Bioinformatics, 35(9), 2019, 1585–1587
doi: 10.1093/bioinformatics/bty816
Advance Access Publication Date: 20 September 2018
Applications Note

Structural bioinformatics

Large-scale prediction of binding affinity in protein–small ligand complexes: the PRODIGY-LIG web server

Anna Vangone *,†,‡, Joerg Schaarschmidt †, Panagiotis Koukos†, Cunliang Geng, Nevia Citro, Mikael E. Trellet, Li C. Xue and Alexandre M. J. J. Bonvin *

Department of Chemistry, Faculty of Science, Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands

*To whom correspondence should be addressed.
†The authors wish it to be known that these authors contributed equally.
‡Present address: Roche Pharmaceutical Research and Early Development, Large Molecule Research, Roche Innovation Center Penzberg, Nonnenwald 2, 82377 Penzberg, Germany
Associate Editor: Alfonso Valencia

Received on March 15, 2018; revised on August 15, 2018; editorial decision on September 15, 2018; accepted on September 19, 2018

Abstract

Summary: Recently we published PROtein binDIng enerGY (PRODIGY), a web-server for the prediction of binding affinity in protein–protein complexes. By using a combination of simple structural properties, such as the residue-contacts made at the interface, PRODIGY has demonstrated a top performance compared with other state-of-the-art predictors in the literature. Here we present an extension of it, named PRODIGY-LIG, aimed at the prediction of affinity in protein-small ligand complexes. The predictive method, properly readapted for small ligand by making use of atomic instead of residue contacts, has been successfully applied for the blind prediction of 102 protein–ligand complexes during the D3R Grand Challenge 2. PRODIGY-LIG has the advantage of being simple, generic and applicable to any kind of protein-ligand complex. It provides an automatic, fast and user-friendly tool ensuring broad accessibility.
Availability and implementation: PRODIGY-LIG is freely available without registration requirements at http://milou.science.uu.nl/services/PRODIGY-LIG.
Contact: a.vangone@gmail.com or a.m.j.j.bonvin@uu.nl

1 Introduction

Functions in cells are controlled by interactions between biomolecules, such as proteins, nucleic acids and small ligands. Understanding such interactions is therefore a crucial step in the investigation of biological systems and in drug design. The binding affinity of a complex, or the Gibbs free energy (ΔG) in thermodynamics words, is a crucial quantity for the study of such systems since it determines whether an interaction will actually occur or not in the cell.

Recently, we have introduced a contact-based method for the prediction of the binding affinity of protein–protein complexes (Vangone and Bonvin, 2015), implemented in the web server PRODIGY (PROtein binDIng enerGY prediction) (Vangone and Bovin, 2017; Xue *et al.*, 2016). We have shown that the binding affinity between proteins can be described by the number and type of interfacial residue–residue contacts in combination with properties of the non-interacting surface (Kastritis *et al.*, 2014). PRODIGY is currently one of the best predictors reported so far, with a Pearson's Correlation coefficient of 0.73 between the predicted and measured binding affinity (P-value < 0.0001) and root mean square error (RMSE) of 1.89 kcal mol^{-1}. It has been trained and validated on a

large and heterogeneous dataset (Kastritis *et al.*, 2011; Vreven *et al.*, 2015) and robust both for rigid and flexible complexes.

Here, we present an extension named PRODIGY-LIG (PRODIGY for LIGands), aimed at the prediction of the binding affinity in protein–small ligand systems. To date many methods have been developed in order to predict the affinity between a protein and a small ligand. These are frequently implemented as scoring function in docking protocols or separate analysis tools. Many of those can remain, however, out of reach for a large community, limited by computational cost and/or required skills and resources needed to properly make use of the software. In fact, despite some valuable online available tools (Brylinski, 2013; Hongjian Li, 2012; Jain and Jayaram, 2005; Jimenez *et al.*, 2018; Labbe *et al.*, 2015, 2017; Pires and Ascher, 2016; Wang *et al.*, 2012), there is still a lack of free web servers to perform fast and automatic prediction of binding affinity in protein-ligand complexes.

