



Article

Tools to Ease the Choice and Design of Protein Crystallisation Experiments

Nicholas Rosa ^{1,†}, Marko Ristic ^{1,‡}, Luke Thorburn ¹, Gabriel J. Abrahams ¹, Bevan Marshall ¹, Christopher J. Watkins ², Alex Kruger ², Alex Khassapov ² and Janet Newman ^{1,*}

¹ Biomedical Manufacturing, CSIRO Parkville, 343 Royal Parade, Parkville Vic 3052, Australia; nicholas.rosa@monash.edu (N.R.); marko.ristic@kit.edu (M.R.); luke.thorburn@gmail.com (L.T.); abrahams.gabi@gmail.com (G.J.A.); bevan.marshall@csiro.au (B.M.)

² Scientific Computing, CSIRO Clayton, Research Way, Clayton Vic 3168, Australia; chris.watkins@csiro.au (C.J.W.); alex-k@live.com.au (A.K.); alex.khassapov@csiro.au (A.K.)

* Correspondence: janet.newman@csiro.au; Tel.: +61-03-9662-7326

† Current address: Department of Engineering, Monash University, Wellington Rd, Clayton ViC 3800, Australia.

‡ Current address: Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany.

Received: 16 January 2020; Accepted: 31 January 2020; Published: 7 February 2020



Abstract: The process of macromolecular crystallisation almost always begins by setting up crystallisation trials using commercial or other premade screens, followed by cycles of optimisation where the crystallisation cocktails are focused towards a particular small region of chemical space. The screening process is relatively straightforward, but still requires an understanding of the plethora of commercially available screens. Optimisation is complicated by requiring both the design and preparation of the appropriate secondary screens. Software has been developed in the C3 lab to aid the process of choosing initial screens, to analyse the results of the initial trials, and to design and describe how to prepare optimisation screens.

Keywords: crystallization cocktail; crystallization screen; crystallization optimisation; similarity

1. Introduction

There are a number of techniques which can produce atomic resolution structures of macromolecules (for example, proteins), but by the far the most productive technique to date is X-ray crystallography [1], whereby a crystal of the sample of interest is irradiated with high-energy X-rays, and the resultant diffraction pattern deconvoluted to give an image of the molecules making up the crystalline sample [2]. The production of ‘diffraction quality’ crystals is required for the generation of atomic resolution structures by X-ray crystallography, and this is the major limitation of the technique [3].

Although there are many techniques used for growing crystals [4], most crystals are grown using the method of vapour diffusion, in which a droplet of pure, concentrated protein is mixed with a droplet of a chemical mixture (also called the *condition*, or the *cocktail*), and the combined droplet is allowed to equilibrate against a larger volume of the cocktail. These experiments can be hanging drop experiments (often set up laboriously by hand, in 24 well arrays) or sitting drop experiments (which are more likely to be set up with the help of automation, in 96 well arrays). The screening experiments most often use standard sets of conditions, ‘commercial screens’ which are widely available through a number of different vendors. Many different commercial screens exist. The variables for the screening step are quite limited—what screen(s) to use, what protein concentration to test, how much of the protein sample to include in the experimental drop, how much of the condition to include (i.e., the ratio

of these two liquids), and the temperature at which to incubate the experiments. Each experimental droplet must be examined, over a number of days or weeks for indications of a crystalline outcome. If a suitable crystal is observed it can be harvested from the screening droplet, but most often several cycles of optimisation are needed before a suitable crystal is obtained.

Optimisation can be done in many different ways [5], but almost all optimisation requires the design and creation of a focused screen around conditions that appear to be promising from the initial screening. This might be a screen using only chemicals found in 'hits' from the screening (often called fine screening), or it might be a single cocktail, which is dispensed 96 times and where each well is tweaked by the addition of a different chemical (additive screening). Optimisation requires more decision making than screening, even putting aside the often very difficult question of what screening condition(s) should be included in subsequent optimisation. Each cocktail consists of one or more chemical factors, that is, a chemical with an associated concentration, concentration unit and possibly pH value. In optimisation fine screening, one has to choose which chemicals to include, what concentration, or concentration range, to use for each chemical, and possibly which pH or range of pH values to be tested. On top of this there are orthogonal considerations that interfere with the design process, such as the solubility and/or buffering range of a chemical, or the possibility of unwanted interactions. For example, mixing solutions of calcium chloride and ammonium sulfate, two commonly used chemicals in protein crystallisation, will inevitably lead to the formation of calcium sulfate crystals, confounding the interpretation of the optimisation outcome.

There are a number of applications that estimate the likelihood of a successful protein crystallisation given a sequence [6–10], but there are few software tools available that help to clarify what to actually do in the laboratory to make the protein crystallise. A common approach is to do a search to see if the target of interest has been crystallised before, and use the published conditions as a starting point. If the target of interest has not been crystallised, then often the advice is to search for (and then start screening with) crystallisation conditions used to crystallise similar proteins (perhaps in the Biological Macromolecule Crystallization Database (BMCD; <http://bmcd.ibbr.umd.edu/>)) [11], although recently the soundness of that approach has been questioned [12]. However, even if the protein target has been crystallised previously, the literature conditions tend only to be a rough guide—subtle differences in the exact construct and/or expression and purification can alter how the protein target behaves in crystallisation experiments. Nevertheless, published conditions can narrow down the screening process—for example, by suggesting that trialling conditions containing mid-range polyethylene glycol (PEG) might be appropriate.

The Collaborative Crystallisation Centre (C3; <http://crystal.csiro.au/>) is a full service protein crystallisation facility operated by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Parkville, Australia. It was set up in the early 2000s to give the local structural biology community access to crystallisation automation, but has extended its mandate since then to provide novel software tools to facilitate the process of growing macromolecular crystals. To make the software tools accessible, the tools are web-based, and include manuals on how to use the various features of the programs. This report aims to give an flavour of what the tools can do, rather than describe in detail how each feature works.

One of the first software tools developed in the C3 was the C6 webtool [13]. C6 (<https://c6.csiro.au/>) was designed over ten years ago to provide a tool that would answer questions such as 'Which commercial screen contains the most PEG 8000?', or 'What range of ammonium sulfate concentrations are found in Crystal Screen HT from Hampton Research?' The web server PICKScreens [14] (<https://registration.cc.biophys.mpg.de/pickscreens/>) had a similar mandate. The C6 tool tries to capture information about all the commercial screens in a format to allow these types of queries. Of course, some of this information could be found by navigating to each vendor's website in turn, but the advantage of the C6 tool was the translation of all of the screen information provided by the vendor into a common vocabulary [15], which enabled the development of a heuristic distance measurement between individual conditions and screens [13]. Formalising the concept of similarity has been found

useful by other groups since as well [16]. There were 220 commercial screens captured in the original C6 tool (compared to the 129 in the PICKScreens database), since then over 100 new commercial screens have been added. A further function of the original C6 tool was to provide a list of what screens were physically present in C3, and thus available for use by the C3 user community. Along with new commercial screens, other features and improvements have been added to the original tool. The most fundamental was a $\approx 1000\times$ increase in the speed of generating the underlying database, with a concomitant increase in the real time utility of the tool. In this paper we describe five of the more novel/useful new features of the C6 program.

More recently, we have developed a viewing, scoring and optimisation tool 'See3'. See3 (<https://see3.csiro.au/>) incorporates a machine learning (ML) tool (MARCO) [17] which runs in real time and returns one of four possible classifications to images of the crystal experiment (*Clear, Precipitate, Crystal* or *Other*). The MARCO tool was generated using a multi-layered neural network (deep learning) using a large set of training images of $\approx 500,000$ human scored collected from a consortium of academic and industrial crystallisation groups. The generated classification is estimated to be about 90% accurate for the images collected in C3. The MARCO scores, along with manual human scoring, can be used to generate fine screens for optimisation.

Overview of the C6 and See3 Programs

The See3 program is a web-based application for the viewing, scoring and optimisation of crystallisation experiments which have been set up within C3. The C6 program is a tool for the analysis and comparison of crystallisation screens, both those commercially available and those designed in the See3 application. The two applications both rely on information from the 'CM' crystallisation database in C3. This is an OracleTM12c database, which started as the 'Crystal' database provided by the (now defunct) company RoboDesign (then Rigaku) with their Minstrel imaging system. Although C3 migrated from the Minstrel imagers in 2016, C3 continues to use the CM database, thus preserving ten years of legacy data, and uses a Dynamic-link library (DLL) to translate between the CM schema and the dataforms required to run the current (Formulatrix RI-1000) imagers. The CM database captures information about the screens (and the cocktails that make up the screens), samples, users, images, scores, as well as tracking the barcoded deepwell blocks and experimental crystallisation plates used in C3. Both See3 and C6 have public information (commercial screens and screens offered by C3) and private information (any data associated with a particular user). The latter is password protected and only visible to the authorised users.

