

Other tools for submitting and using structure reports

Brian McMahon

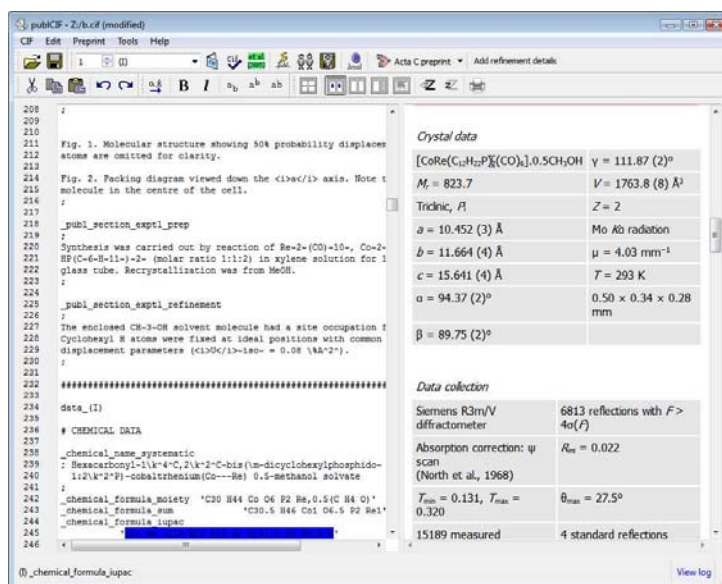
International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, UK

bm@iucr.org

Over the years IUCr journals have created a number of online tools to assist in the submission of structure report articles, though many can also be used to generate nicely formatted tables of data or molecular graphics suitable for use in theses or presentations. In addition, the online published articles, which are always accompanied by CIFs containing full structural and geometric data, and for small-unit-cell structures, structure factors or Rietveld refinement profiles, include a number of valuable visualisation, search and validation tools.

Authoring tools: (1) *publCIF*

Desktop CIF publishing editor, validator and formatter for small-molecule, powder, modulated and incommensurate structure CIFs
<https://journals.iucr.org/services/cif/publCIF/>

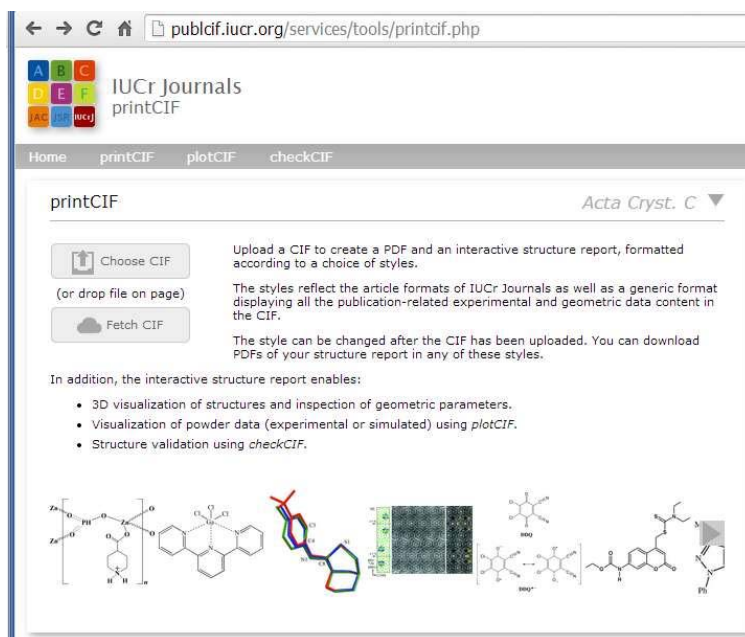
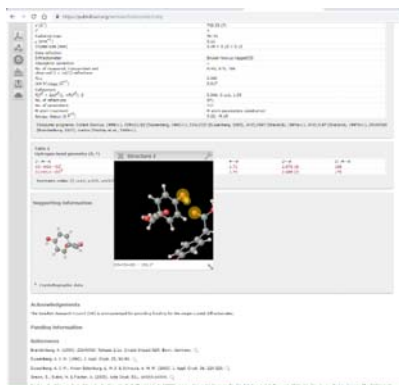


publCIF is maintained by the IUCr specifically to facilitate author submissions of structural reports and to create an efficient publication workflow for the online journals of the IUCr. Its practical use will be demonstrated in the tutorial sessions of this School, but here we mention a few points that are often overlooked. [1] On import of a CIF, the file is checked for syntax and core dictionary compliance, and systematic structural errors (optionally) auto-corrected. [2] The application has an embedded copy of the core and other small-unit-cell CIF dictionaries (allowing for drag-and-drop insertion of missing data names). [3] The current file can be re-validated by the application or be run through online *checkCIF* without exiting the program. In this mode a validation response form can be inserted and edited. [4] For very short structure reports, *publCIF* has tools for writing standard parts of a report based on the crystallographic data and publication metadata included in the CIF. [5] It also includes tools for checking citations and adding standard references. [6] It has tools for inserting images and refinement instruction files. Previous versions have also included an interface for generating interactive figures. While this has been disabled pending re-implementation of the *Jmol* toolkit, it is hoped that it can be restored in future versions.

