



Organizing a crystallization laboratory

Allen Gu, Bevan Marshall, Nicholas Rosa, Marko Ristic and Janet Newman

J. Appl. Cryst. (2018). **51**, 47–54



IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see <http://journals.iucr.org/services/authorrights.html>



Organizing a crystallization laboratory

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

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

Received 26 July 2017
Accepted 20 November 2017

Edited by A. R. Pearson, Universität Hamburg, Germany

Keywords: sample management; stock management; crystallization stocks.

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

Managing chemical stocks and samples in any laboratory is an arduous task; in a crystallization laboratory this is particularly burdensome, given the need for many stocks to facilitate optimization of crystal hits obtained from screening experiments. Although inventory management is widespread in retail and other arenas, most small academic laboratories do not adopt formal stock management systems. Without an overarching system for handling stocks and samples, problems such as stock duplication, inappropriate stock storage and insufficient labelling are rife. Two applications have been developed in the Collaborative Crystallization Centre, the first of which manages the hundreds of stocks used for crystallization, and a second which manages protein (and other) samples stored in the 193 K freezer. Both applications are built around a simple database, with a Python front end that allows samples or stocks to be scanned in or out. Information from a decade of crystallization stock usage allows a good estimation of what chemicals are used (and in what quantities) in a crystallization laboratory.

1. Introduction

Crystallization of biological macromolecules currently involves setting up numerous experiments in which different mixtures of chemicals are mixed with concentrated protein (here used to mean any appropriate biological sample) in small droplets, which are stored for days to months and observed periodically for signs of crystal growth (McPherson & Gavira, 2014). The chemical mixtures (also known as crystallization cocktails) are generally produced in sets (24, 48 or 96 cocktails per set) called screens. Initial testing for crystals may be done with commercially sourced screens (Fazio *et al.*, 2014); however, following up any hits from the initial trials almost assuredly requires the production (or purchase) of stocks of the chemicals found in the successful cocktails, so that they can be recombined to give more focused screens around the area identified in the first round of experiments (a 'hit') (Luft *et al.*, 2014). For a crystallization laboratory with many projects this can lead to the accumulation of a large number of different stocks. In the Collaborative Crystallization Centre ('C3' Melbourne, Australia), chemical stocks are used not only to optimize hits but also to create the initial screens used in C3 (Newman, 2011). Approximately 30% of the 150 or so 96-well blocks made up in C3 each year contain initial screens; the remaining blocks are of fine screens used in optimization of crystal hits. Creating initial screens has the benefit of ensuring that all stocks needed for optimizing a hit obtained from an initial C3 screen are instantly available for fine-screening experiments, but has the downside of requiring a large library of chemical stocks. Currently C3 has close to 400 different chemical stocks. With this number of stocks, controlling inventory becomes challenging; simply being



aware of what stocks are available at any time is difficult. Being proactive about managing this stock collection is harder – for example, knowing what is the likely lifetime of a stock, what indicates that a stock should be discarded and when a new stock of a chemical needs to be produced. C3 maintains independent quality control certification (ISO/IEC 17025; <https://www.iso.org/standard/39883.html>); efficiently managing stock inventory is one of the cornerstones of our accreditation.

Mundane issues plague stock management. Lack of transparency about how stocks are made makes it difficult to have common stocks; every laboratory has stocks that are trusted, and other stocks that are actively avoided. A label created by one laboratory member might not be interpretable by another (assuming that the label even exists), creating a further barrier which hinders stocks from being shared. The pre-existence of stocks can sway the design of an optimization experiment – for example, if two initial crystal hits were found, one containing the buffer MES at pH 6 and the other containing ADA pH 6, and the laboratory has stocks of the MES buffer but not of the ADA buffer, then (more often than readily admitted) the resulting optimization will not contain ADA. Thus the goal of most crystallization laboratories is to have enough of the right stocks on hand so as to limit the activation barrier to creating good optimization experiments.

Along with stocks, managing samples – purified proteins, co-factors, plasmids and so on – is notoriously difficult. A plethora of both commercial and free solutions exist (Perkel, 2015). But (according to a poll of about ten colleagues in the field from Australia and other countries) these are not in wide use in many academic laboratories, with the possible exception of Quartzzy (currently <http://www.quartzzy.com>), which is perhaps better suited to chemical inventory management. The commercial solutions range from complete (costly) solutions (e.g. TTP Labtech's Arctic system), which consist of specific pre-barcoded tubes, associated readers and storage units, to other less complete systems. The latter might be combinations of specific tubes/rack and readers (from e.g. Matrix or FluidX) which would be combined with inventory and tracking software (e.g. RuRo's FreezerPro). These systems are hard to implement retrospectively. Many frozen biological samples degrade with each freeze/thaw cycle, thus precluding thawing existing stocks to transfer them to different storage tubes. Furthermore, in most laboratories samples are acquired as well as created. Samples created in one's own laboratory can be stored in the sample tubes that the laboratory protocols dictate. Acquired samples come in whatever format deemed appropriate by the originator, and often with minimal labelling to boot. The free (academic) inventory and tracking solutions [e.g. LINA (Yousef *et al.*, 2011), OpenFreezer (Olhovskiy *et al.*, 2011), Savanah (List *et al.*, 2017)] seem to have limitations either in what they can store (e.g. limited to nucleic acid/cells) or in how they store samples (grids in freezer boxes).

