

Dark Suite: a comprehensive toolbox for computer-aided drug design

Eleni Papakonstantinou^{1,2}, Vasileios Megalooikonomou³, Dimitrios Vlachakis^{1,2,4}✉

¹Laboratory of Genetics, Department of Biotechnology, School of Food, Biotechnology and Development, Agricultural University of Athens, Athens, Greece

²Lab of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

³Computer Engineering and Informatics Department, School of Engineering, University of Patras, Patras, Greece

⁴Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, Strand, London, United Kingdom
Competing interests: EP none; FB none; VM none; AE none; DV none

Abstract

Dark Suite is a complete interactive software pipeline aiming to exploit the advantages of free software and modern programming. Besides two commands (installation and opening) on the command line, the handling and the program operation is done through a user's friendly interface. This platform has a central graphical interface which allows the user to choose in what computational tool to work. Each computational tool has an interface. Dark Suite combines the functions of other programmes to create a pipeline for high-quality secondary effects through a friendly user interface. It is made to run on GNU / Linux distributions and its interface was built using JAVA to seamlessly integrate scientific tools written in Perl, Java, R and Python.

Introduction

The discovery and development of a new drug is still a time and cost consuming process, estimating for approximately 10-15 years for a new drug to enter the market (Paul *et al.*, 2010; Pan *et al.*, 2013). In pharmaceutical research, computational methods are being employed to reduce the cost and time of drug development, while enabling the synthesis of large component drug libraries (Clark *et al.*, 2010; Szymański *et al.*, 2012). Computer-aided drug design (CADD) is based on the combination of computational techniques that can enable the simulation of molecular interactions between proteins and target molecules and predict the effectiveness of a new lead molecule (Veselovsky *et al.*, 2014). Structure-based and ligand-based drug design are the primary methodologies for drug discovery and lead optimisation; depending on the available biological information. Ligand-based approaches generate SAR models based on known active and/or inactive molecules, whereas structure-based approaches use the structural information of the protein target to discover lead molecules as potent inhibitors. Molecular Dynamics, Quantum Mechanics and Linear Interaction

Energy (LIE) are included in the drug discovery process optimisation to evaluate *in silico* the effectiveness of the lead molecule.

CADD approaches have led to many successful applications, through lead discovery or drug repurposing. However, conventional CADD has its limitations and their results have to be validated in real biological systems, as numerous molecules that have been recognised *in silico*, do not eventually exhibit the predicted activity and efficiency. Apart from the complexity of the biological systems and the difficulty in interpreting and simulating them *in silico*, an additional limitation of CADD applications is that the tools used to discover and design a new drug are based on specific algorithms and each tool has its limitations. For that purpose, it is considered necessary to continually update the tools and algorithms to improve the accuracy and the provision of new drugs. The employment of CADD applications is an integral part of pharmaceutical research and novel approaches need to be implemented (Baig *et al.*, 2018).

Herein, we describe a novel solution for computer-aided drug design through the Dark Suite application.

Article history

Received: 28 December 2019

Accepted: 31 December 2019

Published: 11 March 2020

© 2020 Papakonstantinou *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at <http://journal.embnet.org>.

The suite is designed to facilitate the drug discovery process by introducing freely available tools that have been shown to outperform conventional approaches. In the first step of the structure-based drug design and the discovery of homologous proteins, an enhanced sequence similarity search can be performed by exploiting secondary structural elements and the hydropathy profiles of the proteins. Homology modelling can then be achieved using additional restraints based on the shape and size similarity of the homologous proteins, leading to a refined 3D model structure. In the next step, the lead discovery is approached by a fully integrated platform set to automate the drug design through efficient algorithms for protein preparation, energy minimisation, pharmacophore elucidation, building and growing new drug-like moieties, and molecular dynamics simulations. Additionally, the suite includes tools for protein clustering based on gene ontology annotations, and for enhanced prediction of protein interactions based on protein-protein interaction networks.

