

Online tools for enhancing presentation, understanding, and retention of 3D structural data

Alexander Wlodawer

Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD, USA

Transition of scientific journals from paper to online publication opened new opportunities for conveying ideas and sharing data in novel ways. For manuscripts describing macromolecular structures, in particular, a number of tools are currently available for enhancing presentation and improving visualization, as well as for preservation and sharing of the large amounts of original data. For example, Proteopedia (<http://proteopedia.org>) makes it possible to help visualize and understand the relationship between the three-dimensional structures of biological macromolecules and their functions in novel, interactive ways via Interactive 3D Complements (I3DC) (<http://proteopedia.org/w/I3DC>), Molstack (<http://molstack.bioreproducibility.org/>) gives the authors opportunity to interactively present and compare different coordinates and density maps, whereas Integrated Resource for Reproducibility in Macromolecular Crystallography (IRRM) (<http://proteindiffraction.org>) serves as a hassle-free repository for diffraction images.

Proteopedia is a tool for sharing knowledge about the structure and function of proteins, DNA, and other macromolecules on the web, developed by scientists at the Weizmann Institute of Science in Israel. It is a mature web site that can be viewed on virtually all devices via a web browser, including computers, tablets, and smart phones (IOS and Android). It is a web-based Wiki for crowd sourcing concepts on the 3D structure/function relationships of biological macromolecules [1–4]. As it ties text to 3D interactive images, Proteopedia aims to serve as a forum for the scientific community to share, retrieve, and discuss information related to proteins, macromolecules, small molecules, and chemicals of interest. It also aims to maintain low barriers to contribution.

The procedure for authors to create these I3DCs is very simple. Once a paper is accepted, the author or journal's editorial office should notify Proteopedia with an email to manuscript@proteopedia.org. Proteopedia will then develop, in close collaboration with the authors, an I3DC on Proteopedia, that will enhance the paper and allow it to reach a wider audience.

doi: 10.1111/febs.14316

A link within the publication to an I3DC in Proteopedia enables readers to rotate molecular scenes with the same initial orientation, color schemes, and labeling as figures in the publication. Links to the I3DC in Proteopedia can be placed in the text of the publication, and/or in the online text file included on the journal website as supplementary material. The I3DC has a link directly back to the original journal article and the journal article has a link pointing to Proteopedia. It is not visible to the public until the online version of the article appears on the journal's website.

A full description of the process for creating I3DC is available at <http://proteopedia.org/jrnld>. There are already ~ 90 I3DCs in Proteopedia (see list: <http://proteopedia.org/w/Proteopedia:I3DC>) complementing scientific publications from 13 journals, which serve as complements to published papers in scientific journals. An illustration of the Proteopedia entry for a paper published in FEBS Open Bio (http://proteopedia.org/w/Journal:FEBS_Open_Bio:2) is seen in Fig. 1.

Another tool for enhancement of presentation of structural models is called Molstack. Molstack [5] allows the visualization and comparison of selected sets of coordinates and electron density maps. The idea behind the platform is to give authors the ability to present various combinations of coordinates and density maps in an interactive environment that is more informative than a static figure in a publication. Molstack can be used to supplement published figures or present in 3D various coordinates and maps that would otherwise not be published in any fashion. For example, omit maps that are customarily requested by reviewers for data presenting small molecule binding to the protein are usually published as a single figure in supplementary materials, if they are published at all. However, 2D figures may not only be hard to interpret but may also be impossible to reproduce for more thorough analysis due to variations in computational protocols, lack of data, or missing metadata.

Molstack's flexibility allows authors to upload any coordinate or density data and present a multitude of different models and their supporting crystallographic data. Molstack allows side-by-side views of different model/map combinations with all rotations/translation

Structural evidence for Arabidopsis glutathione transferase AtGSTF2 functioning as a transporter of small organic ligands

Laziana Ahmad, Elizabeth L. Rylott, Neil C. Bruce, Robert Edwards and Gideon Grogan ^[1]

Molecular Tour

Glutathione transferases (GSTs) are involved in many processes in plant biochemistry, with their best characterised role being the detoxification of xenobiotics through their conjugation with glutathione. GSTs have also been implicated in noncatalytic roles, including the binding and transport of small heterocyclic ligands such as indole hormones, phytoalexins and flavonoids. Although evidence for ligand binding and transport has been obtained using gene deletions and ligand binding studies on purified GSTs, there has been no structural evidence for the binding of relevant ligands in noncatalytic sites. Here we provide evidence of noncatalytic ligand-binding sites in the phi class GST from the model plant *Arabidopsis thaliana*, AtGSTF2, revealed by X-ray crystallography. Ligands used in this study: 1 = **Indole-3-aldehyde**; 2 = **Camalexin**; 3 = **Quercetrin**; 4 = **Quercetin**. For comparison was used another GST ligand **S-hexyl glutathione** from (1gnw^[2]). Complexes of the AtGSTF2 dimer were obtained with indole-3-aldehyde, camalexin, the flavonoid quercetrin and its non-rhamnosylated analogue quercetin, at resolutions of 2.00, 2.77, 2.25 and 2.38 Å respectively. Two symmetry-equivalent-binding sites (L1) were identified at the periphery of the dimer, and one more (L2) at the dimer interface. **Structure of AtGSTF2 dimer.** The figure is derived using the complex with indole-3-aldehyde and shows location of ligand-binding sites L1 and L2 labelled

Fig. 1. An example of an I3DC in Proteopedia for a FEBS OpenBio manuscript (http://proteopedia.org/w/Journal:FEBS_Open_Bio:2). The green hyperlinks in the text change the live structures on the right so as to illustrate the points in the text. Users can rotate, zoom and interact with the 3D structures for a full 360° view.