Authoring tools: (2) *printCIF*

Online CIF publishing validator and formatter for small-molecule, powder, modulated and incommensurate structure CIFs

<https://publCIF.iucr.org/services/tools/printcif.php>

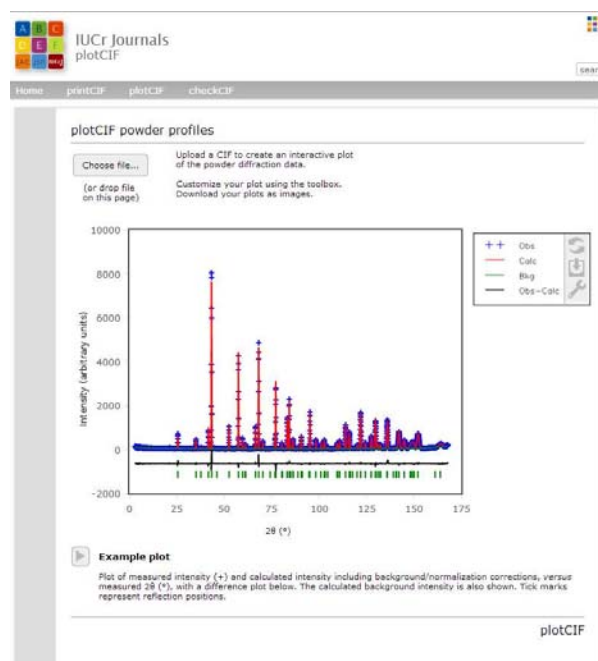
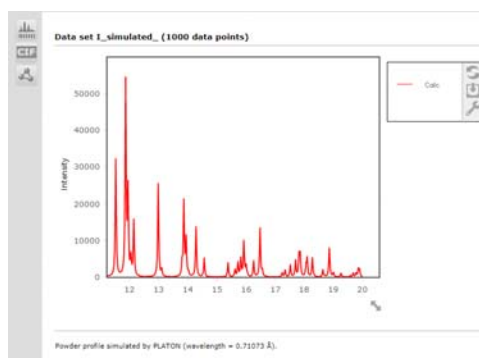
A screenshot of the publCIF website interface. The page title is 'IUCr Journals printCIF'. The navigation bar includes 'Home', 'printCIF', 'plotCIF', and 'checkCIF'. The main content area is titled 'printCIF' and features a 'Choose CIF' button with the instruction '(or drop file on page)'. Below this is a 'Fetch CIF' button. To the right, there is a description of the tool: 'Upload a CIF to create a PDF and an interactive structure report, formatted according to a choice of styles. The styles reflect the article formats of IUCr Journals as well as a generic format displaying all the publication-related experimental and geometric data content in the CIF. The style can be changed after the CIF has been uploaded. You can download PDFs of your structure report in any of these styles.' Below the description, there is a list of features: 'In addition, the interactive structure report enables: 3D visualization of structures and inspection of geometric parameters, Visualization of powder data (experimental or simulated) using plotCIF, Structure validation using checkCIF.' At the bottom, there is a row of chemical structures, including a polymer chain, a complex organic molecule, a crystal structure, and a molecular formula.

This online tool formats a CIF in the style of individual IUCr journals, allowing an author to visualize how an article will look before it is submitted. In this sense it is a valuable proofing tool and complements and extends the proofing capability of *publCIF*. There is also an option to format as a ‘generic’ report, which exposes more of the structural data. This tool also allows easy communication with other publication tools (*checkCIF*, *plotCIF* and the auto-generation of 3D interactive figures). A particularly nice feature is that it allows immediate 3D visualisation of geometric features of interest as selected by the author. Also noteworthy is that it can be loaded with any CIF found on the Internet. For IUCr journals, all CIFs deposited as supporting information have their own DOI – try this with <http://doi.org/10.1107/S2056989019009289/wm5509sup1.cif>. This is a very nice way of visualizing structures in open databases, *e.g.* <http://www.crystallography.net/cod/1518961.cif> from COD

Authoring tools: (3) *plotCIF*

Online visualization of powder data and creation of publication-quality powder profiles

<https://pubcif.iucr.org/services/tools/pdcifplot.php>



plotCIF provides interactive online visualisation of powder diffraction data, including Rietveld refinement difference plots. The user can scale the image, zoom in on portions of the profile (including as an inset within the main diagram) and select different plotting conventions in order to generate publication-quality graphics. It can also be invoked from any single-crystal CIF to construct a simulated powder profile from that structure. The program was developed by the IUCr Chester office in line with the functionality of *pdCIFplot* [Toby, B. H. (2003). CIF applications. XII. Inspecting Rietveld fits from *pdCIF*: *pdCIFplot*. *J. Appl. Cryst.* **36**, 1285-1287]. The latter program is no longer actively supported, but will provide equivalent functionality on a desktop.

Authoring tools: (4) *catCIF*

Simple tool to combine multiple CIFs

<https://publCIF.iucr.org/services/tools/catcif.php>



catCIF is a tool originally created to assist authors to submit multiple CIFs as a single supplementary CIF, because the submission system (which was designed to accommodate a single CIF for article submission only) would not permit the upload of multiple data files. It ensures that there is not duplication of data block names in the composite CIF, autocorrects some common minor problems (*e.g.* case-sensitivity of enumerated values) and gives a parse report. A potentially useful feature is that if any of the CIFs is in CIF2 format, all other CIF1 CIFs will be converted to CIF2 also, so the combined CIF will not be a mixture of CIF1 and CIF2.

