Structure, Vol. 13, 175–182, February, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2004.12.010

Benefits of Automated Crystallization Plate Tracking, Imaging, and Analysis

Technical Advance

Christopher J. Mayo,[1] Jonathan M. Diprose,[1] Thomas S. Walter,[1] Ian M. Berry,[1] Julie Wilson,[2] Ray J. Owens,[1] E. Yvonne Jones,[1] Karl Harlos,[1] David I. Stuart,[1] and Robert M. Esnouf[1,*]

[1]The Division of Structural Biology
The Henry Wellcome Building for Genomic Medicine
Roosevelt Drive
Oxford, OX3 7BN
United Kingdom

[2]The York Structural Biology Laboratory
University of York
Heslington
York, YO10 5YW
United Kingdom

Summary

We describe the design of a database and software for managing and organizing protein crystallization data. We also outline the considerations behind the design of a fast web interface linking protein production data, crystallization images, and automated image analysis. The database and associated interfaces underpin the Oxford Protein Production Facility (OPPF) crystallization laboratory, collecting, in a routine and automatic manner, up to 100,000 images per day. Over 17 million separate images are currently held in this database. We discuss the substantial scientific benefits automated tracking, imaging, and analysis of crystallizations offers to the structural biologist: analysis of the time course of the trial and easy analysis of trials with related crystallization conditions. Features of this system address requirements common to many crystallographic laboratories that are currently setting up (semi-)automated crystallization imaging systems.

Introduction

Over the past few years a number of structural genomics consortia have invested heavily in devising high-throughput technologies, which can be linked to form a complete structural biology pipeline (Lesley et al., 2002; Rupp, 2003a; Adams et al., 2003; DiDonato et al., 2004; O'Toole et al., 2004). This intense, and well-funded, activity would be expected to change the face of structural biology in all its aspects by identifying the points of weakness in the overall process, prompting assessment of alternative strategies and hence finding improved solutions. The general thrust has been to seek generic methodologies that are amenable to automation, miniaturization, and parallelization. In order to optimize the technologies and manage the process effectively, informatics developments are also required; indeed, the proper capture of both positive and negative results and the potential to query and mine the databases may contribute significantly to the efficiency of the process (Hui and Edwards, 2003; Rupp, 2003b; Page et al., 2003; Goh et al., 2004; Page and Stevens, 2004). In the early stages, the major effort in structural genomics was focused on establishing the core technologies and the major stumbling block was usually the production of soluble protein suitable for crystallization studies. While this remains a major hurdle for many interesting proteins, especially those of higher eukaryotic organisms, it is becoming increasingly clear that the fundamental step of producing crystals suitable for high-resolution diffraction analysis remains problematic. Developments led by high-throughput activities are already starting to have a major impact in this area (Hui and Edwards, 2003; DeLucas et al., 2003; Chayen, 2003; Brown et al., 2003; Page and Stevens, 2004; Rupp and Wang, 2004) and in this article we consider an implementation of automated crystallization plate tracking, imaging, and analysis which demonstrates this potential. Highly roboticized systems are under development at a number of industrial sites but there is little information in the public domain about most of this activity. We describe here the strategy of the UK MRC-funded Oxford Protein Production Facility (OPPF), which has developed automated crystallization, and which implements and integrates a number of high-throughput technologies. While none of the individual procedures are unique, we believe that their integration gives one of the first glimpses of the full impact that they can potentially make on crystallography. The workflow and robotics of the OPPF crystallization facility have been described elsewhere (Walter et al., 2003; Brown et al., 2003; Walter et al., submitted). One aim in this facility was to achieve true hands-off automation of crystallization plate storage and imaging under full software control. This has resulted in the design and implementation of a robotic system able to store, retrieve, and image 10,000 96-well crystallization plates in accordance with predetermined imaging schedules. This technology is capable of capturing over 100,000 images of crystallization trials per day (20 terabytes of data per year). The demands placed on image analysis and database management have driven a matched development of a web-service based informatics suite. This has been created as part of the OPPF contribution to the framework of the UK BBSRC-funded eHTPX project, which is establishing a pipeline for distributed and high-throughput protein crystallography. Thus, many of the solutions developed in this high-throughput context will be equally applicable to (and indeed made available to) smaller laboratories, both for use as a remote resource and for local installation. We describe here the underpinning informatics with particular focus on the functionality that allows the typical user to keep track of, analyze, and respond efficiently to the data contained in large, regularly incrementing, sets of images. Finally, we consider the benefits that the systematic collection and organization of images, effectively a col-

*Correspondence: robert@strubi.ox.ac.uk

lection of time-lapse movies of crystallogenesis, has for the experiment of crystallization itself.

