This is the post-print version of the following article published in Bioinformatics:

- L.C. Xue, J.P.G.L.M. Rodrigues, P.L. Kastritis, **A.M.J.J. Bonvin**# and A. Vangone#.PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*, 32, 3676-3678 (2016).

  DOI: https://doi.org/10.1093/bioinformatics/btw514

Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Structural bioinformatics

PRODIGY: a web server for predicting the binding affinity of protein-protein complexes

Li C Xue[1], João PGLM Rodrigues[1A], Panagiotis L Kastritis[1B], Alexandre MJJ Bonvin[1*], Anna Vangone[1*]

[1]Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Faculty of Science – Chemistry, Utrecht University, Utrecht, The Netherlands.

[AB]Current address: A: Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA; B: European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, Germany

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

ABSTRACT

Summary: Gaining insights into the structural determinants of protein-protein interactions holds the key for a deeper understanding of biological functions, diseases and development of therapeutics. An important aspect of this is the ability to accurately predict the binding strength for a given protein-protein complex. Here we present PRODIGY, a web server to predict the binding affinity of protein-protein complexes from their three-dimensional structure. The PRODIGY server implements our simple but highly effective predictive model based on intermolecular contacts and properties derived from non-interface surface.

Availability: PRODIGY is freely available at: http://milou.science.uu.nl/services/PRODIGY.

Contact: a.vangone@uu.nl, a.m.j.j.bonvin@uu.nl

1 Introduction

Biomolecular interactions between proteins are involved in regulation and control of almost every biological process in the cell. Alterations in such interactions are responsible for many diseases, making protein-protein complexes crucial targets for therapeutics development (Petta, Lievens, Libert, Tavernier, & De Bosscher, 2015). In this scenario, identifying the structural determinants of these interactions and their binding energetics is an important step for a better understanding and controlling of such systems. In particular, the binding affinity (or binding free energy), which defines whether complex formation occurs or not in specific conditions, holds the key to control interactions (e.g. engineering high affinity interactions), design new therapeutics (e.g. guiding rational drug design) or predict the impact of mutations at protein interfaces. The prediction of binding affinity has been investigated for decades (Chothia & Janin, 1975; Horton & Lewis, 1992) yielding approaches ranging from exact methods (e.g. free energy perturbation), which are accurate but computationally costly, to empirical approaches (e.g. scoring functions in docking, various regression models), which are fast but less accurate (Kastritis & Bonvin, 2010). Several valuable web servers have been made available to the scientific community, providing a series of different descriptors (energetics, structural features, etc.) of protein-protein interfaces (Moal, Jiménez-García, & Fernández-Recio, 2015; Reynolds, Damerell, & Jones, 2009; Saha, Bahadur, Pal, Mandal, & Chakrabarti, 2006; Tina, Bhadra, & Srinivasan, 2007; Tuncbag, Kar, Keskin, Gursoy, & Nussinov, 2009; Vangone, Spinelli, Scarano, Cavallo, & Oliva, 2011).

Some of these have also been tested as binding affinity predictors. There is, however, a lack of specific online tools for the prediction of binding affinity (Su, Zhou, Xia, Li, & Sun, 2009).

Recently, we introduced a simple and robust descriptor of binding affinity based only on structural properties of a protein-protein complex. Using the protein-protein binding affinity benchmark in Kastritis *et al.*, (2011), we demonstrated that the number of interfacial contacts at the interface of a protein-protein complex correlates with its experimental binding affinity. This information, combined with properties of the non-interacting surface (Kastritis, Rodrigues, Folkers, Boelens, & Bonvin, 2014; Marillet, Boudinot, & Cazals, 2016), has led to one of the best performing predictor reported so far. In terms of accuracy, our method showed a Pearson's Correlation coefficient of 0.73 between the predicted and measured binding affinity on the benchmark (p-value < 0.0001) and Root Mean Square Error (RMSE) of 1.89 kcal mol^{-1} (Vangone & Bonvin, 2015). While our method performs well on average, errors in particular cases may be expected; for instance, some natural ultra high-affinity complexes have average or below-average buried surface area, and PRODIGY may underestimate their affinity. Alternative physical models of entropy, solvent effects, and electrostatics could be taken into consideration to address such cases, although, to date, no such model performs better on average than our simple contact-based approach (see Figure 4 in Vangone & Bonvin, 2015).

