Aquaria: simplifying discovery and insight from protein structures

To the Editor: Since the discovery of the DNA double helix, biologists have been aware that atomic-scale three-dimensional (3D) structures can provide significant insight. The Protein Data Bank[1] (PDB) contains a wealth of structural information, but few biologists take full advantage of it[2]. Thus, we developed Aquaria (http://aquaria.ws), a publicly available web resource that streamlines and simplifies the process of gleaning insight from protein structures.

In contrast to most molecular graphics tools (for example, Astex[3] or Chimera[4]), the user interface of Aquaria is organized primarily by protein sequence, not structure (**Fig. 1**). A user starts by specifying a protein of interest by name and organism (**Supplementary Fig. 1**), by identifier or by URL (for example, http://aquaria.ws/P04637); Aquaria then generates a concise visual summary of all related PDB structures (**Fig. 1** and **Supplementary Methods**), using a precalculated all-against-all comparison of Swiss-Prot[5] and PDB[1] sequences (updated monthly).

The related structures are grouped first by matches to the specified sequence and second by oligomeric state. Structures are then ranked—in both groupings—by sequence similarity to the specified protein. Users can quickly review all known structural information for a protein and find the structures most relevant to them (**Supplementary Video 1**).

Initially, 3D structures are colored to highlight amino acid differences from the specified protein sequence, with bright, saturated colors indicating identical residues and with slightly dark and very dark coloring indicating conserved and nonconserved substitutions, respectively (**Fig. 1**).

Aquaria also allows mapping of InterPro[6] and UniProt[5] sequence features (for example, domains, single-nucleotide polymorphisms or post-translational modifications) onto 3D structures: a simple yet effective way to gain insight into molecular function[2] (**Supplementary Figs. 2** and **3**).

Aquaria is designed for biologists; its user interface creates clear and useful default views that show only the most relevant structural information tightly integrated with sequence, features and text that provide biological context. Aquaria uses a minimal set of mouse-based controls that are intuitive yet powerful[7]. For example, its "Autofocus" feature allows exploration of large complexes by focusing on one molecule at a time. Aquaria can also be controlled via hand gestures using the Leap Motion[8].

Currently, Aquaria contains 46 million precalculated sequence-to-structure alignments, resulting in at least one matching structure for 87% of Swiss-Prot proteins and a median of 35 structures per protein; this provides a depth of sequence-to-structure information currently not available from other resources.

