



Interactive visualization of data as a feature of online crystallography journal articles

Brian McMahon
International Union of Crystallography
5 Abbey Square
Chester CH1 2HU
UK
bm@iucr.org



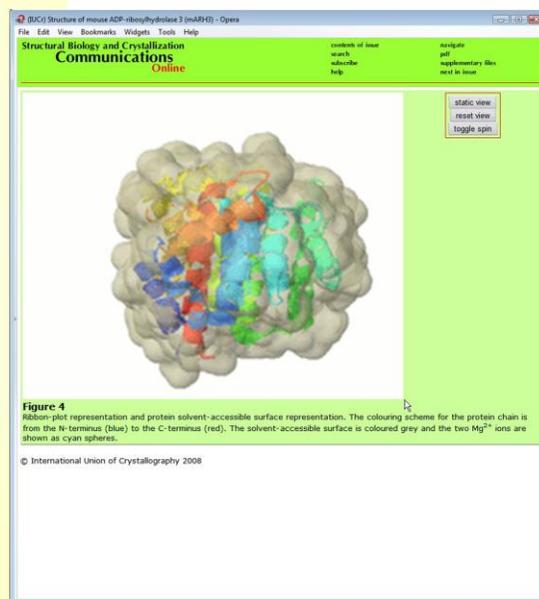
CODATA Conference 2008

Crystallography
Journals
Online

There are many powerful interactive programs allowing chemists, materials scientists and biomolecular scientists to visualize molecular structures. The ability to render three-dimensional structure provides deep insight into atomic and molecular interactions. Research articles can report these insights, but cannot demonstrate them directly. Now the International Union of Crystallography has developed a graphics authoring tool, allowing the easy creation of fully interactive three-dimensional enhanced graphics figures for publication in its journals (*Acta Crystallographica*, *Journal of Applied Crystallography* and *Journal of Synchrotron Radiation*). The tool is completely integrated within the journals' article submission and peer reviewing systems. It requires a data file containing molecular coordinates and crystallographic symmetry operations; in most cases the journal policy already requires the author to deposit such data. It produces a static version of the figure for publication in PDF and hard-copy editions of the journals, but a fully interactive applet using the *Jmol* visualization program (<http://www.jmol.org>) that appears in the online edition. The author may provide menu buttons and options to display a number of carefully selected representations of the molecular or crystal structure; the reader is, however, free to use these as a starting point to explore the structure, change the style of representation, rotate, zoom or orient the molecule in any desired position.

Interactive molecular visualizations to enhance understanding

- Rotate
 - Modify orientation
 - Alternative representations
 - Overlay representations
 - Interrogate
- Long and honourable history:
- *Kinemage*, *Chime*, *RasMol*
 - *First Glance in Jmol*, *PDB*
- but...**
- Platform dependent, require plugins, or supplementary information

