Title
Online interactive analysis of protein structure ensembles with Bio3D-web.

Permalink
https://escholarship.org/uc/item/358382gf

Journal
Bioinformatics (Oxford, England), 32(22)

ISSN
1367-4803

Authors
Skjærven, Lars
Jariwala, Shashank
Yao, Xin-Qiu
et al.

Publication Date
2016-11-01

DOI
10.1093/bioinformatics/btw482

Peer reviewed
Structural bioinformatics

Online interactive analysis of protein structure ensembles with Bio3D-web

Lars Skjærven¹,†,*, Shashank Jariwala²,†, Xin-Qiu Yao² and Barry J. Grant²,*

¹Department of Biomedicine, University of Bergen, Bergen, Norway and ²Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors

Associate Editor: Anna Tramontano

Received on April 11, 2016; revised on June 17, 2016; accepted on July 10, 2016

Abstract

Summary: Bio3D-web is an online application for analyzing the sequence, structure and conformational heterogeneity of protein families. Major functionality is provided for identifying protein structure sets for analysis, their alignment and refined structure superposition, sequence and structure conservation analysis, mapping and clustering of conformations and the quantitative comparison of their predicted structural dynamics.

Availability: Bio3D-web is based on the Bio3D and Shiny R packages. All major browsers are supported and full source code is available under a GPL2 license from http://thegrantlab.org/bio3d-web.

Contact: bjgrant@umich.edu or lars.skjarven@uib.no

1 Introduction

Many structures are now available for homologous proteins determined under different crystallization conditions and oligomerization states. Detailed comparison of these structures can inform on structural dynamic mechanisms critical for protein function including ligand binding, enzymatic catalysis, allosteric regulation and biomolecular recognition. A wide range of bioinformatics tools and online servers enable researchers to explore and analyze individual biomolecular structures. Notable examples include molecular visualization, pairwise structural alignment and biophysics based tools including various normal mode analysis servers (Eyal et al., 2015; Suhe and Sanejouand, 2004; Tiwari et al., 2014). However, existing tools for detailed quantitative analysis of the sequence, structure and dynamics of large heterogeneous protein families often require significant computational expertise and typically remain accessible only to expert users with relevant programming skills. For example, the Bio3D package requires R (Grant et al., 2006), ProDy requires python and Maven requires Matlab knowledge (Bakan et al., 2011; Zimmermann et al., 2011). A recent contribution to making such analysis more widely available online is the PDBFlex database (Hrabe et al., 2016). PDBFlex catalogues the structural variation of the same protein (PDB structures sharing 95% or more sequence identity). However, more in-depth interactive analysis of user defined structure sets across species and diverse protein families is not currently available.

To this end we have developed Bio3D-web—a web application that implements a complete workflow for user customized investigation of protein sequence-structure-dynamic relationships. Bio3D-web provides unparalleled online functionality including inter-conformer relationship mapping with principal component analysis (PCA), and quantitative comparison of predicted internal dynamics across protein families via new ensemble normal mode analysis (eNMA). Together with conventional sequence and structure analysis methods these approaches allow researchers to map the structural dynamic properties of proteins for which PDB structures are available.

Bio3D-web requires no programming knowledge and thus decreases the entry barrier to performing advanced comparative sequence, structure and dynamics analysis. Bio3D-web is powered by the previously described Bio3D R package for structural

Bio3D-web is an online application for analyzing user defined heterogeneous biomolecular structure data. The design of Bio3D-web emphasizes simplicity over exhaustive inclusion of the many analysis methods available in the full Bio3D package. This effectively reduces the required technical expertise and thus facilitates advanced structural bioinformatics analysis for a broader range of researchers. In many cases it is envisaged that researchers will use Bio3D-web to understand general trends in their protein family of interest, which may then inform more specialized analyses. Bio3D-web is therefore designed to quickly explore biomolecular structure datasets and to act as a hypothesis-generating tool with sharable summary reports that capture all users defined analysis choices and optionally enable collaborators to visit previous analysis sessions.

