

Structural bioinformatics

pyGOMoDo: GPCRs modeling and docking with python

Rui P. Ribeiro^{1,2} and A. Giorgetti ^{1,2*}

¹Computational Biomedicine - IAS-5/INM-9, Forschungszentrum Jülich, Jülich, Germany

²Department of Biotechnology, University of Verona, Verona 34134, Italy

*Corresponding author. Department of Biotechnology, University of Verona, Verona, Italy. E-mail: alejandro.giorgetti@univr.it

Associate Editor: Lenore Cowen

Abstract

Motivation: We present pyGOMoDo, a Python library to perform homology modeling and docking, specifically designed for human GPCRs. pyGOMoDo is a python wrap-up of the updated functionalities of GOMoDo web server (<https://molsim.sci.univr.it/gomodo>). It was developed having in mind its usage through Jupyter notebooks, where users can create their own protocols of modeling and docking of GPCRs. In this article, we focus on the internal structure and general capabilities of pyGOMoDo and on how it can be useful for carrying out structural biology studies of GPCRs.

Results: The source code is freely available at <https://github.com/rribeiro-sci/pygomodo> under the Apache 2.0 license. Tutorial notebooks containing minimal working examples can be found at <https://github.com/rribeiro-sci/pygomodo/tree/main/examples>.

1 Introduction

GPCRs Online Modeling and Docking Webserver (GOMoDo) is a G-protein coupled receptors (GPCRs) online modeling and docking web server, developed by Sandal and collaborators (Sandal et al. 2013), and it is publicly available since 2013 (<https://molsim.sci.univr.it/gomodo>). With a very easy user interface, this biocomputing platform allows users to effortlessly model GPCR structures and dock ligands to the model, obtaining biological and pharmacological relevant data, in a consistent pipeline: protein sequence alignment, homology modeling and model quality assessment, and docking (Sandal et al. 2013). One of the novelties that GOMoDo brought to the bioinformatics community was the use of a curated local database of pre-aligned GPCR sequences, plus the possibility of directly docking the ligands into the models. Moreover, the user has the possibility of uploading models generated by other tools. However, at the time of its development and deployment only less than 3% of the human GPCR structures were available. Moreover, the code behind both the back and front-end of the web server has not only been difficult to maintain over the years, but no longer satisfies the necessary requisites for the utmost web performance and security. Despite these limitations, GOMoDo is still used today, which prompted us to revamp it by creating pyGOMoDo. Compared to its predecessor, pyGOMoDo boasts several noteworthy improvements, such as the addition of a new model quality assessment tool [QMEANBrane (Studer et al. 2014)], multiple options for docking programs (including AUTODOCK VINA (Trott and Olson 2010) and rDOCK (Ruiz-Carmona et al. 2014)), enhanced utilities for interaction analysis [ProLIF (Bouysset and Fiorucci 2021)], and up-to-date internal databases (Fig. 1).

GPCRs, which are expressed in every organ system, are one of the most crucial pharmacological targets, playing a

significant role in regulating almost every aspect of physiology. Although many big pharmaceutical and biotech companies lost their interest and expertise in GPCR targets (Mullard 2018), the truth is that not only around 20% of FDA approved drugs each year target these receptors, but also a new opening up in GPCR-based drug discovery has been stated (Mullard 2018). In fact, GPCRs are still offering emerging therapeutic opportunities: peptide therapeutics and modification of natural ligands (Rask-Andersen et al. 2011; Hauser et al. 2017; Demartis et al. 2018), GPCRs-target antibody therapeutics (Hutchings et al. 2017), allosteric sites (Thal et al. 2018), orphan human GPCRs (Cohen et al. 2017), oligomerization (George et al. 2002; Ferré 2015), and biased (Oprea et al. 2018) and bivalent (Pérez-Benito et al. 2018) ligands for GPCRs.

