

Structural bioinformatics

# Michelangelo: sculpting protein views on web pages without coding

Matteo P. Ferla <sup>1,2,\*</sup>, Alistair T. Pagnamenta<sup>1</sup>, David Damerell<sup>2</sup>, Jenny C. Taylor<sup>1</sup> and Brian D. Marsden<sup>2,3</sup>

<sup>1</sup>NIHR Oxford BRC Genomic Medicine, Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK, <sup>2</sup>Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Oxford OX3 7DQ, UK and <sup>3</sup>Kennedy Institute of Rheumatology, University of Oxford, Oxford OX3 7FY, UK

\*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on November 26, 2019; revised on February 3, 2020; editorial decision on February 7, 2020; accepted on February 10, 2020

## Abstract

**Motivation:** The sharing of macromolecular structural information online by scientists is predominantly performed via 2D static images, since the embedding of interactive 3D structures in webpages is non-trivial. Whilst the technologies to do so exist, they are often only implementable with significant web coding experience.

**Results:** Michelangelo is an accessible and open-source web-based application that supports the generation, customization and sharing of interactive 3D macromolecular visualizations for digital media without requiring programming skills. A PyMOL file, PDB file, PDB identifier code or protein/gene name can be provided to form the basis of visualizations using the NGL JavaScript library. Hyperlinks that control the view can be added to text within the page. Protein-coding variants can be highlighted to support interpretation of their potential functional consequences. The resulting visualizations and text can be customized and shared, as well as embedded within existing websites by following instructions and using a self-contained download. Michelangelo allows researchers to move away from static images and instead engage, describe and explain their protein to a wider audience in a more interactive fashion.

**Availability and implementation:** Michelangelo is hosted at [michelangelo.sgc.ox.ac.uk](https://michelangelo.sgc.ox.ac.uk). The Python code is freely available at <https://github.com/thesgc/MichelaNGLo>, along with documentations about its implementation.

**Contact:** [matteo.ferla@well.ox.ac.uk](mailto:matteo.ferla@well.ox.ac.uk)

## 1 Introduction

Communicating the interpretations of macromolecular experimental structures of proteins and their ligands to non-structural biologists is an important but non-trivial task. Providing static 2D images, no matter how visually appealing, inhibits the understanding of the 3D nature of the structural data. Yet, this approach is the current state-of-the-art within contemporary websites and digital publications because, whilst it is relatively trivial to embed two-dimensional images from visualization platforms into modern websites, it is entirely non-trivial to export and embed interactive 3D visualizations without significant experience of web technologies such as JavaScript, HTML and CSS.

A range of software applications that support the visualization of protein structures in three dimensions currently exist, either as commercial pieces of software or as freely available desktop applications (e.g. PyMOL, ICM-Browser etc.; [Martinez \*et al.\*, 2019](#); [Raush \*et al.\*, 2009](#); [Schrödinger LLC, 2015](#)) These standalone platforms are able to create interactive visualizations of structures, but the sharing of these as interactive views is often fundamentally locked into the ecosystem of the creator platform in terms of file formats,

related costs and restrictive licencing terms. Previously, we developed the iSee platform ([Lee \*et al.\*, 2009, 2011](#); [Raush \*et al.\*, 2009](#)) which for the first time made it possible to generate web-based interactive views of macromolecular structures whilst embedding them within text, making it possible for those with limited or no structural biology experience to better understand the information presented in such views. However, this platform depended upon a type of web-browser plugin which is no longer supported by mainstream browsers. More recently, a number of freely available JavaScript web-based protein structure 3D visualization libraries that can be embedded into websites have been created including NGL ([Rose and Hildebrand, 2015](#)), Mol\* ([Sehnal \*et al.\*, 2018](#)), PV ([Guex \*et al.\*, 2009](#)) and Jmol ([Hanson and Lu, 2017](#)). However, there is still a significant barrier associated with the creation of the interactive visualizations for embedding in websites either from the ground-up or via the use of widely adopted desktop applications such as PyMOL.

We present Michelangelo (<https://michelangelo.sgc.ox.ac.uk>), a web application which attempts to lower the barrier to generation and delivery of basic interactive 3D macromolecular visualizations for web pages without the need to write any JavaScript.