### 2 Example application

Providing the PDB ID of a single Rho GTPase (1FTN) in the SEARCH tab identifies 1059 sequence similar structures from which the top 214 are automatically selected as the most related structures for further analysis. Users can optionally (de)select additional structures for further analysis. In this example we proceed with these structures that span RhoA/B/C, Cdc42 and Rac1/2/3 PDB entries. In the ALIGN tab the selected structures are subject to multiple alignment, similarity and conservation analysis. Characterization of superimposed structures is available from the superposition, PCA and eNMA form the basis of the application (Supplementary Fig. S1). Structure and sequence annotations are derived from the RCSB PDB (Berman, 2000) and PFAM databases (Finn et al., 2014). Bio3D-web employs Shiny’s reactive programming and web application framework to provide its interactive online interface (Shiny, http://shiny.rstudio.com).

Collectively these results indicate the existence of three major Rho superfamily conformational states that differ by a collective displacement of two nucleotide-binding site regions, which in turn display significantly distinct flexibilities upon nucleotide binding. In the case of Rho GTPases these regions are known to be key for nucleotide dependent modulation of protein-protein interactions (Wittinghofer and Vetter, 2011). Further example application of Bio3D-web to a range of distinct protein families is presented as interactive demos online (http://thegrantlab.org/bio3d-web).

### 3 Conclusion

Bio3D-web is an online application for analyzing user defined heterogeneous biomolecular structure data. The design of Bio3D-web emphasizes simplicity over exhaustive inclusion of the many analysis methods available in the full Bio3D package. This effectively reduces the required technical expertise and thus facilitates advanced structural bioinformatics analysis for a broader range of researchers. In many cases it is envisaged that researchers will use Bio3D-web to understand general trends in their protein family of interest, which may then inform more specialized analyses. Bio3D-web is therefore designed to quickly explore biomolecular structure datasets and to act as a hypothesis-generating tool with sharable summary reports that capture all users defined analysis choices and optionally enable collaborators to visit previous analysis sessions.

### Acknowledgements

We thank Dr. Guido Scarabelli and Hongyang Li for extensive testing.

### Funding

This work was supported by the University of Michigan, National Institutes of Health (GM070862 to BG) and the University of Bergen.

### Conflict of Interest

None declared.

### References

Bakan, A. et al. (2011) ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*, 27, 1575–1577.

Berman, H.M. (2000) The Protein Data Bank. *Nucleic Acids Res.*, 28, 235–242.

Eyal, E. et al. (2015) The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics*, 31, 1487–1489.

Finn, R.D. et al. (2014) Pfam: the protein families database. *Nucleic Acids Res.*, 42, D222–D230.

Grant, B.J. et al. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, 22, 2695–2696.

Hrabe, T. et al. (2016) PDBFlex: exploring flexibility in protein structures. *Nucleic Acids Res.*, 44, D423–D428.
Skjærven, L. et al. (2014) Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinf.*, 15, 11.

Suhre, K. and Sanejouand, Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.*, 32, W610–W614.

Tiwari, S.P. et al. (2014) WEBnm@ v2.0: Web server and services for comparing protein flexibility. *BMC Bioinf.*, 15, 427.

Wittinghofer, A. and Vetter, I.R. (2011) Structure-function relationships of the G domain, a canonical switch motif. *Annu. Rev. Biochem.*, 80, 943–971.

Yao, X.Q. et al. (2016) Rapid Characterization of Allosteric Networks with Ensemble Normal Mode Analysis. *J. Phys. Chem. B.*, doi:10.1021/acs.jpcb.6b01991.

Zimmermann, M.T. et al. (2011) MAVENs: motion analysis and visualization of elastic networks and structural ensembles. *BMC Bioinf.*, 12, 264.