Journal of Structural Biology 175 (2011) 230–235

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

xtalPiMS: A PiMS-based web application for the management and monitoring of crystallization trials

Ed Daniel [a], Bill Lin [a], Jonathan M. Diprose [b,c], Susanne L. Griffiths [d], Chris Morris [a], Ian M. Berry [b,1], Raymond J. Owens [b,c], Richard Blake [a], Keith S. Wilson [d], David I. Stuart [b,e], Robert M. Esnouf [b,f,*]

[a] CSED, STFC Daresbury Laboratory, Warrington WA4 4AD, UK
[b] Division of Structural Biology, University of Oxford, Henry Wellcome Building for Genomic Medicine, Roosevelt Drive, Oxford OX3 7BN, UK
[c] The Oxford Protein Production Facility UK, Research Complex at Harwell, Rutherford Appleton Laboratory, R92, Harwell Oxford, Didcot OX11 0FA, UK
[d] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK
[e] Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK
[f] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

ARTICLE INFO

Article history:
Available online 14 May 2011

Keywords:
Laboratory Information Management Systems (LIMS)
Protein crystallization
Robotic imagers
Java web application
Data management and databases

ABSTRACT

A major advance in protein structure determination has been the advent of nanolitre-scale crystallization and (in a high-throughput environment) the development of robotic systems for storing and imaging crystallization trials. Most of these trials are carried out in 96-well (or higher density) plates and managing them is a significant information management challenge. We describe xtalPiMS, a web-based application for the management and monitoring of crystallization trials. xtalPiMS has a user-interface layer based on the standards of the Protein Information Management System (PiMS) and a database layer which links the crystallization trial images to the meta-data associated with a particular crystallization trial. The user interface has been optimized for the efficient monitoring of high-throughput environments with three different automated imagers and work to support a fourth imager is in progress, but it can even be of use without robotics. The database can either be a PiMS database or a legacy database for which a suitable mapping layer has been developed.

1. Introduction

There have been rapid advances in the techniques of protein production, crystal growth and structure determination, accelerated in part by structural genomics and structural proteomics initiatives (for recent reviews and overviews of technology platforms see Terwilliger et al., 2009; Weigelt, 2010; Chruszcz et al., 2010; Elsliger et al., 2010; Xiao et al., 2010). A particular aim has been to develop generic techniques which can be applied in parallel, in miniature and on automated platforms. Nowhere have these advances been greater than for protein crystallization by vapor diffusion, where nanolitre-scale pipetting technologies have allowed for at least an order of magnitude reduction in the volume of sample required per crystallization trial coupled with at least an order or magnitude increase in the rate of creation of these trials (reviewed in Berry et al., 2006). This allows many more crystallization trials to be performed, giving a much better sampling of crystallization space and an increase in the number of successful crystallizations. In laboratories with sufficient throughput, a concomitant need for the robotic management of these crystallization trials has been created: storing the trials under controlled conditions coupled with a regular schedule of imaging sessions (also in a controlled environment).

The information management problem associated with the creation of many thousands of crystallization trials is substantial. For laboratories without imaging systems, researchers have to spend long periods at microscope stations and develop more efficient ways of recording the outcomes of trials. For laboratories with access to imaging systems the scale of the problem is much greater, but the scheduled recording of images electronically (*e.g.* as JPEG images) makes it possible to develop management software. Manufacturers of storage and imaging systems usually provide some management software but this may not be suited to academic environments for many reasons, including that

- the user interface may be designed for optimization of known crystallization conditions rather than the discovery of new ones (*i.e.* good for detailed analysis of a few drops rather than quickly scanning many drops for the few good outcomes);

* Corresponding author at: Division of Structural Biology, University of Oxford, Henry Wellcome Building for Genomic Medicine, Roosevelt Drive, Oxford OX3 7BN, UK.
E-mail address: robert@strubi.ox.ac.uk (R.M. Esnouf).
[1] Present address: Evotec (UK) Ltd., 114 Milton Park, Abingdon OX14 4SA, UK.

1047-8477 © 2011 Elsevier Inc. Open access under CC BY license.
doi:10.1016/j.jsb.2011.05.008

- the software may not handle multiple users in a "crystallization service" model with remote clients and accounting;
- the software may only run on the local control computer;
- the licensing terms or development decisions may restrict the number of concurrent connections; and
- the software itself can be very expensive.

Electronic recording of laboratory information has many potential benefits, and some of these have been discussed elsewhere in an article on the Protein Information Management System (PiMS; Morris et al., 2011). However, as we point out, for low-throughput work on the benchtop (where the benefits are often only felt in the distant future or by other people) adoption of electronic information management is still limited. It was noted that in certain areas electronic recording of data is almost essential: for work that is (i) miniaturized, (ii) parallelized and/or (iii) automated. Protein crystallization is a case in point, exploiting all three approaches, and thus is an excellent candidate for electronic information management.

