Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

CLIMS: Crystallography Laboratory Information Management System

Kate F. Fulton,[a] Shaun Ervine,[b] Noel Faux,[a,c,d] Richard Forster,[b] Rachel A. Jodun,[a] Wayson Ly,[b] Lee Robilliard,[b] Jai Sonsini,[b] Dan Whelan,[b] James C. Whisstock[a,c,d] and Ashley M. Buckle[a,c]*

[a]The Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Faculty of Medicine, Monash University, Clayton, Victoria 3800, Australia, [b]Swinburne University of Technology, 12-50 Norton Road, Croydon, Victoria 3136, Australia, [c]Victorian Bioinformatics Consortium, PO Box 53, Monash University, Clayton, Victoria 3800, Australia, and [d]ARC Centre for Structure and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, Australia

Correspondence e-mail: ashley.buckle@med.monash.edu.au

Received 6 May 2004
Accepted 17 July 2004

Macromolecular crystallography requires simple yet effective means of organizing and managing the large amounts of data generated by crystallization experiments. There are several freely available web-based Laboratory Information Management Systems (LIMS) that assist in these tasks. These, however, rely on the limited user interfaces allowed in HTML-based web pages. To address this limitation, a new LIMS for protein crystallization, which features a novel rich graphical user interface (GUI) to a relational database, has been developed. This application, which is called CLIMS (Crystallography LIMS), assists in all aspects of protein-crystallization projects: protein expression, handling, crystallization optimization, visualization of results and preliminary diffraction data. Extensive use of templates, particularly for commercial screens and common optimization grid screens, exploits the redundancy in experimental setups. The crystallization tray is the central focus of the graphical interface, thus facilitating rapid visualization and annotation of results. CLIMS was developed specifically to cater for the needs of individual laboratories requiring an intuitive and robust system for managing crystallization experiments and is freely available.

1. Introduction

Macromolecular crystallography has witnessed astonishing growth over the last 15 years, fuelled largely by developments in methodology and the gathering momentum of structural genomics efforts (Gerstein *et al.*, 2003; Thornton, 2001; Westbrook *et al.*, 2003). Protein crystallization forms the hub of all structural biology projects, yet it is also a major bottleneck (Chayen, 2002; Kimber *et al.*, 2003; McPherson, 2003). Effective management of crystallization projects is now of key importance given the rapidly growing adoption of protein crystallography by molecular-biology laboratories worldwide. It is also a non-trivial task because experimental details are often recorded using a combination of manual and electronic methods, with no standard data structure. In most cases, crystallization results are recorded in notebooks in an unstructured fashion and crystal diffraction data can be fragmented across several computers. This presents a serious challenge to laboratories where a plethora of proteins and variants are undergoing crystallization trials at any one time, involving many researchers. One of the most effective means of managing and organizing scientific research data is the use of a Laboratory Information Management System, or LIMS. These consist of graphical and text-based interfaces to databases that store laboratory data in a highly structured manner. There are currently more than 100 commercially available LIMS packages for protein crystallization (http://www.limsource.com). However, there are far fewer freely available or open-source alternatives. To our knowledge, the only open-source systems currently available are LISA (Haebel *et al.*, 2001), Xtrack (Harris & Jones, 2002), MOLE (Collaborative Computational Project, Number 4, 1994) and HALX (http://halx.genomics.eu.org). All of these systems use a PHP-coded web-based front-end to a relational database, usually MySQL (http://www.mysql.com) or PostgreSQL. There are also similar web-based LIMS for protein crystallography that have been constructed for structural genomics consortia, involving a large number of groups worldwide (Goh *et al.*, 2003). All of these systems are sufficiently detailed in their design and allow data from structural biology experiments to be stored in a highly organized and logical fashion. However, their interfaces and usability are limited by the restricted design imposed by HTML layout on a web page. To meet the need for a LIMS with a rich, responsive interface, we have developed CLIMS (Crystallography Laboratory Information Management System). To maximize usability, we designed a graphical user interface (GUI) into a software application, rather than a web page. The central visual focus point of the GUI is the crystallization tray, because this is often the central physical element of crystallization

experiments. CLIMS was developed specifically to cater for the needs of individual laboratories looking for an intuitive easy-to-use system for managing crystallization experiments. CLIMS also has the added capability of managing protein expression, purification and crystal diffraction data.