Our predictive method, adapted from PRODIGY to address systems with small ligands, makes use of atomic instead of residue contacts. It has been successfully applied for the blind prediction during the D3R Grand Challenge 2 (Gaieb *et al.*, 2018; Kurkcuoglu *et al.*, 2018). In Kurkcuoglu *et al.* (2018), we trained PRODIGY-LIG on 200 protein–small ligand complexes with known experimental binding affinity and structure, retrieved from the 2P2I dataset (Basse *et al.*, 2016). For each entry, we performed a refinement of the interfaces through HADDOCK (our in-house docking software) (van Zundert and Bonvin, 2014) and extracted the inter-molecular electrostatics energy and the type and number of intermolecular atomic contacts (ACs). Precisely, the ACs between the protein and the ligand within a 10.5 Å distance cutoff were calculated and classified according to the atoms involved in the interaction (C = Carbon, O = Oxygen, N = Nitrogen and X = All other atoms). We used this combination of structural- and energy-based terms to train multiple linear regression models. The resulting binding affinity predictor models ΔG_{score} and $\Delta G_{prediction}$ for ranking ligands and predicting the affinity, respectively, are:

$$\Delta G_{score} = 0.343794 * E_{elec} - 0.037597 * AC_{CC} + 0.138738 * AC_{NN} + 0.160043 * AC_{OO} - 3.088861 * AC_{XX} + 187.011384 \quad (1)$$

$$\Delta G_{predicted} = 0.0115148 * Eelec - 0.0014852\ AC_{CC} + 0.0057097 * AC_{NN} - 0.1301806 * AC_{XX} - 5.1002233 \quad (2)$$

where AC_{CC}, AC_{NN}, AC_{OO} and AC_{XX} are the ACs between Carbon–Carbon, Nitrogen–Nitrogen, Oxygen–Oxygen and between all other atoms and polar hydrogens, respectively. E_{elec} is the electrostatic energy calculated through the HADDOCK refinement protocol. In addition, we trained the 'no electrostatic protocol' which only makes use of structural terms (i.e. ACs):

$$\Delta G_{noelect} = 0.0354707 * AC_{NN} - 0.1277895 * AC_{CC} - 0.0072166 * AC_{CN} - 5.1923181 \quad (3)$$

Our predictor was successfully used to predict the binding affinity of 102 protein-ligand targets during the blind D3R Grand Challenge 2—Stage 2. Using exclusively docked models, it reached a correlation score (Kendall's Tau) of 0.37, placing our approach as the ninth best predictor out of over 82 submissions (Kurkcuoglu *et al.*, 2018). However, if the crystal structures made available at the second stage of the D3R challenge would have been used, our approach would have reached a correlation of 0.43, making it the third best ranking method, as reported by the organizers of the Challenge (see page 10 in Gaieb *et al.*, 2018).

We further tested our predictor on an independent dataset of protein–small ligand complexes, reported in the PDBbind database (core_set_v.2013) (Liu *et al.*, 2017). Over a set of 124 entries, for which the Ki has been experimentally determined, PRODIGY-LIG reported an accuracy in terms of Pearson's Correlation coefficient of 0.57 ('no electrostatic protocol'), *P*-values < 0.0001 and RMSE of 2.6 kcal mol^{-1} (see the online page 'Dataset' for further details and data download, info reported below).

In order to facilitate the use of our predictor, we have extended our PRODIGY web-server to support binding affinity prediction for protein–small ligand complexes. Its user-friendly interface allows for automatic, fast and large-scale prediction, supporting submission of ensembles of complexes. Dedicated online pages on the top banner report about the predictive method implemented ('Method' page), the testing on the D3R and PDBbind datasets ('Dataset' page), the step by step submission process ('Manual' page) and a typical example page ('Example' page) and provide a link to the PRODIGY-LIG user forum (http://ask.bioexcel.eu/c/prodigy) ('User Forum' page).

2 The web server

We have implemented PRODIGY-LIG as a user-friendly web server, freely available without registration at http://milou.science.uu.nl/services/PRODIGY-LIG.

Users are required to provide the following information to the server:

1. A 3D experimental or modeled structure of the protein-ligand complex in PDB or mmCIF format. This can be provided as a single structure or as a compressed archive (tar, tgz, zip, bz2 or tar.gz) containing multiple protein-ligand complexes for batch predictions. The archive can contain different modelled/refined poses of the same complex or different protein/ligand systems.
2. The chain identifiers for the protein and the ligand, as well as the residue identifier of the ligand involved in the interaction.
3. (Optional) The electrostatic energy of the complex, calculated by the HADDOCK refinement server.
4. (Optional) An email address to which a link to the result page upon job completion will be sent.

