Structural bioinformatics

AHoJ: rapid, tailored search and retrieval of apo and holo protein structures for user-defined ligands

Christos P. Feidakis¹, Radoslav Krivak², David Hoksza² and Marian Novotny ¹,∗

¹Department of Cell Biology, Faculty of Science, Charles University, Prague 12843, Czech Republic and ²Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague 12116, Czech Republic

*To whom correspondence should be addressed.

Received on July 5, 2022; revised on September 29, 2022; editorial decision on October 19, 2022; accepted on October 24, 2022

Abstract

Summary: Understanding the mechanism of action of a protein or designing better ligands for it, often requires access to a bound (holo) and an unbound (apo) state of the protein. Resources for the quick and easy retrieval of such conformations are severely limited. Apo–Holo Juxtaposition (AHoJ), is a web application for retrieving apo–holo structure pairs for user-defined ligands. Given a query structure and one or more user-specified ligands, it retrieves all other structures of the same protein that feature the same binding site(s), aligns them, and examines the superimposed binding sites to determine whether each structure is apo or holo, in reference to the query. The resulting superimposed datasets of apo–holo pairs can be visualized and downloaded for further analysis. AHoJ accepts multiple input queries, allowing the creation of customized apo–holo datasets.

Availability and implementation: Freely available for non-commercial use at http://apoholo.cz. Source code available at https://github.com/cusbg/AHoJ-project.

Contact: marian@natur.cuni.cz

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

The study of protein–ligand interactions constitutes a prominent field in structural biology. Observing the effects of ligand binding (Brylinski and Skolnick, 2008), or exploring the specificity of a binding site (Ma et al., 2002), involve studying several protein–ligand interactions. Unveiling cryptic binding sites (Cimermancic et al., 2016), assessing the importance and consistency of water molecules (Wlodawer et al., 2018), or transcending the technical limitations of rigid body docking with ensemble docking methodologies (Amaro et al., 2018), also require access to several conformations (preferably apo and holo).

A number of datasets and tools have been built to address this need. ComSin (Lobanov et al., 2010) comprised a database of apo and holo protein pairs which exhibit significant shifts in their levels of intrinsic disorder upon complex formation. AH-DB (Chang et al., 2012) expanded this scope by including small ligands in its repertoire of apo–holo pairs. The BUDDY-system (Mortia et al., 2011) provided a more flexible solution where the user could specify the ligand of interest, and the application would try to pair up the provided holo structure with an apo counterpart. At the time of writing, none of these servers are available. A recent work in preprint (APObind—unpublished data) aims to complement an existing database of protein–ligand complexes, by pairing up the holo complexes with their apo counterparts. LigASite (Dessailly et al., 2008) is a more dated yet surviving resource that features pairs of apo and holo structures for 550 proteins. In both cases however, the ligand cannot be specified by the user.

The available resources appear to be restricted, and in some cases non-existent. The ability to define a ligand, and therefore a binding site, that will guide the search for apo and holo structures is missing altogether. This can be particularly useful as proteins often bind several ligands, and even within the same protein, different structures can bind different ligands in the same or in different binding sites. Therefore, finding pairs of apo and holo structures for a given target structure, requires specifying one or more ligands of interest. A methodology that defines the relevant ligands according to a fixed assumption (i.e. automatically), can restrict a user who wants to focus on a ligand that is deemed irrelevant, or narrow down the search to a single ligand when more bind the same structure. Ultimately, when an application forcefully decides upon the relevance of a ligand, it strips the user of this choice and it is also confronted with the non-trivial matter of biological relevance (Capitani et al., 2016).

Here, we present a web application that enables the user to conduct easy and fast parameterizable searches for apo and holo structure pairs against a target structure, by specifying one or more ligands of interest in this target structure, or letting the application detect the ligands
instead. By tracking the binding site of the user-defined ligand across structures, it can construct a repertoire of ligands that bind the same site and enable studies on binding-site specificity.

2 Materials and methods

AHoJ starts the search by spatially marking the user-defined ligand(s) and identifying their binding residues with PyMOL. Ligands are typically confined to non-protein chemical moieties, however in AHoJ, the concept of ligand can be extended to include water molecules and modified or non-standard residues (e.g. phosphorylated residues or D-residues) as points of interest or candidate ligands (see Supplementary Information for details).

It then compiles a list of candidate structure chains by (i) detecting the UniProt accession number (AC) (UniProt: the universal protein knowledgebase, 2017) of each query chain and (ii) retrieving all other chains that belong to the same UniProt AC. At the same time, it maps the binding residues of the query ligands onto the UniProt sequence by using the residue-level mappings from SIFTS (Dana et al., 2019), and cross-examines each candidate chain to determine how many of the mapped binding residues are present. If a minimum percentage of binding residues is detected, the chain is considered a successful candidate and it is aligned onto the query chain with TM-align (Zhang and Skolnick, 2005). The user can adjust these parameters (see Supplementary Information for details). The candidate’s area around the superimposed query ligand is examined for ligands, and the results are saved along with the aligned chains. This process is repeated for all candidate chains and each one is listed as holo or apo respective to the presence or absence of ligands in the defined binding site(s). The detected ligands along with metrics for the similarity between candidate and query, presence of binding residues and alignment scores, are reported for each apo and holo chain. The overall workflow is depicted in Supplementary Figure S1. Results are visualized in the browser and can be downloaded locally and loaded into PyMOL through an included script.

Acknowledgements

We thank the reviewers for taking the time to review the manuscript and providing valuable feedback.

Funding

This work was supported by the Grant Agency of Charles University [Project No. 1038120] and the ELIXIR CZ Research Infrastructure [ID LM2018131, MEYS CR].

Conflict of Interest: none declared.

References

Amaro,R.E., et al. (2018) Ensemble docking in drug discovery. Biophys. J., 114, 2271–2278.
Brylinski,M. and Skolnick,J. (2008) What is the relationship between the global structures of apo and holo proteins? Proteins, 70, 363–377.
Capitani,G., et al. (2016) Understanding the fabric of protein crystals: computational classification of biological interfaces and crystal contacts. Bioinformatics, 32, 481–489.
Chang,D.T.-H., et al. (2012) AH-DB: collecting protein structure pairs before and after binding. Nucleic Acids Res., 40, D472–D478.
Cimermancic,P., et al. (2016) CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. J. Mol. Biol., 428, 709–719.
Dana,J.M., et al. (2019) SIFTS: updated structure integration with function, taxonomy and sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. Nucleic Acids Res., 47, D482–D489.
DessauWy,B.H., et al. (2008) LigASite—a database of biologically relevant binding sites in proteins with known apo-structures. Nucleic Acids Res., 36, D667–D673.
Lobanov,M., et al. (2010) ComSIn: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder. Nucleic Acids Res., 38, D283–D287.
Ma,B., et al. (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci., 11, 184–197.
Monta,M., et al. (2011) BUDDY-system: a web site for constructing a dataset of protein pairs between ligand-bound and unbound states. BMC Res. Notes, 4, 143.
Schiebel,J., et al. (2018) Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. Nat. Commun., 9, 3559.
The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res., 45, D158–D169.
Wlodawer,A., et al. (2018) Detect, correct, retract: how to manage incorrect structural models. FEBS J., 285, 444–466.
Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res., 33, 2302–2309.