2. Implementation

CLIMS consists of two components: the user interface is a Java desktop application, acting as the graphical interface to the actual data, which is stored in a relational database. The Java application runs on all modern operating systems and the database is hosted by open-source relational database software that exists on a networked computer. CLIMS uses a 'top-down' approach by utilizing a simple hierarchy: Super-project: Project:Experiment:Template. Super-projects are collections of projects that share the same theme. For example, the super-project 'FKBP12 structural biology' contains the projects 'FKBP12 W59 mutants' and 'FKBP12 wt-Rapamycin complex'. Experiments, which are essentially crystallization tray setups, are grouped within a project. The main CLIMS interface is shown in Fig. 1. Templates are configurations of reagents within a tray that can be reused in more than one experiment. For example, a typical template would be a 24- or 48-condition commercial crystallization screen that would appear in many experiments. This exploits the redundancy present in crystallization screening. CLIMS manages both 24- and 96-well crystallization plates and is therefore particularly useful for managing data generated by liquid-handling crystallization robots employing 96-well plates. Templates can be constructed by either choosing reagents from the database to build up each well solution or, if the template consists of concentration gradients of one or more reagents (*e.g.* PEG *versus* salt), by specifying a few parameters such as concentration minima, maxima and step size. In this case, CLIMS creates the template automatically and also provides a protocol for preparing the crystallization tray from reagent stocks. One of the most powerful means of improving upon crystallization success in initial screens is further screening against a large number of additives. CLIMS allows templates to be constructed for an additive screen and these to be overlaid on top of another constructed template. For example, where promising small crystals grow in 1.5 *M* NaCl, 100 m*M* HEPES pH 7.5, a 24-well template of identical conditions can be easily overlaid with a 24-condition additive screen. Relevant protein expression and purification data can also be recorded for each protein entry in the database.

Individual drops in a well can be inspected by clicking on the relevant well of the tray (Fig. 2). CLIMS is currently configured to manage data for 1–4 drops per well. Clicking on the drop icon brings up a dialogue window where the following observations and annotation can be made.

(i) Visual observations: a description of the drop contents is formulated using a combination of slider bars for states such as 'precipitate', 'phase separation', 'needles', 'single crystals' *etc.* (Fig. 3). Because the drops will be inspected on multiple occasions, a list of observations can build up and

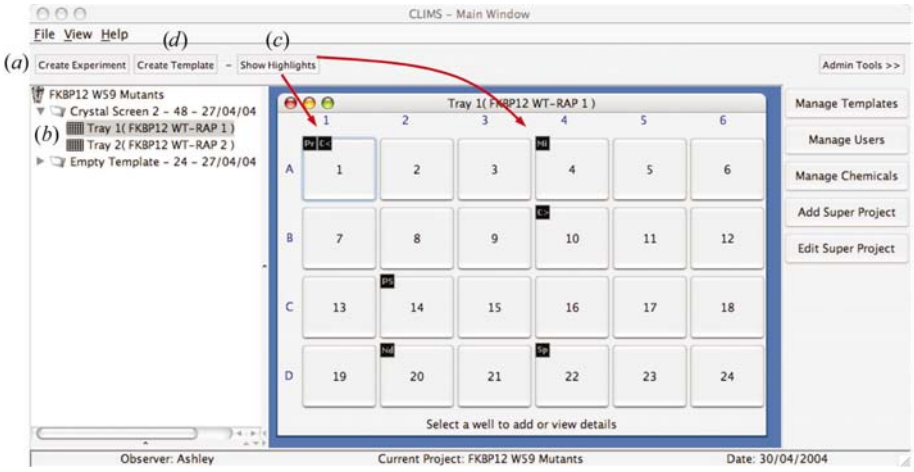

Figure 1
Main CLIMS interface. Experiments are grouped under projects on the left-hand side of the screen. (*a*) New experiments can be created for this project; (*b*) selecting an experiment from this menu displays the relevant crystallization tray, shown in the centre; (*c*) selecting the required parameters from 'Show Highlights' (*e.g* 'Mi' = microcrystals, 'Pr' = precipitate *etc.*) overlays these observations over the relevant well – clicking on these shows the entries in more detail; (*d*) templates can be created and are available to all users. The tools available to administrators are shown on the right-hand side of window.