Description of the program

The Dark Suite is a complete interactive platform that integrates free software and modern programming solutions for a high-quality research pipeline in computational drug design in the drug discovery research. It is a UNIX based software, compatible with all major GNU/LINUX distributions and its interface was built using Java language programming. The integrated tools are written in Perl, Java and bash. It is easily installed through the command line and can be downloaded freely at www.darkdna.gr.

The platform includes 2D, 3D and 4D protein analysis tools, an automated drug design platform and a 3D structure viewer (Figure 1).

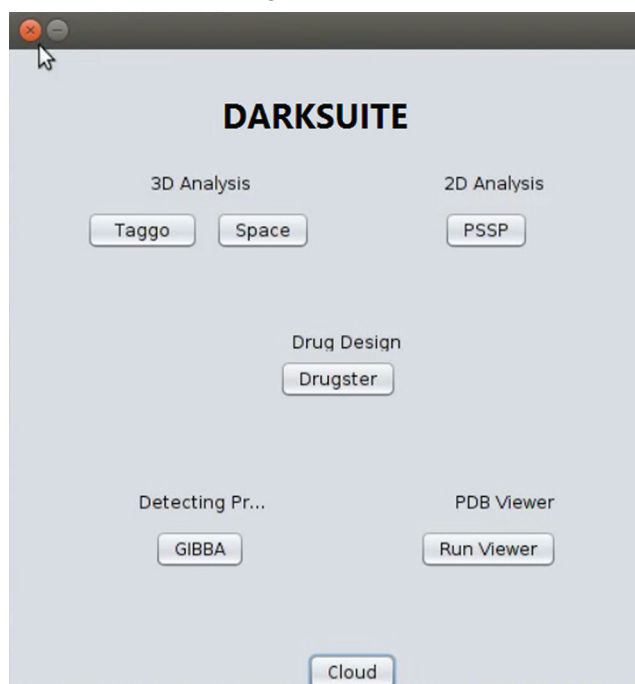


Figure 1. The front-end window of Dark Suite.

Namely, the Dark Suite integrates the PSSP tool (Figure 2), TAGGO, Space, GIBA, Drugster and Jmol that are described below. Each of the tools can be assessed through its autonomous interface after installing and running the suite. Noteworthy, the suite enables the user to run all biological tools in parallel. In such a manner, the drug design pipeline can be adjusted to the user's needs.

Integrated tools

1. 2D analysis

The protein secondary structure profile (**PSSP**) tool is developed to perform fast, efficient similarity searches based not only on the protein sequences but also on the secondary structure profile of the proteins (Vlachakis et al., 2017). PSSP first performs a conventional blast search for the query protein sequence and then uses a custom made hydropathy substitution matrix by which the entries are re-scored and re-ranked. The hydropathy profile is constructed for each protein using the hydropathy index from IMGT (Lefranc *et al.*, 1999). Moreover, PSSP exploits the secondary structural information by searching against the RCSB PDB secondary elements database and is also able to predict secondary elements if the query protein is not crystallographically determined.

2. 3D analysis

TAGGO module is designed for an automated clustering of a set of proteins based on Gene Ontology (GO) resources (Roubelakis *et al.*, 2009). The module has a 5-step interface, in which the user has to select: the path for the input file that includes a list of proteins to be annotated, the appropriate GO file and its respective format, the organism for which the clustering is performed, the Evidence Codes to be considered, and the path to the output directory. TAGGO then evaluates the Information Content (IC) of each ontology term assigned to each protein of the input set for all three GO aspects; Molecular Function (MF), Cellular Component (CC) and Biological Process (BP). The most general terms (with lower information content) are considered and the proteins in the dataset are assigned to categories. The overall output of the process is the percentage of the annotated GO categories of the protein dataset, for each GO aspect. A directory called "Results" is created that includes Venn lists for each aspect, a visualisation of the results in pie and bar charts, and text files describing the parameters and annotation steps throughout the process.

