. Along with viewing the images, all of these tools allow the assignment of one or more classifications or scores to each image. Each of these tools has an underlying database which captures the scoring information; most often, information about the score, the scorer, the image and the time that the score was assigned are captured. The viewing software will display any scores along with the images, often using colour as a quick visual for associated scores (Fig. 1). These viewing tools make reviewing and classifying crystallization images relatively fast and easy, but it is even faster and easier to use the tools to scan quickly through a plate worth of images, locating interesting images and then using the scoring tools to classify only these interesting images. Although this approach can identify crystals readily, it does not generate the complete sets of classified images that would be needed for more in-depth analyses. Furthermore, the viewing tools currently require access to a computer, which can limit how often a scoring session is initiated.

Even in the early days of the development of imaging automation suitable for crystallization trials the problem of scoring the resulting images was recognized, and the groups that were developing in-house imaging systems (pre-dating the commercially available systems) also spent time developing scoring algorithms in order to automate the assignment of a value to a crystallization image (Buchala & Wilson, 2008; Cumbaa & Jurisica, 2005; Spraggon *et al.*, 2002). Although a large amount of work was put into this area, the programs developed in the early part of the century for autoscoring mostly remained within a single institute (where the development took place) or stopped being used at all. A number of factors influenced this, with the most prominent factor likely being that the programs were not sufficiently useful to overcome their significant barriers to implementation. Most of the programs that were developed were machine classifiers, which used some combination of feature extraction and supervised learning to classify images into two or more classes.

Viewing crystallization trials is certainly required to locate crystals for harvesting and X-ray analysis. More commonly, when perfect crystals do not grow out of screening conditions, the analysis is used to tease out which crystallization factors might be positive influencers of crystal growth and which are negative influencers (Ng *et al.*, 2014). For example, if all crystallization conditions that contain zinc result in a brown, 'denatured-looking' precipitate, one hypothesis might be that zinc ions have a negative effect on this protein and should be excluded from future trials. This use of crystallization trials (to develop hypotheses about crystallization space) suggests that any classification scheme, human or machine, should capture

Table 1

Scoring scheme as offered by Hampton Research (<http://www.hamptonresearch.com>).

There are nine classes, so that only a single digit is needed to record a score. The score number reflects the importance of the category for finding harvestable crystals.

Score	Description	Superclass
1	Clear Drop	Clear
2	Phase Separation	Other
3	Regular Granular Precipitate	Precipitate
4	Birefringent Precipitate or Microcrystals	Crystal (or Precipitate)
5	Posettes or Spherulites	Other
6	Needles (1D Growth)	Crystal
7	Plates (2D Growth)	Crystal
8	Single Crystals (3D Growth < 0.2 mm)	Crystal
9	Single Crystals (3D Growth > 0.2 mm)	Crystal

sufficient information to enable the development of optimization strategies.

There is no community-wide consensus on what scores (or even how many classes) are appropriate for both human and machine scoring. Perhaps the most widely used human classification system is that suggested by Hampton Research (<http://www.hamptonresearch.com>; Table 1), which is a nine-class scheme, with most of the classes having to do with subcategories of crystals. *TexRank*, a program developed in the Structural Genomics Consortium in Oxford (Ng *et al.*, 2014), ranks according to the likelihood of crystals (irrespective of the shape or number of crystals) and is thus essentially a two-class system (crystal/no crystal). Another pertinent question is how reliable automatic scoring needs to be in order for it to be useful. A scientist looking down a microscope for crystals has no objective sense of how many crystals they may have missed, and perhaps it is this observational bias which underpins the distrust of a machine-learning system that can 'only' classify 80% or so of crystal-containing droplets successfully. As crystals are rare events, many crystallographers feel that missing crystals (false negatives, or recall <1) is a worse flaw than scoring crystal-free drops as containing crystals (false positives, or precision <1). Work by both Wilson (2006) and DeTitta and coworkers (Snell *et al.*, 2008), in which a number of structural biologists were given the same set of images to score and where the image set contained some deliberate internal duplication, suggests that (even experienced) humans were about 75–85% consistent in their scoring (although the consistency improves for crystal-line results), so any program that can score with an accuracy of 85% or greater (the 'Wilson limit') would be advantageous. In order to obtain the recall and precision that are needed to perform at least as well as the Wilson limit, there is the requirement of having a well scored training set for machine learning. Viewing crystallization images and recognizing the relevant features within the droplet takes practice, as anyone who has watched novices flounder as they observe crystal trials knows. Some of the things which are 'obvious' to a seasoned crystallographer but not to neophytes are shown in Fig. 2.

