Structural bioinformatics

EasyVS: a user friendly web based tool for molecule library selection and structure-based virtual screening

Douglas E. V. Pires[1,2,3,#,*], Wandré N. P. Veloso[4,#], YooChan Myung[2,3,#], Carlos H. M. Rodrigues[2,3,#], Michael Silk[2,3,#], Pâmela M. Rezende[5,#], Francislon Silva[5], Joicymara S. Xavier[5,6], João P. L. Velloso[5], Carlos H. da Silveira[4], David B. Ascher[2,3,7,#,*]

[1]School of Computing and Information Systems, University of Melbourne, 3010, Australia;

[2]Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, 3004, Australia;

[3]Department of Biochemistry and Molecular Biology, Bio21 Institute, University of Melbourne, 3010, Australia;

[4]Institute of Technological Sciences, Universidade Federal de Itajubá, Itabira, 35903-087, Brazil;

[5]Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, 30190-002, Brazil;

[6]Instituto de Ciências Agrárias, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Unaí, 38610-000 ,Brazil

[7]Department of Biochemistry, University of Cambridge, Cambridge, CB2 1GA, United Kingdom.

[#]These authors contributed equally to this paper.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: EasyVS is a web-based platform built to simplify molecule library selection and virtual screening. With an intuitive interface, the tool allows users to go from selecting a protein target with a known structure and tailoring a purchasable molecule library to performing and visualising docking in a few clicks. Our system also allows users to filter screening libraries based on molecule properties, cluster molecules by similarity and personalise docking parameters.

Availability and implementation: EasyVS is freely available as an easy-to-use web interface at http://biosig.unimelb.edu.au/easyvs

Contact: douglas.pires@unimelb.edu.au, david.ascher@unimelb.edu.au.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Structure-based virtual screening has been widely and successfully used in early stages of drug development, aiding in the identification of potential hits and guiding further experimental validation (Cheng *et al.*, 2012). Molecular docking is one of the most widely used virtual screening approaches, which uses the three-dimensional structure of a target protein to predict the predominant binding mode of a small molecule with the target of interest. In this way, docking can be used to evaluate a large library of molecules and identify those most likely to interact with the target in the desired manner. This has been a powerful tool in the

identification of initial hits, significantly reducing the chemical space to experimentally test, and increasing the proportion of positive compounds being screened.

Significant improvements in docking protocols (Di Muzio *et al.*, 2017), scoring functions (Pires *et al.*, 2016), and molecule libraries (Sterling *et al.*, 2015), and the greater availability of computational power, has made virtual screening a more tractable and reliable hit identification strategy. Despite this, current virtual screening approaches typically require specialist computational and technical expertise.

In order to make virtual screening more friendly and accessible to a wider audience, we propose EasyVS (http://biosig.unimelb.edu.au/easyvs), a web-based, efficient and intuitive system that allows users to go from defining a protein structure and molecule library to performing docking in a few clicks. Our system allows users to optimize their screening library based on their properties (and define a chemical space of interest). Through a molecule clustering approach, we can more rapidly screen a larger chemical space, and present the top solutions to the user.

2 Platform description

Initially users are asked to define their target of interest by either uploading a structure of interest, or using the biological assembly of a previous experimental structure deposited in the RCSB by providing the PDB accession code.

The selected structure is then analyzed using Ghecom (Kawabata, 2010) to identify druggable pockets. While by default the largest pocket is chosen for docking, users may select another pocket of interest for screening, and can refine the boundaries and docking parameters used (Supplementary Figure S1). These parameters include box size and position (which can be set to any of the identified pockets and finely adjusted by the user) and depth of the search.

In the next step, users are asked to define the set of molecules to be assessed via docking. EasyVS currently supports the small molecule databases ChEMBL 24_1 (Gaulton *et al.*, 2017), HMDB 4.0 (Wishart *et al.,* 2018), Drugbank 5.0 (Wishart *et al.*, 2018), Maybridge[1], Super Natural II (Banerjee *et al.*, 2015), Chembridge (Desai *et al.*, 2004) and Zinc15 (Sterling *et al.*, 2015), which together comprise over 16 million molecules. There are many filters available for refining these molecular libraries, including by molecular weight, number of acceptors or donors of hydrogen, *logP* (or only selecting Lipinski's Ro5 molecules), fragments or natural products (Fig. 1A). Once the molecule library has been selected, users have the option to cluster molecules by similarity to improve screening performance. If users opt to perform clustering, one representative molecule from each group is randomly selected for docking. Users have the option to select the level of similarity used during clustering, which will change the number of clusters and, therefore, the number of molecules that will proceed for docking stages.