Both stocks and samples are often created and used by more than one person, so any inventory application needs to be a simple and robust multiuser system. We have developed two applications that are used to manage stocks and samples,

with the specific workflow of a crystallization laboratory in mind. The first, *StockAide*, manages the chemical stocks and tracks what stocks are available, how much of each stock is available, when the stock was made and how often the stock is used. The second, *GPSamplr*, manages the samples found in the 193 K freezer. This application tracks when samples are placed into the freezer, where they are found, when they are removed and who owns them, and allows for comments and other information to be associated with each sample.

2. StockAide

2.1. Stocks and stock labelling

As all of the C3 screens (with the exception of additive screens) are made in-house, all the stocks required to create the 20 or so initial screens offered to users of the service are also available in C3. The screens and their contents can currently be found at <https://c6.csiro.au> under the 'Available screens' report. The 21 initial screens require maintaining an inventory of close to 300 stocks, which are documented in the Stocks table of the C3 crystallization database (which is based on the Oracle CM database from Rigaku Automation). Each stock in the Stocks table has a unique identifier (the Stock_ID), along with a concentration, a unit, a stock name, (potentially a pH), an associated chemical, a storage location and hazard information. The list of stocks maintained in C3 is found at <https://c6.csiro.au> under the 'Chemical stocks report'.

Some of the stocks contain the same chemical at different concentrations; for example, sodium chloride is available in stocks of 5 M, 2 M and 1 M. The need for multiple stocks of the same chemical is mandated by our desire to make our crystallization conditions as reproducible as possible – a larger dispense of a lower-concentration stock is more reliable than a small dispense of a higher-concentration stock. However, we try to ensure that the concentration jumps between stocks of the same chemical are significant, to reduce the number of stocks we maintain. To limit the number of stocks needed for pH buffer chemicals such as Tris chloride or sodium MES, C3 uses two stocks of the same concentration, set 1 pH unit above and below the relevant pK_a of the chemical, then uses a Henderson–Hasselbach approximation to create the desired pH on the fly. For buffer systems where this is inappropriate, such as citric acid/trisodium citrate, a lookup curve is used. A spreadsheet that performs the pH interpolations can be downloaded from C3's information web site (currently <https://research.csiro.au/crystal/about/c3-technology-infrastructure/c3-cool-stuff/>).

Both a consistent set of stocks and a method of labelling those stocks are needed to manage C3 stocks. The consistent set of stocks has been winnowed down by a decade's experience, yet we still have some efficiencies to make – for example, currently we have both 2 M and 2.5 M stocks of magnesium sulfate. C3 stock labels capture the concentration, units and chemical (and pH if appropriate) of the stock, along with the initials of the scientist who made up the stock, the date the stock was created, the Stock_ID, the storage rack for the stock

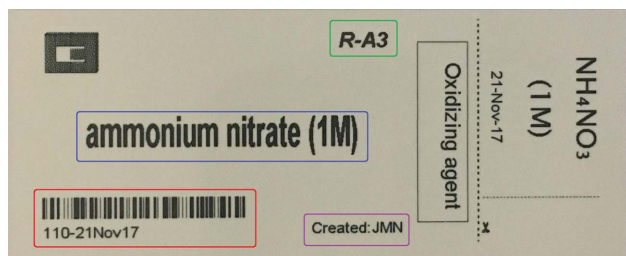


Figure 1

A stock label produced for the 1 M ammonium nitrate stock used in C3. The printed label has both a main label (to be placed lengthwise on the side of a 50 ml centrifuge tube) and a smaller label (for the lid of the centrifuge tube). The label captures the Stock_ID (and date created, both human and barcode readable – red box), stock name (blue box), storage location (green box) and creator (purple box). Two lines are available for any hazard information (black box).

and up to two abbreviated safety notifications. C3 uses commercially available printing software (*Bartender*, Seagull Scientific, USA) and a generic label printer (QL-500, Brother, Australia) to generate stock labels. The *Bartender* software is linked to the CM database, and a stock label template has been created within the software. Printing a stock label requires the selection of one or more stocks from the Stocks table displayed from the *Bartender* template. The template prompts for the initials of the scientist tasked with making the stocks. This produces a standard label (shown in Fig. 1) where the unique Stock_ID and create date of the stock are included on the label as both a human readable string and a code128 barcode.