Any training set used for machine learning should consist of large numbers (many thousands) of crystallization images that have been scored or classified by an experienced crystallographer, or even better by more than one experienced crystallographer. There are a number of factors that go into a high-quality training set. Certainly the images must be well scored, but the images must also be diverse, *i.e.* capture most of the different outcomes that might reasonably appear in crystallization screening; this generally sets the requirement for a large number of images. Finally, some thought needs to be applied to the problem of balancing the data: training data must consist of examples of all of the classes, but whether there should be an equal number of each class or not is still an open question (Weiss & Provost, 2001).

One way of generating training sets would be to harvest human scores from one or more crystallization laboratories that have both automatic imagers and large numbers of different samples. This is the approach that we have taken in C3, but we found that there simply were not enough scores in our system. In the 12-year history of C3, over 3.6 million trials have been dispensed and there are over 45 million associated visible-light images. However, fewer than 585 000 images have been scored through our web interface: less than 2% of the images. Although half a million scores sounds like a suitably large data set, the scoring in the C3 database has not been

curated at all: noisy data require that more data are needed for effective training (Sukhbaatar *et al.*, 2015). Furthermore, the scoring in C3 is strongly biased towards crystal scores, as C3 users almost invariably use the *See3* viewing software to cherry-pick interesting images.

2. *Cinder*

To ease the scoring of crystallization images, we have developed a mobile application that uses swipe technology to allow our users to swiftly classify images. The application, *Cinder*, is free and is available for both Android (downloadable from GooglePlay; <https://play.google.com/store/apps/details?id=au.csiro.cinder>) and iOS (downloadable from the Apple App Store; <https://itunes.apple.com/au/app/cinder/id1074115966?mt=8>) mobile devices. *Cinder* can be used to classify images of a static data set, or can be used by the facility user to scan recently imaged crystallization trials. The app has inbuilt training to allow novice users to produce reasonable scores rapidly.

We envisage that *Cinder* can be used by different groups in different ways. Laboratory heads and principal investigators could use *Cinder* to help to teach new students how to critically view and score crystallization images. Current users of a facility can use the app to 'pre-score' images whenever they

Figure 2

Some of the features seen in images of crystallization trials which may confound less experienced viewers. (a) Only features within a droplet are likely to be interesting (this image shows the hyphae which are the result of fungal contamination of the experimental droplet). (b) Plates may have knit lines or other features that are easy to confuse with crystals. (c) Reflections from lights may look like interesting features. (d) Many crystallization droplets are quite dusty and may contain fibres or other artefacts. (e) Salt crystals might look just like protein crystals. (f) Collapsed bubbles can look like a possible crystal

have a spare moment, and then use the coarse *Cinder* scoring to guide a more nuanced scoring session on a fully featured viewing platform. Finally, the app could be used to engage the broader community (other crystallization scientists, or just interested ‘citizen scientists’) to help to produce a well scored image data set for machine-learning applications.

3. *Cinder* application

Cinder is a mobile application that has three functions or modes (Fig. 3): *Cinder Solo* (score your own images), *Cinder Community* (score a general set of images) and *Cinder Kinder* (learn to score images). Each mode shows the user a single image and provides the option of classifying it into one of four classes: ‘Clear’, ‘Precipitate’, ‘Crystal’ or ‘Other’. The score is assigned by swiping in one of four directions.