2. C6

The C6 program is a webtool which catalogues and compares crystallisation conditions. The tool contains (most of) the commercial crystallisation screens, along with screens designed by users of the C3 crystallisation facility. The data in the C6 SQLite database is derived from the screen information in the CM database; along with the chemical description of each cocktail in every screen obtained from the CM database, the C6 database contains a matrix of distances between every pair of crystallisation cocktails [13]. The matrix of distances allows the calculation of screen (dis)similarity. This, for example, allows for the trivial identification of commercial screens which are essentially the same, which is hard to do by hand, as each vendor may have the same set of conditions in a screen, but ordered (or even spelled) differently (Figure 1).

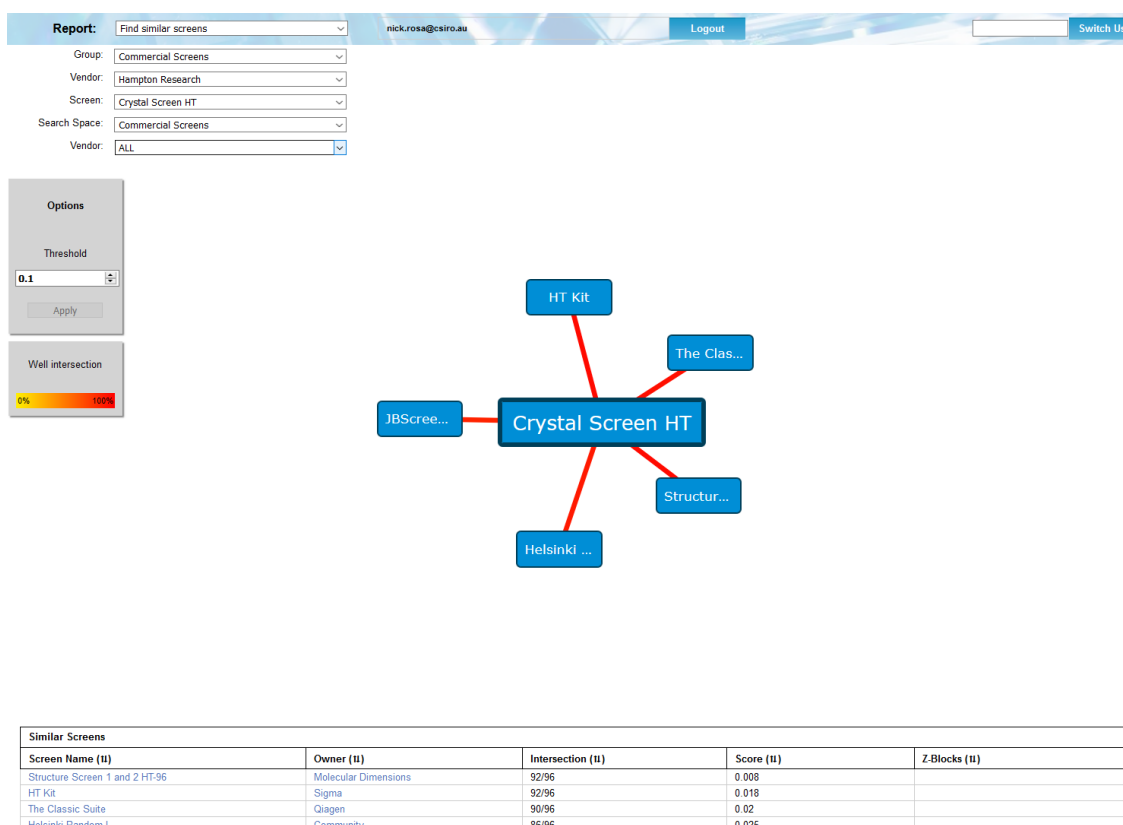


Figure 1. Using the ‘Find similar screens’ report, it can be seen that the screens *Crystal Screen HT* (Hampton Research, Aliso Viejo, CA, USA), *Structure Screen 1 and 2 HT-96* (Molecular Dimensions), *HT kit* (Sigma, St. Louis, MO, USA), *The Classics Suite* (Qiagen, Hilden, Germany), *JBScreen Basic HTS* (Jena BioSciences, Jena, Germany) along with the *Helsinki Random I* screen (a non-commercial ‘community’ screen from Helsinki University) are all essentially the same, as they each have pairwise screen scores (‘Score’) of < 0.1. A screen (dis)similarity score between two screens can be generated by using the minimum paired distance of all the conditions in two screens. Using this approach, two screens with every condition in common will have a screen (dis)similarity score of 0. A pair of screens where none of the conditions in either screen have any chemicals in common will have a screen pair score of 1.

As in the earlier version [13], each screen in C6 can be displayed as a set of cocktails, or as a list of the component chemicals, with usage count, concentration and pH ranges. The interface has been updated, and now groups the available reports into logical classes. Under each section one can choose from a predefined subset of all screens, e.g., ‘C3 screens’, or ‘Commercial Screens from Hampton Research’. C6 also allows for the creation of both persistent custom subsets or on-the-fly custom subsets of screens.

Within each section the queries can be filtered by subsets of screens, such as only those screens available within C3, or only those screens sold by the vendor Hampton Research. For example, if one wanted to see which screens are currently available from Molecular Dimensions, one would select the report ‘Screens Lists’, and choose the filter ‘Commercial Screens’, then the filter ‘Molecular Dimensions’.

2.1. Interface

C6 now breaks down the available reports into four sections: ‘Screens and Stocks’, ‘Compare’, ‘Search’ and ‘Tools’ (Figure 2).

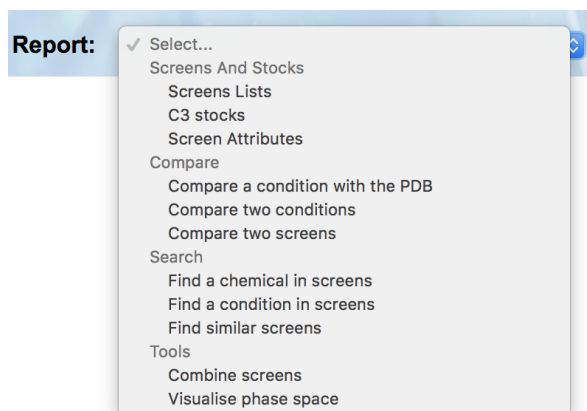


Figure 2. The C6 reports are now grouped into four logical sections: 'Screens and Stocks', 'Compare', 'Search' and 'Tools'.

Once the appropriate filters have been selected a list of screens is returned. Each screen name is selectable, and clicking on a screen name will open a new browser tab containing a header containing information about the screen, followed by a listing of the contents of each well of the screen (Figure 3).

The screenshot shows a web interface with a 'Report:' dropdown set to 'Screens Lists' and a 'Group:' dropdown set to 'C3 Screens'. Below this is a table titled 'Screens (PROPERTIES AVAILABLE)'. The table has columns for 'Screen Name', 'Screen Owner', 'Create date', and 'Barcodes'. The table contains 25 rows of screen data.

Screen Name	Screen Owner	Create date	Barcodes
Buffer Screen 10g	c3@csiro.au	2019-12-16	Z0002439
Buffer_Screen_10	c3@csiro.au	2019-08-30	Z0002430, Z0002370
c3_1	c3@csiro.au	2018-11-01	Z0002324, Z0002219
c3_2	c3@csiro.au	2019-04-10	Z0002292, Z0002291
c3_3	c3@csiro.au	2019-03-15	Z0002381, Z0002275
c3_4	c3@csiro.au	2019-12-11	Z0002435, Z0002318
c3_5	c3@csiro.au	2019-11-25	Z0002429, Z0002353
c3_6	c3@csiro.au	2019-05-07	Z0002434, Z0002313
c3_7	c3@csiro.au	2019-10-03	Z0002393, Z0002391, Z0002235
c3_8	c3@csiro.au	2019-10-01	Z0002392, Z0002256
HELIX_C3	c3@csiro.au	2018-08-17	Z0002191
Index_C3	c3@csiro.au	2019-10-04	Z0002395, Z0002293
LCP_C3	c3@csiro.au	2019-09-16	Z0002384, Z0002146
Lysozyme QA Screen	c3@csiro.au	2019-07-10	Z0002441, Z0002349, Z0002264
MemGold2_C3	c3@csiro.au	2019-05-10	Z0002316, Z0002302
MemGold_C3	c3@csiro.au	2019-02-15	Z0002261, Z0001785
Morpheus_C3	c3@csiro.au	2019-05-20	Z0002320, Z0002140
PACT_C3	c3@csiro.au	2019-06-20	Z0002336, Z0002335
Paydirt_1	c3@csiro.au	2019-07-03	Z0002345
PI_PI2_C3	c3@csiro.au	2014-11-24	Z0001576, Z0001575
ProPlex_C3	c3@csiro.au	2019-09-09	Z0002380, Z0002299
PS gradient	c3@csiro.au	2019-06-24	Z0002337, Z0002317
PS gradient - mid range	c3@csiro.au	2019-02-01	Z0002252
Shotgun	c3@csiro.au	2019-11-21	Z0002440, Z0002426, Z0002425
Solubility and Stability Screen 3	c3@csiro.au	2016-08-22	Z0001876

Figure 3. The screens returned by selecting 'C3 screens' as a filter in the report 'Screens Lists.' The names of the screens in the list are links to the screen content. If the name 'Shotgun' (second screen from the bottom) is clicked, then a new tab is opened and the wells of the Shotgun screen are shown (Figure 4).