## 2 Implementation

MichelANGLo supports the creation of shareable, customizable and secure web pages that use the NGL library (Rose and Hildebrand, 2015) to display the protein interactively alongside text annotated with hyperlinks that control the visualization so that it may reflect the context of the text it is linked to, thereby providing the ability to provide an interactive narrative alongside a responsive 3D view. As sources of initial structural data, MichelANGLo is able to convert PyMOL PSE files to NGL implementations, preserving the majority of graphical features. Alternatively, PDB files or PDB codes can be used as starting points. Furthermore, gene symbols or protein names can be entered, providing a summary of structural coverage for the associated proteins (featureViewer code from Zahn-Zabal *et al.* 2020) thereby allowing the user to make informed decisions on which PDB structures to pick to form the basis of the required visualization.

Basic tools to modify the representation and orientation of the protein and its ligands are available within the interface, but it is important to note that MichelANGLo is not intended to be a replacement for fully featured applications such as PyMOL. Once a visualization is generated, instructions are provided on how to implement the interactive view in a web page (e.g. departmental webpage or blogposts) and a complete HTML file can also be downloaded for use as supplementary material, e.g. The title, description (in markdown format), layout, security and editing rights of the resulting page can be customized.

An important feature within MichelANGLo is the ability to create ‘prolinks’, protein-view-controlling links that are implemented as hyperlinks within user-provided text. These links are specified via HTML syntax and can be created via the user interface (Fig. 1). Examples of the capabilities of prolinks include changing the camera orientation, highlighting residues or regions, changing which chains are visible and additionally the detection and visualization of clashes as spinning spiky balls, similarly to Foldit (Kleffner *et al.*, 2017) or ICM-Browser.

To support the interpretation of non-synonymous genetic variants on protein structure and function, MichelANGLo is able to take a list of point mutations, generating different models showing their location in the structure and combining these visualizations into a single web page.

Static 2D images of proteins in publications and websites are often annotated and sometimes contain inset images or plots. MichelANGLo can be used in combination with such a ‘front’ image whereby the web page loads with the given image but when clicked the image is replaced with the interactive 3D view. This allows MichelANGLo to be easily incorporated into existing web pages or digital publications with no reduction in fidelity, whilst still providing the option for 3D views to be activated. Additionally, different

models and their representations can be combined into a single page wherein prolinks optionally control which model to show.

The use of MichelANGLo as a tool to share interactive visualizations does not depend upon access to a website under one’s own control. MichelANGLo is able to host completed visualizations which can then be shared privately by using the provided URL which contains a unique identifier and is well-suited even for unpublished structures. To support this, users must register for an account allowing them to edit the description, keep track of the pages they create and to control who can edit them. The permissions can also be altered, ranging from making the page publicly listed to encrypting the model and description without storing the password anywhere on the server.

A significant challenge associated with the publishing of digital content is its sustainability and fidelity over time. To combat this problem, a monitoring system for pages flagged by the user as requiring protection is in place wherein every month, and at every codebase update, the visualizations generated from each link in the protected page are compared with a reference image in order to guarantee that the generated views will remain unchanged, therefore allowing pages in MichelANGLo to be safely cited in publications.

The code for MichelANGLo is publicly hosted at [github.com/thesgc/MichelaNGLo](https://github.com/thesgc/MichelaNGLo), under an MIT licence, allowing others to use and improve the platform. Tutorials on usage can be found in the pages and video descriptions are available in the documentation.

## 3 Conclusions

The aim of MichelANGLo is to facilitate the creation of interactive web pages by geneticists and structural biologists. Several examples can be found at [michelangelo.sgc.ac.uk/gallery](https://michelangelo.sgc.ac.uk/gallery). Within NIHR Oxford BRC, the use of MichelANGLo has proven far more effective than text and static images to describe the effects of a range of mutations found in a next generation sequencing cohort, WGS500 (Taylor *et al.*, 2015). A case example is LZTR1 ([michelangelo.sgc.ac.uk/data/LZTR1](https://michelangelo.sgc.ac.uk/data/LZTR1)). In this protein, the mutations result in Noonan syndrome, but the location of these result in different modes of inheritance: mutations on one face of the protein—presumably where its target (hRAS) binds—are dominantly inherited, whilst mutations elsewhere are recessive and destabilizing (Motta *et al.*, 2019; Pagnamenta *et al.*, 2019). The ability to interact with the protein’s structure in three dimensions better imparts this observation, especially being able to focus on recessive mutations to see the clashes (likely destabilization) or dominant mutations to see whether they cluster on the same face or close to a known phosphorylation site, for example.