- (i) Clear: right to left swipe.
- (ii) Precipitate: bottom to top swipe.
- (iii) Crystal: left to right swipe.
- (iv) Other: top to bottom swipe.

The swipe directions are shown as arrows over the image-display area; the swipe-direction indication arrows are also active buttons that can be pressed to achieve the same result as swiping. Zooming in to look at the fine details of an image is possible, but requires pressing the ‘DETAILS’ button, as having four active swipe directions precludes the use of the

well established ‘pinch zooming’ unless the swipe functions are disabled. There is also a ‘SKIP’ button, which allows a user to move on to the next image without scoring the currently loaded image. As we are trying to nudge people to score all images rather than just selecting images of drops containing crystals, we have deliberately added a small (2 s) time delay to this function. Thus, it is quicker to score an image than to skip it, although the time penalty is trivial enough that an image that really does defy classification can be skipped easily enough.

3.1. *Cinder Kinder*

In this mode, *Cinder* is a training tool: the user is presented with an image from a library of pre-annotated images. The user decides to which of the four categories the image belongs and swipes (or presses) appropriately. If the classification is incorrect, the incorrectly chosen button is coloured red and the message ‘Incorrect’ is given. When the right classification is chosen, the correct button is coloured green and a short description of why that category was correct is presented to the user (Fig. 3c). Pressing ‘NEXT EXAMPLE’ will load a new annotated image. Currently there are about 200 annotated images, some from Rigaku Minstrel imagers and some from Formulatrix ROCK IMAGERs, and a small number of images of LCP sandwich droplets collected using a Formulatrix imager. The LCP sandwich droplets were set up in

Figure 3
Cinder options. (a) Screenshot of options: the setup menu (the cog in the upper right corner) is where different APIs can be selected. (b) The *Cinder* main screen, showing buttons and some associated metadata. The buttons which indicate the swipe directions are coloured to match the colour palettes associated with these four superclasses. (c) *Cinder Kinder* example, showing the screen after the correct score has been chosen: the correct score (Crystal) is highlighted in green and a description of the droplet is displayed below the image.

Laminex plates (glass bases, 100 μm spacers, 200 μm plastic seal; all from Molecular Dimensions, UK) and were set up with 100 nl LCP surrounded by 1 μl crystallization solution. The other images are of sitting-drop experiments set up in SD-2 plates (Molecular Dimensions, UK), with initial drop volumes of 300–500 nl.

3.2. *Cinder Community*

In this mode the user is presented with images from a static library of images and the scores are sent back into an associated database table (see Supplementary Fig. S1). The images are chosen randomly and presented to the user. Currently this function is coupled to a library of about 12 500 images of sitting-drop trials.

3.3. *Cinder Solo*

The *Cinder Solo* option is available for users of the C3 facility to view recent images of droplets set up for them in C3. Generally, only the user that submits a sample to C3 can view the results (images) from that sample; access to the *See3* viewing software requires the user to provide both a username and password. The *Cinder Solo* option requires the user to provide their *See3* username and password. On successful authentication the user is presented with unscored images from their inspections (an inspection is a set of images collected from the same crystallization plate at the same time) acquired in the last week. If no new images have been collected within the last 7 d, older unscored images are loaded for scoring. Scores assigned by the user swiping (or pressing) are pushed back into the C3 crystallization database and are associated with the user's images in the *See3* viewing software. The scores are identified as coming from the *Cinder* application by defining a separate score type of 'ManualMobile' which is associated only with *Cinder* scores.

4. *Cinder* technology

There are three parts of the *Cinder* application: the application itself, an application programming interface (API) and a local (server-side) database where images and associated metadata for the *Cinder Community* and *Cinder Kinder* functions are stored. The information for *Cinder Solo* is pulled and pushed from the Oracle database which underpins the whole C3 facility. The authentication API is only used for the *Cinder Solo* component, and manages user authentication and other interactions with the C3 crystallization database.