Results

The Population of the Crystallization Database

The OPPF workflow has been reported in detail elsewhere (Walter et al., submitted). In brief, nanoliter scale crystallization screens are set-up as sitting drops in 96-well format using separate liquid dispensing robots to prepare stock plates, dispense reservoirs, and dispense crystallization drops onto the flat bottom platforms of Greiner plates. The majority of these plates are stored in an automated vault which is integrated with an imaging system, the rest being stored in a cold room with a semiautomated imaging system (see Experimental Procedures). As of December 16, 2004, there were 123 registered users of the facility who had set up 8481 crystallization plates, and 17,645,760 images had been recorded into the OPPF crystallization database (PlateDB, see Experimental Procedures) covering approximately two-thirds of a million crystallization trials. The day-to-day operation of the facility depends on efficient data capture at each stage in the set up and subsequent history of each crystallization plate. Where possible, barcoding is used to minimize user input (e.g., of stock solutions in addition to crystallization plates). Data capture requires seamless data exchange between a series of databases in order (1) to feed data into and out of the protein production database, (2) to integrate automated plate storage with scheduled plate imaging, and (3) to store all data relevant to the crystallization experiment in the main database, PlateDB. A schematic of the OPPF crystallization IT infrastructure is presented in Figure 1 and details of hardware and software are given in the Experimental Procedures section. This network has provided a robust and adaptable solution that runs with minimal hands-on intervention to maintain a detailed and reliable crystallization database. After the initial stages of crystallization plate set-up, the handling of plates by users is restricted to occasional inspection using a standard laboratory microscope with attached digital camera (this can be very useful for example, to record an image at higher magnification, using different illumination, or to assess and digitally record birefringence), and harvesting of crystals for X-ray diffraction data collection. Images recorded during manual inspection are associated with the appropriate crystallization well in the PlateDB. Thus, the central, automatically updated PlateDB database contains all the data relating to users' crystallization experiments and the tools developed for systematic interrogation of this database are crucial in determining the final effectiveness of the whole system. The following sections, therefore, focus on a description of these web-based tools from the user perspective and an assessment of the added value resulting from their use.

A Web-Based Interface to the Crystallization Database

Essentially all crystallization screening by groups within the Division of Structural Biology in Oxford is carried out using the OPPF facility (i.e., samples feed in from projects carried out in the standard laboratories as well as from the OPPF protein expression and purification pipeline). The OPPF crystallization facilities are also used by guests from other local, UK, and SPINE (*S*tructural *P*roteomics *in E*urope) project partner laboratories (of the 123 registered users the majority are external to the Division of Structural Biology). A web-based form is available for initiating such collaboration. The crystallization database is constructed to function as a standalone resource. For crystallization plates set up at the OPPF, information on the plate ownership and contents is pulled from the OPPF protein production laboratory information management system (LIMS; Nautilus) using standard SQL queries when the plate enters the vault. At another laboratory, any other database (or indeed user interface) capable of supplying this information would be equally applicable. Since each user potentially requires efficient access to the crystallization database from a variety of local and remote locations (including for example synchrotrons), we have developed an interface for database interrogation based around a dynamically produced web interface protected by secure log-on procedures.

When a registered user logs on to the OPPF web site, (s)he is given a "My Vault" link which goes to a personalized home page (Figure 2A) the entry point allowing rapid browsing, easy management of many images (and, hence, crystallization trials), and annotation of their own crystallization images. (For a sample session navigate to http://www.oppf.ox.ac.uk/vault.) The page layout is deliberately simple, based around tables and forms, so that navigation is rapid without unnecessary data transfer overheads. The My Vault home page provides general navigation buttons along the top, short cuts to manually classified images and ways to customize the interaction with PlateDB to the individual user's preferences, such as whether the user wants to be notified of each imaging session by email. The "Plates" navigation button goes to a table of all the users plates, their contents, set-up date, and last imaging date (Figure 2B), whereas the "OPPF" navigation button provides a list of projects and for a selected project shows the plates associated with it (Figure 2C). Within the OPPF, a project is defined as the protein product of a unique expression construct or, in the case of complexes, constructs, and is denoted by a unique OPPF number. On any of these pages, clicking on an OPPF number produces a list of plates associated with it, clicking on a plate number gives a list of all imaging sessions for the plate, and clicking on an imaging session goes to the view described below. The home page lists a user-created catalog of well images of special interest; clicking on a catalog entry provides the most recent full-sized image of that well.

The bioinformatics and target tracking sections of the OPPF informatics suite are also directly linked to the My Vault plate list pages for individual OPPF numbers, and since these databases are controlled by a single log-on/authorization component, these navigations are seamless (Figure 2D). Thus, from looking at a crystallization well image it only requires one mouse click to access construct information for the protein in the well (e.g., count of number of methionine residues), one more click to access the constructs derived from