Authoring tools: (5) *publBio publisher*

Online service for authoring and editing structural biology or crystallization communications

<https://publbio.iucr.org>

The screenshot shows the publBio PUBLISHER interface. At the top, there's a navigation bar with links: my articles, paper options, save, save as, PDF, share, and submit to Acta F. The main content area is titled 'structural communications' and includes a 'Save' button. The form is divided into several sections: 1. Metadata: Acta Crystallographica Section F, Structural Biology and Crystallization Communications, ISSN 1744-3091. 2. Authors: C.-W. Tseng, T.-P. Ko, R.-T. Guo and J.-R. Liu. 3. Keywords: Tim Barrel, hydrolase, Carbohydrate/Sugar Binding. 4. PDB reference: 3AYR. 5. Synopsis: One or two sentences of the main findings for use in the journal contents listing. 6. Abstract: Up to 150 words stating as specifically and as quantitatively as possible the principal results obtained. Do not use 'we' or 'I', or refer to specific tables or figures. 7. Introduction: Include brief information about the background to the study, the source organism, the structure and function of the macromolecule, and related structures. 8. Table 1: Protein production information. This section includes a table with fields for Source organism (Piromyces rhizinifatus), DNA source, Forward primer, Reverse primer, Cloning vector, Expression vector, Expression host, and Complete amino acid sequence of the construct produced. 9. Materials and methods: Text in this section should supplement or complete information provided in the tables. 10. Footer: Data collection was performed using Blotoc Control Software. The program HKL-2000 (Otwinowski & Minor, 1997) was used for data reduction. The structure was solved with CN3 (Brünger et al., 1998) and refined with REFMAC 5.5.0109 (Murshudov et al., 2011). MolProbity (Chen et al., 2010) was used for Ramachandran analysis.

This is the latest iteration of an online publishing tool for biological macromolecular structures. It does provide a template for writing a short structure report without any input data file, but it is primarily designed to create a skeleton paper by extracting relevant structural information in an input mmCIF or PDB data file. It can import data direct from the Protein Data Bank, which makes it a very convenient tool for writing a structure report article after structure deposition. Recently it has been upgraded to import entries also from the Small-Angle Scattering DataBank (SASDB) and to populate a table of experiment SAS data accordingly. It has the facility to collect together illustrations, supporting information and PDB validation reports as required. The tool is designed as an online authoring tool integrated into the IUCr journals publication system, with an option to submit immediately to *Acta Cryst. F*. It therefore requires author credentials to access the service, because it allows joint authorship with ‘trusted’ colleagues, but provides a personalised workspace with your own publication history.

Authoring tools: (6) *publBio* annotator

Online service for authoring and editing structural biology or crystallization articles using macromolecular CIFs (mmCIF)

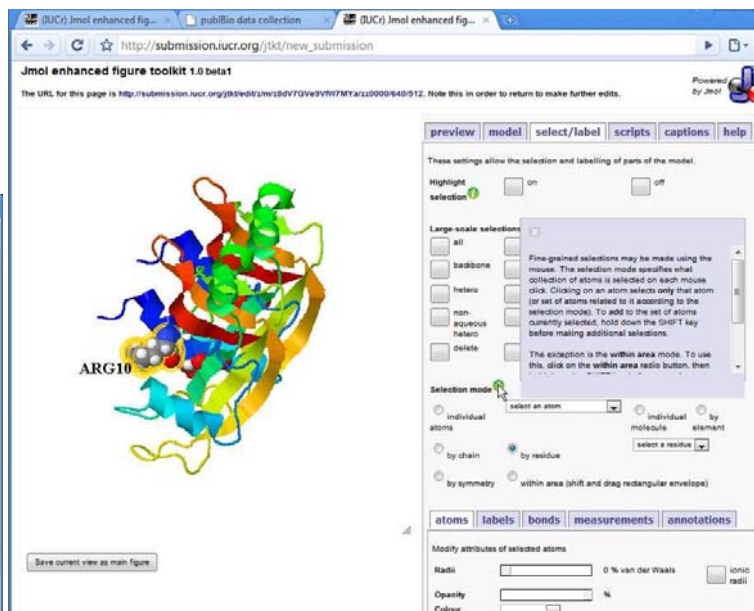
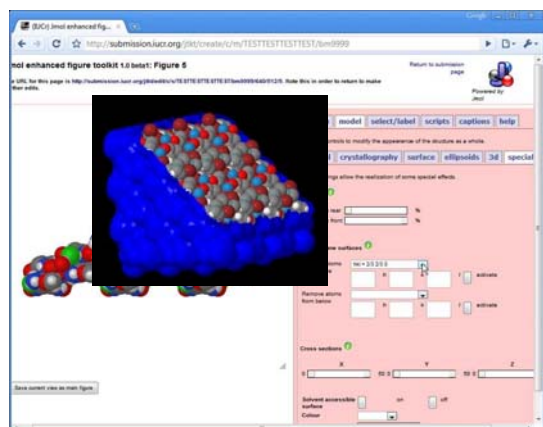
<https://publbio.iucr.org>

The *publBio* annotator is an earlier version of a tool for authoring macromolecular structure articles or crystallization papers. It is currently somewhat hidden from view (select 'about' on the *publBio* home page to find a link to it, or use directly the URL <http://publbio.iucr.org/publbio/publbio.php?version=1>) and development has been frozen. However, it offers greater functionality in retrieving data from a fully populated mmCIF input file, and for populating crystallization detail by lookup of online databases. It involves building a publication by working through a number of distinct pages ('macromolecule', 'crystallization', 'crystal data', 'data collection', 'data-collection statistics', 'phasing' and 'refinement') which correspond to the underlying mmCIF dictionary categories or category groups. Once mastered, it offers an efficient and relatively painless way to populate a full PDB entry, but at present there is still too great an initial learning curve to attract authors used to the conventional way of preparing an article. I hope that the IUCr will continue to offer it, and in future renew its development efforts, as authors come to understand its benefits.