The application was written with the goal of having other (non-C3) laboratories access the app: the images/annotations that support the *Cinder Kinder* function are likely to be general enough to be used directly by most crystallization laboratories, and the current images used for the *Cinder Community* function could potentially be replaced or extended with other images. The *Cinder Solo* function would require that the default calls to the central C3 database are replaced by the appropriate calls to the non-C3 crystallization

database, and the *Cinder* API architecture has been designed with this use case in mind.

4.1. *Cinder* application

The application is written as a website in HTML, CSS and JavaScript and utilizes the freely available, open-source Apache Cordova (<https://cordova.apache.org/>) tool to target multiple platforms. Apache Cordova runs a contained website as an application on Android or iOS. The application in its website form can easily be run and tested on a computer web browser before being exported as a phone application. Apache Cordova utilizes the existing phone webpage-rendering engine, and thus ensures that the display of the webpage is as efficient and consistent as the tested and reliable underlying Android or iOS rendering engine.

The code has been structured to maximize readability and maintainability, and has a distinct separation between views, styling and computation (HTML, CSS and JavaScript), and an appropriate grouping of JavaScript modules which perform similar tasks. Extensibility is similarly ensured with a consistent module structure and low module coupling.

4.2. *Cinder* API

The application obtains the appropriate images and their associated data by making calls to a HTTPS-based API. The API is written in C# and uses the Microsoft.NET web framework (<https://www.microsoft.com/net/>). The .NET framework allows easy deployment to a Microsoft Windows server running Microsoft Internet Information Services (IIS). The API gathers the appropriate data for the three different modes and presents either 1000 random images from the *Cinder Community* set for the *Cinder Community* mode, all of the images from the *Cinder Kinder* set in a randomized order for the *Cinder Kinder* mode, or images from the most recent inspections within the last week for the *Cinder Solo* mode. Images/scores are not cached and *Cinder* requires connectivity to function.

The *Cinder* API uses the 'Model View Controller' (MVC) design-pattern approach which should enable any future addition of more functionality without the need to change existing functions. Access to both the *Cinder* database, the central C3 database and the image-storage location is abstracted to separate helper classes, which will allow access to be easily updated if any database or storage changes are made. The API specification is included as Supporting Information, along with a diagram of the *Cinder* tables (Supplementary Fig. S1) and a diagram of the relevant subset of the central C3 database (Supplementary Fig. S2).

4.3. *Cinder* database

The data for the *Cinder Solo* mode are obtained by querying the central C3 database; however, the data for the *Cinder Community* and *Cinder Kinder* modes are stored in a standalone *Cinder* database. The *Cinder* database uses a PostgreSQL database server, which was chosen because it is both powerful and freely available. It contains tables with

information about the images used in the *Cinder Community* and *Cinder Kinder* modes, as well as a table for the *Cinder Kinder* responses and a scores table that captures the scores generated by users in the *Cinder Community* mode.

5. Results and discussion

Initially, *Cinder* was written with the aim of using crowd-sourcing to generate a set of consensus scores for a set of images that were going to be the definitive C3 training set for machine learning. This set of images (the ‘well scored’ data set) consists of about 12 500 images that had been manually scored by one experienced crystallographer, with the rationale behind consensus scoring being to remove any scoring bias from the set introduced by having only one person score all of the images. *Cinder* (which borrowed the idea of swiping as a sorting mechanism from a popular dating application, thus ‘*Crystallographic Tinder*’) would hopefully generate 10–100 scores for each image, and the ‘true’ score would be that most often picked by a large set of random scorers. There were two problems with this: firstly, random scorers had no idea how to score at all, and secondly, we underestimated how dull the process of assigning scores to images really is. We addressed the first problem by introducing the *Cinder Kinder* mode, which is a training tool for teaching novices the fundamentals of the interpretation of crystallization images. The second