Space represents a novel, efficient methodology for homology modelling that improves the quality and credibility of the resulting model compared to conventional approaches (Vlachakis *et al.*, 2013). Space can be implemented in cases of low sequence identity by considering the proteins' shape and size similarity. Based on the fact that structure is more conserved than the sequence in nature, Space performs 3D modelling

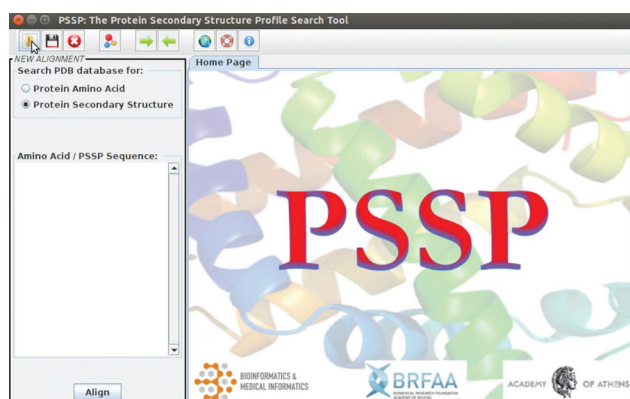


Figure 2. The interface of the PSSP tool which can be run autonomously.

by setting additional constraints determined by the conformational shape of the template protein.

3. 4D analysis

GIBA is an effective and user-friendly tool for the identification of accurate protein complexes through clustering of protein-protein interaction (PPI) networks (Moschopoulos *et al.*, 2009). GIBA encompasses two different clustering algorithms, the MCL (Enright *et al.*, 2002) and the RNSC algorithm (King *et al.*, 2004) and a set of different user-defined parameters for clustering and filtering. Once the workflow is executed, the output is a set of the final clusters with the interacting proteins predicted. GIBA surpasses other methods in quality approximations of protein complexes.

4. Drug Design

Drugster is a freeware platform aimed to assist scientists in the field of computer-aided drug design (Vlachakis *et al.*, 2013). Drugster integrates the algorithms of PDB2PQR v.1.8 (Dolinsky *et al.*, 2004; Dolinsky *et al.*, 2007), Ligbuilder v.1.2 and v.2.0 (Yuan *et al.*, 2011), Gromacs v.4.5.5 (Hess *et al.*, 2008) and Dock v.6.5 (Lang *et al.*, 2009). It is designed to automate the process of structure-based drug design and lead optimization through an easy to use interface. The complete workflow consists of five steps:

1. Input preparation; the 3D structure of a PDB file is refined and problems are automatically fixed;
2. energy minimization; the receptor is optimized using the Gromacs suite;
3. *de novo* structure-based drug design; the Ligbuilder module is used for ligand building after a pharmacophore preparation and the determination of a 3D scaffold to generate novel moieties;
4. ligand optimization; all the candidates are docked to the receptor using the Dock module and ranked after energy minimization, and
5. complex optimization; the ligand-receptor complex is energetically minimized and the system undergoes molecular dynamics simulations.

5. PDB viewer

The structures can be visualised through the PDB viewer that uses **Jmol**¹ (Jmol: an open-source Java viewer for chemical structures in 3D), a free, open-source molecule viewer.

Platform compatibility:

Dark Suite has been tested on the following GNU / Linux distributions:

- Ubuntu 14.04.4 LTS (32&64 bit);
- Ubuntu 15.04 LTS (32&64 bit);
- Ubuntu 16.04 LTS (32&64 bit);
- Kubuntu 14.04.4 LTS (32&64 bit);
- Debian 8.0 (32&64 bit).

Dark Suite Installation steps:

Open a terminal and initially go in the Dark Suite directory and then in the medical directory located inside the Dark Suite that contains all files.

e.g. `cd DarkSuite/medical/`

Within the medical directory, you can locate the installation file.



Run the command: `<./install.sh>`

Installation process:

Run `./install.sh` and your password will be prompted. Press “enter” to begin the installation. You will be asked again to press “enter” after a bit to continue. Installation continues normally for some time and all necessary tools and libraries are being installed. Drugster application will be installed through a dedicated interface, after following instructions. You will be asked to choose the Linux distribution you are using. Choose and press “enter”, and you will be prompted to the graphical installation environment of Drugster. Follow the instructions until completion.