Authoring tools: (7) *Enhanced figure toolkit*

Online service for creating *Jmol* enhanced figures

<https://submission.iucr.org/jtkk>

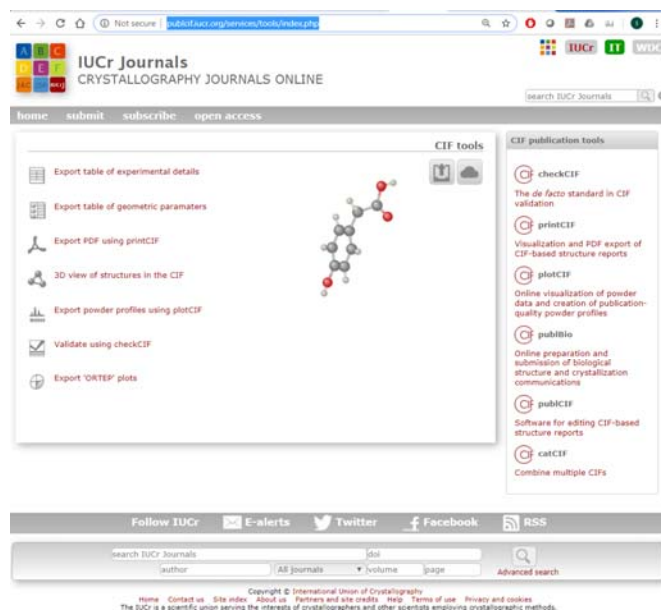
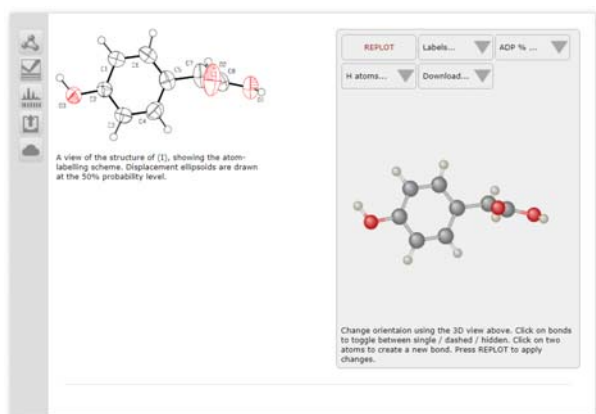


The enhanced figure toolkit is another project that is relatively dormant at the moment, but is ripe for future development. *Jmol* is an incredibly rich visualization program with advanced scripting abilities. I think it is still unique in the breadth of structural data that it can handle and represent – inorganic polyhedra, molecular displacement ellipsoids, chemical structures, proteins rendered as cartoons or space-filling structures, cell packing, intermolecular interactions, incommensurate modulation, magnetic structures. The project was mothballed because *Jmol* is a Java program, and web browser support for inline Java applets has been discontinued in recent years. However, a JavaScript version of the program, *Jsmol*, will run in any modern browser, and I am now on working on reviving the toolkit as an interface to *Jsmol*.

Authoring tools portal

An index of these tools and some others:

<https://publCIF.iucr.org/services/tools/index.php>



IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

home submit subscribe open access

CIF tools

- Export table of experimental details
- Export table of geometric parameters
- Export PDF using printCIF
- 3D view of structures in the CIF
- Export powder profiles using plotCIF
- Validate using checkCIF
- Export 'ORTEP' plots

CIF publication tools

- checkCIF
The de facto standard in CIF validation
- printCIF
Visualization and PDF export of CIF-based structure reports
- plotCIF
Online visualization of powder data and creation of publication-quality powder profiles
- publibio
Online preparation and submission of biological structure and crystallography communications
- publicIF
Software for editing CIF-based structure reports
- catCIF
Combine multiple CIFs

Follow IUCr E-alerts Twitter Facebook RSS

search IUCr Journals
author journal volume page
Advanced search

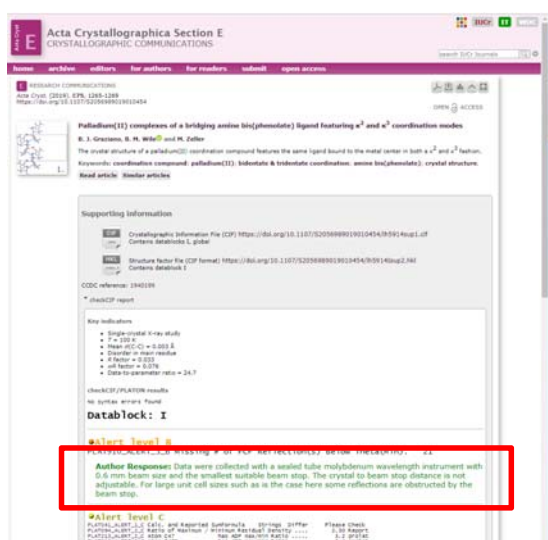
Copyright © International Union of Crystallography
The IUCr is a scientific union serving the interests of crystallographers and other scientists employing crystallographic methods.

It is not widely advertised, but we do provide a portal to access all of the small-unit-cell utilities from a single page, so that you can upload a CIF once and perform the various formatting and visualization procedures in sequence. There are also a few other useful utilities here, like the ability to generate *ORTEP* plots in any orientation simply by manipulating the 3D-molecular image.