Clicking on the screen name will open up a new browser tab containing a listing of the chemicals found in each well. If the cocktail can be made with C3 stocks (i.e., the chemicals are named appropriately, and C3 has stocks of appropriate concentration and/or pH) then a recipe of how that condition can be made is also provided (Figure 4).

NB: Buffers are shown in <i>italic font</i>			
Well	Concentration Units Chemical, pH	pH	Recipe
A1 (1)	0.2 M magnesium chloride; 30 w/v polyethylene glycol 4000; 0.1 M tris chloride, pH=8.5;	8.1	> A1 200.0 ul of: magnesium chloride (1M) 600.0 ul of: poly(ethylene glycol) 4000 (50% w/v) 21.2 ul of: tris chloride pH 7.0 (1M) 78.8 ul of: tris chloride pH 9.0 (1M) 100.0 ul of: H2O
A2 (2)	2 M ammonium sulfate;	5.7	> A2 571.4 ul of: ammonium sulfate (3.5M) 428.6 ul of: H2O
A3 (3)	20 w/v polyethylene glycol 3350; 0.2 M sodium acetate;	7.3	> A3 400.0 ul of: poly(ethylene glycol) 3350 (50% w/v) 200.0 ul of: sodium acetate (1M) 400.0 ul of: H2O
A4 (4)	2 M ammonium sulfate; 0.1 M tris chloride, pH=8.5;	7.9	> A4 21.2 ul of: tris chloride pH 7.0 (1M) 571.4 ul of: ammonium sulfate (3.5M) 78.8 ul of: tris chloride pH 9.0 (1M) 328.6 ul of: H2O
A5 (5)	20 w/v polyethylene glycol 3350; 0.2 M trisodium citrate;	8.1	> A5 400.0 ul of: poly(ethylene glycol) 3350 (50% w/v) 125.0 ul of: (tri)sodium citrate (1.6M) 475.0 ul of: H2O

Figure 4. The contents of the Shotgun screen. The left hand cells show each condition in terms of concentration, and the right hand wells show how the condition can be made using C3 stocks. The ‘pH’ column contains the average pH value of the condition; the pH is measured each time a new screen is created [18].

The report ‘C3 Stocks’ found under the section ‘Screens and Stocks’ gives a listing of all the stocks currently available in C3 (Figure 5).

ammonium acetate	ammonium acetate (5M)	5	M	C2H7NO2
ammonium chloride	ammonium chloride (1M)	1	M	NH4Cl
ammonium chloride	ammonium chloride (5M)	5	M	NH4Cl
ammonium dihydrogen phosphate	ammonium dihydrogen phosphate (2.5M)	2.5	M	NH4H2PO4
ammonium fluoride	ammonium fluoride (1M)	1	M	NH4F
ammonium formate	ammonium formate (1M)	1	M	NH4CHO2
ammonium formate	ammonium formate (4M)	4	M	NH4CHO2
ammonium iodide	ammonium iodide (1M)	1	M	NH4I
ammonium iodide	ammonium iodide (5M)	5	M	NH4I
ammonium nitrate	ammonium nitrate (1M)	1	M	NH4NO3
ammonium nitrate	ammonium nitrate (8M)	8	M	NH4NO3
ammonium sulfate	ammonium sulfate (1M)	1	M	(NH4)2SO4
ammonium sulfate	ammonium sulfate (3.5M)	3.5	M	(NH4)2SO4
ammonium thiocyanate	ammonium thiocyanate (1.5M)	1.5	M	NH4SCN

Figure 5. The stocks report shows the stocks available in C3, shown here are just the subset of stocks containing ammonium. The chemical name in the stocks list is a link, clicking on a chemical name will show the C3 screens that contain that chemical.

Another new feature of the interface is the inclusion of an updated User Manual and Frequently Asked Questions (FAQ) list, but more importantly, there are inbuilt tools that support the facile generation of new entries into the FAQ list. The landing page of the C6 tool, like most of the outward facing web-tools from C3 includes a Twitter feed, where tweets from the C3 Twitter account (@CSIROC3) are displayed. This provides a mechanism for notifications in real time.

2.2. Screen Attributes

One of the screen metrics that was available in the original C6 was the concept of “Internal Diversity” (ID). This is a global measure of how different each condition in a screen is from the other conditions in the same screen. ID is normalised to give a value between 0 and 1; screens with an ID of 0 have each condition in the screen identical to the others, screens with an ID of 1 consist of conditions with no common chemicals in them. The ID value of a screen can be used to identify what type of screen it is, and thus where to use it in a crystallisation campaign. Screens with low ID (<0.5) are often grid screens, or screens with few chemicals in them; these screens are most often optimisation tools (fine screens) and are used once initial conditions have been identified. Screens with very high ID (≈ 1) are most likely to be additive screens; these are used in the early stages of optimisation. Screens

with a high ID (≈ 0.9) are often the initial “sparse matrix” [19] screens and are used at the start of a crystallisation project. Other useful global features of a screen might be the pH range it samples, or the relative amount of salty or PEG-based conditions it contains. All these global indicators are available through the Screens Attribute report. This report returns a sortable table, as shown in Figure 6.

Table of Screen attributes [Right click and 'Save link as...' to download a csv file](#)

Internal diversity estimates the difference between the conditions within a screen: 0 if all conditions are identical, 1 if all conditions are completely different. % PEG gives percentage of conditions where the primary factor is a polyethylene glycol % Salt gives percentage of conditions where the primary factor is a salt							
(%)							
Screen Name (%)	Owner Name (%)	Internal Diversity (%)	% PEG (%)	% Salt (%)	Av. pH (%)	Wells within Av. pH +/- 1 (%)	Distinct chemicals (%)
PL_P12_C3	c3@csiro.au	0.685	100	0	6.4	71	67
Index_C3	c3@csiro.au	0.908	49	34	6.7	67	49
PACT_C3	c3@csiro.au	0.848	100	0	6.8	55	30
PS gradient	c3@csiro.au	0.775	0	94	7	68	7
PS gradient - mid range	c3@csiro.au	0.739	0	100	7	67	7
Lysozyme QA Screen	c3@csiro.au	0.758	50	50	6	0	11
ProPlex_C3	c3@csiro.au	0.92	69	21	6.6	57	42
Shotgun	c3@csiro.au	0.912	76	22	6.6	71	50
MemGold_C3	c3@csiro.au	0.924	86	7	6.8	48	59
MemGold2_C3	c3@csiro.au	0.94	91	5	6.5	66	62
Solubility and Stability Screen 3	c3@csiro.au	0.978	0	50	6.3	59	28
Morpheus_C3	c3@csiro.au	0.88	50	0	7.2	72	46
c3_1	c3@csiro.au	0.93	100	0	6.7	48	46
c3_2	c3@csiro.au	0.935	100	0	6.7	60	63
c3_3	c3@csiro.au	0.848	100	0	6.5	70	53
c3_4	c3@csiro.au	0.928	91	0	6.6	52	49
c3_7	c3@csiro.au	0.949	0	0	6.6	52	65
c3_5	c3@csiro.au	0.872	0	100	6.6	46	46
c3_8	c3@csiro.au	0.954	0	15	6.6	61	63
c3_6	c3@csiro.au	0.934	0	98	6.8	47	52
HELIX_C3	c3@csiro.au	0.892	54	15	6.5	84	34
LCP_C3	c3@csiro.au	0.881	91	3	6.4	60	74
Paydirt_1	c3@csiro.au	0.806	13	0	7.2	60	17
Buffer_Screen_10	c3@csiro.au	0.674	0	0	6.7	66	14
Buffer Screen 10g	c3@csiro.au	0.545	0	0	6.7	60	15

Figure 6. The Screen Attributes table for the crystallisation screens available in C3. The table can be sorted on any column. For the ‘% polyethylene glycol (PEG)’ and ‘% Salt’ values only the primary factor (the primary factor is the chemical which is the most abundant in the condition) for each condition is considered. The ‘Av. pH’ column reports the average of the measured pH for each condition in the screen, if no measured pH is available a calculated estimated pH value (paper in preparation) is used. The ‘Wells within Av. pH +/- 1’ estimates the breadth of the pH sampling, where a larger number indicates a narrow range of pH values. Finally, the ‘Distinct Chemicals’ column counts the number of stocks required to make the screen.