Even if the number of three-dimensional structures of GPCRs have been rapidly increasing thanks to the constant improvement in techniques for structure determination like high-throughput X-ray crystallography, multidimensional NMR spectroscopy, Cryo-Electron microscopy (cryo-EM), or small angle X-ray scattering (SAXS) (Pandey et al. 2016), the supply-demand gap in GPCR structures is still enormous. This is mainly due to the hydrophobic nature of the transmembrane domains of GPCRs that make their purification and crystallization challenging (Doerr 2010). Thus, computational protein modeling methods are a valid alternative. Recently, neural networks-based modeling techniques have been shone through, especially since the release of AlphaFold (Jumper et al. 2021) and RoseTTAFold (Baek et al. 2021) (an AlphaFold-based neural network) (Lee et al. 2022). However, even if such algorithms have achieved substantial success, GPCRs modeling template-based methods, pyGOMoDo in particular, still remains as a valid alternative to build high

Received: December 2, 2022. Revised: March 27, 2023. Accepted: May 2, 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

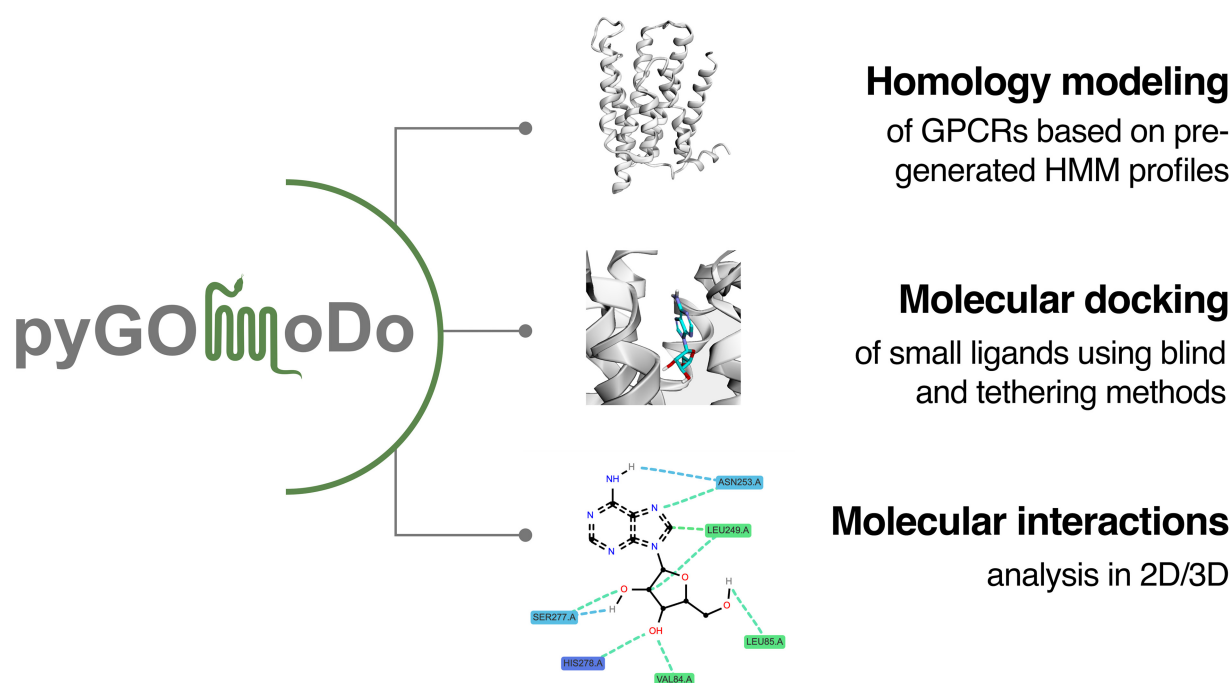


Figure 1. Structure of the pyGOMoDo package.

quality structural models (Fierro et al. 2017, 2019; Capaldi et al. 2018; Schneider et al. 2018; Meyer et al. 2022).

When working with modeling and docking programs, it is typically necessary to have a thorough understanding of the software's internal workings. To address this challenge, object-oriented libraries that encapsulate essential functionalities may provide a consistent working interface, enabling users to easily create analysis tools without having to spend time and effort re-implementing standard functionalities. Here, we present the pyGOMoDo: a python library to perform homology modeling and molecular docking of GPCRs. We have built the library from scratch having in mind the key features of the original GOMoDo. Although pyGOMoDo is not a web server, we have chosen to keep its original name to facilitate recognition and establish a connection with its earlier versions.