Structure validation: *checkCIF*

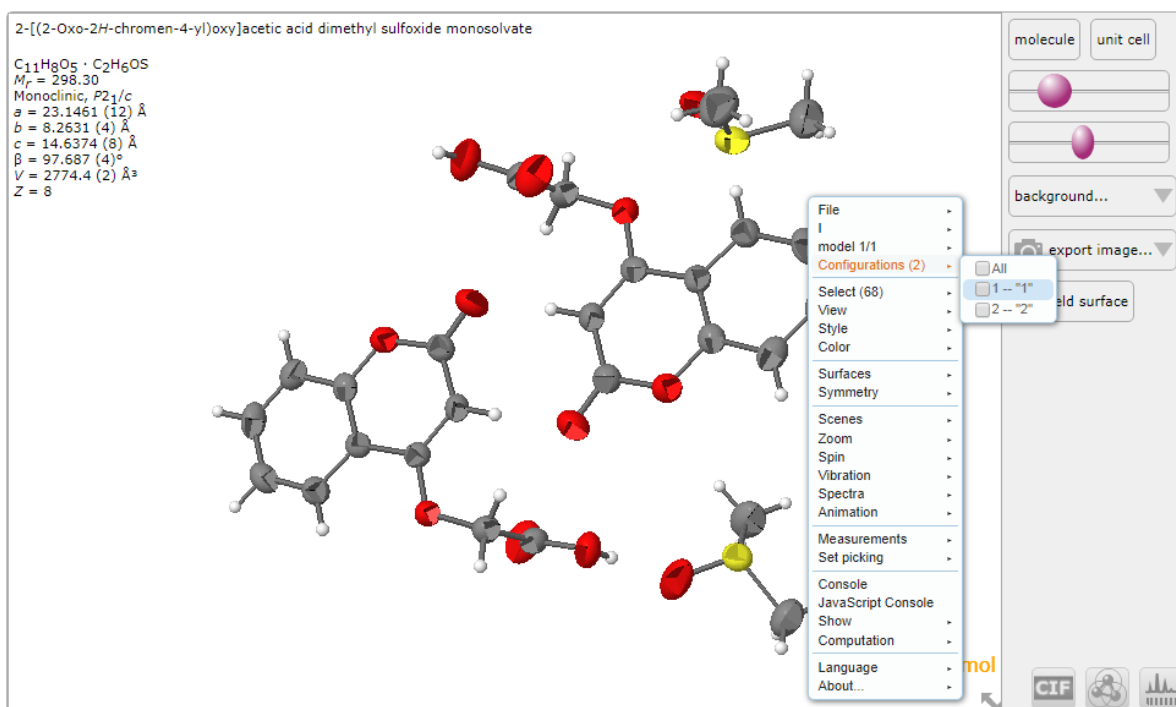
- Check internal consistency of data dependencies (CIF dictionary)
- Check scientific reasonableness of model
- Check completeness of experimental metadata
- Check quality of derived structural model

<https://checkcif.iucr.org>



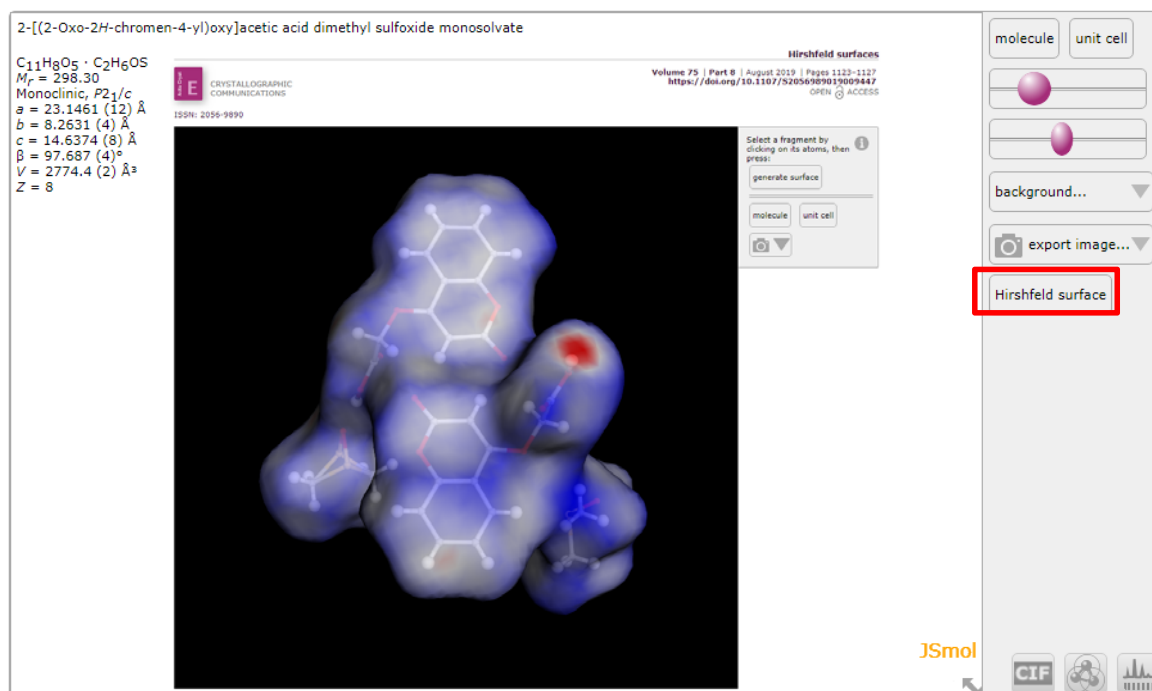
checkCIF is well known to small-unit-cell crystallographers, and is being explored in considerable detail in this School. We mention it here for completeness, reminding users that it is also available within *publCIF*, and that it is available for validating any structural CIF (not only those intended for publication in IUCr journals). *checkCIF* itself has two main components. One checks the syntactical integrity of the CIF and how well specific publication requirements are met (the exact criteria used vary from journal to journal). The other, based on *PLATON*, performs an exhaustive chemical and crystallographic analysis. The *PLATON* checks used by the IUCr checkCIF service are detailed at <http://journals.iucr.org/services/cif/datavalidation.html>. Users can download *PLATON* and run a subset or superset of these specific tests by changing a configuration file. A particularly important aspect of the use of *checkCIF* in the IUCr publishing workflow is the ability to record author responses to flagged alerts, as illustrated here.