problem, that scoring is intrinsically very dull, was harder to solve. One approach would be to ‘gamify’ *Cinder*, so that the *Cinder* community essentially competes against each other. This would add quite a lot of complexity to the app: collecting user information, maintaining stats and working out what the rewards would be for being a good scorer. At the same time, we started to appreciate some of the weaknesses of the ‘well scored’ data set that we were presenting to the *Cinder* community. Essentially, the images were simply not diverse enough: all of the 12 500 images had been collected with Rigaku Minstrel imagers from the same plate type using the same type of camera. All were greyscale images. All of the images were collected within a year of each other, and probably were images from no more than 50 different protein samples and a limited set of crystallization conditions. Even in C3 the ‘well scored’ data set became less relevant; after the C3 Rigaku imagers had been replaced by ROCK IMAGERS (Formulatrix) the C3 crystallization images looked quite different from the images in the Rigaku-based training set. Given that the ‘well scored’ data set was never going to be the ultimate training set for these (and probably other) reasons, we decided to change our approach to obtaining training data and just use any human-scored image for our machine-learning studies. Our current approach to obtaining training data is to scrape up all of the scores that have been given to images by C3 users and hope that any scoring inconsistencies

Score	Colour	Description	Superclass
Shoot me		In plate screening marker	Crystal
Crystals ***		Worth harvesting	Crystal
Crystals **		Needs some work, but very promising	Crystal
Crystals *		too many, too small, needs a lot of optimising	Crystal
Crystalline		Microcrystalline, optimise only if desperate	Crystal
Cinder Crystal			Crystal
Spherulites		Things that look like meteors	Other
Other		Unknown	Other
Cinder Other			Other
Salt xtal?		Putative salt crystals	Crystal
Good precipitate		Grainy, shiny, birefringent	Precipitate
Precipitate		All other precipitates	Precipitate
Cinder Precipitate			Precipitate
Bad precipitate		Brown, smoky, denatured	Precipitate
Phase separation		Oily, immiscible	Other
Clear		Clear droplet	Clear
Clear with stuff		Clear drops with dust, fibres, or other non-crystalline matter	Other
Cinder Clear			Clear
Confirmed Salt		Confirmed by UV or diffraction	Crystal
Skin		Wrinkled surface due to protein or polymer skin	Other
Dried		Drop remains after harvesting crystals	Other
Null experiment		This experiment didn't work (dried, not put down well)	Other

Figure 4

Scoring scheme as offered in C3. There are many more scores, but they can all be collapsed into the four superclasses Clear, Precipitate, Crystal or Other. There are colours associated with each score. The Crystal superclass is represented by the warm colours yellow to red. The Precipitate class is represented by the blue palette, the Other superclass by greens and the Clear superclass is shown in grey.

will be diluted out by using more data. This approach of using as much data as possible is a particularly valid approach for nonlinear learning algorithms, for example many-layered neural nets (Halevy *et al.*, 2009), and recent work has shown that there is a logarithmic improvement (in particular with vision tasks) with an increasing volume of training data (Sun *et al.*, 2017).

5.1. Scores and learning from classified data

As yet there is no defined vocabulary for macromolecular crystallization experiments in general: no standards for chemical naming, volumes, units, experimental details or experimental outcomes (Luft *et al.*, 2011; Newman *et al.*, 2013, 2014). Although the need to categorize images is widely understood (Newman *et al.*, 2012), there is no similar consensus on the appropriate number of classes, what those classes are and how to define each class. Of the scoring schemes that are recognized, such as the nine-class scheme suggested by Hampton Research (Table 1), the majority of the classes describe crystals, with 'better' crystals being larger and more three-dimensional. In reality this is only partially true, as it is the quality of the diffraction from a crystal when exposed to X-rays which is the true determinant of crystal quality.

Further, visual 'goodness' of crystals is highly user- and project-dependent. A crystallographer with decades of experience in harvesting crystals from droplets for X-ray diffraction experiments might be much more generous in their definition of which crystals are suitable for X-ray studies than a shaky-handed beginner. Many crystallographic projects simply do not produce the large, three-dimensional single crystals that are commonly grown from the standard test proteins (for example lysozyme, as shown in Figs. 1 and 3*b*).