2.3. Persistent Screen Sets

Many of the C6 reports require the selection of a subset of screens on which to perform the report action. For example, when using either the ‘Find a chemical in screens’ or ‘Find a condition in screens’ report, the user is prompted to select a pre-defined subset of screens, such as ‘C3 screens’ or ‘Commercial screens’ in which to search for the chemical (or condition). Initially C6 only offered pre-defined screen subsets, and although these pre-defined sets were helpful, there are cases where the user might wish to choose their own subset of screens. Being able to define one’s own specific subset of screens is particularly useful for users outside the C3 user community, who can create a subset of commercial screens corresponding to those they have in their home laboratory. The creation of screen sets is done through the report ‘Create a set of screens’, found in the ‘Tools’ section of the top menu. The user-defined sets persist in the C6 database and can be edited as needed. Any user-defined screen sets are found under the ‘My Screen Sets’ filter option (Figure 7).

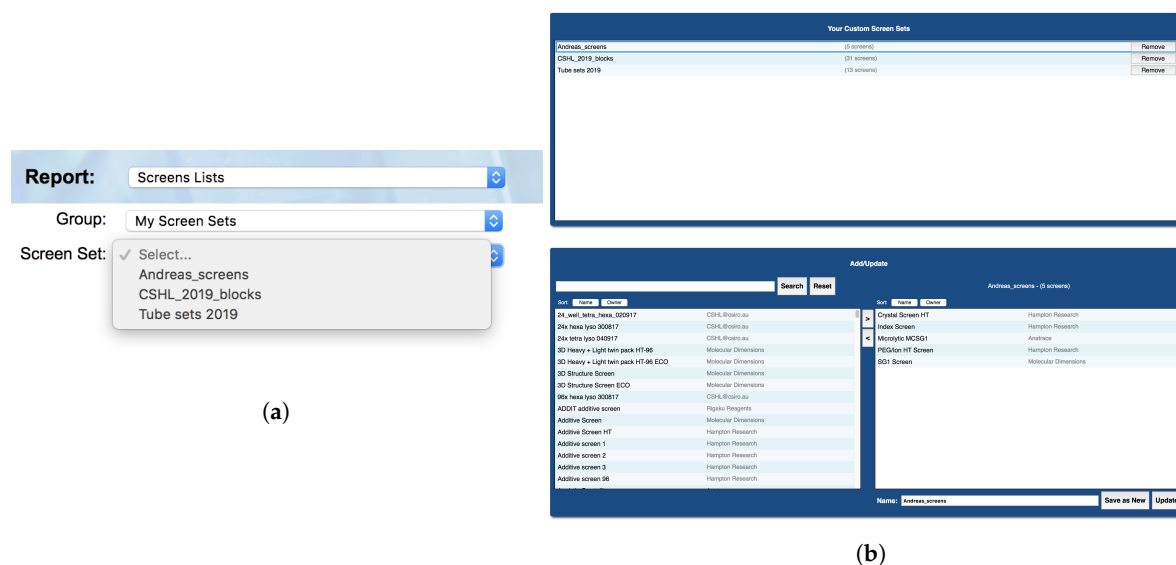


Figure 7. (a) The second level of filtering that is required to select a custom screen set for a report. (b) The top part of the ‘Create Custom Sets’ report shows sets that are available for the user. The sets are created from the list of screens (bottom right) available to that user: commercial screens and screens owned by the user. The interface for creating a custom set allows for screens to be moved into and out of the custom set by dragging into (or from) the selected screens list (bottom right).

The screens able to be added to a user’s screen set include all private screens belonging to the user, and all public screens both commercial and in-house (C3 screens). These are presented in a list, and can be searched via the search bar for ease of use. From the same report, existing screen sets can be modified or deleted as well as created, and all changes take effect immediately and will update accordingly in the screen group selection boxes in the other C6 reports.

2.4. Recipes

Most often a crystallisation screen is described as a set of conditions, where each condition contains one or more chemical factors. This way of documenting a screen does not capture how the condition was made—for example, 1 mL of a simple condition containing only 1 M sodium chloride could be made by pipetting out 1 mL of a 1 M sodium chloride stock, or by pipetting out 0.2 mL of a 5 M sodium chloride stock and adding 0.8 mL water. Generally, we refer to the description of a crystallisation condition in terms of concentration as a ‘design’ and the description of the the same solution in terms of stocks and volumes as a ‘recipe’. For every condition that can be created with C3 stocks the recipe for producing 1 mL of that condition is given alongside the design of that condition (see Figure 4). If every condition in the screen can be created with the stocks available in C3 then an option appears in the header of that screen to export a file containing the recipes for every well. The final volume required can be stipulated (it defaults to 1 mL), and the recipe can be exported in different formats. To date the export formats include an xml format and a human-readable csv format suitable for import into the Dragonfly (SPT Labtech) liquid dispensing robot (Figure 8).

Screen (click for stats)	Shotgun
Export as	CSV-ROW CSV-CELL TEXT XML MMCIF RECIPE DRAGONFLY
Owner	c3@csiro.au
Create date	2019-11-21
Description	This is a collection of the 96 most successful commercial conditions as determined by mining the PDB
Barcodes	Z0002440, Z0002426, Z0002425
Well count	96
Number of distinct chemicals	50
Internal diversity (mean dist)	0.912

Recipe Generator	Volume (mL) <input type="text" value="1"/> <input checked="" type="checkbox"/> Minimize stocks <input type="button" value="Recipe"/> <input type="button" value="Summary"/>
------------------	---

NB: Buffers are shown in <i>italic font</i>			
Well	Concentration Units Chemical, pH	pH	Recipe
A1 (1)	0.2 M magnesium chloride; 30 w/v polyethylene glycol 4000; 0.1 M tris chloride, pH=8.5;	8.1	> A1 200.0 ul of: magnesium chloride (1M) 600.0 ul of: poly(ethylene glycol) 4000 (50% w/v) 21.2 ul of: tris chloride pH 7.0 (1M) 78.8 ul of: tris chloride pH 9.0 (1M) 100.0 ul of: H2O

Figure 8. The row in the header ‘Recipe Generator’ allows the user to set the required volume for the screen. ‘Minimize stocks’ retains only the most concentrated stock of a chemical, if more than one stock of the same chemical was used in the initial screen recipe. The ‘Summary’ button opens a report that shows which stocks (and how much of them) will be used to create the recipe for the screen.

2.5. Phase Space Visualisation

Both fine screening and additive screening methods of optimisation rely on the user having a clear understanding of patterns in the outcomes of previous crystallisation experiments. In practice, such understanding is often sought by unsystematically eyeing a table of conditions in the hope of spotting commonalities between successful experiments. With this approach, one is prone to missing important positive correlations between a particular condition and the formation of crystals, or conversely, not grasping which areas in the condition space have been explored thoroughly without proving fruitful. While progress is being made towards automation of this analysis using chemical ontologies, human expertise will continue to be valuable, so there is a need for tools that augment human abilities to perceive which points in condition space have been tested for a given protein, and what the outcomes of those experiments were.

To this end, a ‘Visualise phase space’ report has been added to C6, which, for any user-selected combination of experiment barcodes, plots a subset of the conditions used including temperature, pH, chemical species and concentration on an interactive parallel coordinates plot (a common plot-type for high-dimensional data). Each experiment is represented by a line intersecting each of the condition axes at the relevant value, and the colour of each line indicates the outcome of the experiment it represents (as scored by the user in See3). Axes can be reordered, inverted or removed to make key trends clearer. Experiments can be easily filtered by any of the axis dimensions, as well as screen, well or outcome. To aid new users, a draggable card summarising the available interactions can be displayed on top of the interface, which also links to a video walk-through of how to use it.

An example plot is included in Figure 9.