Tools for readers (1): *Jsmol* structure explorer



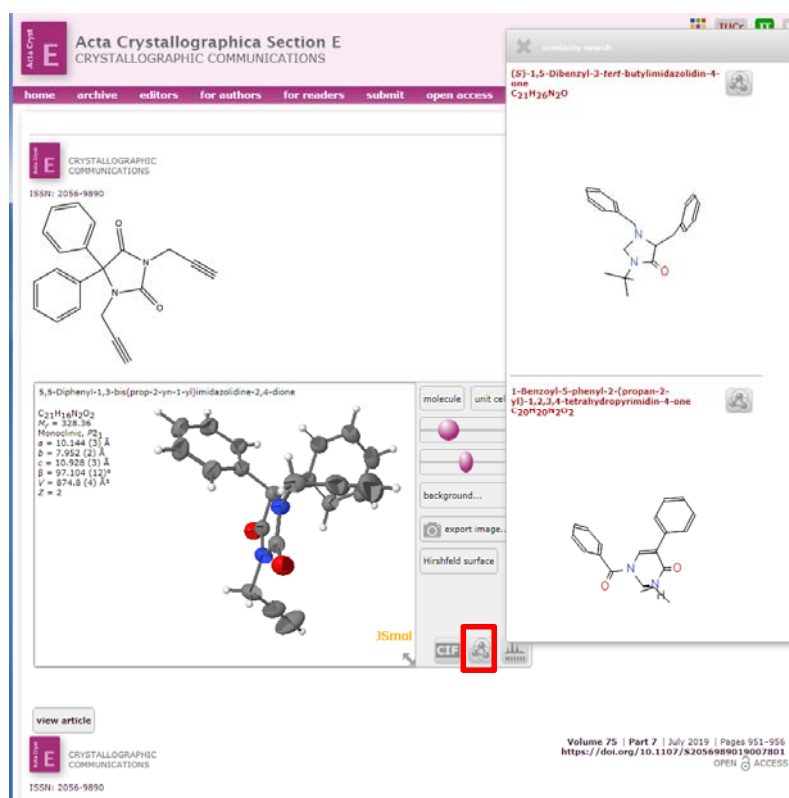
The 3D view button from any IUCr article brings up a web page with a variety of functions. The primary one is a representation of the structure in *Jsmol*, configured as appropriate for the type of structure (polyhedral for inorganics, ellipsoid plots for organics, cartoons for proteins). There are a small number of controls in the right-hand panel that allow some simple changes to the default view (show a single molecule or the unit cell, change the size of displacement ellipsoids or atom spheres, change the colour of the background). There is also a button for creating an image (static or animated GIF). But more importantly, and not always immediately appreciated by users, right-clicking in the main display frame gives access to the full set of *Jmol* features, including crystal packing, geometry measurement, or (as in this example) selection of alternative conformations of disordered molecules. Note also the buttons at the bottom of the right-hand panel for downloading the CIF data file (on the left) and synthesising a powder profile using *plotCIF*. (The function of the other button will be described shortly.)