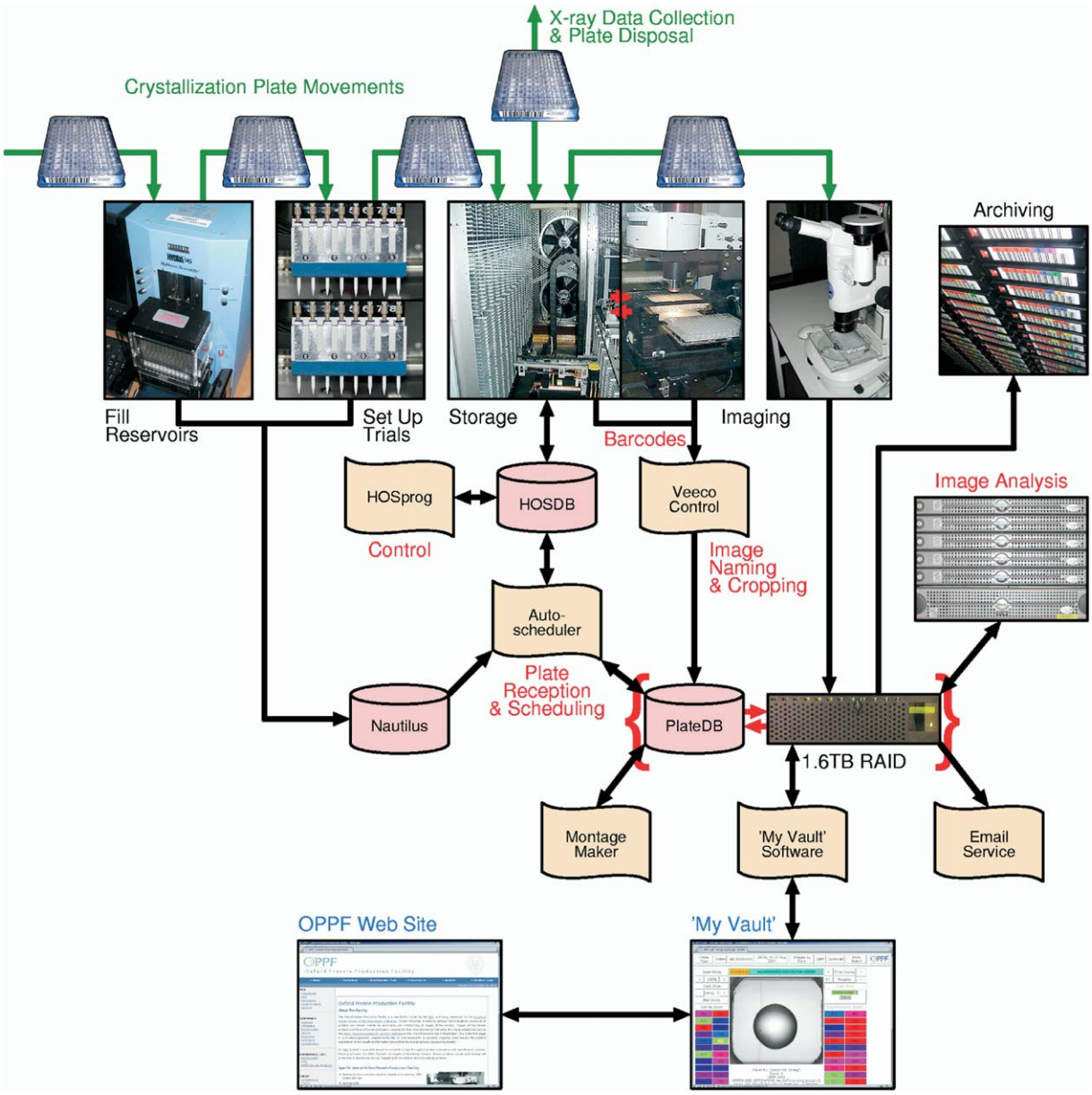

Figure 1. Schematic Showing the Process and Operation of the OPPF High-Throughput Crystallization Facility

Green arrows show the transfer of 96-well crystallization plates between robots, shown by photographs. The flow of images and control data is shown by black arrows. Databases are indicated by pale red "disk cylinders" and specific sections of the control software are represented by orange "paper". Curly braces and arrows indicate that PlateDB and the RAID storage work together as a single information resource. The method of interaction with the web interface is indicated by sample web pages.

the full-length protein and then a final click to get the automated bioinformatics analysis for the full-length protein summarized graphically.

The Plate-Level View Web Page

The plate-level view is reached by clicking on the link for a specific imaging session for a plate either from the individual plate page or else from the chronological list of imaging sessions for all the user's plates. Most commonly, users will wish to inspect newly imaged plates and a direct link to the relevant plate-level view is provided in the optional notification email. At the plate level, the user is presented with a top bar of handy navigation links below which is a montage of 96 thumbnail images of wells matching the physical plate format (Figure 2E). Each image is given a border color-coded to correspond to the automatic image analysis score (Wilson, 2002, 2004). Empty drops are coded in red with intermediate levels of interest in purples and dark blues through to cyan for a drop containing good crystals. If no drop is detected in the image (for example, only a subset of the positions in the 96-well plate