We have implemented our contact-based method as a web server, PRODIGY (PROtein binDIng enerGY prediction), a user-friendly online tool for the prediction of binding affinity in protein-protein complexes.

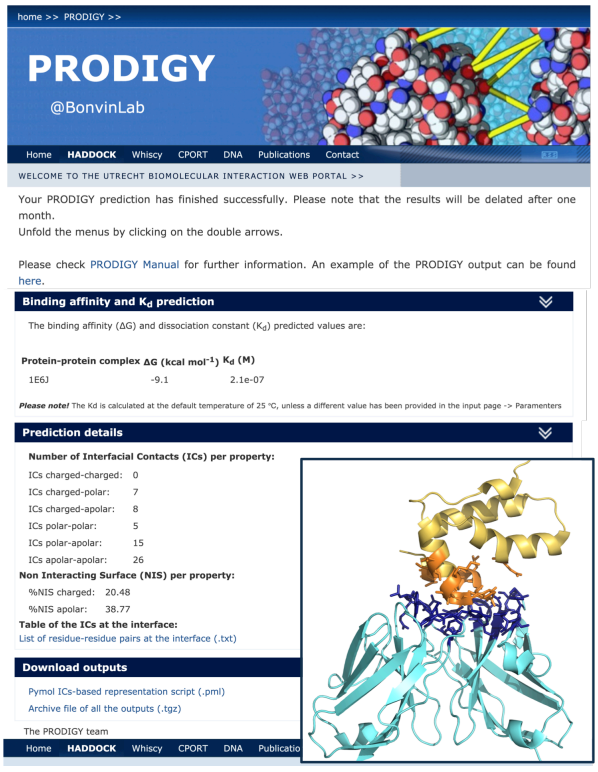

Fig 1. Example output of PRODIGY for the complex between the FAB and the HIV-1 capsid protein p24 (PDB code: 1E6J). A three-dimensional representation of the complex interface is shown in the inset figure with the color coding of the PRODIGY script (.pml) for Pymol: aquamarine and yellow for Interactor 1 (chains L and H in this example) and Interactor 2 (chain P), respectively. The interacting residues are represented in blue and orange for Interactor 1 and Interacting 2, respectively, and side-chains showed in sticks.

2 The web server

The PRODIGY server requires as input the 3D structure of a protein-protein complex, which can be provided in three different manners:

- upload of the 3D structure in PDB or mmCIF format;
- automatic download from the protein databank;
- upload as an archive file (.tar, .tgz, .zip, .bz2, .tar.gz) for analyzing multiple structures at the same time (with a limit of 50MB).

The user is required to specify the chain identifiers for the molecules involved in the interaction. It is also possible to specify the temperature, at which once can calculate the dissociation constant (25 °C by default) and an email address, where a link to the results page will be sent. When an ensemble of models of a NMR-determined complex is submitted as a single input PDB file, only the first model will be used for the prediction. The results (downloadable for two weeks) include:

1. the predicted value of the binding free energy (ΔG) in kcal mol^{-1};
2. the predicted value of the dissociation constant (K_d) in M calculated from $\Delta G = RT \ln(K_d)$ where R is the idea gas constant (kcal K^{-1} mol^{-1}) , T the temperature (K).
3. the number and type of intermolecular contacts within the 5.5 Å distance cutoff (for details see Vangone & Bonvin, 2015));
4. the percentages of charged and polar amino-acids on the non-interacting surface;
5. a downloadable table (.txt) of all residues occurring at the interface and a ready-to-run Pymol script (.pml) (www.pymol.org);
6. A compressed file with all the result files.