Tools for readers (2): Hirshfeld surface viewer



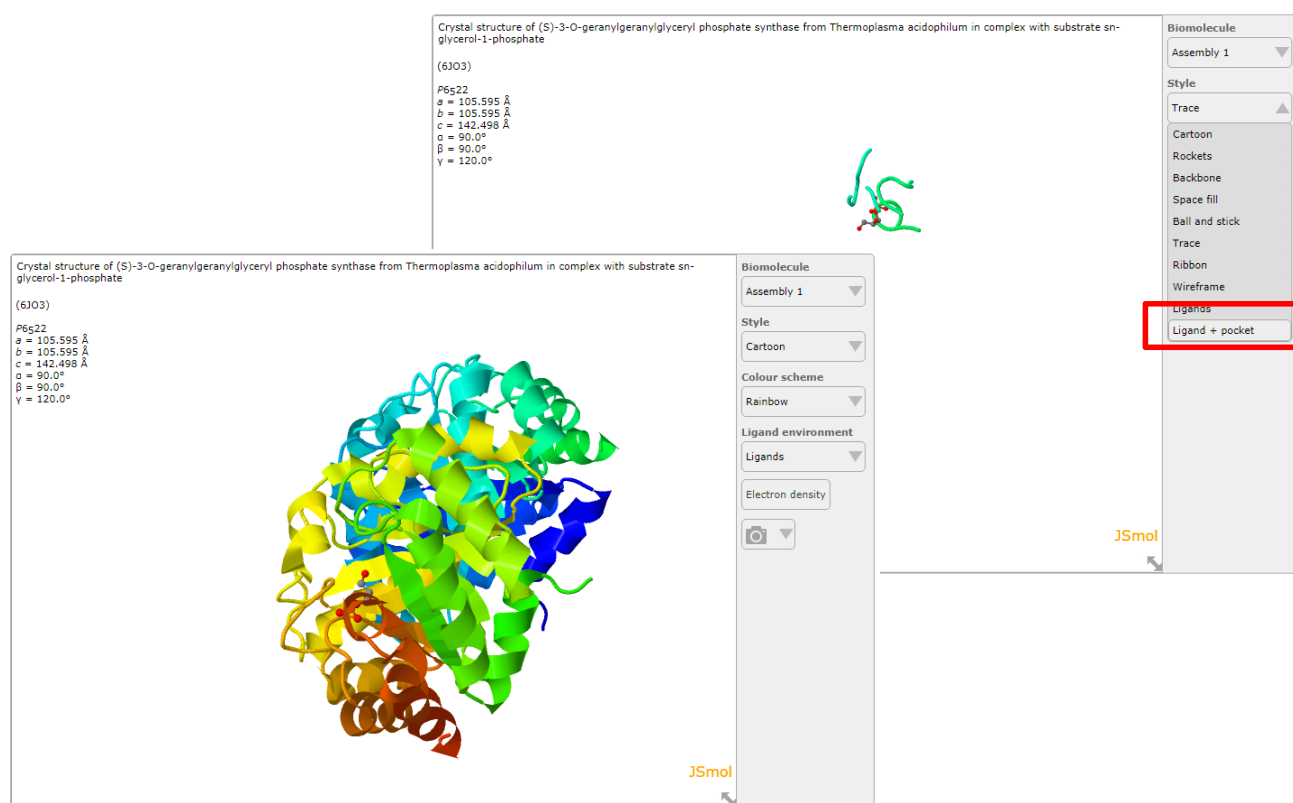
Clicking on the ‘Hirshfeld surface’ button generates another page where the user can collect portions of a molecule or molecular assembly and generate the associated Hirshfeld surface. The Hirshfeld surface data is generated by *Tonto* (as used by *CrystalExplorer*): http://physical-chemistry.scb.uwa.edu.au/tonto/wiki/index.php/Welcome_to_Tonto! The surface data are rendered by WebGL.

Tools for readers (3): structure similarity search



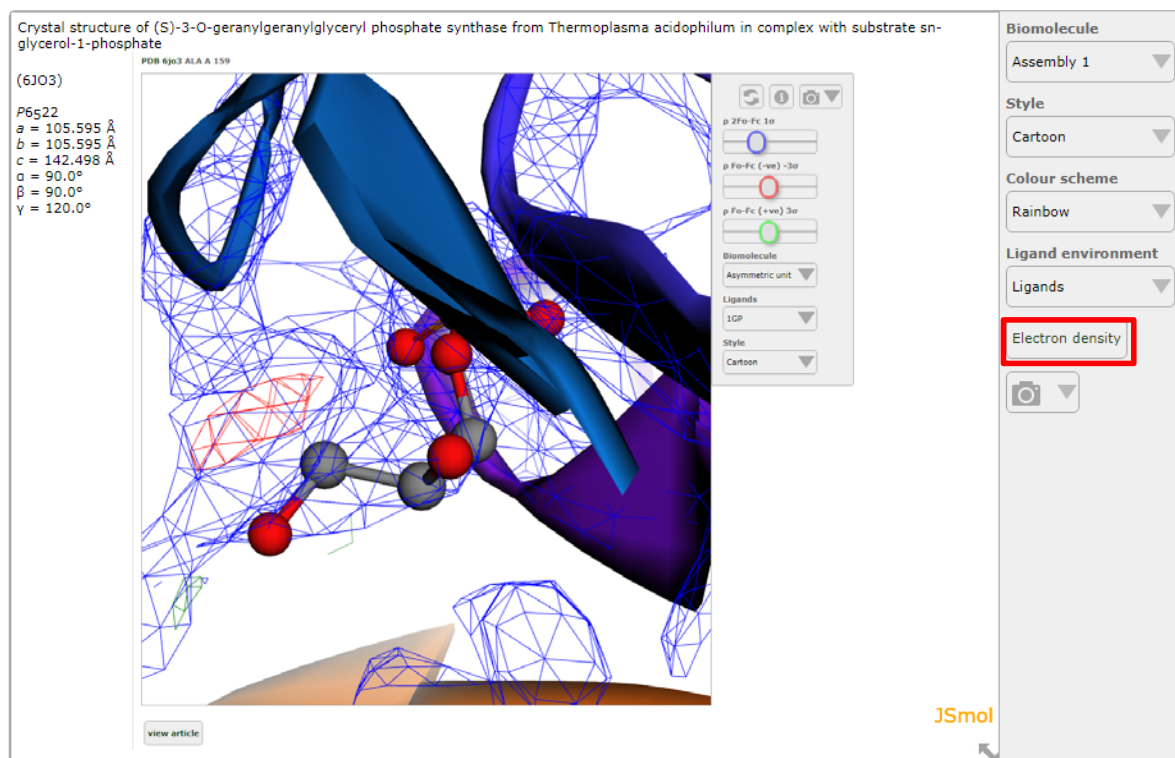
The middle button in the right-hand panel of the 3D view page launches a chemical similarity search for the current molecule. This search is against other structures published in IUCr journals (and so is not a comprehensive chemical database search), although it does also provide a link to the PubChem database. The retrieved structures are rendered in the pop-up window as Jsmol three-dimensional chemical representations, and are themselves manipulable in 3D. Each hit links directly to the publication reporting it and to another similarity search using that molecule as the starting point for a new similarity search.