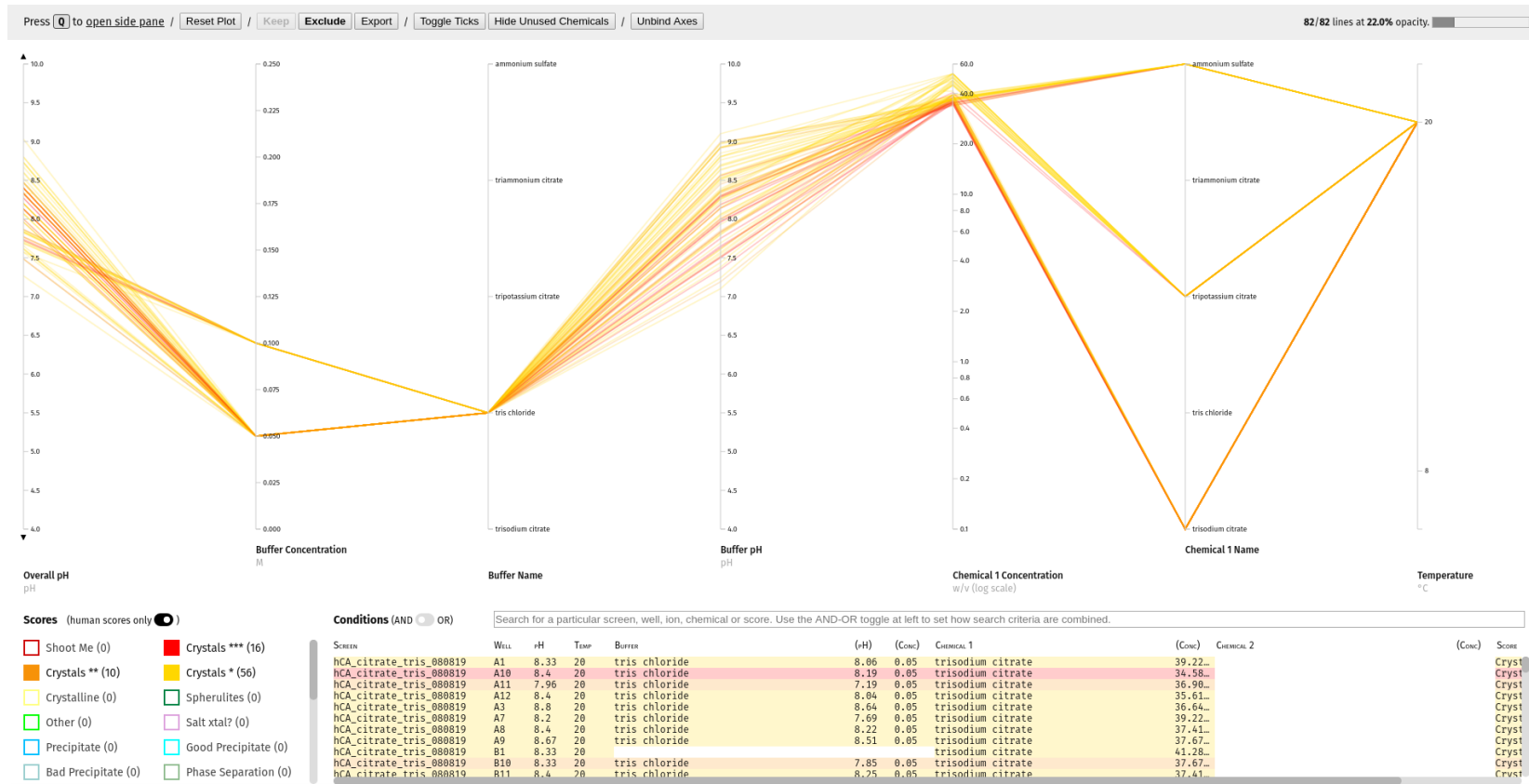


Figure 9. The parallel coordinates plot of the crystallisation phase space for a particular sample, generated using the ‘Visualise Phase Space’ report. By default lines are coloured by simplified set of experimental outcomes (*Crystal*, *Clear*, *Precipitate*, and *Other*) but, as shown here, the user can toggle on the use of the full set of classifications used in See3 [20] for greater discrimination. The legend is in the bottom-left corner of the interface. The full set includes several classes for human scoring of crystals: *Crystalline*, *Crystals **, *Crystals ***, *Crystals **** and *Shoot Me*. In this example, the darker oranges/reds used for higher quality crystals suggest that lower buffer concentration, lower buffer pH, and the presence of trisodium citrate tend to produce higher quality crystals. The reasoning behind the use of different numbers of asterisks (*), rather than explicit descriptions to distinguish the different crystal scores is that crystal quality is highly project dependent—one project’s crystals assigned the visual score of *Crystal **** might only be rated *Crystal ** in another system.

3. See3

See3 has three major sections—a viewing and scoring section, a design section and an administration section, which is not visible to the general user. The viewing and design sections are accessible to all users; a user will see only plates owned by themselves. The viewing and scoring section of See3 provides the user with a selection grid of existing crystallisation plates. After a plate is selected See3 shows a grid of the most recent images associated with that plate (Figure 10).

Position	Type	Date	Score
13	Auto	08/12/20	100
10	Auto	02/01/21	100
9	Auto	26/12/20	100
8	Auto	18/12/20	100
7	Auto	12/12/20	100
6	Auto	06/12/20	100

Figure 10. On opening See3, the user sees a searchable grid on the left, and once a plate is selected, it will appear as a grid on the right hand side of the screen. Each droplet image is surrounded by a narrow band of colour, indicating any score associated with the droplet. The pale pink border around many of the droplets indicate that they have been MARCO autoscored as *Crystal*.

By default, the thumbnail images are arrayed in a grid corresponding to their position on the crystallisation plate, however the images may also be arrayed by scores. A MARCO autoscore for each image is generated automatically and associated with an image, and human scores can also be associated with each image. A visual indication of the score for a well is given in the coloured border around the well [20]. Sorting by score will display images with *Crystal* scores first, followed by images with *Other* scores, then *Precipitate* scores and finally those with *Clear* scores, with human scores in each category being shown before autoscored images with the same classification. Double-clicking on a thumbnail will open up a larger image, and display associated information, including the sample name, the cocktail name, and what chemicals they respectively contain.

Users may place images from one or more plates on to a clipboard, this is a way of cherry-picking conditions for cycles of optimisation. The images on the clipboard can be used to generate a hit report (see *Supplementary Information* for an example of a hit report for a crystallisation plate set up with the protein Lysozyme). The hit report contains the details of conditions and samples in the droplets contained in the images, listed for each image on the clipboard. The hit report concludes with a summary of the chemical factors (the ‘Chemistry Range Table’) found in the conditions on the clipboard (Figure 11).

Chemistry Range Table

Class	Avg Conc	Conc Range	Units	Name	Count
Buffer	100		mM	sodium HEPES (pH 7)	1
Buffer	0.1		M	tris chloride (pH 8.5)	1
Buffer	0.1		M	sodium acetate-acetic acid (pH 4.5)	1
Polymer	0.5		w/v	jeffamine ED-2001 (pH 7)	1
Salt	1.1		M	sodium malonate-malonic acid (pH 7)	1
Salt	1		M	lithium chloride	1
Salt	0.2		M	magnesium chloride	1
Precipitant	20		w/v	polyethylene glycol 6000	1
Precipitant	20		w/v	polyethylene glycol 8000	1
Precipitant	1		M	sodium chloride	1
Fundamental Mixtures	0.1		M	trisodium citrate-citric acid (pH 4)	1

Figure 11. The ‘Chemistry Range Table’ from the Hit Report shown in Appendix A. This shows the chemical factors for the four conditions in the Hit Report and some initial clustering into chemically equivalent groups has been attempted.

If appropriate images were placed on the clipboard, the information in the ‘Chemistry Range Table’ is a reasonable starting point for optimisation through fine screening. The optimisation tool allows one to work with the information from the clipboard to design new screens. Once designed and saved, the screens will appear in the next build of C6, generally within an hour. By default, the nascent design in See3 is checked against the stocks available in C3 before being saved, thus any design that was saved successfully in See3 will be able to be produced with C3 stocks, and the recipe for its creation will be available from C6, along with the screen description.

4. Features in See3

The See3 application initially provided a modern viewing, scoring and reporting tool for images collected in C3. Many of the features included in the viewing component of See3 are standard, and are found in many viewing programs—for example, measuring tools, drop history, drop components etc. Along with the expected functions, See3 has some tools which were built after user feedback. These include a tool for adjusting the zoom or centre of images, the ability to see the recipe for a single condition and the ability to select and write out regions of interest within crystallisation images. Beyond the viewing capability, See3 allows users to design optimisation experiments. Optimisation screens can be automatically generated from information on the clipboard, or can be created from a blank slate by the user. Screens can be in 24, 48 or 96 well formats.

4.1. Recentering Images

The very first image of a drop is at a zoom level that encompasses the entire subwell in which the droplet is located. This image is segmented by the RockImager software and subsequent images are taken at a zoom level and with centre coordinates so that the experimental droplet fills the field of view. Mostly this process works well, but it fails sufficiently often that a tool that allows the user to reset the centre and zoom level (to be used on the subsequent images of that droplet) has proven very useful (Figure 12).