Molecule docking of the selected compound library is then performed using Autodock Vina (Trott *et al.*, 2010). Users can also rapidly rescore selected poses using NNScore (Durrant, *et al.*, 2011) or CSM-Lig (Pires and Ascher, 2016), analyse the intramolecular interactions using Arpeggio (Jubb *et al.*, 2016) and predict pharmacokinetic properties of top hits using pkCSM (Pires *et al.*, 2015). While the docking is running, users can view the results in real time to analyze best poses of selected molecules (Fig. 1B) as well as include additional molecules for docking. Further exploration of the chemical space of top docked ligands is available by an additional round of virtual screening, using compounds that are structurally similar to the ligand of interest.

The EasyVS docking protocol was validated using two different benchmarks. We performed a redocking procedure for a selection of eight GPCR-ligands complexes with available crystallographic structures. Ligands have been successfully redocked with an average RMSD of 0.98 Å (Supplementary Table S1). We have also created decoy libraries using DUD-e (Mysinger et al. 2012) for the same set of proteins considered for redocking. The docking scores for real ligands were considerably higher than those obtained for the decoys (p-value < 0.001, Supplementary Table S2), demonstrating the robustness of the EasyVS docking protocol. We have evaluated the system's ability to process multiple submissions, demonstrating the system's responsiveness (Supplementary Fig. S2).