The results are kept on our server for download for 2 weeks. Upon successful validation of the input data, users are redirected to the job page, which displays the status of the job during execution and the results upon completion. The summarized results include the predicted binding affinity in kcal/mol and information about interfacial ACs, classified by atom type. In case of multiple complexes, the information is summarized in a sortable table. On the server, users can find detailed information about the predictive approach and how to use it, in the 'Method' and 'Manual' pages, respectively. In order to seamlessly integrate PRODIGY-LIG, we have migrated the PRODIGY server to the Python framework Flask (version 0.12.2) and added the new functionalities to this instance. The server is running on a dedicated Linux server.

In addition to the web-server, we also provide a standalone version of the PRODIGY-LIG code for local use. Information about use, installation and download are reported in our GitHub repository at https://github.com/haddocking/prodigy-lig.

3 Conclusion

We have presented here an expansion of PRODIGY, namely PRODIGY-LIG, for the prediction of binding affinities in protein–

small ligand complexes. Its user-friendly interface and free accessibility will contribute to a broad usage, reaching researchers with different background and limited computational resources and boosting the development of new binding affinity predictors and scoring functions.

Funding

This work was supported by the European H2020 e-Infrastructure grants West-Life [grant no. 675858] and BioExcel [grant no. 675728], by the Dutch Foundation for Scientific Research (NWO) [TOP-PUNT grant 718.015.001] and the ASDI eScience grant [027016G04]. LX acknowledges financial support from the Netherlands Organization for Scientific Research [Veni grant 722.014.005]. C.G. acknowledges financial support from the China Scholarship Council [Grant No. 201406220132]. AV was supported by the Marie Skłodowska-Curie Individual Fellowship H2020 MSCA-IF-2015 [BAP-659025].

Conflict of Interest: none declared.

References

Basse,M.J. *et al*. (2016) 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. *Database (Oxford)*, **2016**, baw007.

Brylinski,M. (2013) Nonlinear scoring functions for similarity-based ligand docking and binding affinity prediction. *J. Chem. Inf. Model*, **53**, 3097–3112.

Gaieb,Z. *et al*. (2018) D3R Grand Challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des*., **32**, 1–20.

Hongjian Li,K.-S.L. and Man-Hon,W. (2012) *idock: A Multithreaded Virtual Screening Tool for Flexible Ligand Docking*. IEEE.

Jain,T. and Jayaram,B. (2005) An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. *FEBS Lett*., **579**, 6659–6666.

Jimenez,J. *et al*. (2018) KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J. Chem. Inf. Model*, **58**, 287–296.

Kastritis,P.L. *et al*. (2011) A structure-based benchmark for protein-protein binding affinity. *Protein Sci*., **20**, 482–491.

Kastritis,P.L. *et al*. (2014) Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. *J. Mol. Biol*., **426**, 2632–2652.

Kurkcuoglu,Z. *et al*. (2018) Performance of HADDOCK and a simple contact-based protein-ligand binding affinity predictor in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des*., **32**, 175–185.

Labbe,C.M. *et al*. (2017) AMMOS2: a web server for protein-ligand-water complexes refinement via molecular mechanics. *Nucleic Acids Res*., **43**, W448–W454.

Labbe,C.M. *et al*. (2015) MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Res*, **43**, W448–W454.

Liu,Z. *et al*. (2017) Forging the basis for developing protein-ligand interaction scoring functions. *Acc. Chem. Res*., **50**, 302–309.

Pires,D.E. and Ascher,D.B. (2016) CSM-lig: a web server for assessing and comparing protein-small molecule affinities. *Nucleic Acids Res*., **44**, W557–W561.

van Zundert,G.C. and Bonvin,A.M. (2014) Modeling protein-protein complexes using the HADDOCK webserver "modeling protein complexes with HADDOCK". *Methods Mol. Biol*., **1137**, 163–179.

Vangone,A. and Bonvin,A.M. (2015) Contacts-based prediction of binding affinity in protein-protein complexes. *Elife*, **4**.

Vangone,A. and Bonvin,A. (2017) PRODIGY: a Contact-based Predictor of Binding Affinity in Protein-protein Complexes. *Bio-Protocols*, **7**.

Vreven,T. *et al*. (2015) Updates to the integrated protein-protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *J. Mol. Biol*., **427**, 3031–3041.

Wang,J.C. *et al*. (2012) idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res*., **40**, W393–W399.

Xue,L.C. *et al*. (2016) PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*, **32**, 3676–3678.