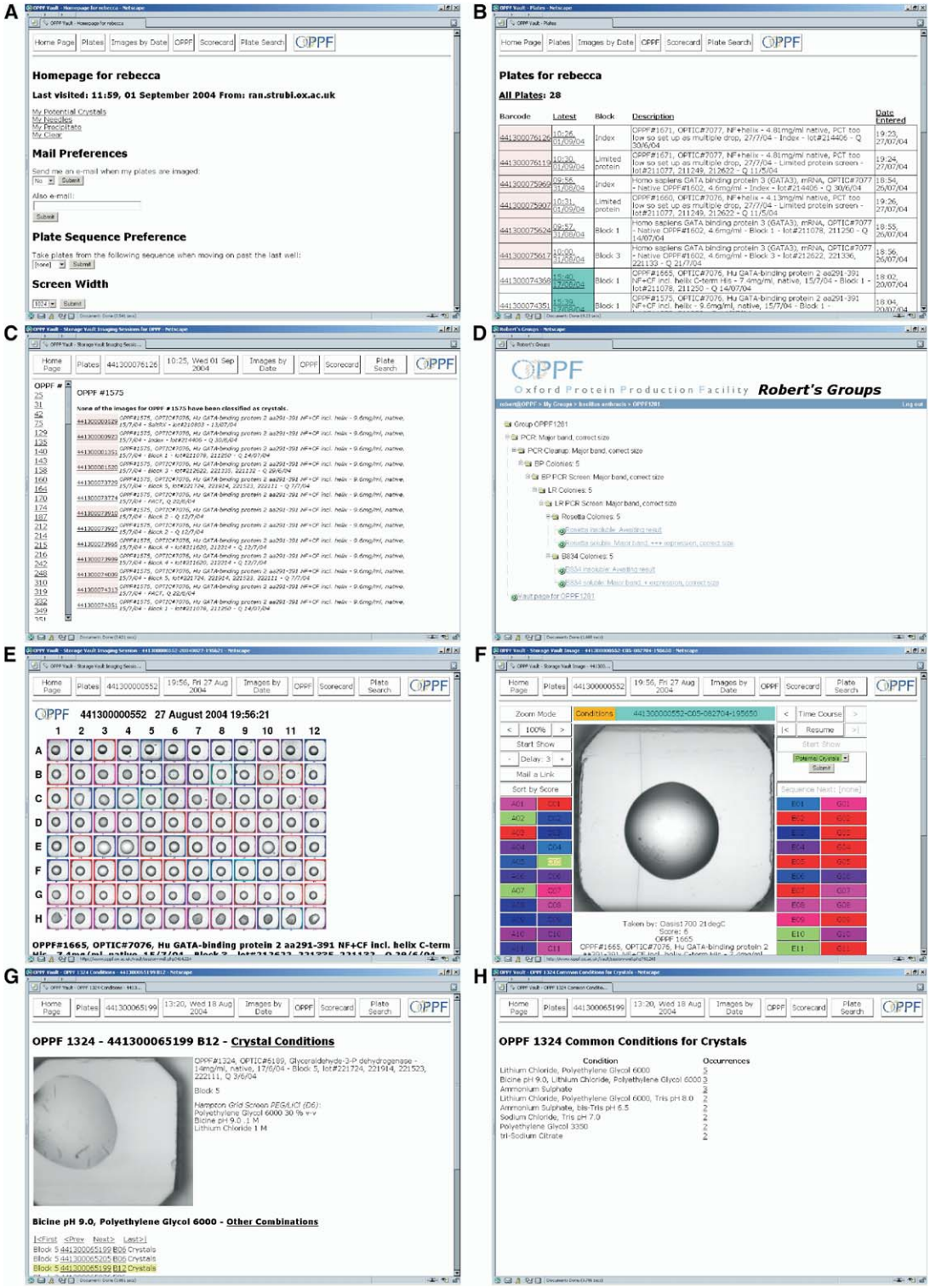

Figure 2. Sample Output for the Web Interface to the Crystallization Database (PlateDB)

(A) A user's home page allowing setup of preferences.

(B) List of plates belonging to a user, showing contents, time of creation, and time of last imaging.

(C) List of plates for a particular target (denoted by OPPF number) showing contents of each plate.

(D) Link from bioinformatics pages to pages on protein production and crystallization data.

(E) A plate-level view web page showing a montage of all 96 wells for a plate.

(F) A well-level view web page showing links to other wells and color coding resulting from automated and manual classification. To the top left and right are tools for well navigation, examination, and annotation.

(G) Tool for viewing well images grouped by crystallization conditions (and combinations of conditions).

(H) Output of database query for showing crystallization conditions (and combinations of conditions) that have produced crystals.

are used in the standard OPPF crystal optimization screen) then orange is used. Details of the expression construct and the type of crystallization trial (typically one of the standard set of 96 condition screens, Walter et al., 2003) are also given. The montage of thumbnails is an image map and clicking with the pointer over any image takes the user to the well-level view containing the full-sized image.

The Well-Level View Web Page

The well-level view (Figure 2F) is the main page for user interaction. In its standard mode, it avoids oversophistication allowing the user to work through thousands of images in a session in an ergonomic manner, but also provides options that allow the user trivially to drill down for additional information on any one image. The image has a title bar, color-coded according to classification (as in the plate-level view), and potentially containing three links: one to the crystallization condition search pages (see below), one to a display of any manually acquired images for the well, and one linking to any example images of salt crystals that have been observed with this reservoir condition. Below the image is annotation, largely generated automatically, which minimally describes the expression construct, the protein sample concentration, the temperature at which the plate is stored, the reservoir solution for the well, and the score assigned by automated image analysis. This annotation can be supplemented at any time with user-entered details such as additives and special drop ratios. Two image modes are supported: "Measure" and "Zoom". In (default) Measure mode, the well image is itself an image map and a pair of clicks on different places in the image will define a line and give its length (in μm), providing a quick-and-easy way to measure the size of crystals. If the user selects Zoom mode then a click over an image will produce a 2× enlarged image. Cropped well images are normally displayed full size (i.e., 750 × 700 pixels which is suitable for display on SXGA [1280 × 1024 pixels] resolution displays), but it is possible to change to a 75% view (562 × 525 pixels) for lower-resolution displays.

To the left and the right of the image are columns containing links for each well in the plate colored according to automatic or manual image classification (Figure 2F). The links within the columns can be arranged either in "plate" order (the row-by-row order of the wells within the plate) or in classification "score" order, ordered by "interest" (as judged by crystal classification).