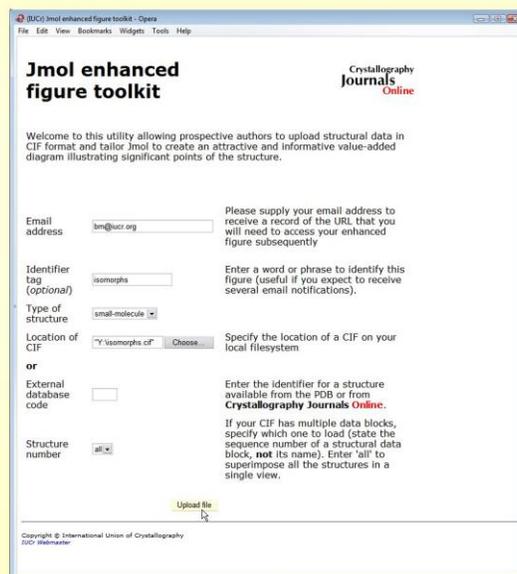


The research literature in structural science has long been attracted by the idea of using software to provide dynamic or interactive visualizations of molecular and crystal structures. Structural biology journals experimented with CD-ROM supplements offering protein visualizations as 'kinetic images', or 'kinemages', as long ago as 1993. Structures of small-molecule chemicals were shown in interactive Netscape plugins using *Chime* in a number of pilot experiments during the 1990s. *RasMol*, a similar program able to visualise small-molecule and macromolecular structures, was used as a browser helper application by a number of journals (including those of the IUCr) to represent three-dimensional structural views of molecules described by supplementary data sets. More recently the structural biology community has provided services based on the Java application *Jmol* to allow investigation of protein structural data sets. Journals such as *Nature* provide links to *First Glance in Jmol* to allow readers to visualize cited structures.

These approaches have all been very valuable and provide additional insight into reported structures. But they all have limitations: they require particular software (often proprietary), are operating-system or browser dependent, allow access only to supplementary data sets, and do not necessarily reflect the features of the data that the author wishes to emphasise.

Requirements for routine publication of enhanced figures

- Platform independence
- Web access for authors
- Serving visualization application and data
- Integration into submission/review procedures
- Integration into journal production workflow
- Automated generation of static copy (for failsafe/PDF edition/archiving)
- Authoring tools

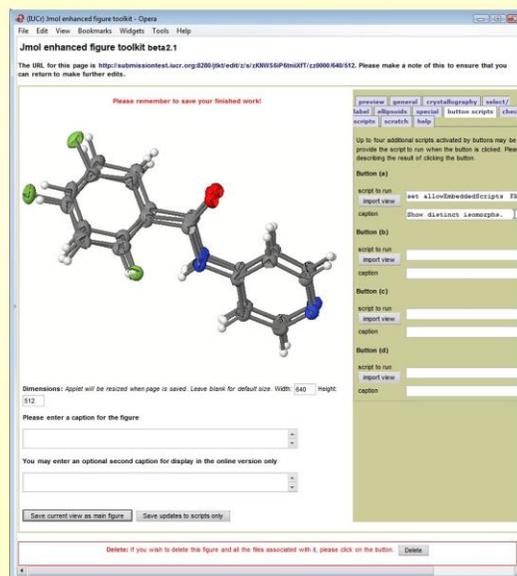


The screenshot shows a web browser window titled "Jmol enhanced figure toolkit" from "Crystallography Journals Online". The page contains a form for uploading structural data. The form fields include: "Email address" (with "jmol@iucr.org" entered), "Identifier tag (optional)" (with "isomorphs" entered), "Type of structure" (a dropdown menu set to "small-molecule"), "Location of CIF" (with "Y:\isomorphs.cif" entered and a "Choose" button), "External database code" (an empty checkbox), and "Structure number" (a dropdown menu set to "all"). There are several instructional text blocks explaining the purpose of each field. At the bottom of the form is an "Upload file" button. The footer of the page reads "Copyright © International Union of Crystallography IUCr Webmaster".

Several requirements are needed to provide an environment for routine publication of enhanced figures involving interactive software. The enhanced figures must be viewable on any common computer platform. The relevant materials must be able to be transferred via the Web. The publisher's server must be able to supply the software and the data it requires transparently, as embedded applets. The enhanced figure must be visible to the referees before publication to allow critical peer review, and must be able to be edited in response to referees' comments. The management of the figures must be an automated part of the standard journal production workflow (*Acta Crystallographica Section E: Structure Reports Online* handles over 300 articles per month, all of which could, in principle, have enhanced figures). There must be a mechanism to create corresponding static images for each enhanced figure. These appear as the illustration of record in the PDF or hard-copy edition. They also guarantee preservation of the scientific content of the article if the software cannot be supported in future. They also appear in browsers lacking JavaScript and Java capability. Finally, authors must learn how to create meaningful figures in this new medium.