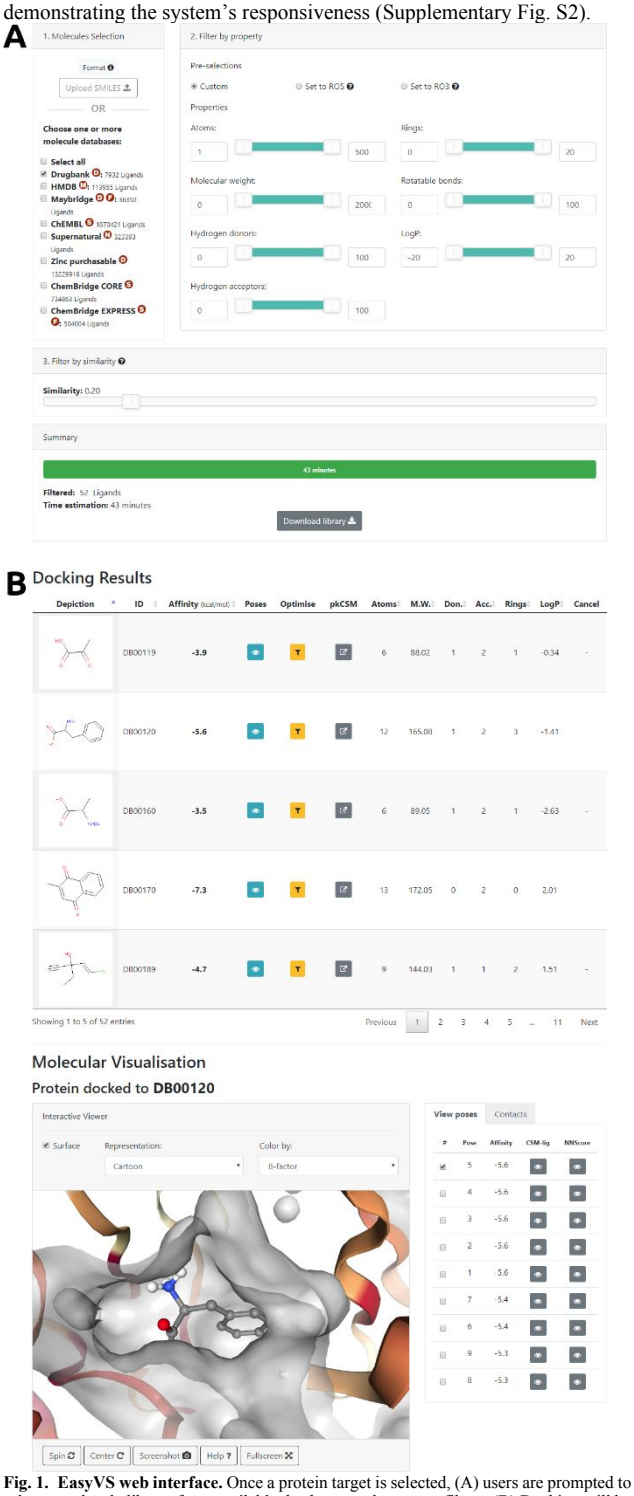

Fig. 1. EasyVS web interface. Once a protein target is selected, (A) users are prompted to select a molecule library from available databases and property filters. (B) Docking will be performed on the selected target/library set and best poses shown as an interactive molecule visualization.

3 Conclusions

Here we present EasyVS, a freely available, user-friendly platform for simplifying molecule library construction and docking. EasyVS allows users to choose molecules from well-established and diverse databases, including fragments, approved drugs and natural products, and perform

[1] https://www.maybridge.com

the docking with just a few clicks. We also show EasyVS was successful in identifying GPCR ligands (Supplementary Materials) as a case study. We believe this will be an invaluable tool for the exploratory stages of hit identification, allowing for the selection either stringent or very diverse sets of molecules for virtual screening and the intelligent assessment of different small molecule chemical spaces.

Funding

This work has been supported by Melbourne Research Scholarships [to C.H.M.R and Y.M.]; Medical Research Council (MRC) [MR/M026302/1 to D.B.A., D.E.V.P.]; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001 [to C.H.S.]; National Health and Medical Research Council of Australia [APP1072476 to D.B.A.]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to D.E.V.P. and C.H.S.]; Universidade Federal de Itajubá [to W.N.P.V. and C.H.S.]. Supported in part by the Victorian Government's OIS Program.

Conflict of Interest: none declared.

References

Banerjee, P., Erehman, J., Gohlke, B. O., Wilhelm, T., Preissner, R., & Dunkel, M. (2015). Super Natural II—a database of natural products. *Nucleic Acids Res.*, **43**(D1), D935-D939.

Cheng, T., Li, Q., Zhou, Z., Wang, Y., & Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.*, **14**(1), 133-141.

Desai, P. V., Patny, A., Sabnis, Y., Tekwani, B., Gut, J., Rosenthal, P., ... & Avery, M. (2004). Identification of novel parasitic cysteine protease inhibitors using virtual screening. The ChemBridge database. *J. Med. Chem.*, **47**(26), 6609-6615.

Di Muzio, E., Toti, D., & Polticelli, F. (2017). DockingApp: a user friendly interface for facilitated docking simulations with AutoDock Vina. *J. Comput. Aided Mol. Des.*, **31**(2), 213-218.

Durrant, J. D., & McCammon, J. A. (2011). NNScore 2.0: a neural-network receptor–ligand scoring function. *J. Chem. Inf. Model.*, **51**(11), 2897-2903.

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., *et al*. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.*, **45**(D1), D945-D954.

Jubb, H. C., Higueruelo, A. P., Ochoa-Montano, B., Pitt, W. R., Ascher, D. B., Blundell, T. L. (2017). Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol*., **429**(3), 365-371.

Kawabata, T. (2010). Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins: Struct., Funct., Bioinf.*, **78**(5), 1195-1211.

Mysinger, M. M., Carchia, M., Irwin, J. J., Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem.*, **55**(14), 6582-6594.

Pires, D. E., Blundell, T. L., Ascher, D. B. (2015). pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J Med Chem*., **58**(9), 4066-72.

Pires, D. E. V., & Ascher, D. B. (2016). CSM-lig: a web server for assessing and comparing protein–small molecule affinities. *Nucleic Acids Res.*, **44**(W1), W557-W561.

Sterling, T., & Irwin, J. J. (2015). ZINC 15–ligand discovery for everyone. *J. Chem. Inf. Model.*, **55**(11), 2324-2337.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry, **31**(2), 455-461.

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., *et al*. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, **46**(D1), D608-D617.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... & Assempour, N. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**(D1), D1074-D1082.