2.2. Keeping track of stocks

The dispensing technology used to create initial and optimization screens in C3 uses an EVO100 liquid-handling robot (TECAN), which is set up with racks for 50 ml centrifuge tubes as the stock containers. This is an enormous simplification to stock management, as every C3 stock is found in the same type of 50 ml centrifuge tube, so that it can be placed on the TECAN deck. Heavily used stocks [50% (w/v) polyethylene glycol 3350, for example] are made up in larger quantities, but are then aliquoted into the 50 ml tubes. Stocks needed only in small quantities can be made up in 15 ml centrifuge tubes that are then held in 50 ml tubes with foam spacers. Each time a stock is made up, the process is handwritten into the laboratory's common stock book, where the date, the chemical (Vendor, catalogue number, lot number), the amount of chemical needed to make up the desired volume and details (e.g. 'gentle heat required

to dissolve') are recorded. The tubes are labelled as described above. When a 50 ml tube of a stock is exhausted, the tube is discarded. A general guide as to how to make up stocks for C3 is part of the documentation underpinning C3's ISO 17025 Quality Assurance certification, and this information is available through our web site (currently <https://research.c-siro.au/crystal/faqs/sops/stock-preparation-c3/>). This guide contains general instructions about how to make up *M*, w/v and v/v stocks, and describes which acids and bases are used to make up buffer stocks. All stocks in C3 are made up following these general guidelines. One of the topics covered in the standard operating procedure is our pH naming convention. A stock where the pH has been set explicitly has that in its name: 'sodium HEPES pH 6.5'. A stock where the pH has been measured (or simply assumed), but not changed, has the pH in double quotes '(tri)sodium citrate "pH 8.1" (1 M)'.

There are some stocks in the C3 laboratory which are not intended to be used by C3 users, but are in our database so that labels can be created for them. These stocks are distinguished by having the prefix 'TEMP' in the stock name, allowing these to be filtered out of lists presented to users. For example, although C3 has a MES stock at pH 3.8 as the low-pH component of the MES–imidazole buffer system used in the Morpheus screen (Gorrec, 2009), we do not allow users to use this stock outside that buffer system; thus it is given the name 'TEMP MES "pH 3.8" (1 M)'. Users who wish to use the buffer system would select the chemical 'MES–imidazole', which indicates to our software to find the appropriate mix of

name	No. Tubes	Av. Lifetime	Last scanned out	Days until Empty
(di)sodium hydrogen phosphate "pH 9.3" (0.5M)	1	327 (9)	21 Apr 16	0
2-methyl-2,4-pentanediol (100% v/v)	1	163 (18)	02 Nov 16	0
citric acid "pH 2.2" (0.5M)	1	400 (5)	27 Apr 16	0
glycerol (80% v/v)	1	170 (17)	27 Sep 16	0
lithium sulfate (2.5M)	1	120 (23)	29 Nov 16	0
potassium dihydrogen phosphate (1.7M)	1	280 (7)	15 Aug 16	0
propan(2)ol (80% v/v)	1	194 (16)	06 Jul 16	0
sodium malonate - malonic acid pH7 (3.4M)	1	59 (54)	16 Feb 17	0
magnesium chloride (1M)	1	90 (35)	10 Apr 17	17
magnesium formate (1M)	1	376 (6)	06 Jul 16	25
sodium formate (7M)	1	643 (4)	15 Oct 15	27
sodium malonate (3.4M)	1	400 (4)	21 Jun 16	34
ammonium sulfate (3.5M)	2	22 (145)	13 Jun 17	35
lithium chloride (10M)	1	459 (6)	05 May 16	46
sodium formate (4M)	1	299 (11)	24 Oct 16	58
sodium chloride (1M)	2	81 (39)	17 Mar 17	65
dioxane	1	371 (5)	04 Feb 16	67
nickel(II) chloride (1M)	1	713 (4)	21 Sep 15	73
cesium chloride (5M)	1	645 (4)	30 Nov 15	75
sodium DL-malate pH7 (3M)	1	536 (5)	24 Mar 16	81
(tri)sodium citrate "pH 8.1" (1M)	2	155 (19)	18 Nov 16	94
HEPES pH 8.5 (1M)	1	125 (24)	07 Jun 17	110
sodium chloride (5M)	2	87 (35)	21 Apr 17	112
poly(ethylene glycol) 3350 (50% w/v)	11	12 (253)	14 Jun 17	124
malonate-imidazole-borate pH4.0 (100% v/v)	1	569 (4)	04 Apr 16	125
poly(ethylene glycol) 200 (80% v/v)	3	207 (13)	07 Mar 16	149
HEPES pH 6.5 (1M)	2	96 (31)	07 Jun 17	181
sodium chloride (2M)	2	131 (23)	03 Apr 17	182
tartrate-bis-tris-glycol pH 4.0 (100% v/v)	1	494 (4)	15 Aug 16	183
calcium chloride (1M)	1	259 (10)	11 Apr 17	187
imidazole pH 8.0 (1M)	2	271 (6)	06 Jul 16	191
sodium morpholinoethanesulfonate pH 7.2 (1M)	2	216 (14)	24 Oct 16	191
bis-tris pH 5.5 (1M)	1	217 (14)	29 May 17	193
poly(ethylene glycol) 400 (80% v/v)	6	35 (90)	08 Jun 17	196
ammonium sulfate (1M)	2	148 (20)	29 Mar 17	211
poly(ethylene glycol) 600 (60% w/v)	5	281 (7)	21 Mar 14	216
poly(ethylene glycol) 1500 (50% w/v)	6	89 (42)	07 Dec 16	217
potassium sodium tartrate (2M)	1	294 (10)	07 Apr 17	218
(tri)sodium citrate (1.6M)	3	101 (30)	18 Apr 17	238
poly(ethylene glycol) monomethyl ether 2000 (50% w/v)	5	73 (42)	10 Mar 17	261

Figure 2

StockAide view of current stocks, here sorted by 'Days until Empty'. The 'Days until Empty' value is calculated by (No. tubes × lifetime) – days since last scanned out, where lifetime is the average number of days between scanning out tubes. If 'Days until Empty' is less than –365, then no value is shown in the Current Stocks view; if 'Days until Empty' is negative, but greater than –365, then '0' is shown in the Current Stocks view.