The authoring environment

- The author uploads a data file (CIF)
- The system provides different default styles according to the type of structure
- The author edits and annotates the view
- The author may supply additional scripts
- The author saves the result as an enhanced figure + publication-quality static figure

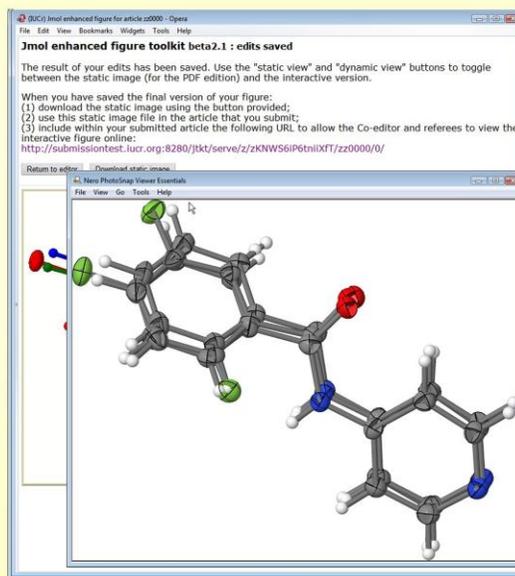


The service provides the author with a toolkit to create and edit the enhanced figure in the *Jmol* applet window. The initial representation depends on the type of structure. Inorganics are shown as space-filling bonded lattices. Small molecules are shown individually with displacement ellipsoids representing the mobility of their constituent atoms. Macromolecules are drawn in cartoon representation of secondary structure (alpha helices, beta sheets *etc.*) The author is free to modify these views, although the defaults are intended to match closely the usual journal requirements for each type of structure. The toolkit provides a tabbed user interface allowing access to different sets of editing functions. Very good results can be achieved without any knowledge of the *Jmol* scripting language. As authors become familiar with the toolkit, they can provide *Jmol* scripts to increase the functionality of the enhanced figures. *Jmol* gurus can import arbitrarily complex scripts to run animations and complex interactive views. The author may also supply additional scripts that become available to the reader through form buttons and checkboxes. The result is saved as an enhanced interactive figure and accompanying static image.

In this example, the author has uploaded a data file containing coordinates of three isomorphous compounds. The initial view overlays all three to show how very similar indeed they are.

Saving the enhanced figure

- Interactive applet
- Active scripts provided by the author
- High-resolution static image
- Option to view dynamic or static image online
- Link to allow peer review



When the author saves, the result is: the scripts and data needed to populate the applet are stored on the journals server in a form that can be used directly in online publication; the author previews the applet with any additional scripts created; the author may download a high-resolution TIFF image of the initial view (for incorporation in the manuscript that will be submitted, and also for subsequent publication in the PDF/hard-copy editions); the enhanced figure page comes complete with facilities to swap in the static image, to reset the view to the author's original representation, or to rotate the view; there is also a link that should be included in the submitted manuscript. This will allow a referee to evaluate the enhanced figure online during peer review.

Submission/review

- Author may prepare enhanced figure ahead of publication
- Simply enter URL of edit workspace when asked to 'upload source files'
- Presented alongside other conventional figures
- Available for peer review
- Can be edited in response to referee comments

The screenshot shows a web browser window titled "upload files for submission gx5130". The page has a green header with "Stage 1 2 3 4 5" and "Article in production". Below the header, the title of the submission is "Crystal structure of a Trypanosoma brucei α/β -hydrolase fold protein with unknown function". The authors listed are Ethan A Herdt, Margaret Holmes, Frederick S Buckner, Wesley C Van Voorhis, Erin Quarterly, Eric M Phizicky, Angela Lauricella, Joseph Luft, George DeTitta, Helen Neely, Frank Zucker and Wim G J Hol.