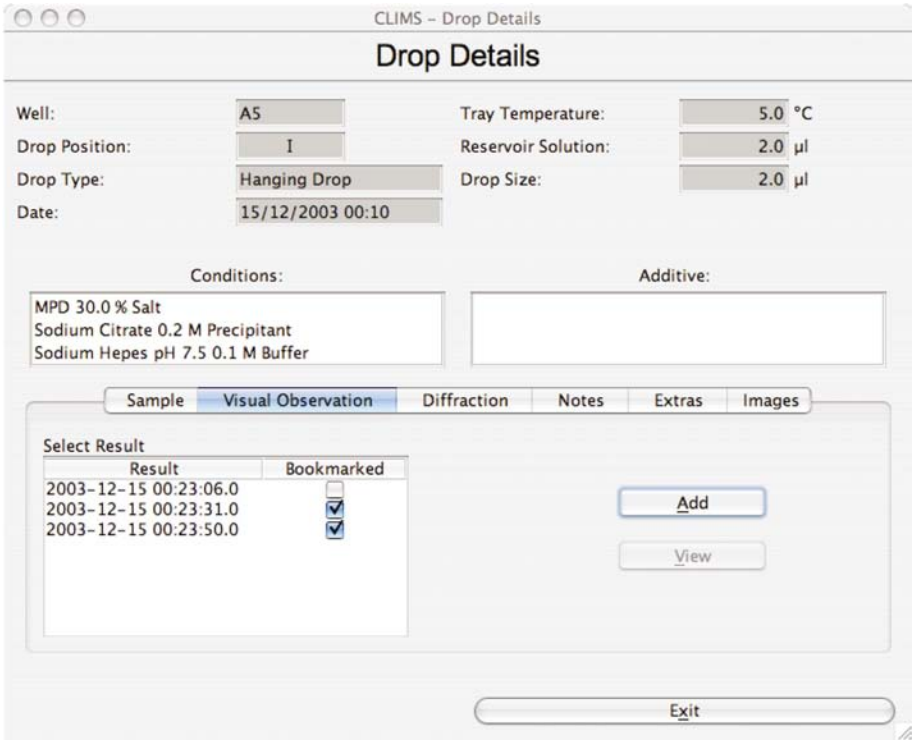

Figure 2
Selecting a well allows data annotation. Visual observations are time-stamped and notable observations can be bookmarked. Details of the drop conditions and other parameters are also shown. Diffraction data and further annotation is also possible, as well as uploading images of crystals.

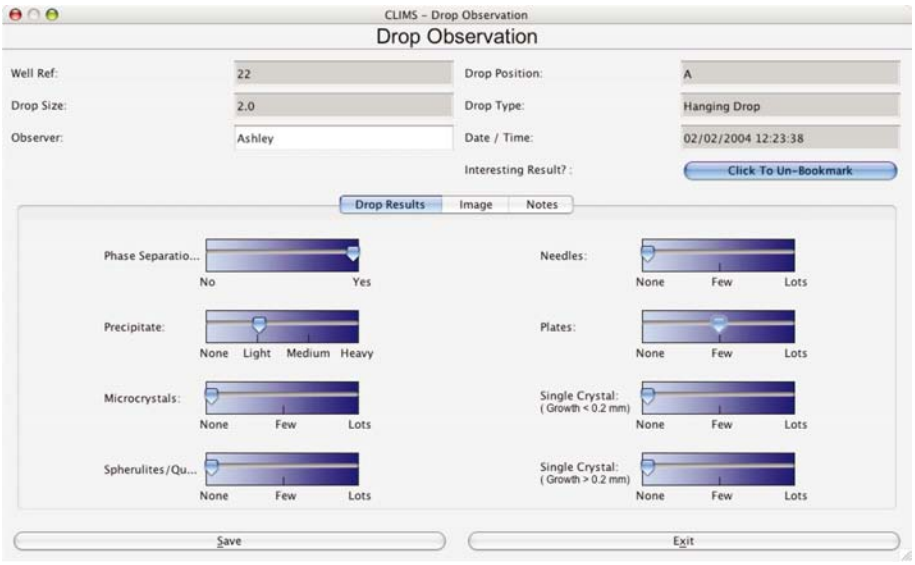

Figure 3
Visual observations can be made using a selection of sliding scales to accurately describe the physical characteristics of the drop, such as 'Phase separation', 'Needles', 'Micro-crystals' *etc*. Each observation can be easily bookmarked for later inspection.

may become unwieldy. To circumvent this, noteworthy results can be bookmarked and thus retrieved easily. Classification of drops according to variables such as 'needles', 'precipitate' *etc.* facilitates rapid inspection of a whole crystallization tray: icons for each 'state' appear over the relevant drops, allowing the user to explore encouraging results without extensive searching throughout the tray.