MES “pH 3.8” (1 M) and imidazole “pH 9.4” (1 M) for the pH requested.

2.3. StockAide functions

The *StockAide* application has a very simple four-option dialog which allows the user to scan stocks in or out, look at current stocks, or quit *StockAide*.

2.3.1. Scan Stock In. This dialog allows one to enter a stock into the system, either by using a barcode scanner (reading the Stock_ID and create date from the barcode on the stock label) or by typing the information in by hand. The interface defaults to a single tube, containing 50 ml, but both the number of tubes and the volume in each tube can be varied.

2.3.2. Scan Stock Out. This dialog scans out a tube, using a barcode scanner, or by typing the information by hand. The reason for finishing a tube can be varied; the default reason is ‘Empty’, but other options including ‘Contaminated’, ‘Precipitated’ or ‘Old’ may be selected from a pull-down menu. When a stock is scanned out *StockAide* will alert the user as to how many tubes of that stock are left.

2.3.3. Current Stocks. This option opens a spreadsheet-type listing of the stocks, showing the stock name, number of remaining tubes and average lifetime of a tube (and over how many tubes this average is obtained) along with an estimate of how long we have until the stock is exhausted (Days until Empty) (Fig. 2). Clicking on a row of the Current Stocks screen opens a more detailed report for that stock, which gives the same information as the Current Stocks spreadsheet but also includes a breakdown of each tube of the chemical: date scanned in; date scanned out; reason for scanning out; and tube volume.

3. GPSamplr

3.1. Freezer organization

There are many biological samples that might be frozen and stored in a protein crystallization/crystallography laboratory. Often the samples are aliquots of purified protein, but they might also be plasmids, cells, co-factors *etc.* Each of these types of samples might be found in any number of different formats, from beads created by slowly dripping a protein solution into liquid nitrogen to a ziplock bag containing a mass of cell paste. Although in theory C3 is a crystallization service where samples are provided and plates set up in real time, in reality samples are provided when they are made, or otherwise become available. Initially samples provided to the C3 were stored in 0.2 ml (PCR) tubes in 96-position freezer blocks. These frozen aliquots were tracked using a simple notebook which described the sample and the freezer block in which the aliquots were placed. PCR tubes are very difficult to label individually, so the notes were often cryptic: ‘blue PCR tubes in freezer block C3_2 contains protein X’. Laboratory members were requested to update the notebook when aliquots were added or removed from the freezer. Samples that were in containers other than PCR tubes often ended up in plastic bags, and the notebook was rarely updated. The

Table 1
Guide for storing aliquots.

Sample aliquot container	No. of aliquots	Storage box
1.6 ml Eppendorf tubes	Single aliquots	Box with 9 × 9 grid (each aliquot occupies one grid position)
0.2 ml PCR tubes	≤10	Box holding 15 ml centrifuge tubes
0.2 ml PCR tubes	11–40	Box holding 50 ml centrifuge tubes
Other (<i>e.g.</i> racks, ziplock bags)	1	Blob – one per freezer box

fragility of this system prompted the development of an application, loosely based on *StockAide*, to manage the frozen samples. From the beginning it was recognized that this system (unlike *StockAide*) would have to be able to handle different types of containers, as frozen samples come in many different formats. In response to this we have designed a hierarchical system where standard (130 × 130 × 50 mm) cardboard freezer boxes are the basic units of organization, and ultimately every sample must fit into one of these. There are four levels of organization: freezers, racks, boxes and tubes. A box corresponds to a position in a freezer rack and is defined by three location keys – freezer, rack and position. The position in the rack is identified by a row letter and column number; for example, if a freezer rack suitable for an upright freezer has five boxes stacked vertically and four horizontally, then it would have box positions A1–E4, where position A1 is the topmost box closest to the front of the rack, mimicking the position labelling of an SBS (Society for Biomolecular Sciences) footprint plate. The boxes are conceptually indistinguishable from their position and are thus physically labelled with freezer, rack and position information.