The Oxford Protein Production Facility (OPPF), established with the support of the Medical Research Council (MRC) in 2001, has been influential in developments in high-throughput crystallization screening. During 2002, the OPPF commissioned a custom-integration of a storage system (The Automation Partnership, Royston, UK) and an imaging station (Veeco-Optimag, San Diego, USA) with capacity for 10,000 SBS-format 96-well plates (Walter et al., 2003,2005; Brown et al., 2003) and developed the PHP-based in-house *Vault* software for web-based management of these trials (Mayo et al., 2005). In complete contrast to the tardy adoption of wet-laboratory Laboratory Information Management Systems (LIMS), there was an (almost) overnight switch to complete dependence on this software across both the OPPF and the Division of Structural Biology (STRUBI) in Oxford. Support was subsequently added for the Oxford lab's RI1000 and RI182 imagers (Formulatrix, Waltham, USA), but the user interface was not developed and the database schema remained closely aligned to the original storage system. Other web-based viewing software has been developed for crystallization facilities, *e.g.* the CRIMS application primarily in use at EMBL sites (http://embl.fr/htxlab/). CRIMS is a PHP-based application offering a hierarchical view of User → Plate → Thumbnails/Well Images but compared to the *Vault* software it lacks project grouping, zooming, movie/timecourse modes and support for automated analysis.

This article describes xtalPiMS, a web-based application for the management and monitoring of crystallization trials. xtalPiMS has a user-interface layer based on the standards of PiMS and a database layer which links the crystallization trial images to the meta-data associated with a particular crystallization trial. The user interface has been optimized for the efficient monitoring of high-throughput environments. However, it can also be used without robotic imagers: at the Oxford lab an external digital camera has already been linked into xtalPiMS using a customized TWAIN driver to ensure systematic naming of images. The database can either be an instance of a PiMS database or (as is the case at Oxford) a legacy database for which a suitable mapping layer has been developed. xtalPiMS is in service in Oxford, on the UK's new national protein production facility at the Rutherford-Appleton Laboratory site (RAL; using Formulatrix imagers) and at the York Structural Biology Laboratory (YSBL) to support its BioTom imager. Having already been adapted to support three imagers, work is now underway to support Rhombix (Thermo Scientific) systems at the RAL site, Helsinki and Oulu (Finland) and a feasibility study indicates that xtalPiMS could be integrated with the Minstrel (Rigaku) system. Our model is that xtalPiMS interfaces with the imager databases only to read the crystallization trial meta-data into PiMSdb and then independently scans the image repositories for new imaging sessions. Thus, developers can integrate new imagers with xtalPiMS in a matter of weeks since the dependence on imager software is kept to a minimum and manufacturers have been willing to share either database schemas or APIs, as appropriate.

2. Methods

2.1. Data models and databases

The PiMS LIMS is based on a sophisticated and complex data model developed from the Protein Production Data Model (PPDM; Pajon et al., 2005) and embodied in the PiMS database (PiMSdb). However, this complexity is largely hidden from the end user (Morris et al., 2011). The model is generalized enough to cope with all eventualities of protein crystallization and a requirement for xtalPiMS was for it to work with PiMSdb and in a complementary manner to PiMS. Thus, PiMS and xtalPiMS can be considered to be two views onto the same data store. Indeed, since they share a common access control mechanism, share many software components and can share the same database, the user need not be aware that they are distinct. The decision to maintain separation was made primarily for developer, administrator and user convenience.

With xtalPiMS development, however, there were additional requirements. First, the software had to be potentially usable with alternative databases, for example the large amount of data already stored in the Oxford crystallization database (PlateDB; Mayo et al., 2005) or that of an equipment manufacturer. Second, xtalPiMS needs to have a flexible software interface to abstract and support different (and even multiple) imagers. Third, the large number of crystallization images means that the images themselves are not stored directly within the database, rather they are recorded to an external image store and xtalPiMS has to detect the creation of these image files and link them to the crystallization meta-data. Finally, xtalPiMS must allow users to browse very large numbers of images quickly and ergonomically.

Database performance is a significant issue for xtalPiMS. The requirement not to assume a standard way of working means that PiMSdb is complex and much effort has been spent to optimize its responsiveness for xtalPiMS. However, within one laboratory most crystallization trials are very standardized. For example, the PlateDB database retained simplicity by exploiting the Oxford standard crystallization protocols: only one protein per 96-well plate and sets of standard screens (Walter et al., 2003).

2.2. Implementation issues in the crystallization trial browser

The user experience of xtalPiMS is focused on one page, the viewer for a single image of a crystallization trial (an example is shown in Fig. 1). The features of this page from a user-perspective are described in Section 3.2. However, the over-riding design goals have been:

- to allow the user to browse conveniently and rapidly through thousands of images in a single session;
- to offer efficient keyboard shortcuts for commonly used functions;
- to display images as large as possible, and to allow easy zooming and repositioning of images;
- to allow the user easily to locate related images, such as from previous imaging sessions or other plates containing the same protein.

xtalPiMS achieves these goals with heavy dependence on AJAX technologies to allow background fetching of crystallization