To the top-left of the well image is a cluster of tools for (1) toggling Zoom and Measure modes; (2) toggling image size; (3) navigating forward and backward through the well links; (4) providing a way of emailing a link to an image to a collaborator; and (5) toggling between plate order and classification score order. This set of tools also contains a slide show facility that automatically displays each of the well images in turn (in plate or score order) with a user-controllable delay between images. The slide show can be stopped at anytime to allow for manual annotation of images.

To the top-right of the image is a second cluster of tools. The user can navigate backward or forward through a time series of images for an individual well using navigation buttons similar to those commonly found on CD players etc. It is also possible to run through the time series for a well as an automated slide show or to show a full series of images on a single page. Finally, there is a manual classification section that allows the user to supplement the automatically generated crystal image classification. The automated image analysis has not yet achieved a sufficient level of reliability to trust entirely, although it is sufficiently accurate at identifying clear drops that some users skip manual inspection of well images with this classification (Wilson, 2002, 2004). For the majority of the images manual (i.e., user) assessment and classification remains essential. Images can be given manual classifications ranging from clear drops through precipitation and crystals of various qualities. In practice, it is the crystal annotations that are primarily used, ranging from potential crystals, through crystals for optimization and needles to crystals flagged for immediate examination at X-ray sources such as synchrotrons. These manual annotations are the ones that are also presented on the My Vault home page allowing the user easily to keep track of promising experiments.

Crystal Condition Web Pages

Clicking on the "Conditions" button at the well-level view takes the user to a list of the components of the reservoir condition along with all combinations of these components. Next to each combination is the number of crystallization trials set up containing those components. Clicking on one of these goes to a list of those crystallization trials, with trials which contain crystals listed first (Figure 2G). The displayed image can be selected from the list or by using navigation buttons. These pages also link to a crystal conditions page, which lists combinations of reservoir components for drops which have been annotated as containing crystals and the number of trials containing these components which have produced crystals, sorted in order of decreasing frequency (Figure 2H). Clicking on one of these produces an expanded list of relevant trials for inspection.

Web Page Access

The whole PlateDB web interface can run quickly: setting a delay of zero for a slide show can deliver the well images to a local user at the rate of up to three a second (in addition to handling several other users and accepting new entries into PlateDB). The web interface is controlled from a Linux machine running Apache situated on a DMZ of the OPPF firewall (a segment of network with strictly controlled access from both inside and outside the firewall), which is allowed access to the PlateDB database. The web pages can, therefore, be externally accessed by users who have logged into the OPPF website using a secure (SSL) connection. This allows users at synchrotrons to browse crystal images and other OPPF databases to inform decisions on data collection, as well as allowing external collaborators to track their crystallization experiments.

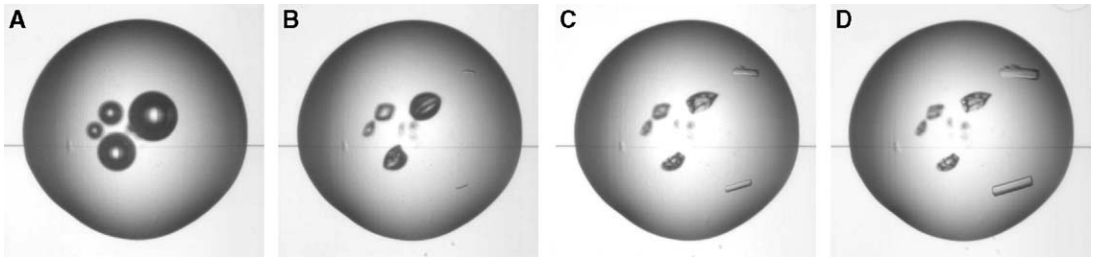

Figure 3. Images Taken at Different Times Showing the Growth of a Crystal in a Single Crystallization Experiment after (A) 0 Hours; (B) 3 Days, 3 Hours; (C) 9 Days, 21 Hours; and (D) 48 Days, 23 Hours after Set Up

The full set of images from this experiment has been deposited as Supplemental Material both as 24 separate images and as a time-lapse movie in .AVI format (available with this article online).

Discussion

We, and others, have discussed elsewhere the advantages of using small drops for crystallization trials, both in terms of economy in amount of protein required and in terms of success rate for crystal growth (Brown et al., 2003; Walter et al., submitted). The use of liquid handling robots and miniaturization also allows a very significant increase in the number of crystallization trials that can be conducted by each investigator. Automation of crystallization plate storage and imaging facilitates exploitation of this increase in capacity, but requires a well integrated IT infrastructure for its efficient management (Figure 1). The OPPF crystallization web pages and the underlying PlateDB database provide informatics tools to present these data to the user in a practical way such that (s)he is able to keep track of, and extract key information from, a large number of regularly imaged crystallization trials. They also allow all aspects of the crystallization process to be analyzed far more readily than in traditional investigator-imaged, notebook-based systems. One key issue is the extent to which such automated systems are taken up by users and the nature of scientific added value that accrues (and drives user uptake). In our case, user acceptance of the crystallization database and associated tools was rapid and complete. This reflects the convenience of the system and also the additional scientific value. Naturally, there are opportunities for further development of the software but already two areas of added value stand out:

(1) Time course information. Many crystals are only quasi-stable and decay or disappear over time. The series of automatically captured images allows the systematic analysis of the course of any crystallization condition (Figure 3). Tools developed to display time courses allow the more rational design of repeat experiments. In addition, time course images can help object classification in several ways, for instance: (a) objects which do not change over time are unlikely to be crystals; (b) small air bubbles initially introduced into a crystallization drop during the setup of the trial collapse over time, often leaving a small area of disrupted skin which, at the later time point could easily be mistaken for crystals; and (c) noncrystalline objects often move around a drop quite dramatically during the course of an experiment (Figure 4).