(ii) Images of crystals captured digitally can also be uploaded.

(iii) Crystal diffraction: details of diffraction data for crystals can be stored, such as resolution limit, cryocooling conditions, data-collection statistics, details of crystal storage *etc.*

All observations and annotations are time-stamped.

There are two levels of login security access built into CLIMS, based upon a username/password combination: most users login using a single normal user account, allowing them to create projects, templates and experiments; administrators have additional power to make certain projects 'secure', meaning that they cannot be viewed by normal users. Also, certain administration tools are available to manage data, including templates, super-projects, chemical reagents and users.

3. Technical aspects

CLIMS was created using the following cross-platform tools: Java Development Kit version 1.4.1 (Sun Corp), MySQL relational database server software v.4.016 (http://www.mysql.com), the JGoodies plastic look and feel 'skin' (http://www.jgoodies.com), JUnit (http://www.junit.org) for testing and the NetBeans integrated development environment (http://www.netbeans.org). CLIMS has been tested under the Mac OS X 10.2 and 10.3, Windows NT/2000/XP and Linux operating systems. All data are stored in a MySQL relational database, consisting of 27 tables. The Java application runs on most PC, Mac and Unix environments that have Java 2 runtime environment (JRE) installed. Additionally, CLIMS can be used over the internet using the cross-platform Java WebStart technology (http://java.sun.com/products/javawebstart): it is deployed on a standard web server and accessed using any web browser. This is particularly useful when collecting diffraction data at synchrotrons. Data is backed up automatically daily, weekly and monthly onto removable media, using the 'mysqldump' utility built into the MySQL software and the Unix cron command, and stored off-site.

4. Current use and future development

CLIMS is used in our laboratory to manage a substantial amount of crystallization data generated by many projects and involving several medium-size groups. There are several modifications to CLIMS currently under development, notably (i) integration into a laboratory-wide LIMS for molecular biology and (ii) interface improvements and additions such as the ability to store diffraction images, perform detailed database searching (including raw SQL queries), printing formatted hard copies of database entries, a fast visual template creator and keyboard customization for faster data entry and visualization. There are several enhancements that we envisage developing in the long-term: integration with *XY* platforms and USB microscopes (and subsequent interfaces to image-analysis technologies), data-mining tools that facilitate crystal optimization, integration with liquid-handling robots, electronic signatures for cases where intellectual property matters arise and built-in voice recognition to enhance usability.

5. Documentation and availability

CLIMS can be accessed at http://clims.med.monash.edu.au/.

We thank Ruby Law for helpful discussions and Jamie Rossjohn for enthusiasm and support. AMB wishes to thank the Apple University Consortium Australia for financial support. JCW is a National Health and Medical Research Council Senior Research Fellow and Monash University Logan Fellow. The work was supported by the NHMRC, the ARC, the Monash University Research Fund as well as the State Goverment of Victoria.

References

Chayen, N. E. (2002). *Trends Biotechnol.* **20**, 98.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Gerstein, M., Edwards, A., Arrowsmith, C. H. & Montelione, G. T. (2003). *Science*, **299**, 1663.
Goh, C. S., Lan, N., Echols, N., Douglas, S. M., Milburn, D., Bertone, P., Xiao, R., Ma, L. C., Zheng, D., Wunderlich, Z., Acton, T., Montelione, G. T. & Gerstein, M. (2003). *Nucleic Acids Res.* **31**, 2833–2838.
Haebel, P. W., Arcus, V. L., Baker, E. N. & Metcalf, P. (2001). *Acta Cryst.* D**57**, 1341–1343.
Harris, M. & Jones, T. A. (2002). *Acta Cryst.* D**58**, 1889–1891.
Kimber, M. S., Vallee, F., Houston, S., Necakov, A., Skarina, T., Evdokimova, E., Beasley, S., Christendat, D., Savchenko, A., Arrowsmith, C. H., Vedadi, M., Gerstein, M. & Edwards, A. M. (2003). *Proteins*, **51**, 562–568.
McPherson, A. (2003). *J. Struct. Biol.* **142**, 1–2.
Thornton, J. (2001). *Trends Biochem. Sci.* **26**, 88–89.
Westbrook, J., Feng, Z., Chen, L., Yang, H. & Berman, H. M. (2003). *Nucleic Acids Res.* **31**, 489–491.