

---

## Towards a structural biology work bench

---

### Synopsis

The trends to investigate larger complexes and more transient phenomena pose some challenges to software developers.

**Keywords:** software; integrative structural biology; structural biology.

### Queries and comments

Please supply or correct as appropriate all **bold underlined** text. In describing corrections please refer to line numbers where appropriate: these are shown in grey.

### Author index

Authors' names will normally be arranged alphabetically under their family name and this is commonly their last name. Prefixes (*van, de etc.*) will only be taken into account in the alphabetization if they begin with a capital letter. Authors wishing their names to be alphabetized differently should indicate this below. **Author names may appear more than once in this list; it is not necessary to mark this correction on your proofs.**

Morris, C.

## Chris Morris

STFC Daresbury Laboratory, Warrington,  
EnglandCorrespondence e-mail: [chris.morris@stfc.ac.uk](mailto:chris.morris@stfc.ac.uk)

## Towards a structural biology work bench

This is an introduction to four papers based on presentations given at a workshop entitled Integrated Software for Integrative Structural Biology. The use of hybrid techniques, and other trends in structural research, pose new challenges to software developers. A structural biology work bench that meets these needs would provide seamless data transfer between processing steps, and accumulate archival data and metadata without intruding into the scientist's work process.

Four papers in this issue of *Acta Crystallographica Section D* are based on presentations given at a workshop entitled Integrated Software for Integrative Structural Biology organised by the Computational Centre for Integrative Structural Biology (CCISB) on 21–23 May 2012.

This workshop discussed some new challenges which structural biologists are accepting. They are addressing larger macromolecular machines rather than single gene products (Perrakis *et al.* **Please give reference in full.**); investigating movement of molecules rather than snapshots; and processing data that is heterogeneous, noisy and incomplete. Karaca & Bonvin (2013) say 'understanding how a single cell functions is the fundamental quest of life sciences'. This can only be comprehensively addressed once the structure–function relationships of biomolecular complexes occurring in that particular cell will have been explored thoroughly. There are two main classical experimental techniques that can reveal the structure of the biomolecular complexes in atomistic detail: X-ray crystallography and NMR spectroscopy. Although these have helped immensely to shed light on the mechanical and functional world of biomolecules, they are faced with many challenges when the biomolecular systems under study become very large, comprise flexible or unstructured regions, exist in very tiny amounts, are membrane associated, or when their constituents interact only transiently.

Most structural biologists are expert in one or two techniques. Many now spend part of the year working as a novice, using a technique which they have not yet mastered to provide supplementary evidence about their target. Indeed, the Instruct visit mechanism is designed to encourage such work. The paper by Karaca & Bonvin provides an example of software support for such hybrid work, as they 'have integrated low-resolution shape data obtained from either ion mobility mass spectrometry (IM-MS) or SAXS experiments, into the conventional scoring function of our information-driven docking program *HADDOCK*'. The software described by Marabini *et al.* (2013) offers, among other things, 'the visualization of maps obtained from 3D-EM, together with annotations provided by other techniques'.

Received 21 December 2012  
Accepted 7 January 2013

## introduction

Biasini *et al.* (2013) explain that ‘research projects in structural biology increasingly rely on combinations of heterogeneous sources of information, *e.g.* evolutionary information from multiple sequence alignments, experimental evidence in the form of density maps, or proximity constraints from proteomics experiments ... new methods in computational structural biology often rely on custom-made *ad hoc* combinations of command-line tools built to perform specific tasks’. They present current developments in the computational structural biology framework OpenStructure, which supports the integration of information from a variety of origins.

Structural biologists want to deliver their results to other life scientists, including systems biologists and medicinal chemists. Gutumanas *et al.* (2013) explain ‘it seems unavoidable that the role of structural biology archives will change from being a pure repository of historic data into becoming an indispensable resource for the wider biomedical community. As part of this transformation, it will be necessary to validate the biomacromolecular structure data and ensure the highest possible quality for the archive holdings; to combine structural data from different spatial scales into a unified resource; and to integrate structural data with functional, genetic and taxonomy data as well as other information available in bioinformatics resources’.