Dark Suite Operation steps:

1. Run the command: `<java -jar darksuite.jar>` and Dark Suite will load;
2. press the button TAGGO to execute TAGGO;
3. press the button SPACE and choose the pdb file you want; a separate window is used to display information about file which was selected;
4. press the button PSSP to execute PSSP;
5. press the button Drugster to execute Drugster;
6. press the button GIBBA to execute GIBBA;
7. press the button Run Viewer to open a window to insert the pdb file you want to view; Then after selecting and pressing “ok”, you will see the 3D protein structure hat selected;
8. press the button Cloud to create a new directory in the directory with the executable file on a cloud file location.

¹<http://www.jmol.org/>

Key Points

- An innovative solution for computer-aided drug design to facilitate and refine the drug discovery process.
- Quickly installed, fully interactive platform integrating state of the art protein analysis tools.
- Efficient pipeline and novel approaches for structure-based drug design.
- Dark Suite is compatible with all major GNU / Linux distributions and is freely downloadable at darkdna.gr

Conclusions

Dark Suite is a stand-alone application for computer-aided drug design performing all fundamental steps for the process of lead discovery and optimisation, in a user-friendly environment. The Dark Suite pipeline introduces novel and efficient methods for surpassing the limitations on homologous protein discovery based on the primary structural information, traditional homology modelling and drug design approaches. Moreover, it encompasses tools for refined gene ontology annotations and predicted protein-protein interactions. Dark Suite is freely available for the scientific community and allows for a user-defined workflow by selecting the appropriate tools for use.

Acknowledgements

DV would like to acknowledge funding from: i. Microsoft Azure for Genomics Research Grant (CRM:0740983) ii. FrailSafe Project (H2020-PHC-21-2015 - 690140) "Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions", co-funded by the European Commission under the Horizon 2020 research and innovation program. iii. Amazon Web Services Cloud for Genomics Research Grant (309211522729). iv. AdjustEBOVGP-Dx (RIA2018EF-2081): Biochemical Adjustments of native EBOV Glycoprotein in Patient Sample to Unmask target Epitopes for Rapid Diagnostic Testing. A European & Developing Countries Clinical Trials Partnership (EDCTP2) under the Horizon 2020 "Research and Innovation Actions" DESCA. EP would like to acknowledge funding by the State Scholarships Foundation (IKY) - European Union (European Social Fund-ESF) and Greek national funds through the action entitled "Strengthening Human Resources Research Potential via Doctorate Research" in the framework of the Operational Program "Human Resources Development Program, Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) 2014 – 2020.