Figure 5

An image of a crystallization trial which shows light precipitate, spherulites and microcrystals, and has been scored as containing all three. The coloured border shows this: yellow indicates microcrystals, green indicates spherulites and blue indicates precipitate. The coloured border is drawn starting at 9 o'clock, with the colours arranged according to a defined scheme (captured in the *score_type* table in the C3 database). Currently the arrangement shows any Crystal result before any score that falls into the Other category, followed by Precipitate categories. Thus, a glance at the top left-hand corner of a multi-scored image shows the most relevant results for locating crystals.

In C3 we offer a limited set of classes that users can associate with an image of a crystallization trial through the *See3* web tool. All of the scores can be mapped back to the four more fundamental scores which can be assigned through the *Cinder* app (see Fig. 4). Limited scoring sets often means that the scores represent an assessment of the outcome, not just a description of the image. Thus, an image that shows a clear droplet with a fibre overlaid is likely to be classified by a crystallographer as being 'Clear', as the experienced scorer will appreciate that the fibre is extraneous to the experiment. However, in terms of machine learning the image is not clear (it has a fibre in it); thus, the C3 score 'clear with stuff' is mapped to the 'Other' superclass rather than to the 'Clear' superclass. In the same way, a droplet that contains both precipitate and crystals would be scored by a crystallographer as 'Crystal', even though the crystal is only a minor component of the overall outcome of the trial. To overcome this, the C3 users may associate one or more score classes with an image: this enables nuanced scoring without a concomitant explosion in the number of outcome classes. All of the scores associated with an image are indicated in the border colour of the drop image in the *See3* interface, allowing the granular scoring to be recognized very quickly (Fig. 5).

We know that much of the scoring performed by C3 users is to identify crystals, and thus the rate of crystal growth cannot be estimated by simply comparing the number of scores that reflect a crystal outcome with the number of scores that do not. This certainly holds true for scores which are associated with images through the *See3* interface, where the image data presentation makes it easy to cherry-pick which images are scored. The *Cinder* app makes it much harder to cherry-pick 'interesting' images, so we believe that the ratio of crystal scores assigned through *Cinder* is a better representation of the real proportion of droplets containing crystalline matter. Fig. 6 compares the scoring outcomes through the *See3* interface (where the scores have been mapped to the four fundamental classes using the mapping shown in Fig. 4) and the *Cinder* scores. There are almost three times as many crystal outcomes in the *See3* data as in the *Cinder* scores (28.4% versus 11.3%, respectively). Filtering the data to look at scores that have only been associated with initial experiments, the bias towards scoring crystals becomes even clearer. We believe that the 4% of images scored as 'Crystal' using the *Cinder* app is the best estimate that we have of the initial success rate of crystallization trials in C3, and this number agrees well with values from the literature, which give estimates of <1 to 10% for the initial hit rate (Newman *et al.*, 2012; Ng *et al.*, 2016).

Overall, as of March 2018 there were ~600 000 user scores (non-*Cinder*) and ~20 000 *Cinder* scores in the C3 database. In the seven months since the *Cinder* app was released there have been ~34 000 scores from the *See3* web tool, so that the overall rate of scoring has increased by 60% by the addition of the *Cinder* tool as a scoring interface. The introduction of the *Cinder* application does not seem to have negatively affected the rate of scoring through the web interface; over the same seven-month period in the previous two years there were

Figure 6

The ratio of four different classes of scores (Clear, Precipitate, Crystal and Other) for scores in C3 assigned through (a) *See3* (a web-based interface) and (b) the *Cinder* mobile application. These graphs represent all of the scoring data in our crystallization database, so that the *See3* data represent 12 years of user scoring and the *Cinder* data represent less than three months of scoring. The graphs shown in (c) for *See3* and (d) for *Cinder* show only those scores associated with initial screening experiments. We postulate that the large difference in the score distribution between the *See3* scores and the *Cinder* scores reflects the selective scoring of 'interesting' images through the web viewing tool and thus the over-representation of crystal scores from that source. The pie charts are coloured in the standard C3 colouring system, where Crystal outcomes are coloured in warm colours (yellow to red), Precipitate is coloured blue, Clear is grey and Other scores are green. The slight differences in the colours between the two pie charts [(a) and (c) versus (b) and (d)] reflect the colours associated with the different scoring systems (*See3* versus *Cinder*) and which are used to colour the borders of images in the *See3* web application. The colour variation allows the user to tell how the image was scored at a glance, as well as what the score was.