3.2. Freezer box organization

A box may contain four different types of sample storage containers (Table 1) – 50 ml centrifuge tubes, 15 ml centrifuge tubes, 80 single tubes or a single object (blob). A box may contain only one type of sub-component – 50 ml tubes, for example. The maximum number of any particular tube type that can be placed into a box is set in a configuration file. The 50 and 15 ml centrifuge tubes are assumed to contain aliquots (0.2 ml PCR tubes) of the same sample. The only difference between the two tube sizes is the number of aliquots which each holds. A box set up to contain 80 tubes has a standard 9 × 9 grid inside, and the orientation of the grid is dictated by including an empty tube as an orientation guide. This empty tube is labelled ‘A1’ and is placed in one of the four corner positions of the grid. This filled corner then becomes the reference point for labelling the positions in the grid – the positional identifier is assumed to always be found in the top left-hand corner of the 9 × 9 array. The rows are then lettered and the columns numbered, following the conventions of a standard SBS plate. Thus a 9 × 9 grid with a positional marker has 80 positions available for use, from A2 to I9. When a new tube or sample is defined, the tube is automatically assigned to the first free slot in an appropriate box. If two or more tubes of

Table 2

Information about samples captured by *GPSamplr* – rows in bold contain mandatory information.

Field name	Field datatype	Example data	Notes
Sample name	String	Integrase 3H	
Alternative names	String	–	Alternative/common name
Concentration	Real	5.4	mg ml ⁻¹
Aliquot number	Integer	4	–
Aliquot volume	Real	50	µl
Owner/origin	String	Alice	Person/place of origin of the sample
Stored by	String	Bob	Laboratory scientist responsible for protein input
Batch date	Date	20/04/2017	Date of sample creation
Date stored	Date	21/04/2017	Date of sample storage
Tag	String	N-term his (cleaved)	Optional information about tags
Buffer conditions	String	50mM NaCl; 25mM DTT; pH 7.5	Storing information about buffers, semicolon delimited
General comments	String	Inactivated; solved	Miscellaneous information; semicolon delimited
Peptide/nucleotide sequence	String	>ExampleProtein MACDEFGHIKL...	FASTA format
Barcode	String	T00004	Assigned automatically
Location	Multiple strings	F1, R1, A1	Assigned automatically, but manual editing possible

the same sample are entered into *GPSamplr* at the same time, then they are always placed next to each other in a grid, even if this requires leaving gaps in the grid or starting a new box. The content type for each box is set interactively, and may only be reset if the box is empty.

Each unique position in a box (a tube or a grid reference point) has information associated with it. Some of this information must be entered when a sample is defined; other data fields are optional and can be either left empty or filled in after a sample has been entered into the system. Table 2 shows the data fields associated with a sample.

3.3. Working with samples

Samples need to be put into the freezer and taken out of the freezer. It is assumed that most samples are put into the freezer once only, but may be taken out more than once if the sample comes in multiple aliquots. Maintaining an accurate inventory of what stocks are available requires that the process of journaling sample input and egress be as simple as possible. The *GPSamplr* user interface allows users to either enter sample information through the ‘New Sample’ button or browse through a table of information about the samples that are in the system *via* the ‘Current Samples’ button.

3.3.1. Sample input. The final use of the sample needs to be considered to decide what the appropriate input format will be. Samples that can undergo only one freeze–thaw cycle (*e.g.* proteins) are stored in single-use PCR tubes. More robust samples that can be successfully refrozen (*e.g.* stock solutions of nucleotides) would be stored in 2 ml cryovials. In C3, aliquots of protein (50–100 µl) are placed in 0.2 ml PCR tubes and flash frozen in liquid nitrogen. The filled and frozen PCR tubes are placed into a labelled and pre-chilled centrifuge tube (15 ml tubes can hold up to ten PCR tubes comfortably, 50 ml tubes can hold up to 40 PCR tubes), which is then placed in the appropriate box. Given an estimation of the number of aliquots, the user opens *GPSamplr*, where the form for entering data about a sample will be pre-populated with the

next available tube number. Pull-down menus ease the process of filling in the mandatory information described in Table 2. The *Bartender* software template for *GPSamplr* tubes uses the information in the *GPSamplr* database to create a label for the tube. Given that general-use labels do not work very well at 193 K, these labels are not stuck on the side of the tube but are positioned inside the tube, visible through the transparent wall. The labelled tube is pre-chilled either on dry ice or in the 193 K freezer. Once the sample is defined in the system, *GPSamplr* dictates the location in the freezer: the first free spot in an appropriate freezer box. For individual samples, stored in 1.6 ml Eppendorf tubes or similarly sized cryo-vials, the same process of entering information about the sample is followed, but the *GPSamplr* software will direct the tube to be stored in a free grid point of a gridded box. If many Eppendorf tubes of the same sample are entered into *GPSamplr* simultaneously then all the tubes are placed together in the first grid with enough free, contiguous grid positions.

3.3.2. Sample inventory. Clicking the ‘Current Samples’ button on the *GPSamplr* graphical user interface (GUI) opens a spreadsheet-style table consisting of all the information about the samples in the system. This table can be searched with a (partial) string search function, and the table can be sorted by any column. Selecting a row defining a sample opens a details page (Fig. 3) and gives the user the option of removing the sample, or one or more aliquots of a sample. Only final removal of a sample is recorded – for vials which are taken from the freezer and then returned, no action is taken.