(2) Rapid correlation of crystallization results with reagents is possible. For each crystallization screen condition, the components of the solution have been consistently and systematically defined in the database. Thus, for any protein it is trivial to group the trials by any component or combination of components (e.g., precipitant or buffer). This provides a straightforward method for database mining, an important advantage being the ability to group and compare interactively subsets of images. The user can thus quickly and intuitively explore which reagents and parameters are significant for crystal growth.

The increase in throughput of the crystallization process has also led to a concomitant increase in the number of crystals grown with more that 7500 wells (manually) annotated as crystal containing. We have developed protocols for cryoprotecting and harvesting crystals from small drops (to be described elsewhere), but there are also issues in managing data collection. Data are collected as soon as possible after crystal growth and often by experimenters not directly involved in each project. Thus, we are making heavy use of block allocations and rapid-access facilities at synchrotrons, particularly BM14 at the ESRF (the UK-funded MAD beamline; http://www.bm14.ac.uk/), and developing distributed information management systems to allow informed data collection in a “service” mode with secure bidirectional exchange of data. These software developments are the result of software collaboration between the OPPF, BM14, and the other partners of the eHTPX project (http://www.e-htpx.ac.uk/) and will also benefit less highly automated facilities.

The OPPF provides a working example of a high-throughput crystallization facility. Data management and presentation to individual users is the key to the effective functioning of this facility and realizing the potential advances resulting from the miniaturization and automation of the crystallization workflow. The database and software described here handles the regular production of up to 100,000 images per day. While some aspects of the software are specific to the hardware and methodological choices made at the OPPF, most of the features in the design and implementation

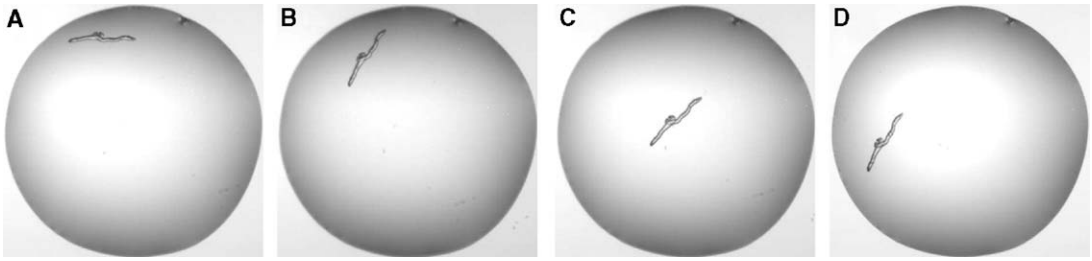

Figure 4. Images Taken at Different Times Showing the Movement of a Noncrystalline Object around a Drop from a Single Crystallization Experiment after (A) 0 Hours; (B) 1 Day, 11 Hours; (C) 3 Days, 23 Hours; and (D) 12 Days, 7 Hours after Set Up

The full set of images from this experiment has been deposited as Supplemental Material both as 12 separate images and as a time-lapse movie in .AVI format.

address requirements common to many crystallography laboratories that are currently setting up (semi-) automated crystal imaging systems. The considerable benefits (both in terms of science and convenience) of such a system means that, at a suitable scale and cost it would be justified for all but small occasional crystallographic laboratories and we expect further benefits to become evident as the software improves and data mining methods are implemented. The software and documentation are freely available to academic sites.

Experimental Procedures

Crystallization Plate Setup and Barcoding

The OPPF crystallization procedure has been described elsewhere (Walter et al., 2003, Brown et al., 2003, Walter et al., submitted), but the key stages are summarized here. Initial crystallization screening uses a panel of 480, or more, conditions selected from standard (commercially available) crystallization kits. The kits are reformatted into 96-deep well "master blocks" (2 mL Masterblock-PP; Greiner Bio-One Ltd., United Kingdom) by a Qiagen Biorobot 8000. Prebarcoded 96-well crystallization plates (reference number 609.101; Greiner Bio-One Ltd., United Kingdom) are used for the trials, the precipitant being transferred from the master blocks to the reservoirs using a Hydra-96 microdispenser (Matrix Technologies Ltd., United Kingdom). The barcode uses Code128C symbology encoding twelve digits that describe its provenance (the OPPF), the type of plasticware, a serial number, and a (human-readable) TAT checksum. The barcode is read using a Gryphon D100 handheld barcode reader (DataLogic UK Ltd.) as the reservoir wells are filled, triggering the creation of a crystallization-plate record describing the reservoir conditions in the OPPF protein production LIMS (to be described elsewhere), which is built around the Nautilus LIMS (ThermoInformatics Ltd., United Kingdom). The plate is then placed on a Cartesian Technologies Microsys MIC400 (Genomic Solutions Ltd., United Kingdom) where a 100 nL drop of protein solution is placed on the central position of each crystallization shelf and mixed with 100 nL of the corresponding reservoir. During this process, the barcode is verified using a PSC LM520 barcode reader (PSC Bar Code Ltd., United Kingdom) mounted on the MIC400 as an in-house custom modification, and a description of the protein sample (including any cofactors and additives) with a link back to the sample's production history is added to the LIMS description of the plate.