~20 000 and ~26 000 scores recorded (and these necessarily were associated with the *See3* tool).

Inevitably, the crystallography community will adopt machine scoring of images and move away from human inspections for crystal identification. However, looking at the data (rather than just relying on machine classification) will still be important to find the most appropriate crystals for further analysis or optimization. Rapid scoring (or perhaps

'conscious viewing') through *Cinder* may be a way of experiencing image data even in the era of machine learning.

6. Conclusions

A free application has been created for both Android and iOS mobile devices which allows the rapid scoring of crystallization images. The *Cinder* application offers a common training mode, as well as a solo mode which allows users to score their own images. The application has been designed and built to allow it to be harnessed to any crystallization database through an intermediate API. Early results show that the *Cinder* app increases the rate of human scoring compared with the more traditional viewing platform by an estimated 60%.

Acknowledgements

We thank the users of the C3 facility for testing the *Cinder* application.

References

- Buchala, S. & Wilson, J. C. (2008). *Acta Cryst.* **D64**, 823–833.
- Cumbaa, C. & Jurisica, I. (2005). *J. Struct. Funct. Genomics*, **6**, 195–202.
- Halevy, A., Norvig, P. & Pereira, F. (2009). *IEEE Intell. Syst.* **24**, 8–12.
- Luft, J. R., Wolfley, J. R. & Snell, E. H. (2011). *Cryst. Growth Des.* **11**, 651–663.
- Newman, J., Bolton, E. E., Müller-Dieckmann, J., Fazio, V. J., Gallagher, D. T., Lovell, D., Luft, J. R., Peat, T. S., Ratcliffe, D., Sayle, R. A., Snell, E. H., Taylor, K., Vallotton, P., Velanker, S. & von Delft, F. (2012). *Acta Cryst.* **F68**, 253–258.
- Newman, J., Burton, D. R., Caria, S., Desbois, S., Gee, C. L., Fazio, V. J., Kvasnakul, M., Marshall, B., Mills, G., Richter, V., Seabrook, S. A., Wu, M. & Peat, T. S. (2013). *Acta Cryst.* **F69**, 712–718.
- Newman, J., Peat, T. S. & Savage, G. P. (2014). *Aust. J. Chem.* **67**, 1813.
- Ng, J. T., Dekker, C., Kroemer, M., Osborne, M. & von Delft, F. (2014). *Acta Cryst.* **D70**, 2702–2718.
- Ng, J. T., Dekker, C., Reardon, P. & von Delft, F. (2016). *Acta Cryst.* **D72**, 224–235.
- Snell, E. H. *et al.* (2008). *Acta Cryst.* **D64**, 1123–1130.
- Spraggon, G., Lesley, S. A., Kreusch, A. & Priestle, J. P. (2002). *Acta Cryst.* **D58**, 1915–1923.
- Sukhbaatar, S., Bruna, J., Paluri, M., Bourdev, L. & Fergus, R. (2015). *arXiv:1406.2080*. <https://arxiv.org/abs/1406.2080>.
- Sun, C., Shrivastava, A., Singh, S. & Gupta, A. (2017). *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 843–852. Piscataway: IEEE.
- Weiss, G. M. & Provost, F. (2001). *The Effect of Class Distribution on Classifier Learning: An Empirical Study*. Technical Report ML-TR-44, Department of Computer Science, Rutgers University, New Jersey, USA.
- Wilson, J. (2006). *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, edited by P. Perner, pp. 450–473. Berlin, Heidelberg: Springer.