References

1. Baig, M. H., K. Ahmad, G. Rabbani, M. Danishuddin and I. Choi (2018). "Computer Aided Drug Design and its Application to the Development of Potential Drugs for Neurodegenerative Disorders." *Curr Neuropharmacol* **16**(6),740-748. <http://dx.doi.org/10.2174/1570159X15666171016163510>
2. Clark, R. L., B. F. Johnston, S. P. Mackay, C. J. Breslin, M. N. Robertson and A. L. Harvey (2010). "The Drug Discovery Portal: a resource to enhance drug discovery from academia." *Drug Discov Today* **15**(15-16), 679-683. <http://dx.doi.org/10.1016/j.drudis.2010.06.003>
3. Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker (2007). "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations." *Nucleic Acids Res* **35**(Web Server issue): W522-525. <http://dx.doi.org/10.1093/nar/gkm276>
4. Dolinsky, T. J., J. E. Nielsen, J. A. McCammon and N. A. Baker (2004). "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations." *Nucleic Acids Res* **32**(Web Server issue): W665-667. doi: <http://dx.doi.org/10.1093/nar/gkh381>
5. Enright, A. J., S. Van Dongen and C. A. Ouzounis (2002). "An efficient algorithm for large-scale detection of protein families." *Nucleic Acids Res* **30**(7), 1575-1584. <http://dx.doi.org/10.1093/nar/30.7.1575>
6. Hess, B., C. Kutzner, D. van der Spoel and E. Lindahl (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." *J Chem Theory Comput* **4**(3): 435-447. <http://dx.doi.org/10.1021/ct700301q>
7. King, A. D., N. Przulj and I. Jurisica (2004). "Protein complex prediction via cost-based clustering." *Bioinformatics* **20**(17), 3013-3020. <http://dx.doi.org/10.1093/bioinformatics/bth351>
8. Lang, P. T., S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James and I. D. Kuntz (2009). "DOCK 6: combining techniques to model RNA-small molecule complexes." *RNA* **15**(6), 1219-1230. <http://dx.doi.org/10.1261/rna.1563609>
9. Lefranc, M. P., V. Giudicelli, C. Ginestoux, J. Bodmer, W. Muller, R. Bontrop, M. Lemaître, A. Malik, V. Barbie and D. Chaume (1999). "IMGT, the international ImMunoGeneTics database." *Nucleic Acids Res* **27**(1), 209-212. <http://dx.doi.org/10.1093/nar/27.1.209>
10. Moschopoulos, C. N., G. A. Pavlopoulos, R. Schneider, S. D. Likothanassis and S. Kossida (2009). "GIBA: a clustering tool for detecting protein complexes." *BMC Bioinformatics* **10** Suppl 6: S11. <http://dx.doi.org/10.1186/1471-2105-10-S6-S11>
11. Pan, S. Y., S. F. Zhou, S. H. Gao, Z. L. Yu, S. F. Zhang, M. K. Tang, J. N. Sun, D. L. Ma, Y. F. Han, W. F. Fong and K. M. Ko (2013). "New Perspectives on How to Discover Drugs from Herbal Medicines: CAM's Outstanding Contribution to Modern Therapeutics." *Evid Based Complement Alternat Med* **2013**: 627375. <http://dx.doi.org/10.1155/2013/627375>
12. Paul, S. M., D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg and A. L. Schacht (2010). "How to improve R&D productivity: the pharmaceutical industry's grand challenge." *Nat Rev Drug Discov* **9**(3), 203-214. <http://dx.doi.org/10.1038/nrd3078>
13. Koubelakis, M. G., P. Zotos, G. Papachristoudis, I. Michalopoulos, K. I. Pappa, N. P. Anagnostou and S. Kossida (2009). "Human microRNA target analysis and gene ontology clustering by GOMir, a novel stand-alone application." *BMC Bioinformatics* **10** Suppl 6: S20. <http://dx.doi.org/10.1186/1471-2105-10-S6-S20>
14. Szymanski, P., M. Markowicz and E. Mikiciuk-Olasik (2012). "Adaptation of high-throughput screening in drug discovery-toxicological screening tests." *Int J Mol Sci* **13**(1), 427-452. <http://dx.doi.org/10.3390/ijms13010427>
15. Veselovsky, A. V., M. S. Zharkova, V. V. Poroikov and M. C. Nicklaus (2014). "Computer-aided design and discovery of

- protein-protein interaction inhibitors as agents for anti-HIV therapy." SAR QSAR Environ Res **25**(6), 457-471. <http://dx.doi.org/10.1080/1062936X.2014.898689>
16. Vlachakis, D., A. Armaos and S. Kossida (2017). "Advanced Protein Alignments Based on Sequence, Structure and Hydrophathy Profiles; The Paradigm of the Viral Polymerase Enzyme." Mathematics in Computer Science **11**(2), 197-208. <http://dx.doi.org/10.1007/s11786-016-0287-8>
 17. Vlachakis, D., D. G. Kontopoulos and S. Kossida (2013). "Space constrained homology modelling: the paradigm of the RNA-dependent RNA polymerase of dengue (type II) virus." Comput Math Methods Med 2013: 108910. <http://dx.doi.org/10.1155/2013/108910>
 18. Vlachakis, D., D. Tsagrasoulis, V. Megalooikonomou and S. Kossida (2013). "Introducing Drugster: a comprehensive and fully integrated drug design, lead and structure optimization toolkit." Bioinformatics **29**(1), 126-128. <http://dx.doi.org/10.4018/ijsbtt.2013070105>
 19. Yuan, Y., J. Pei and L. Lai (2011). "LigBuilder 2: a practical de novo drug design approach." J Chem Inf Model **51**(5), 1083-1091. <http://dx.doi.org/10.1021/ci100350u>