Movement of Crystallization Plates into and out of the Automated Storage Vault

The majority of crystallization trials are stored at 21°C (Walter et al., submitted) in an automated storage vault, the OPPF HomeBase storage vault (The Automation Partnership Ltd., United Kingdom). After set up of crystallization plates, they are entered on a tray into the vault through an access port controlled by the HomeBase control software that manages all plate movements into, out of, and within the vault. Commands to start and stop the vault robotics are given directly through a local graphical interface. Requests for individual movements (e.g., to image or retrieve a plate) are written into a set of tables in an external Oracle instance (HOSDB). The control software logs all plate movements to these tables and also scans them periodically to receive requests. External application software can only interact with the HomeBase through this mechanism, meaning that the vault cannot be forced to respond in real time. The interface database has been extended with both a GUI and automated scripts to allow plates to be selected for imaging or retrieval. Up to twelve plates at a time can be introduced into the vault: each plate in turn is removed from the tray, its barcode is scanned, and the plate is stored in a free position. The HomeBase creates a record for the plate in its internal database and in the interface database. When the server script reads a record for a plate that has never been picked for imaging or retrieval it assumes that it is a new plate, verifies that the plate is correctly recorded in the OPPF LIMS (e.g., its owner is known), creates a record in the main crystallization database, PlateDB, and sends an imaging schedule back to the HomeBase via HOSDB (Figure 1).

Automated Crystallization Plate Imaging

Crystallization plates stored in the OPPF HomeBase vault are scheduled for imaging as soon as possible after setup and regularly thereafter (in hours after set up: 0, 5, 15, 35, 55, 75, 95, 115, 135, 175, 215, 255, 295, 375, 455, 535, 615, 695, 855, 1015, 1175, 1495, 1655, 1815, 2455, 3095, 3735, 5015, 6295, and 7575). Imaging occurs as soon after the scheduled times as possible with priority given to the most overdue session. Imaging is performed by an Oasis 1700 automatic imaging system (Veeco, United Kingdom), which is housed in an annex to the storage vault with a shared air handling system. In addition to the main access port, the vault has an alternate access port that can be used to transfer plates into the imaging annex. To effect imaging, the vault transfers the crystallization plate on the alternate output tray, where a pick-and-place robot moves it onto the input stage of the imaging system. The plate is then imaged and placed on the imaging system output stage where a second pick-and-place robot returns it on the vault's alternate access port. A 96-well plate can be imaged in about 40 s with an overhead of about 20 s for plate transfers. The plate barcode along with necessary handshaking and error messages are passed between the control computers of the vault and imaging system using a DCOM interface. Well images are 1024 × 1024 pixels with 8-bit grayscale, and these are cropped to 750 × 700 pixels before being transferred (using a name incorporating the plate barcode, well position, date, and time) to the input directory of a 1.6 terabyte JAD ADV RAID storage system (JAD Logic, United Kingdom) as a MS Windows bitmap (.BMP) file.

A subset of OPPF crystallization trials are carried out at 4°C (Walter et al., submitted), the plates being stored in home-built racking within a standard laboratory cold room. Imaging is carried out by an Oasis 1750 imaging system (Veeco, United Kingdom; a development stage of the LS-3) manually loaded with a cassette con-

taining up to 32 plates with each plate barcode being scanned as it is imaged. The Oasis 1750 images are of the same type and size, are named using the same convention and are transferred to the same network file store as those from the Oasis 1700.

Manual imaging of specific drops by the experimenter can be used to supplement the database (e.g., to characterize further potential data-collection quality crystals). Plates are removed from the vault, imaged on a Nikon SMZ1500 microscope (Nikon UK Ltd.) fitted with a Pixera 120es digital camera (Digital Imaging Systems Ltd., United Kingdom) and then returned to the vault. A graphical interface developed at the OPPF is integrated directly with the Pixera camera driver, presents the user with the latest automated image of a well, associates it with the barcode obtained using a Gryphon D100 handheld barcode reader (DataLogic UK Ltd.) and the well position, and transparently stores the new image to the network file store, where it is integrated into the PlateDB database.