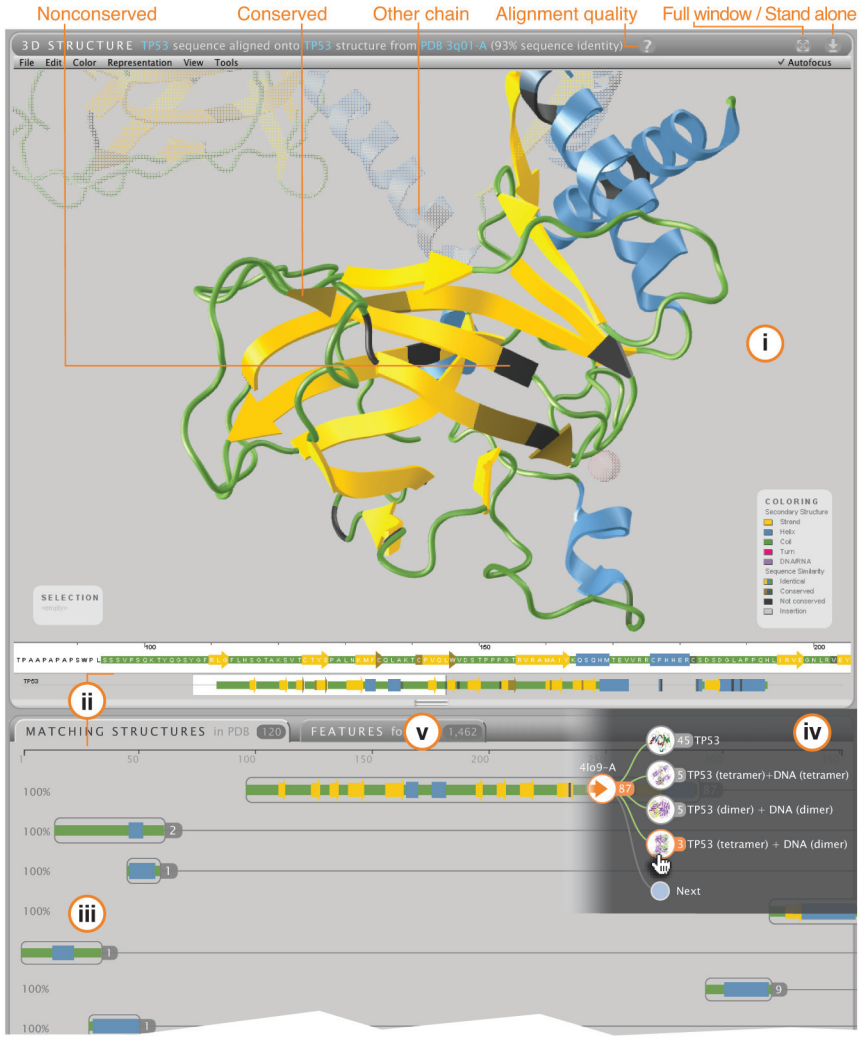

Figure 1 | Aquaria page for human tumor suppressor protein p53. (**i**) Initially, the PDB structure estimated to be most relevant is shown. Dark and very dark residues indicate conserved and nonconserved substitutions, respectively, between the structure and the wild-type p53 sequence (**ii**). Aquaria also shows all related PDB structures, grouped by region of match (**iii**). Clicking on a group loads the top-ranked structure; clicking on a group number shows a tree view of structures organized by oligomeric state (**iv**). InterPro and UniProt features (**v**) can also be mapped onto structures.

Although much of the Aquaria workflow could be performed using a combination of other resources, Aquaria greatly reduces the time and effort required.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.3258).

ACKNOWLEDGMENTS

We thank C. Hammang and V.P. Satagopam for helpful discussions. This work was supported by CSIRO's OCE Science Leader program and its Computational and Simulation Sciences platform, as well as the Alexander von Humboldt Foundation. We gratefully acknowledge the support of Amazon AWS in hosting the Aquaria server and in making PSSH2 available as a public data set.

AUTHOR CONTRIBUTIONS

S.I.O. and A.S. jointly conceived of the project. S.I.O. oversaw the project and wrote the manuscript with contributions from A.S. and K.S.S. Software implementation was led by K.S.S. with contributions from S.I.O., C.S., V.H., N.P., F.A.B. and J.H. Interface design was led by C.S. with contributions from S.I.O. The development of PSSH2, the database underlying Aquaria, was led by A.S. with contributions from M.K., B.W., M.R. and B.R.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Seán I O'Donoghue[1–3], Kenneth S Sabir[2,4], Maria Kalemanov[5], Christian Stolte[1], Benjamin Wellmann[5], Vivian Ho[2], Manfred Roos[5], Nelson Perdigão[6,7], Fabian A Buske[2,8], Julian Heinrich[1], Burkhard Rost[5] & Andrea Schafferhans[5]

[1]Digital Productivity, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sydney, Australia. [2]Division of Genomics and Epigenetics, Garvan Institute of Medical Research, Sydney, Australia. [3]School of Molecular Bioscience, The University of Sydney, Sydney, Australia. [4]School of Information Technologies, The University of Sydney, Sydney, Australia. [5]Department for Bioinformatics and Computational Biology, Technische Universität München, Munich, Germany. [6]Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal. [7]Instituto de Sistemas e Robótica, Lisboa, Lisbon, Portugal. [8]St. Vincent's Clinical School, The University of New South Wales (UNSW), Sydney, Australia. e-mail: sean@odonoghuelab.org

1. Berman, H.M. *et al. Nucleic Acids Res.* **28**, 235–242 (2000).
2. O'Donoghue, S.I. *et al. Nat. Methods* **7**, S42–S55 (2010).
3. Hartshorn, M.J. *J. Comput. Aided Mol. Des.* **16**, 871–881 (2002).
4. Pettersen, E.F. *et al. J. Comput. Chem.* **25**, 1605–1612 (2004).
5. The UniProt Consortium. *Nucleic Acids Res.* **42**, D191–D198 (2014).
6. Hunter, S. *et al. Nucleic Acids Res.* **40**, D306–D312 (2012).
7. O'Donoghue, S.I., Meyer, J.E.W., Schafferhans, A. & Fries, K. *Bioinformatics* **20**, 2476–2478 (2004).
8. Sabir, K., Stolte, C., Tabor, B. & O'Donoghue, S.I. *Proc. IEEE Symp. Biol. Data Vis.* **3**, 49–56 (2013).