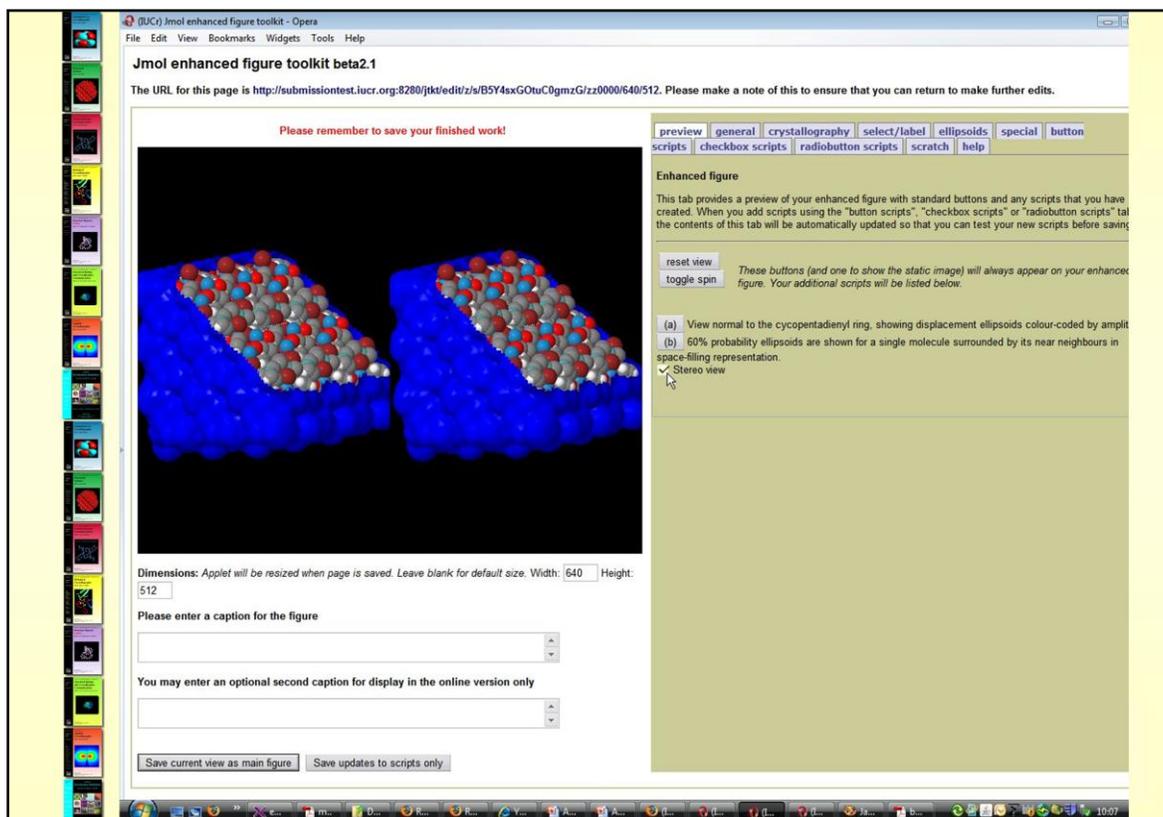
The following files have been deposited for this submission (Last revision seen by referees (For reference only)):

Name	Component	Size	Upload time	Delete
gx5130.Rx	Article body	30259	Mon Apr 21 18:34:23 2008	Delete
gx5130fig1.ps	Figure	230280	Mon Apr 21 18:37:03 2008	Delete
gx5130fig2.ps	Figure	120373	Mon Apr 21 18:37:36 2008	Delete
gx5130fig3.tif	Enhanced figure	308439	Mon Apr 21 18:39:03 2008	Delete Edit View

Below the table is a 3D molecular structure visualization of the protein. A text box next to the structure reads: "One monomer of the dimeric T. brucei protein (Pdb:2O26) is shown in grey. The other monomer is colored from blue at the N-terminus to red at the C-terminus. The right strands making up the core β -sheet of the α/β -hydrolase fold in this monomer are arranged left to right as in Figure 2. Secondary structure elements between $\beta 6$ and $\beta 7$ (dark green in the figure) form a capping domain at the top of the structure. The position of Ser 116 in each monomer is indicated by a yellow sphere."