Arising from these scientific challenges, there are several challenges to the software developers whose work supports structural biology. One is to improve the collection of data and metadata for archiving. Another is to improve the automatic processing pipelines used in protein crystallography, and to create pipelines in disciplines that do not yet have them, to support the work of ‘visiting scientists’ from other subdisciplines. Complementary to this, in cases where automated processing is insufficient, it is necessary to better record the information that scientists provide, to make the processing truly reproducible.

Even more challenging is the development of algorithms that can combine data from different techniques, including improved integration of experimental methods with modeling; and to provide some indication of the reliability of the results, represented by an effective visualization.

A structural biology work bench that meets these needs would provide seamless data transfer between processing steps, and accumulate archival data and metadata without intruding into the scientist’s work process.

For a crystallographic group, the effort required to install the *CCP4* suite is repaid many times. If one of them needs to process a few SAXS results, the effort of installation will be a more significant overhead. The work bench must support a very broad range of techniques, and so it should be available without any installation at all: all its functionality should be available through a website.

The internal organisation of the work bench must be open to the addition of new algorithms, without imposing particular architectural choices on the algorithm developers. The architectural design should also take good note of the point that Marabini *et al.* (2013) make that ‘developers can never quite know what requirements will be needed to accommodate the different algorithms of each software package, so high flexibility is mandatory. As the data become more complex and heterogeneous, an ontological database is probably the best option’.

The scientific community includes a broad group of life scientists who use structural results without much attention to the underlying physics, and a small number of pioneers who develop new structural algorithms. There is also a significant intermediate group: there are many structural biologists who are highly competent in computing and who write scripts to plumb together the executables they use in novel ways. The work bench must serve all three groups.

Facebook supports its users in developing and sharing new applications (or apps). ‘Google widgets’ provide similar functionality. The work bench should similarly make it easy to develop and share new functional web pages, to provide novel structural and bioinformatic functionality. Important steps have already been taken in this direction, for example by weNMR, Scipion, and the archiving performed at Diamond Light Source. Any future collaboration must build on existing achievements, not duplicate them.

Developing infrastructure software that supports the daily work of a broad community needs a different approach to developing new algorithms. Experience shows that success depends on an effort to understand the context of use and then more detailed requirements, to take into account the user experience, and then design a software architecture (Morris & Segal, 2012). This will take a significant effort, which will provide structural biologists the software tools they need to continue their extraordinary progress in rate of discovery.

Thanks to CECAM, Instruct, CCP4, and weNMR for funding the workshop.

## References

- Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., Johner, N., Andreas Schenk, A., Philippsen A. & Schwede T. (2013). *Acta Cryst. D* **69**, XXX–XXX.
- Gutmanas, A., Oldfield, T. J., Patwardhan, A., Sen, S., Velankar, S. & Kleywegt, G. J. (2013). *Acta Cryst. D* **69**, XXX–XXX.
- Karaca, E. & Bonvin, A. M. J. J. (2013). *Acta Cryst. D* **69**, XXX–XXX.
- Marabini, R., Macias, J. R., Vargas, J., Quintana, A., Sorzano C. O. S. & Marabini, J. M. C. (2013). *Acta Cryst. D* **69**, XXX–XXX.
- Morris, C. & Segal, J. (2012). *IEEE Software*, **29**, 9–12.
- Perrakis **Please give reference in full.**

## INTERNATIONAL UNION OF CRYSTALLOGRAPHY

### Proof instructions

These proofs should be returned **within 7 days of 28 January 2013**. After this period, the Editors reserve the right to publish articles with only the Managing Editor's corrections.

Please

- (1) Read these proofs and assess if any corrections are necessary.
- (2) Check that any technical editing queries have been answered.
- (3) Return any corrections **immediately** by e-mail to **lj@iucr.org** giving a full description of the corrections in plain text and indicating the line numbers where appropriate. Please do not make corrections to the pdf file electronically and please do not return the pdf file.