MichelANGLo, by lowering the barrier to implementation and adoption of 3D interactive visualizations, is a useful basic tool for bridging the gap between web developers and researchers. We hope that it will encourage the move away from static images as the default communications modality and towards interactive protein visuals by default in digital publications.

## Acknowledgements

We thank the SGC code review group, in particular Elliot Nelson, Rachael Skyner, Mihaela Smilova and Conor Wild, for their invaluable help in beta testing the site.

## Funding

This work was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre Programme and a Wellcome Trust Core Award [203141/Z/16/Z]. B.M. is supported by the Kennedy Trust for Rheumatology Research. The SGC is a registered charity [1097737] that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA Darmstadt Germany, MSD, Novartis Pharma AG, Ontario Ministry of Economic Development and

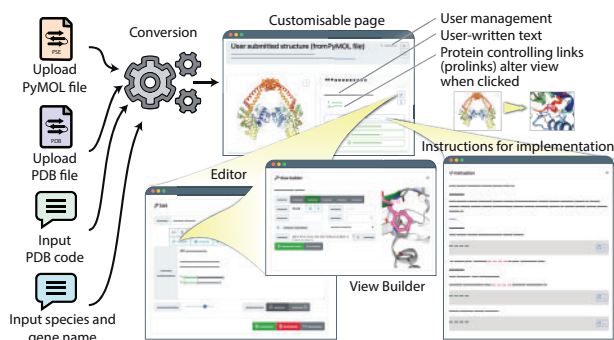


Fig. 1. From model to webpage. There are four input routes (PyMOL file, PDB file, PDB code and protein name), which generate a page that can be further customized by editing the description and by adding protein controlling links within the description via the view builder. Once a satisfactory page has been made, the user can share the URL of the page, download a static copy or follow the detailed instructions on implementing the view on their own site

Innovation, Pfizer, Takeda and Wellcome Trust [106169/ZZ14/Z]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Wellcome Trust.

*Conflict of Interest:* none declared.

## References

- Guex,N. *et al.* (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*, **30**, S162–S173.
- Hanson,R.M. and Lu, X.J. (2017) DSSR-enhanced visualization of nucleic acid structures in Jmol. *Nucleic Acids Res.*, **45**, W528–W533.
- Kleffner,R. *et al.* (2017) Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics*, **33**, 2765–2767.
- Lee,W.H. *et al.* (2009) SGC—structural biology and human health: a new approach to publishing structural biology results. *PLoS One*, **4**, e7675.
- Lee,W.H. *et al.* (2011) Interactive JIMD articles using the iSee concept: turning a new page on structural biology data. *J. Inherit. Metab. Dis.*, **34**, 565–567.
- Martinez,X. *et al.* (2019) Molecular graphics: bridging structural biologists and computer scientists. *Structure*, **27**, 1617–1623.
- Motta,M. *et al.* (2019) Activating MRAS mutations cause Noonan syndrome associated with hypertrophic cardiomyopathy. *Hum. Mol. Genet.*, doi: 10.1093/hmg/ddz108.
- Pagnamenta,A.T. *et al.* (2019) Delineation of dominant and recessive forms of LZTR1-associated Noonan syndrome. *Clin. Genet.*, **95**, 693–703.
- Raush,E. *et al.* (2009) A new method for publishing three-dimensional content. *PLoS One*, **4**, e7394.
- Rose, A.S. and Hildebrand, P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.
- Schrödinger LLC. (2015) The PyMOL Molecular Graphics System, 1.8. [Software].
- Sehna,D. *et al.* (2018) Mol\*: towards a common library and tools for web molecular graphics. In: *MolVa: Workshop on Molecular Graphics and Visual Analysis of Molecular Data, Brno, Czech Republic*. Eurographics.
- Taylor,J.C. *et al.* (2015) Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat. Genet.*, **47**, 717–726.
- Zahn-Zabal,M. *et al.* (2020) The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res.*, **48**, D328–D334.