2 Software implementation

The library is divided into two main modules: the modeling and docking modules. The modeling module follows the same framework pipeline as the original web server: protein sequence alignment and hidden Markov model profile generation with the hh-suite (Steinegger et al. 2019), and homology modeling with the Modeller software (Webb and Sali 2016). We have included in pyGOMoDo a new model quality assessment. The models can now be assessed with the QMEANBrane program (Studer et al. 2014) through the Swiss-Model web server API (Waterhouse et al. 2018). The HMM profiles are generated against the most updated version of the UniRef30 (Mirdita et al. 2017) database, and the templates are obtained by aligning the generated HMM profile against a database of HMM profiles for the GPCRs whose structures are known. We constructed the latter database from the ground up by gathering all human GPCR structures available on the Protein Data Bank (rcsb.org) (Berman et al. 2000). All the structures were parsed, cleaned, and uniformed, and their HMM profiles were built also against the

UniRef30 database (Mirdita et al. 2017). We have computed the coverage (the percentage of solved amino acids in the sequence) for each PDB structure and, whenever feasible, obtained information about the receptor's conformational state. Thus, the choice for the best templates for modeling can be done according to the sequence identity shared with the template and considering the “solved coverage” of the structure and/or its conformational state (when possible).

Differently from the original GOMoDo web server, the docking module of pyGOMoDo works independently from the modeling module. The user can, therefore, perform molecular docking over a structure modeled with the modeling module, another software or web server, or even over an experimental structure.

The docking module implements the Autodock Vina (Trott and Olson 2010) and rDOCK (Ruiz-Carmona et al. 2014) docking programs in three different docking protocols: (i) a typical Autodock Vina protocol, but with automatic selection of the orthosteric binding site (all the Autodock Vina setting values can be toggled by the user), (ii) docking of molecules on the binding site of crystallographic ligand with rDock, and (iii) tethered docking of molecules on a crystallographic ligand also with rDOCK. The docking module also implements the Protein-Ligand Interaction Fingerprints (ProLIF) (Bouysset and Fiorucci 2021) tool that allows the analysis of all the ligand-protein interactions.

The pyGOMoDo library was specifically designed having in mind its usage through Jupyter notebooks providing the sweet-spot between performance, flexibility, and visualization [tables visualization are rendered by the pandas (McKinney 2010) library and protein visualization by the py3Dmol (Rego and Koes 2015) library]. The pyGOMoDo code and its documentation and tutorials can be accessed through <https://github.com/rribeiro-sci/pygomodo/tree/main/examples>. However, since all the software dependencies are already precompiled inside a docker container, we recommend using pyGOMoDo through the latter, available at <https://hub.docker.com/r/rpribeiro/pygomodo>.

3 Conclusion

Taken as a whole, here we present an updated tool, pyGOMoDo, to perform homology modeling and docking on GPCRs. By developing a python library instead of a static web server, we are offering a much more flexible tool, so users can create their own modeling and docking pipelines according to their needs. The user has the flexibility to intervene at various stages of the process. For example, for the docking component, the software allows the utilization of models or structures generated via alternative methods, such as AlphaFold2 (Jumper et al. 2021) and RoseTTAFold (Baek et al. 2021). Indeed, as for homology models, AlphaFold2 now allows the prediction of different conformational sites of GPCRs (del Alamo et al. 2022; Stein and Mchaourab 2022; Pándy-Szekeres et al. 2023). Thus, models created by independent modeling programs/algorithms can be uploaded and funneled to the different docking protocols offered by pyGOMoDo.

Acknowledgements

The authors deeply acknowledge Margherita de Riu and Francesco Gualdi for their preliminary work in the early phase of pyGOMoDo development.

Conflict of interest

None declared.

Funding

This work was supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3); and by #NEXTGENERATIONEU (NGEU) and the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS [PE0000006]—A Multiscale integrated approach to the study of the nervous system in health and disease [DN. 1553 11.10.2022].

References

del Alamo D, Sala D, Mchaourab HS *et al.* Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* 2022;11:e75751.

Baek M, DiMaio F, Anishchenko I *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;373:871–6.

Berman HM, Westbrook J, Feng Z *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.