Each time a sample is removed, the aliquot count is decremented appropriately. If the final sample is removed the user is given the option of removing the empty tube as well. A record of all removal transactions is kept, including the date, who removed it, what it was used for, and if there are any remaining aliquots or tubes of the same sample. Once the final tube of a sample has been removed, information about that sample is still maintained in the database but is excluded from the ‘Current Stocks’ table by default. The information about all stocks is available through the interface by selecting an option to show all samples.

4. StockAide and GPSamplr technology

Both *StockAide* and *GPSamplr* are Python (Oliphant, 2007) applications which use *Tkinter* (Grayson, 2000) to create simple GUIs for sample entry and exit. Python was chosen as the development environment as it is familiar to many developers and is freely available on many platforms. Python also has the advantage of having large libraries of open-source functions (e.g. *SciPy*; Jones *et al.*, 2001), and a helpful online community (e.g. Stack Overflow, currently <https://stackoverflow.com>). *StockAide* uses a SQLite database which is populated with current stocks. In C3 this information is retrieved from the Stocks table of the CM crystallization database. Updating the *StockAide* local database from an external database is optional, thus allowing *StockAide* to be used as a stand-alone application. *StockAide* is not distributed software; in C3 there is a single machine that is used to manage the stock information.

GPSamplr data are stored in a PostgreSQL database, which in C3 runs on a virtual Windows server. This architecture was selected as planned future extensions to this program would allow for web-based sample search/checkout functionality.

Data from both *StockAide* and *GPSamplr* are backed up through regular scheduled backups of the virtual server or the local machine running *StockAide*, and the code is maintained in a GIT-based repository (Atlassian Bitbucket).

The sample/stock management tools described herein are the visible layer of a process-driven laboratory: without the underlying structure created by having and enforcing rules about stock and sample creation and naming, the use of any

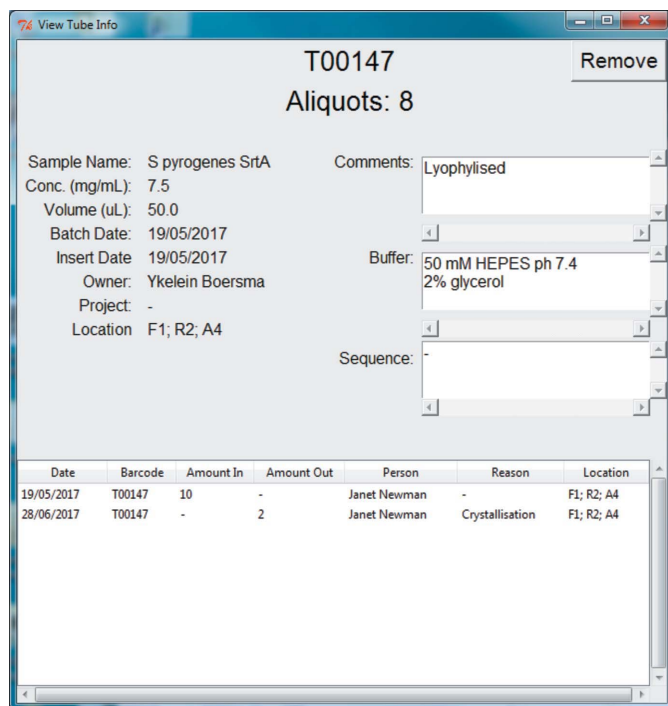


Figure 3
GPSamplr dialogue showing the fields that can be associated with each sample. A transaction history is provided, and the optional fields (e.g. Sequence) can be edited at any time.

software-based management tool is limited. The goal is to create tools which are useful enough to make up for the effort required to keep them current. Both *StockAide* and *GPSamplr* have been designed to be appropriate for use outside C3 as well. Code is available on GitHub, and implementation instructions are available on request.

5. Screening in C3

C3 currently offers a suite of 21 screens which might be suitable for initial screening of a protein sample (Fig. 4a). Each of these screens is created from the C3 stock library and requires between 40 and 85 individual stocks. Some of the C3 screens are based on commercially available screens, and the

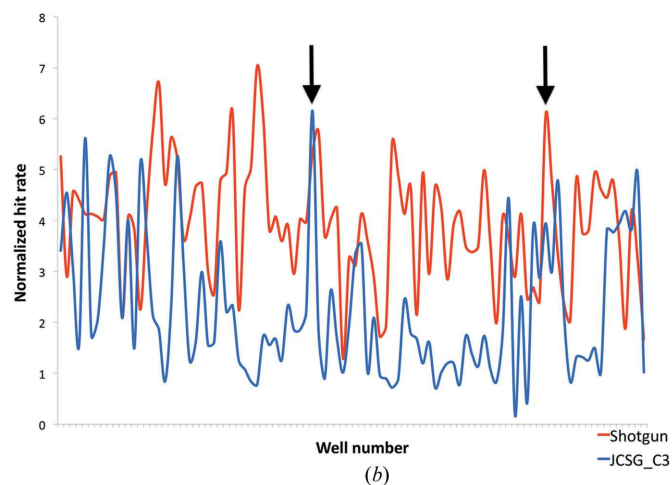
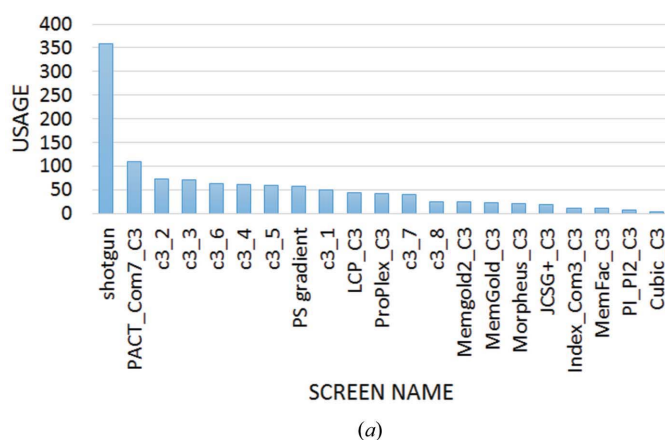