Image Processing, Storage, and Display

The system runs on six dedicated GNU/Linux (Debian) servers with, in addition, the web interface being hosted on the main OPPF web server. One server hosts the network file store and runs control scripts. The arrival of images in the input directory corresponding to an imaging session is detected by Perl scripts which scan the network file store. Each well image is classified using the York University crystal image analysis software (Wilson, 2002, 2004) on an openMosix cluster comprising three dual-processor 1.26 GHz Pentium III PowerEdge 1650 servers (Dell Computer Corporation, United Kingdom) and each analysis takes about 2–5 CPU seconds depending on image complexity. The fifth server runs scripts to convert sets of 96-well images for an imaging session into a montage image laid out in plate format, with each well image given a color-coded frame based on the result of the analysis program (red through blue in order of increasing score, or orange if no drop is detected), and converts well images into JPEG format for the web site. Links to the images are stored in the PlateDB database (PostgreSQL) on the sixth server. This server bears the real load of managing the database and is currently a dual 2.8 GHz Xeon PowerEdge 2650 server (Dell Computer Corporation, United Kingdom) with two pairs of mirrored 36 GB U320 SCSI disks (RAID1) and 4 GB memory. The original bitmap images are eventually migrated offline using Tivoli Storage Manager (IBM UK Ltd.) controlling an Scalar 1000 tape library (Adic UK Ltd., United Kingdom) with a pair of LTO tape drives. Access to the crystallization database is through a dynamically created web interface written in PHP.

Implementation at Other Sites

Hardware requirements: (1) a web server (not necessarily dedicated to this application); (2) a dedicated database server with sufficient memory (see specification above); (3) access to a large file system for image storage; and (4) a Linux machine or small cluster/farm to perform automated crystal detection (see above). Software requirements: all our software is freely available, in addition standard freely available operating system, database management, web server, and clustering (for image analysis) software are required. We use Linux, PostgreSQL, Apache, PostNuke, and OpenMosix. Integration of alternative imaging hardware would only require software development if it is to use our automated image scheduling protocols.

Supplemental Data

Supplemental Data include two figures and two movies and can be found with this article online at http://www.structure.org/cgi/content/13/2/175/DC1/.

Acknowledgments

The OPPF is supported by the UK Medical Research Council (MRC) and SPINE (QLRI-CT-2002-00988). C.J.M. is supported by the eHTPX project funded by the UK Biotechnology and Biological Sciences Research Council. I.M.B. is supported by the Wellcome Trust, J.W. is a Royal Society University Research Fellow, E.Y.J. is a Cancer Research UK Principal Fellow, and D.I.S. is an MRC Research Professor.

Received: October 29, 2004
Revised: December 16, 2004
Accepted: December 17, 2004
Published: February 8, 2005

References

Adams, M.W., Dailey, H.A., DeLucas, L.J., Luo, M., Prestegard, J.H., Rose, J.P., and Wang, B.C. (2003). The southeast collaboratory for structural genomics: a high-throughput gene to structure factory. Acc. Chem. Res. *36*, 191–198.

Brown, J., Walter, T.S., Carter, L., Abrescia, N.G.A., Aricescu, A.R., Batuwangala, T.D., Bird, L.E., Brown, N., Chamberlain, P.P., Davis, S.J., et al. (2003). A procedure for setting up high-throughput nanolitre crystallization experiments. II: crystallization results. J. Appl. Crystallogr. *36*, 315–318.

Chayen, N.E. (2003). Protein crystallization for genomics: throughput versus output. J. Struct. Funct. Genomics *4*, 115–120.

DeLucas, L.J., Bray, T.L., Nagy, L., McCombs, D., Chernov, N., Hamrick, D., Cosenza, L., Belgovskiy, A., Stoops, B., and Chait, A. (2003). Efficient protein crystallization. J. Struct. Biol. *142*, 188–206.

DiDonato, M., Deacon, A.M., Klock, H.E., McMullan, D., and Lesley, S.A. (2004). A scaleable and integrated crystallization pipeline applied to mining the Thermotoga maritima proteome. J. Struct. Funct. Genomics *5*, 133–146.

Goh, C.S., Lan, N., Douglas, S.M., Wu, B., Echols, N., Smith, A., Milburn, D., Montelione, G.T., Zhao, H., and Gerstein, M. (2004). Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. J. Mol. Biol. *336*, 115–130.

Hui, R., and Edwards, A. (2003). High-throughput crystallization. J. Struct. Biol. *142*, 154–161.

Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H.E., McMullan, D., Shin, T., et al. (2002). Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. USA *99*, 11664–11669.

O'Toole, N., Grabowski, M., Otwinowski, Z., Minor, W., and Cygler, M. (2004). The structural genomics pipeline: insights from global target lists. Proteins *56*, 201–210.

Page, R., Grzechnik, S.K., Canaves, J.M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R.C., and Lesley, S.A. (2003). Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. Acta Crystallogr. *D59*, 1028–1037.

Page, R., and Stevens, R.C. (2004). Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens. Methods *34*, 373–389.

Rupp, B. (2003a). High-throughput crystallography at an affordable cost: the TB structural genomics consortium crystallization facility. Acc. Chem. Res. *36*, 173–181.

Rupp, B. (2003b). Maximum-likelihood crystallization. J. Struct. Biol. *142*, 162–169.

Rupp, B., and Wang, J. (2004). Predictive models for protein crystallization. Methods *34*, 390–407.

Walter, T.S., Diprose, J., Brown, J., Pickford, M., Owens, R.J., Stuart, D.I., and Harlos, K. (2003). A procedure for setting up high-throughput nanolitre crystallization experiments. I: Protocol design and validation. J. Appl. Crystallogr. *36*, 308–314.

Wilson, J. (2002). Towards the automated evaluation of crystallization trials. Acta Crystallogr. *D59*, 1619–1627.

Wilson, J. (2004). Automated evaluation of crystallization experiments. Cryst. Rev. *10*, 73–84.