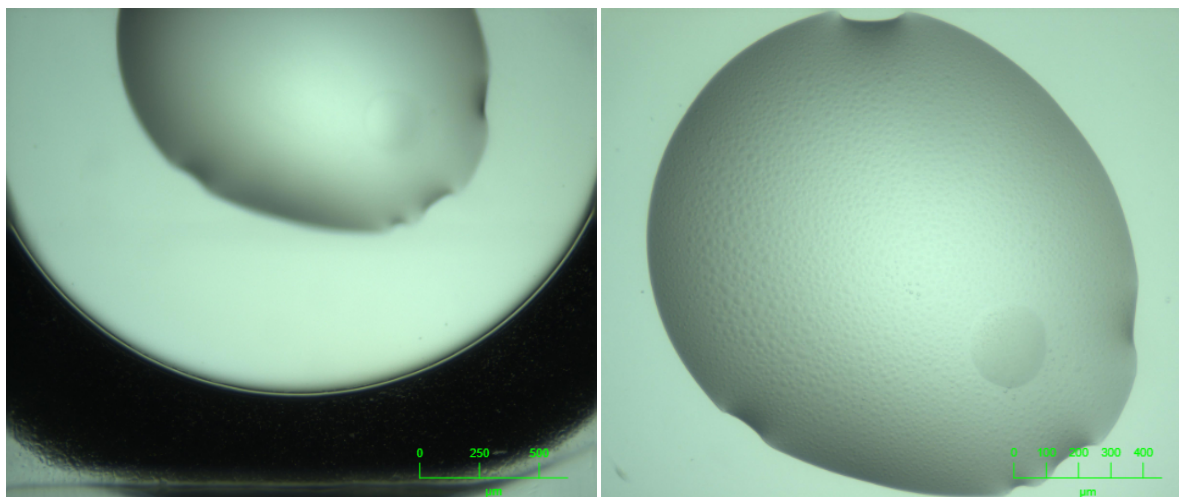


Figure 12. Images before (left) and after (right) manual recentering.

4.2. ROI Setting and Export

Being able to locate a subregion within a droplet can be used for a number of purposes: to earmark particular crystals for harvesting, or for *in-situ* X-ray interrogation. Within the See3 viewing software one can mark one or more regions of interest (ROI), and the coordinates of these can be exported. The (x, y) coordinates of the ROI are relative to the centre of the subwell. A Z-offset is also included which gives the curved surface depth of the subwell (assuming the plate is an MRC sitting drop plate) at the ROI position.

4.3. Optimisation

Optimisation is the refinement of initial screening conditions with the goal of producing a well diffracting crystal. Aside from microseeding [21], the most common methods of optimisation are fine screening, using information gleaned from any previous experiments, and additive screening, where a single promising condition is doped with many different chemicals to test their effect on the crystallisation process. The optimisation function of See3 has two distinct modes for optimisation; an *Automatic* and a *Manual* mode. In essence, the *Automatic* mode considers the entire screen as an entity, whereas the *Manual* mode considers the individual wells as components that together make up a screen.

4.3.1. Automatic Optimisation

The Automatic optimisation tool allows the generation of a new optimisation screen based on the currently selected conditions on the clipboard. By default, the Automatic mode generates a screen by populating each well with a random pick of a chemical from a number of Factor Groups, and assigns a random concentration (from a range) for each chemical. The first step is the clustering of the chemicals found on the clipboard into Chemical Factor groups. This grouping of chemicals is based on their class or role, and is similar to how the chemicals are grouped in the 'Chemistry Range Table' (Figure 11). The Primary Factor Group contains the chemicals that are found at the highest concentration in each of the conditions selected on the clipboard. Other default Chemical Factor Groups include Buffers, Polymers, Organics and Salts. Each chemical within a Factor Group is given four attributes: a flag which decides if the concentration or the pH is to be varied (conc/pH flag), a lower limit, an upper limit, and a weight. The attributes are set from the values seen in the 'Chemical Range Table.' For chemicals with the pH/conc flag set to 'conc,' the upper and lower values are set according to the average concentration of that chemical found in the 'Chemical Range Table.' For chemicals in the Primary Factor group (and for chemicals found at greater than 0.5 M), the default

limits are $0.8 \times$ average concentration for the lower limit, and $1.1 \times$ average concentration for the upper limit. For other Factor groups the concentration limits are $0.1 \times$ average concentration (lower limit) and $2 \times$ average concentration (upper limit). The pH range is set by default to the (nearest) pK_a for the buffer ± 1 pH unit. The weight of the chemical depends on how many clipboard conditions contained that chemical. The Factor Groups for the 'Chemical Range Table' shown in Figure 11 are shown in Figure 13.

Factor Group					Color	Vary By				Location	Steps	
Primary						Gaussian Random				Random	96 / 96	Delete
Name	Avg Conc	Start Conc	End Conc	Units	Avg pH	Start pH	End pH	Weight	Vary	Steps	Distribut...	
sodium malonate-malonic...	1.1	0.88	1.21	M	7	7	7	0.25	Conc	96	24	Delete
polyethylene glycol 6000	20	16	22	w/v				0.25	Conc	96	24	Delete
polyethylene glycol 8000	20	16	22	w/v				0.25	Conc	96	24	Delete
sodium chloride	1	0.8	1.1	M				0.25	Conc	96	24	Delete
Buffer						Gaussian Random				Random	96 / 96	Delete
Name	Avg Conc	Start Conc	End Conc	Units	Avg pH	Start pH	End pH	Weight	Vary	Steps	Distribut...	
sodium HEPES	100	100	100	mM	7	6.55	8.55	0.25	pH	96	24	Delete
trisodium citrate-citric acid	0.1	0.1	0.1	M	4	3.76	5.76	0.25	pH	96	24	Delete
tris chloride	0.1	0.1	0.1	M	8.5	7.1	9.1	0.25	pH	96	24	Delete
sodium acetate-acetic acid	0.1	0.1	0.1	M	4.5	3.76	5.76	0.25	pH	96	24	Delete
Polymer						Gaussian Random				Random	24 / 96	Delete
Name	Avg Conc	Start Conc	End Conc	Units	Avg pH	Start pH	End pH	Weight	Vary	Steps	Distribut...	
jeffamine ED-2001	0.5	0.05	0.75	w/v	7	7	7	0.25	Conc	96	24	Delete
Salt						Gaussian Random				Random	48 / 96	Delete
Name	Avg Conc	Start Conc	End Conc	Units	Avg pH	Start pH	End pH	Weight	Vary	Steps	Distribut...	
lithium chloride	1	0.8	1.1	M				0.25	Conc	96	24	Delete
magnesium chloride	0.2	0.02	0.3	M				0.25	Conc	96	24	Delete

Figure 13. The Automatic optimisation function in See3 is based around random picks from a Factor Group Table. This Factor Group table corresponding to the 'Chemistry Range Table' from the Hit Report shown in Appendix A.

The weighting and upper and lower limits can be edited, chemicals can be added or deleted from Factor groups, and Factor Groups can be added or deleted. Users also have some control over where different chemicals can be found on the plate, and how the concentration or pH values are chosen. The user can choose to include the original hit conditions (those placed on the clipboard) in the final design. The screen generated from the Factor Group table can be saved, also the Factor Group table can be saved. Of course, if a saved Factor Group is re-loaded, it is unlikely that exactly the same screen will be generated again, due to the random selection of concentrations.

After the optimisation screen has been generated, it undergoes a number of checks before it can be saved. If the screen cannot be created with C3 stocks it will not be saved, and the user will be alerted. The most common reason for failing to save is that the available stocks are not sufficiently concentrated to make up the conditions. This is called 'overflowing.' If there are fewer than 20 overflowing wells, the program will try to correct this by reducing the concentration of the chemical from the Primary Factor group. The user will also be warned if incompatibilities such as divalent metals in the same condition as a phosphate salt are detected, but the incompatibility warnings do not preclude the screen from being saved.

4.3.2. Manual Optimisation

The *Manual* optimisation tool also enables the creation of a new optimisation screen but with greater customisability. Users are given a blank design and are free to create the experiment as they wish. Portions of existing designs or clipboard conditions can be copied, sections of the optimisation design can be repeated, individual chemicals can be added modified or removed, and chemical gradients can be placed across the new design (Figure 14).