Bouysset C, Fiorucci S. ProLIF: a library to encode molecular interactions as fingerprints. *J Cheminform* 2021;13:72.

Capaldi S, Suku E, Antolini M *et al.* Allosteric sodium binding cavity in GPR3: a novel player in modulation of A β production. *Sci Rep* 2018;8:11102.

Cohen LJ, Esterhazy D, Kim S-H *et al.* Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* 2017;549:48–53.

Demartis A, Lahm A, Tomei L *et al.* Polypharmacy through phage display: selection of glucagon and GLP-1 receptor Co-agonists from a phage-displayed peptide library. *Sci Rep* 2018;8:585.

Doerr A. NMR and the elusive GPCR. *Nat Methods* 2010;7:581.

Ferré S. The GPCR heterotetramer: challenging classical pharmacology. *Trends Pharmacol Sci* 2015;36:145–52.

Fierro F, Suku E, Alfonso-Prieto M *et al.* Agonist binding to chemosensory receptors: a systematic bioinformatics analysis. *Front Mol Biosci* 2017;4:63.

Fierro F, Giorgetti A, Carloni P *et al.* Dual binding mode of “bitter sugars” to their human bitter taste receptor target. *Sci Rep* 2019;9:8437.

George SR, O'Dowd BF, Lee SP *et al.* G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat Rev Drug Discov* 2002;1:808–20.

Hauser AS, Attwood MM, Rask-Andersen M *et al.* Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* 2017;16:829–42.

Hutchings CJ, Koglin M, Olson WC *et al.* Opportunities for therapeutic antibodies directed at G-protein-coupled receptors. *Nat Rev Drug Discov* 2017;16:787–810.

Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9.

Lee C, Su B-H, Tseng YJ *et al.* Comparative studies of AlphaFold, RoseTTAFold and modeller: a case study involving the use of G-protein-coupled receptors. *Brief Bioinf* 2022;23:bbac308.

Mckinney W. Data structures for statistical computing in python. *Proc 9th Python Sci Conf* 2010;445:56–61.

Meyer M, Jurek B, Alfonso-Prieto M *et al.* Structure-function relationships of the disease-linked A218T oxytocin receptor variant. *Mol Psychiatry* 2022;27:907–17.

Mirdita M, von den Driesch L, Galiez C *et al.* Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* 2017;45:D170–D176.

Mullard A, Fiona Marshall. *Nat Rev Drug Discov* 2018;17:158–9.

Oprea TI, Bologna CG, Brunak S *et al.* Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov* 2018;17:317–32.

Pandey A, Shin K, Patterson RE *et al.* Current strategies for protein production and purification enabling membrane protein structural biology. *Biochem Cell Biol* 2016;94:507–27.

Pándy-Szekeres G, Caroli J, Mamyrbekov A *et al.* GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Res* 2023;51:D395–D402.

Pérez-Benito L, Henry A, Matsoukas M-T *et al.* The size matters? A computational tool to design bivalent ligands. *Bioinformatics* 2018;34:3857–63.

Rask-Andersen M, Almén MS, Schiöth HB *et al.* Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* 2011;10:579–90.

Rego N, Koes D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* 2015;31:1322–4.

Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N *et al.* rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* 2014;10:e1003571.

Sandal M, Duy TP, Cona M *et al.* GOMoDo: a GPCRs online modeling and docking webserver. *PLoS One* 2013;8:e74092.

Schneider J, Korshunova K, Musiani F *et al.* Predicting ligand binding poses for low-resolution membrane protein models: perspectives from multi-scale simulations. *Biochem Biophys Res Commun* 2018;498:366–74.

Stein RA, Mchaourab HS. SPEACH_AF: sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLoS Comput Biol* 2022;18:e1010483.

Steinegger M, Meier M, Mirdita M *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 2019;20:473.

Studer G, Biasini M, Schwede T *et al.* Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBran). *Bioinformatics* 2014;30:i505–511.

Thal DM, Glukhova A, Sexton PM *et al.* Structural insights into G-protein-coupled receptor allostery. *Nature* 2018;559:45–53.

Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31:455–61.

Waterhouse A, Bertoni M, Bienert S *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;46:W296–W303.

Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinf* 2016;54:5.6.1–5.6.37.