An important element of the new service is the integration of the enhanced figure into the peer review process. The way this works is normally as follows. While preparing the manuscript, the author will create an enhanced figure. (Linking the graphical representation to the scientific argument is an important part of composing the article.) The IUCr system provides a URL for the author to work on the figure, editing it as the writing of the article progresses. It also provides a 'read-only' URL that will allow reviewers (or others) to see the enhanced figure but make no changes. The author also is given a high-resolution graphic of the initial view. The author includes the static image as one of the figures in the article, and also includes the URL for the read-only view so that the referee can see and interact with the figure online. If the referee requests changes to the enhanced figure, the author can return to make further edits. When the article is accepted, the author must upload source files for publication. At this point, the URL of the enhanced figure is simply entered in the source file upload dialogue, and the enhanced figure is transferred to the journal production workflow. It can be accessed alongside other source materials.

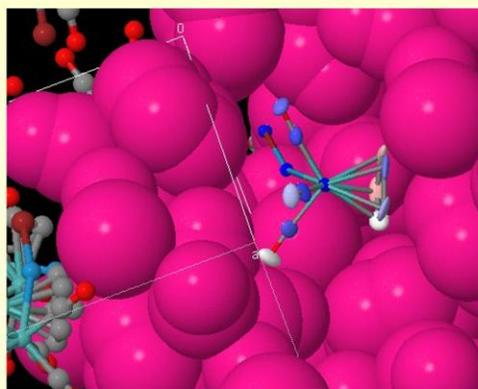
For small-molecule structure reports (*Acta Crystallographica Sections C and E*) there is also an alternative workflow. For these journals, the article submission itself is in CIF format. It is therefore possible to submit a CIF, and subsequently create an enhanced figure within the submission system interface *using the primary data set (CIF) already uploaded*. This is often efficient where the author wishes to do relatively little work in editing the enhanced figure, but still considers it useful to provide a basic interactive three-dimensional visualization.



The editing window is essential in allowing authors to participate fully in the creation and modification of enhanced figures, even if they are not expert users of the visualization software. A tabbed interface allows them to work through the various functions available, focusing on features that are most relevant to the science they wish to emphasise: crystal packing and symmetry; atomic displacement ellipsoids; coordination polyhedra; protein secondary structure etc. Only those tabs that are relevant to the type of structure are displayed. Other tabs allow them to add custom *Jmol* scripts that will appear as buttons, checkboxes or radio buttons on the final enhanced figure. If they know the *Jmol* scripting language they can write the scripts directly; otherwise, they can use the graphical tools to set up a desired view and import it as a complete script with a single button click. The toolkit provides a scratch pad for testing new scripting ideas; and provides live previews of the effects of the script, in order to provide immediate feedback to novice users.

Objectives

- Enhance the scientific value of illustrations
- Increase educational function
- Allow readers to interact with the data
- Preserve the essential aspects of the scientific argument
- Enhance visual impact
- Extend the scholarly authoring paradigm



'...future generations [of scientific authors] will increasingly tell the scientific story using the talents of animators, of movie directors - and even of designers of interactive videogames.'

Crystallography
Journals
Online

As we noted at the outset, online visualization of molecular structures is not new. What *is* new in what we offer is the absorption of this technology into the paradigm of writing scholarly research articles in the literature of science. Crystallographers now have the tools to enhance the scientific and educational value of their illustrations and to tie their written description directly to visualization and manipulation of the actual data they are discussing. Although the aesthetic impact is undeniable, what is more significant is the ability to deepen understanding of the information bound up in the data, to provide more profound, and, ultimately, new insights in science. The technology is still in its infancy, and so we offer this in the conservative context of preserving static 'snapshots' of the illustrations for archival purposes. But in practice, we are anticipating the development of scientific discourse from literature to multimedia.