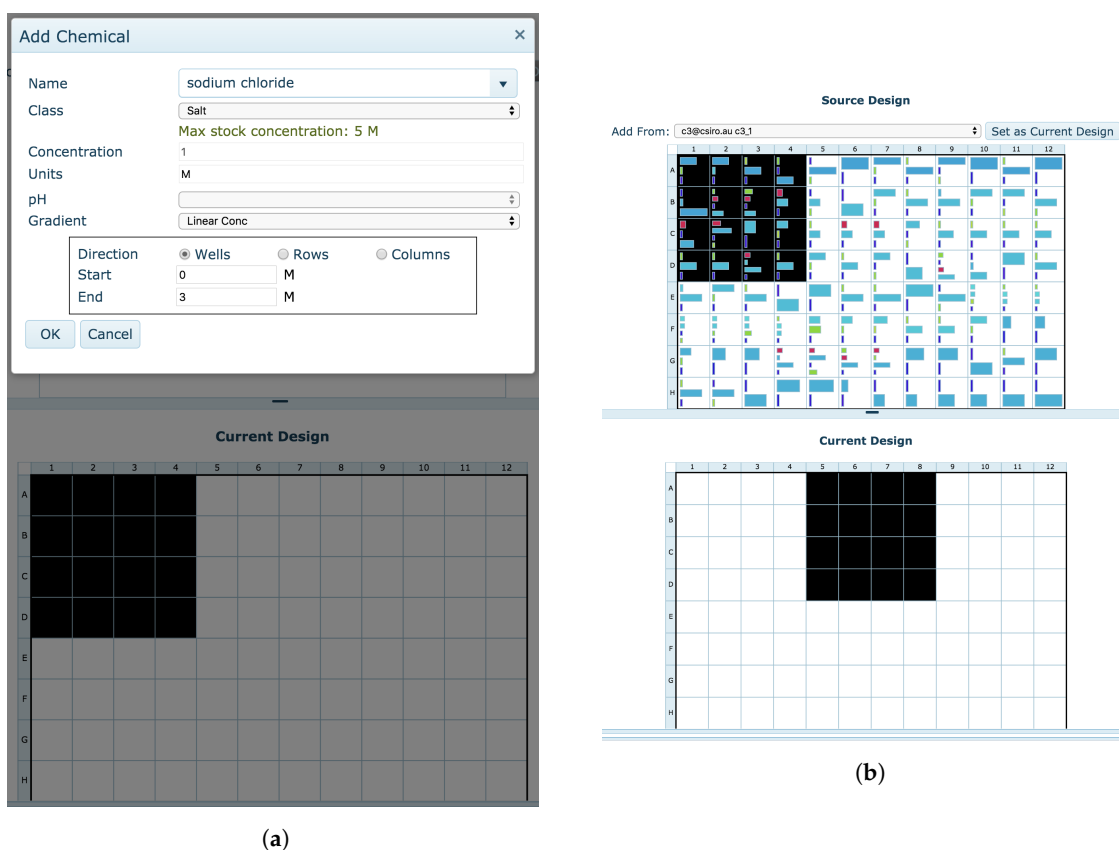


Figure 14. In *Manual* optimisation users select a subset of wells, and can either add chemicals either at a single concentration or as a gradient, as shown in (a), or can add conditions from other screens (b).

When switching from the *Automatic* design to the *Manual* design, users are asked if they wish to begin their *Manual* optimisation design with their current *Automatic* optimisation design. This allows for the modification of specific conditions within an automatically generated screen.

5. Discussion

In the heyday of the Structural Genomics initiatives it was believed that simply by having sufficient protein, automation and appropriate screens any desired crystal—and thus structure—could be generated. More recent analyses suggest that screening alone is insufficient in about 80% of cases [22]. Once the truism that optimisation is required is accepted, then the pathway of crystallisation becomes less obvious. Which screens are best? What is a hit? Which conditions gave hits? How do I optimise the hit(s)? What is noticeably lacking is a set of tools to guide a structural biologist through this complicated process. We require tools for helping select initial screens, and for analysing the outcomes of those screens. Following that, tools are needed both to create designs for optimisation screens as well as to produce the recipes for those designs. Of course automation aids the process by enabling plate preparation, plate imaging and the production of screens. However, automation does not provide an answer to the question of *what* to do; without tools to help the researcher decide what is the best (or at least a reasonable) way forward, automation just speeds up the process of doing the wrong thing.

Our experience from working in a crystallisation core facility for over a decade suggests that one of the notable benefits of a facility is the expertise that builds up within the centre which then becomes available to the users of the facility. Most of the individuals using a crystallisation facility only spend a small percentage of their time thinking about crystallisation, as structure is only part of a bigger picture for understanding any protein system. Almost invariably, the users ask questions about what screens to use, “if this is a hit,” and then how to optimise. The development of the software tools

described here is a result of our interactions with users, and is an effort to distill our experience so that it persists and can be shared. In particular, the hope that screening would solve the problem of crystal production has led to an explosion in the number of screens available. Counter-intuitively, this has made it harder to work out where to start.

The C6 tool cannot give answers about what is the 'best' screen with which to start a crystallisation campaign. However, given some information about the protein (e.g., it has been crystallised before in PEG) it can help select which screens to use. Further, C6 allows one to see if a further screen is testing a novel area of chemical space or is replicating conditions that have already been set up, and allows one to check a group of screens for condition redundancy. The visualisation tool provides an interactive interface for building up hypotheses about what parts of crystallisation space might be used to produce crystals, which can either guide subsequent screen selection or optimisation strategies.

The See3 viewing platform, coupled with reliable, real-time autoscoring, enormously speeds up the process of analysing the results of crystallisation trials. The link between finding a hit and creating optimisation from any hit(s) is aided by the automatic extraction of chemical factors from the hits. See3 optimisation tools are intuitive, and sufficiently sophisticated to allow for the design of almost anything that can be set up with current automation. The combination of automatic and manual screening modes brings the power of combinatorial optimisation which allows for the simultaneous optimisation of many conditions, with the specificity that comes from having control over each well. The most practical aspect of the optimisation is tying it in with stock availability, so that only screens which can be easily created can be designed.

Most importantly, these tools are web-based, and are available to the wider structural biology community. Both C6 and See3 can be accessed with a guest account (username `guest@c3`, password `vegemite`), or by registering as a C3 user. Any screens developed in See3 as a guest will be visible and editable by any other guest user, screens designed by a C3 user are only accessible by that user.

6. Materials and Methods

6.1. C6 Implementation

The C6 back-end was predominantly written in Python 2, which has now been migrated to Python 3, while the front-end is written in HTML/CSS and JavaScript. The similarity calculation comparison code has been implemented in C. C6 uses an internal SQLite database (<https://www.sqlite.org/>), which currently takes two days or so to build from scratch (pairwise comparisons of all screens), and roughly 7 minutes to update hourly, once initially built. The interactive data visualisation in the 'Visualise Phase Space' C6 report uses v3.5.17 of the d3.js JavaScript visualisation library (<https://d3js.org/>), with additional use of the jQuery library (<http://jquery.com/>) for various UI elements and keyboard shortcuts.

6.2. See3 Implementation

See3 is written as an ASP MVC .NET Framework project which is written with HTML/CSS and JavaScript for the front-end user interface, and C# for the back-end. The front-end JavaScript user interface is written using the Kendo UI framework. Back-end database connections to the Oracle CM database were written using the the Nhibernate framework.

6.3. MARCO Implementation

The MARCO autoscoring pipeline (C4) consists of a number of steps, and is controlled using a Python script, for a complete description see [23]. Briefly, Cron is used to provide lists of recently inspected plates; images from these are copied to a GPU cluster using scp. Images in batches of 32 are passed through a SLURM workflow manager where they are processed through the MARCO tensorflow AI. The output from this is post-processed to extract the single most likely classification

and returned via SQL into the CM database. The C4 workflow is available from <https://doi.org/10.4225/08/5a97375e6c0aa>.

Author Contributions: conceptualization, J.N.; software, N.R., M.R., G.J.A., L.T., C.J.W., A.K. (Alex Kruger) and A.K. (Alex Khassapov); validation, B.M. and J.N.; writing—original draft preparation, J.N.; writing—review and editing, J.N., L.T.; visualization, L.T., N.R., M.R., G.J.A., A.K. (Alex Kruger) and A.K. (Alex Khassapov); supervision, J.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Acknowledgments: We thank the Undergraduate Research Opportunity Program (UROP) administered by BioMedVic, and Rigaku Automation for providing the initial XtalTrak code upon which See3 has been developed. We thank Tom Peat for helpful discussions and proofreading.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PEG	Polyethylene glycol
C3	Collaborative Crystallisation Centre
CSIRO	Commonwealth Scientific and Industrial Research Organisation
ID	Internal Diversity

Appendix A

A hit report is generated from images placed on the clipboard in See3. Images can be from one or more plates, and the report includes details about the drops, and concludes by summarising all the chemical factors found collectively in the drops included in the hit report.