Information about the predictive model and the training dataset can be found online in the "Method" and "Dataset" page of PRODIGY, respectively, accessible through the main page. RODIGY has been written in Python and Perl. The solvent accessible surface area is calculated with open-source tool freeSASA (Mitternacht, 2015) using the default NACCESS (Hubbard & Thornton, 1993) parameters for atomic radii. The server is fast, performing the prediction in few seconds for the largest complex examined in the benchmark (1DE4). An example output page of PRODIGY is shown in Figure 1.

In conclusion, the PRODIGY server should contribute to speeding-up development of new predictive approaches and facilitate its use within various fields of biology. PRODIGY is freely accessible at http://milou.science.uu.nl/services/PRODIGY. A standalone version to run locally is freely available from our GitHub repository (see http://www.bonvinlab.org/software).

Funding

AV was supported by a Marie Skłodowska-Curie Individual Fellowship H2020 MSCA-IF-2015 [BAP-659025]. LCX was supported by the Netherlands Organisation for Scientific Research [Veni grant 722.014.005].

Conflict of Interest: none declared.

References

Chothia, C., & Janin, J. (1975). Principles of protein-protein recognition. *Nature*.
Horton, N., & Lewis, M. (1992). Calculation of the free energy of association for protein complexes. *Protein Science*, *1*(1), 169–181.
Hubbard, S. J., & Thornton, J. M. (1993). Naccess. Computer Program.
Kastritis, P. L., & Bonvin, A. M. J. J. (2010). Are Scoring Functions in Protein−Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. *Journal of Proteome Research*, *9*(5), 2216–

2225.
Kastritis, P. L., Moal, I. H., Hwang, H., Weng, Z., Bates, P. A., Bonvin, A. M. J. J., & Janin, J. (2011). A structure-based benchmark for protein–protein binding affinity. *Protein Science*, *20*(3), 482–491.
Kastritis, P. L., Rodrigues, J. P. G. L. M., Folkers, G. E., Boelens, R., & Bonvin, A. M. J. J. (2014). Proteins Feel More Than They See: Fine-Tuning of Binding Affinity by Properties of the Non-Interacting Surface. *Journal of Molecular Biology*, *426*(14), 2632–2652.
Marillet, S., Boudinot, P., & Cazals, F. (2016). High-resolution crystal structures leverage protein binding affinity predictions. *Proteins: Structure, Function, and Bioinformatics*, *84*(1), 9–20.
Mitternacht, S. (2015). FreeSASA: A Free C Library for Solvent Accessible Surface Area Calculations.
Moal, I. H., Jiménez-García, B., & Fernández-Recio, J. (2015). CCharPPI web server: computational characterization of protein–protein interactions from structure. *Bioinformatics*, *31*(1), 123–125.
Petta, I., Lievens, S., Libert, C., Tavernier, J., & De Bosscher, K. (2015). Modulation of Protein|[ndash]|Protein Interactions for the Development of Novel Therapeutics. *Molecular Therapy*.
Reynolds, C., Damerell, D., & Jones, S. (2009). ProtorP: a protein–protein interaction analysis server. *Bioinformatics*, *25*(3), 413–414.
Saha, R. P., Bahadur, R. P., Pal, A., Mandal, S., & Chakrabarti, P. (2006). ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Structural Biology*, *6*(1), 1.
Su, Y., Zhou, A., Xia, X., Li, W., & Sun, Z. (2009). Quantitative prediction of protein–protein binding affinity with a potential of mean force considering volume correction. *Protein Science*, *18*(12), 2550–2558.
Tina, K. G., Bhadra, R., & Srinivasan, N. (2007). PIC: Protein Interactions Calculator. *Nucleic Acids Research*, *35*(suppl 2), W473–W476.
Tuncbag, N., Kar, G., Keskin, O., Gursoy, A., & Nussinov, R. (2009). A survey of available tools and web servers for analysis of protein–protein interactions and interfaces. *Briefings in Bioinformatics*, *10*(3), 217–232.
Vangone, A., & Bonvin, A. (2015). Contacts-based prediction of binding affinity in protein-protein complexes. *eLife*, *4*, 291.
Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., & Oliva, R. (2011). COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, *27*(20), 2915–2916.