**Substantial alterations should be avoided wherever possible as they may delay publication.** Where alterations are unavoidable every effort should be made to substitute words or phrases equal in length to those deleted.

You will be informed by e-mail when your paper is published and you may then download an electronic offprint of your paper from the author services page of **Crystallography Journals Online** (<http://journals.iucr.org>)

If you wish to make your article **open access**, please fill out the attached order form.

Printed offprints may be purchased using the attached form which should be returned as soon as possible.

YOU WILL AUTOMATICALLY BE SENT DETAILS OF HOW TO DOWNLOAD  
AN ELECTRONIC OFFPRINT OF YOUR PAPER, FREE OF CHARGE.  
PRINTED OFFPRINTS MAY BE PURCHASED USING THIS FORM.

Please scan your order and send to [lj@iucr.org](mailto:lj@iucr.org)

**INTERNATIONAL UNION OF CRYSTALLOGRAPHY**

5 Abbey Square  
Chester CH1 2HU, England.

VAT No. GB 161 9034 76

Article No.: D130276-ME0493

Title of article Towards a structural biology work bench

Name Dr Chris Morris

Address STFC Daresbury Laboratory , Warrington, UK

E-mail address (for electronic offprints) [chris.morris@stfc.ac.uk](mailto:chris.morris@stfc.ac.uk)

**OPEN ACCESS**

IUCr journals offer authors the chance to make their articles open access on **Crystallography Journals Online**. For full details of our open-access policy, see <http://journals.iucr.org/services/openaccess.html>. The charge for making an article open access is **1000 United States dollars**.

I wish to make my article open access.

**OFFPRINTS**

I wish to order . . . . . paid offprints

**These offprints will be sent to the address given above. If the above address or e-mail address is not correct, please indicate an alternative:**

[Empty box for alternative address]

**PAYMENT**

Charge for open access . . . . . USD    Charge for offprints . . . . . USD    Total charge . . . . . USD

A cheque for . . . . . USD payable to **INTERNATIONAL UNION OF CRYSTALLOGRAPHY** is enclosed

I have an open-access voucher to the value of . . . . . USD

Voucher No. [ ]

An official purchase order made out to **INTERNATIONAL UNION OF CRYSTALLOGRAPHY**  is enclosed  will follow

Purchase order No. [ ]

Please invoice me

Date | Signature

## OPEN ACCESS

The charge for making an article open access is **1000 United States dollars**.

A paper may be made open access at any time after the proof stage on receipt of the appropriate payment. This includes all back articles on **Crystallography Journals Online**. For further details, please contact support@iucr.org. Likewise, organizations wishing to sponsor open-access publication of a series of articles or complete journal issues should contact support@iucr.org.

## OFFPRINTS

An electronic offprint is supplied free of charge.

Printed offprints without limit of number may be purchased at the prices given in the table below. The requirements of all joint authors, if any, and of their laboratories should be included in a single order, specifically ordered on the form overleaf. All orders for offprints must be submitted promptly; it will not be possible to supply offprints for orders received after the journal is printed.

Please note that normally offprints are sent about one month after publication of the article.

Prices for offprints are given below in **United States dollars** and include postage.

Number of offprints required	Size of paper (in printed pages)				
	1–2	3–4	5–8	9–16	Additional 8's
50	56	81	113	170	75
100	83	115	159	243	99
150	108	149	208	315	124
200	134	184	256	390	154
Additional 50's	26	34	49	74	25

## PAYMENT AND ORDERING

Cheques should be in **United States dollars** payable to **INTERNATIONAL UNION OF CRYSTALLOGRAPHY**. Official purchase orders should be made out to **INTERNATIONAL UNION OF CRYSTALLOGRAPHY**.

Orders should be returned by email to lj@iucr.org

## ENQUIRIES

Enquiries concerning offprints should be sent to support@iucr.org.