Re-refinement - ligand identification

Sign Up Sign In

Saved views

Disaccharide 1

Mannobiose is present in multiple conformations, while single conformation of trehalose can explain the electron

Disaccharide 2

The second binding site is most likely occupied by trehalose as well

Overview

Reset the view to show whole scene

514F (MANNOBIOSE) NO LIGAND TREHALOSE MALTOSE

514F (MANNOBIOSE) NO LIGAND TREHALOSE MALTOSE

Created by Minor lab at University of Virginia

Fig. 2. Screenshot from Molstack. The users of the Molstack are presented with dual, synchronized side-by-side windows which make Molstack particularly useful for comparisons of density maps and their different interpretations. Users can navigate the data using links to author-predefined scenes and are presented with short descriptions.

linked between the two stacks and controlled by the user. The data may represent alternative ligands, ligands in alternative conformations, or speculative interpretations of densities. Molstack can be used to present reinterpretations, re-refinements, or corrections of existing structures (Fig. 2) or present supporting data, such as anomalous difference maps. Collections of related structures, for example, structures with different ligands or different mutations, can be shown for comparison.

Providing easy access to experimental data is a necessary condition for assuring reproducibility of published results. This is particularly important for macromolecular crystallography, where original datasets are quite large. Currently, the largest public repository of raw diffraction data is the Integrated Resource for Reproducibility in Macromolecular Crystallography (IRRM), available at <http://proteindiffraction.org> [6]. The process of deposition of images into the proteindiffraction.org repository is straightforward. A potential depositor can sign-up for a repository account via a simple registration form or one can proceed using an existing ORCID or LinkedIn account. Diffraction datasets can be uploaded using an intuitive 'drag-and-drop' interface. Uploaded data are handled differently according to their category. For datasets associated with released PDB depositions, the only required information is the PDB ID—other metadata are acquired automatically. Data can be also uploaded before being released by the PDB and can be kept 'on hold' until PDB deposit is released. Another category comprises the diffraction data that did not lead to structure determination. Such data are immediately public and it is assumed that authors of the data should be contacted and credited before a corresponding structure is deposited to PDB. Deposition of data implies acknowledgement that it can be freely used according to the Creative Commons CCO license (Public Domain Dedication Waiver).

Deposition of original diffraction data to a specialized repository offers several benefits. Together with PDB deposit or metadata submitted together with diffraction images, data in a repository provide a full description of diffraction experiment in the form that would allow others to easily perform data reduction and structure redetermination. All datasets uploaded

to proteindiffraction.org are assigned Digital Object Identifiers (DOIs) and can be cited in publications. Once uploaded to a repository, the raw data are safely archived, freeing the local storage and freeing the crystallographer from the necessity to track where the data currently resides. Recent papers showing that old diffraction data may be processed to higher resolution or provide a better structural model [7,8] offer a nice illustration of benefits of such a resource.

References

- 1 Hanson RM, Prilusky J, Renjian Z, Nakane T & Sussman JL (2013) JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Isr J Chem* **53**, 207–216.
- 2 Prilusky J, Hodis E & Sussman JL (2012) Proteopedia: Exciting Advances in the 3D Encyclopedia of Biomolecular Structure. Springer, Dordrecht.
- 3 Prilusky J, Hodis E, Canner D, Decatur WA, Oberholser K, Martz E, Berchanski A, Harel M & Sussman JL (2011) Proteopedia: a status report on the collaborative, 3D web-encyclopedia of proteins and other biomolecules. *J Struct Biol* **175**, 244–252.
- 4 Hodis E, Prilusky J, Martz E, Silman I, Moulton J & Sussman JL (2008) Proteopedia – a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome Biol* **9**, R121.
- 5 Porebski PJ, Sroka P, Zheng H, Cooper DR & Minor W (2017) Molstack-Interactive visualization tool for presentation, interpretation, and validation of macromolecules and electron density maps. *Protein Sci.* <https://doi.org/10.1002/pro.3272>, in press.
- 6 Grabowski M, Langner KM, Cymborowski M, Porebski PJ, Sroka P, Zheng H, Cooper DR, Zimmerman MD, Elsliger MA, Burley SK *et al.* (2016) A public database of macromolecular diffraction experiments. *Acta Crystallogr D* **72**, 1181–1193.
- 7 Shabalin I, Dauter Z, Jaskolski M, Minor W & Wlodawer A (2015) Crystallography and chemistry should always go together: a cautionary tale of protein complexes with cisplatin and carboplatin. *Acta Crystallogr D* **71**, 1965–1976.
- 8 Rupp B, Wlodawer A, Minor W, Helliwell JR & Jaskolski M (2016) Correcting the record of structural publications requires joint effort of the community and journal editors. *FEBS J* **283**, 4452–4457.