Tools for readers (4): macromolecular ligand environments



As with the small-molecule structure explorer page, there are a small number of presets to allow different representations of a macromolecular complex to be easily generated; and right-clicking in the explorer window gives full access to all *Jmol* commands. In this example, the immediate structural environment of bound ligands can be displayed.

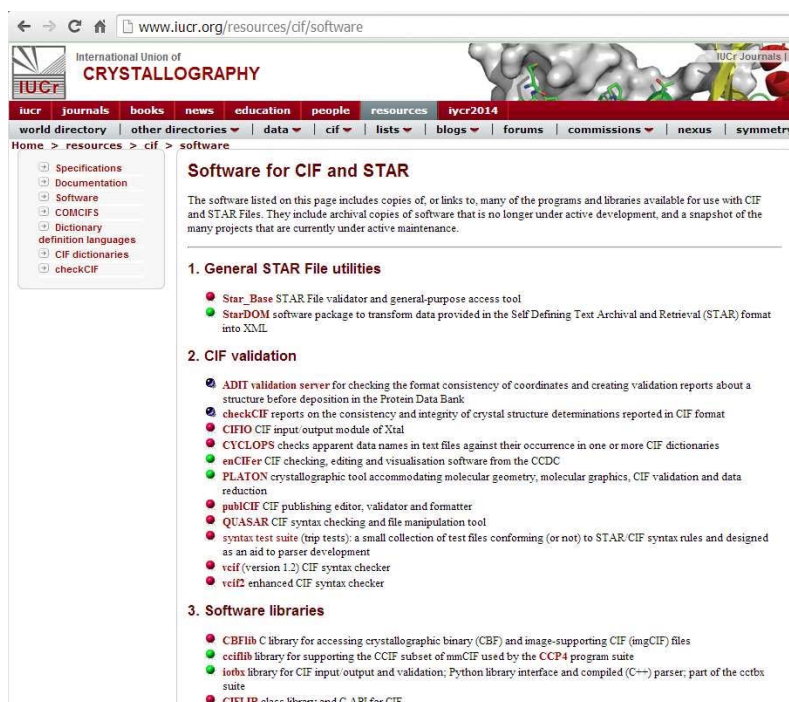
Tools for readers (5): electron density mesh



There is also an option to show the electron-density mesh around selected portions of the structure. This functionality, as with most of the enhancements provided through these tools, is available elsewhere, and the user may have more sophisticated tools to carry out a really detailed analysis. However, the idea behind all the online article tools is to provide access to the data in interactive and meaningful ways as a part of the reading process. To state yet again the main theme of the School: the science is in the data, the data are available in accordance with FAIR principles; and the information presented through the IUCr journals has been tested rigorously by human peer review, computational validation, and openness to inspection.

Tools for working with CIF

<https://www.iucr.org/resources/cif/software>



The screenshot shows the IUCr website's 'Software for CIF and STAR' page. The page has a red header with the IUCr logo and navigation links. A left sidebar contains a tree menu with categories like Specifications, Documentation, Software, COMCIFS, Dictionary, and CIF dictionaries. The main content area is titled 'Software for CIF and STAR' and includes an introductory paragraph. It is organized into three numbered sections: 1. General STAR File utilities, 2. CIF validation, and 3. Software libraries. Each section contains a list of tools and libraries with brief descriptions.

Software for CIF and STAR

The software listed on this page includes copies of, or links to, many of the programs and libraries available for use with CIF and STAR Files. They include archival copies of software that is no longer under active development, and a snapshot of the many projects that are currently under active maintenance.

- 1. General STAR File utilities**
 - Star_Base STAR File validator and general-purpose access tool
 - StarDOM software package to transform data provided in the Self Defining Text Archival and Retrieval (STAR) format into XML
- 2. CIF validation**
 - ADIT validation server for checking the format consistency of coordinates and creating validation reports about a structure before deposition in the Protein Data Bank
 - checkCIF reports on the consistency and integrity of crystal structure determinations reported in CIF format
 - CIFIO CIF input/output module of Xtal
 - CYCLOPS checks apparent data names in text files against their occurrence in one or more CIF dictionaries
 - enCIFer CIF checking, editing and visualisation software from the CCDC
 - PLATON crystallographic tool accommodating molecular geometry, molecular graphics, CIF validation and data reduction
 - publCIF CIF publishing editor, validator and formatter
 - QUASAR CIF syntax checking and file manipulation tool
 - syntax test suite (trip tests): a small collection of test files conforming (or not) to STAR/CIF syntax rules and designed as an aid to parser development
 - vcif (version 1.2) CIF syntax checker
 - vcif2 enhanced CIF syntax checker
- 3. Software libraries**
 - CBFlib C library for accessing crystallographic binary (CBF) and image-supporting CIF (imgCIF) files
 - cciflib library for supporting the CCIF subset of mmCIF used by the CCP4 program suite
 - ioth library for CIF input/output and validation, Python library interface and compiled (C++) parser; part of the cctbx suite
 - CIFLIB class library and C API for CIF

Finally we provide a link to the list of known software packages for CIF and STAR on the IUCr web site. These packages contain some obsolete tools, and alternative editors and viewers. But they also include programming libraries that are available to software developers, and we encourage everyone to use these in the development of new software for CIF applications. Please also let us know of any software you are aware of that does not appear on this list!