Figure 4
(a) The 21 initial screens offered in C3, with the count of the number of two-subwell vapor diffusion plates that have been set up with each screen in the year prior to November 2017. The contents of each screen can be found at <https://c6.csiro.au>. (b) Count of the normalized number of hits found in each well of the JCSG_C3 (blue) and Shotgun (red) screens. The normalized number of hits is the percentage of user-generated scores of 'crystal' for that well, divided by the number of times the screen has been set up (1674 setups of JCSG_C3, 1865 setups of Shotgun). The black arrows mark the position of the condition containing 0.1 M Tris chloride pH 8.5, 0.2 M magnesium chloride, 20% (w/v) polyethylene glycol 8000 in the two screens (well G8 in Shotgun, D6 in JCSG_C3). The most successful condition from the Shotgun screen (well C9) from this analysis contains 0.1 M Tris chloride pH 8.5, 0.2 M magnesium chloride, 25% (w/v) polyethylene glycol 3350.

commercial screen upon which the C3 screen is based is included in the name. Generally there are only minor differences between the C3 and its eponymous commercial screen; the changes have been made to streamline the process of creating the screen in-house. The Shotgun screen, which was developed in C3 (Fazio *et al.*, 2014), is the most popular initial screen by a large margin (Fig. 4a). The Shotgun screen, introduced in August 2014, has largely replaced JCSG_C3 as the ‘go to’ initial screen: both screens have been set up over 1500 times. An analysis of the success of Shotgun *versus* the JCSG_C3 screen has been done, based on user-assigned scores of the ‘crystal’ associated with images of the droplets. We calculate a ‘normalized hit rate’, which is the percent of unique ‘crystal’ annotations divided by the number of setups. User scoring is neither consistent nor complete, so we compared the hit rates of the 23 conditions in common between the two screens. There is on average a difference of 0.6 percentage points between the identical conditions in the two screens. As many images are taken of the same droplet, only one ‘crystal’ annotation was counted for each set of images of the same droplet. On average, there were 3.9 hits per well in Shotgun *versus* 2.2 hits per well in JCSG_C3. From this analysis we see that every condition in both screens has produced at least one crystal score (Fig. 4b). Overall, the JCSG_C3 screen gave a hit rate of 38% (*i.e.* 38% of the samples tested in the screen gave at least one user-defined ‘crystal’ hit), and the Shotgun screen gave a hit rate of 45%.

6. Crystallization stock usage

Although *StockAide* has been implemented for less than two years in C3, earlier records (the printed labels from exhausted tubes) for the past eight years had been kept. Once *StockAide* had been developed and tested, all the information from the labels retrieved from past stock tubes was entered into the system. These data are somewhat noisy, as there is little chance that every tube of every stock was captured, but they do give some indication of what stocks are used in the crystallography laboratory, rather than what chemicals are found in crystallization screens. Two main pieces of information are available from *StockAide*; the total number of 50 ml tubes that have been used for that chemical, and how long it takes to use up 50 ml of the stock. A list of the *StockAide* records for C3 stocks (in .csv format) from July 2017 is attached as supplementary information.

The most commonly used stock in C3 is 50%(w/v) polyethylene glycol 3350, with 245 × 50 ml aliquots of this being used since 2007. We also have a stock of 20%(w/v) polyethylene glycol 3350 (first created in 2012) that has had 15 × 50 ml aliquots used. The second most commonly used stock is 3.5 M ammonium sulfate. The association of a reason for discarding a stock tube allows us to see that the majority of the stocks are discarded because they were exhausted, but it is notable that some stocks are routinely discarded for other reasons. The nine instances of a 4 M potassium chloride stock were always found to be discarded because of precipitation; as a result, 3 M potassium chloride is now the standard C3 stock.