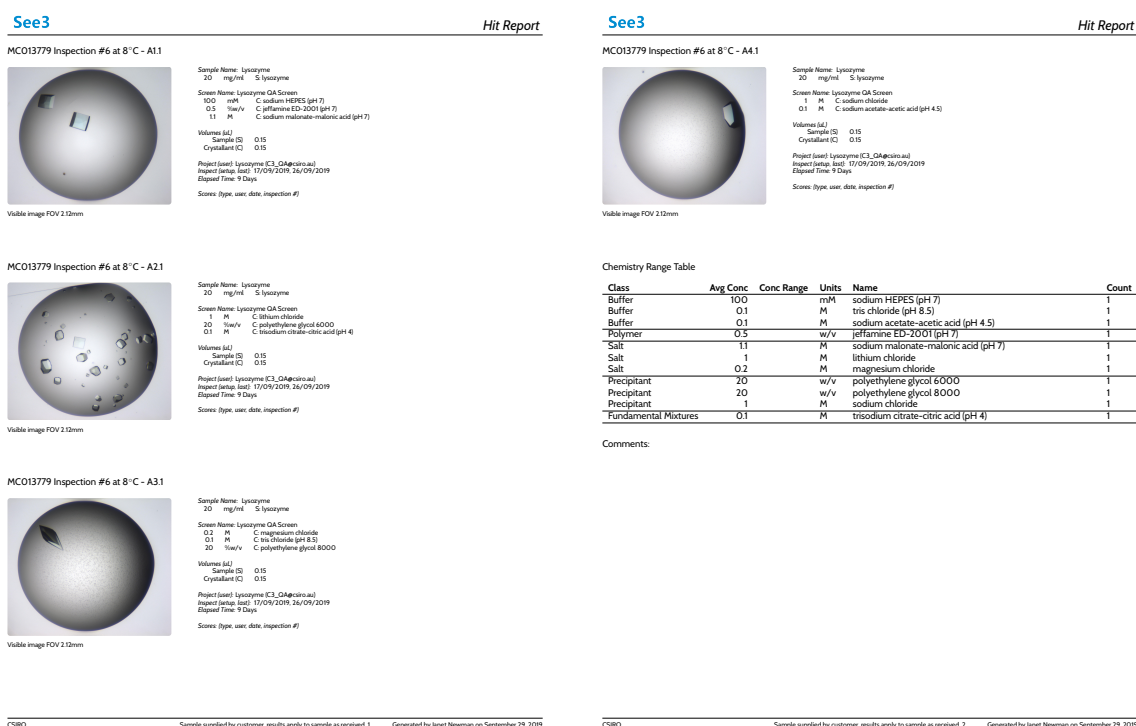


Figure A1. A ‘Hit Report’ generated from four drops of a plate containing Hen Egg White Lysozyme (coloured blue) as the protein sample. The latest image from each drop is included, along with information about the chemical conditions, the drop size, the temperature of incubation, and the elapsed time since setup.

References

1. Burley, S.K.; Berman, H.M.; Kleywegt, G.J.; Markley, J.L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. In *Protein Crystallography*; Wlodawer, A., Dauter, Z., Jaskolski, M., Eds.; Springer: New York, NY, USA, 2017; Volume 1607, pp. 627–641, doi:10.1007/978-1-4939-7000-1_26. [[CrossRef](#)]
2. Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, 1st ed.; Garland Science: New York, NY, USA, 2009.
3. Derewenda, Z.S. It's all in the crystals.... *Acta Cryst. D Biol. Crystallogr.* **2011**, *67*, 243–248, doi:10.1107/S0907444911007797. [[CrossRef](#)] [[PubMed](#)]
4. McPherson, A.; Gavira, J.A. Introduction to protein crystallization. *Acta Cryst. F Struct. Biol. Commun.* **2014**, *70*, 2–20, doi:10.1107/S2053230X13033141. [[CrossRef](#)] [[PubMed](#)]
5. Luft, J.R.; Newman, J.; Snell, E.H. Crystallization screening: The influence of history on current practice. *Acta Cryst. F Struct. Biol. Commun.* **2014**, *70*, 835–853, doi:10.1107/S2053230X1401262X. [[CrossRef](#)] [[PubMed](#)]
6. Overton, I.M.; Padovani, G.; Girolami, M.A.; Barton, G.J. ParCrys: A Parzen window density estimation approach to protein crystallization propensity prediction. *Bioinformatics* **2008**, *24*, 901–907. [[CrossRef](#)] [[PubMed](#)]
7. Kurgan, L.; Razib, A.; Aghakhani, S.; Dick, S.; Mizianty, M.; Jahandideh, S. CRYSTALP2: Sequence-based protein crystallization propensity prediction. *BMC Struct. Biol.* **2009**, *9*, 50. [[CrossRef](#)] [[PubMed](#)]
8. Jahandideh, S.; Jaroszewski, L.; Godzik, A. Improving the chances of successful protein structure determination with a random forest classifier. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 627–635, doi:10.1107/S1399004713032070. [[CrossRef](#)] [[PubMed](#)]
9. Meng, F.; Wang, C.; Kurgan, L. fDETECT webserver: Fast predictor of propensity for protein production, purification, and crystallization. *BMC Bioinform.* **2017**, *18*, doi:10.1186/s12859-017-1995-z. [[CrossRef](#)]
10. Wang, H.; Feng, L.; Webb, G.I.; Kurgan, L.; Song, J.; Lin, D. Critical evaluation of bioinformatics tools for the prediction of protein crystallization propensity. *Brief. Bioinform.* **2018**, *19*, doi:10.1093/bib/bbx018. [[CrossRef](#)] [[PubMed](#)]
11. Tung, M.; Gallagher, D.T. The Biomolecular Crystallization Database Version 4: Expanded content and new features. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *65*, 18–23, doi:10.1107/S0907444908035440. [[CrossRef](#)] [[PubMed](#)]
12. Abrahams, G.J.; Newman, J. BLAST ing away preconceptions in crystallization trials. *Acta Cryst. F Struct. Biol. Commun.* **2019**, *75*, 184–192, doi:10.1107/S2053230X19000141. [[CrossRef](#)] [[PubMed](#)]
13. Newman, J.; Fazio, V.J.; Lawson, B.; Peat, T.S. The C6 Web Tool: A Resource for the Rational Selection of Crystallization Conditions. *Cryst. Growth Des.* **2010**, *10*, 2785–2792, doi:10.1021/cg1004209. [[CrossRef](#)]
14. Hedderich, T.; Marcia, M.; Köpke, J.; Michel, H. PICKScreens, A New Database for the Comparison of Crystallization Screens for Biological Macromolecules. *Cryst. Growth Des.* **2011**, *11*, 488–491, doi:10.1021/cg101267n. [[CrossRef](#)]
15. Newman, J.; Peat, T.S.; Savage, G.P. What's in a Name? Moving Towards a Limited Vocabulary for Macromolecular Crystallisation. *Aust. J. Chem.* **2014**, *67*, 1813, doi:10.1071/CH14199. [[CrossRef](#)]
16. Bruno, A.E.; Ruby, A.M.; Luft, J.R.; Grant, T.D.; Seetharaman, J.; Montelione, G.T.; Hunt, J.F.; Snell, E.H. Comparing Chemistry to Outcome: The Development of a Chemical Distance Metric, Coupled with Clustering and Hierarchical Visualization Applied to Macromolecular Crystallography. *PLoS ONE* **2014**, *9*, e100782, doi:10.1371/journal.pone.0100782. [[CrossRef](#)] [[PubMed](#)]
17. Bruno, A.E.; Charbonneau, P.; Newman, J.; Snell, E.H.; So, D.R.; Vanhoucke, V.; Watkins, C.J.; Williams, S.; Wilson, J. Classification of crystallization outcomes using deep convolutional neural networks. *PLoS ONE* **2018**, *13*, e0198883, doi:10.1371/journal.pone.0198883. [[CrossRef](#)]
18. Newman, J.; Sayle, R.A.; Fazio, V.J. A universal indicator dye pH assay for crystallization solutions and other high-throughput applications. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 1003–1009, doi:10.1107/S0907444912018768. [[CrossRef](#)] [[PubMed](#)]
19. Jancarik, J.; Kim, S.H. Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Cryst.* **1991**, *24*, 409–411. [[CrossRef](#)]
20. Rosa, N.; Ristic, M.; Marshall, B.; Newman, J. Cinder: Keeping crystallographers app-y. *Acta Crystallogr. F Struct. Biol. Commun.* **2018**, *74*, 410–418, doi:10.1107/S2053230X18008038. [[CrossRef](#)] [[PubMed](#)]

21. D’Arcy, A.; Bergfors, T.; Cowan-Jacob, S.W.; Marsh, M. Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallogr. F Struct. Biol. Commun.* **2014**, *70*, 1117–1126, doi:10.1107/S2053230X14015507. [[CrossRef](#)] [[PubMed](#)]
22. Newman, J.; Burton, D.R.; Caria, S.; Desbois, S.; Gee, C.L.; Fazio, V.J.; Kvensakul, M.; Marshall, B.; Mills, G.; Richter, V.; et al. Crystallization reports are the backbone of Acta Cryst. F, but do they have any spine? *Acta Cryst. F Struct. Biol. Commun.* **2013**, *69*, 712–718, doi:10.1107/S1744309113014152. [[CrossRef](#)] [[PubMed](#)]
23. Watkins, C.J.; Rosa, N.; Carroll, T.; Ratcliffe, D.; Ristic, M.; Russell, C.; Li, R.; Fazio, V.; Newman, J. A Crystal/Clear Pipeline for Applied Image Processing. In *Supercomputing Frontiers*; Abramson, D., de Supinski, B.R., Eds.; Springer International Publishing: New York, NY, USA, 2019; pp. 19–37.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).