6.1. Chemicals used in optimization

Using the initial screens currently available in C3, we can estimate how much of our stock usage was needed to simply make up these screens. Subtracting these estimates from our total usage gives an estimate of which chemicals are used most often in optimization in C3. Fig. 5 shows the 20 most popular stocks and chemicals in C3, according to use both in initial screens and in optimization. From the data shown graphically in Fig. 5 we can see that overall usage trends of stocks in initial screens *versus* optimizations mirror each other, as expected. However, there are some chemicals which are over-screened, particularly 2-methyl-2,4-pentanediol (MPD), and some which were under-screened – polyethylene glycol 400, for example. The curious negative value for the volume of MPD optimized results from the coarseness of the ‘used’ measure, as only completely empty 50 ml tubes are scanned out. We had already noted the over-screening of MPD in earlier work (Fazio *et al.*, 2014), as although MPD was in the top ten chemicals used in commercial screens it was absent from the top ten list of successful conditions obtained from the Protein Data Bank. This prompted C3’s move away from JCSG+ as our primary sparse matrix screen (11 conditions containing MPD) to the Shotgun screen (1 condition containing MPD). Of the ten chemicals that were found in the most successful conditions, eight of the ten are also found in our top 20 most optimized chemicals. Oddly, the two sodium acetates found frequently in successful conditions from the PDB did not make our top 20 most optimized list. This could be a result of the different metrics – ours are based on the number of grams of the chemical that we have used, rather than the number of conditions in which the chemicals are found. Sodium acetate is quite light (molecular mass 82 g mol⁻¹) and is often used at low concentrations of 0.1 M (when used as a buffering chemical) or 0.25 M when used as a salt (data from <https://c6.csiro.au/>, analysing sodium acetate use in commercial screens).

These data would suggest that a sparse matrix screen based around the polyethylene glycols 400, 6000, 5000 MME and

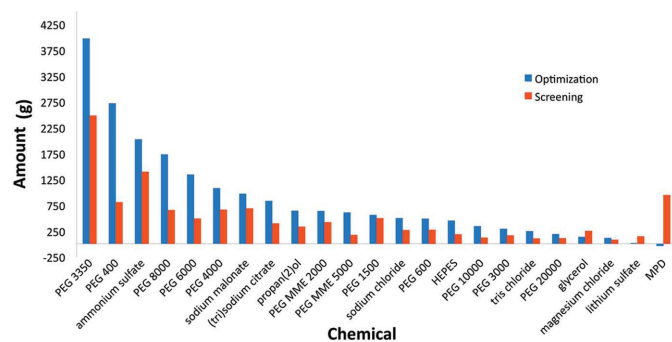


Figure 5

Estimation of the chemical usage in C3 for optimization (blue bars) and in the initial screens offered in C3 (red bars). The values were calculated by estimating what volume of a chemical stock would be needed to make up the initial screens in C3, over all the instances of the screens (blue bars). This volume was subtracted from *StockAide*'s record of the total stock volume to give an estimation of the volume of stock used in optimization.

8000 would be a useful addition to the stable of initial screens offered in C3.

7. Conclusions

Using a simple approach to stock management has enabled C3 to have a common repository of chemical stocks, which are used to create both initial screens and directed optimization screens. Having data about stock usage from almost a decade of operation allows the refinement of screening strategies, as the data can give information about the chemicals most used in optimization experiments. The organizational tool designed to manage the crystallization stocks has been used as a starting point to manage a more diverse sample set, as found in the 193 K freezer. Both of these software tools developed in C3 are freely available on request for academic use.

Acknowledgements

Our thanks to Sumeet Bal and Christine Robinson who have spent so much time making up stocks for C3. Thanks also to Pat Pilling, for sharing the logic of her protein storage system. Allen Gu is an undergraduate scholar supported by the

Biomedical Research Victoria Undergraduate Research Opportunities Program (BioMedVic UROP).

References

- Fazio, V. J., Peat, T. S. & Newman, J. (2014). *Acta Cryst.* **F70**, 1303–1311.
- Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.
- Grayson, J. E. (2000). *Python and Tkinter Programming*. Greenwich: Manning.
- Jones, E., Oliphant, T. E., Peterson, P. *et al.* (2001). *SciPy: Open Source Scientific Tools for Python*, <https://www.scipy.org/>.
- List, M., Elnegaard, M. P., Schmidt, S., Christiansen, H., Tan, Q., Mollenhauer, J. & Baumbach, J. (2017). *SLAS Discov. Adv. Life Sci. RD*, **22**, 196–202.
- Luft, J. R., Newman, J. & Snell, E. H. (2014). *Acta Cryst.* **F70**, 835–853.
- McPherson, A. & Gavira, J. A. (2014). *Acta Cryst.* **F70**, 2–20.
- Newman, J. (2011). *Methods*, **55**, 73–80.
- Olhovsky, M., Williton, K., Dai, A. Y., Pasculescu, A., Lee, J. P., Goudreault, M., Wells, C. D., Park, J. G., Gingras, A.-C., Linding, R., Pawson, T. & Colwill, K. (2011). *Nat. Methods*, **8**, 612–613.
- Oliphant, T. E. (2007). *Comput. Sci. Eng.* **9**, 10–20.
- Perkel, J. M. (2015). *Nature*, **524**, 125–126.
- Yousef, A. F., Baggili, I. M., Bartlett, G., Kane, M. D. & Mymryk, J. S. (2011). *J. Lab. Autom.* **16**, 82–89.