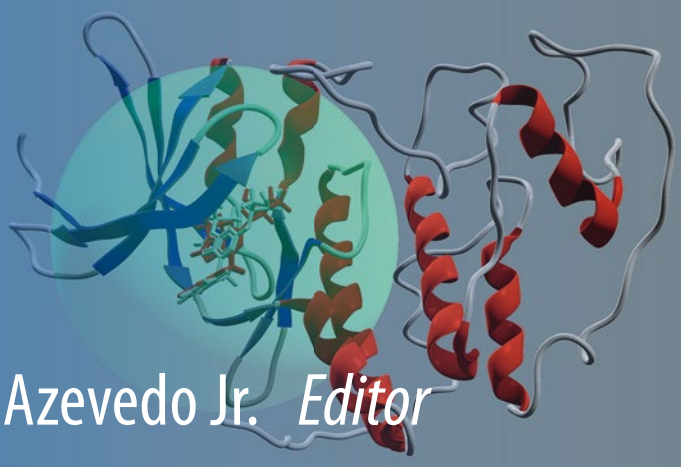
Methods in Molecular Biology 2053

Springer Protocols



Walter Filgueira de Azevedo Jr. *Editor*

Docking Screens for Drug Discovery

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, UK

For further volumes:
http://www.springer.com/series/7651

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Docking Screens for Drug Discovery

Edited by

Walter Filgueira de Azevedo Jr.

Escola de Ciências da Saúde, Pontifícia Universidade Católica do Rio Grande do Sul—PUCRS, Porto Alegre, Ria Grande do Sul, Brazil

Humana Press

Editor
Walter Filgueira de Azevedo Jr.
Escola de Ciências da Saúde
Pontifícia Universidade Católica do Rio Grande do Sul—PUCRS
Porto Alegre, Ria Grande do Sul, Brazil

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-9751-0 ISBN 978-1-4939-9752-7 (eBook)
https://doi.org/10.1007/978-1-4939-9752-7

Dedication

This book is dedicated to my beloved mother Marion de Fátima Pereira de Azevedo and my darling wife Maria do Carmo Dantas de Santana Azevedo.

Preface

The data explosion in the number of biological macromolecules deposited in the Protein Data Bank (PDB) [1–3] opened the possibility to investigate the correlation of these experimentally determined structures with biological information, which is a favorable scenario for the application of computational systems biology approaches to develop a mathematical model to predict ligand-binding affinity for this target protein. It is also possible to use these three-dimensional structures to study target proteins employed in the development and design of drugs [4–10]. The use of structural information for a target protein makes it possible to apply virtual screening methodology to identify new hits and guide the future development of new medicines. The primary approach to investigate potential new hits for a target protein is the methodology of protein-ligand docking simulation [11].

Docking is a simulation method that predicts the structure of a receptor-ligand complex, in which the receptor is a protein and the ligand is a small molecule [12–16]. This simulation is equivalent to the key-lock theory of enzyme specificity [17, 18], in which the lock is the receptor and the key is the ligand. The goal in any protein-ligand docking simulation is to adjust the position of the key (ligand) in the lock (ligand-binding pocket in a protein). From the computational view, we see the protein-ligand docking as an optimization problem, where our goal is to find the best solution (right position for the ligand) from a set of possible locations. Protein-ligand docking often makes use of one or more of the following computational methodologies: genetic algorithm, differential evolution, Lamarckian genetic algorithm, fast shape matching, incremental construction, distance geometry, simulated annealing, and others [19]. Protein-ligand docking methodology can produce several positions for the key in the lock. Therefore, we need a scoring function that will allow evaluations of all possible positions of the key, and then a selection can be carried out for the best location. For general reviews of the principles underlying molecular docking programs, see references [12–16].

Also, to evaluate the ligand-binding affinity for a specific target protein, we can employ a scoring function to compute scores that resemble ligand-binding energy functions. For both approaches, experimental information is vital to validate protein-ligand docking simulations and the ability of scoring functions to estimate ligand-binding affinity [20].

For protein-ligand docking simulations, it is common to start investigating if the computational approach is capable of reproducing an experimental 3D structure for a complex involving a protein and at least one ligand. If such structure is available, we employ it to check whether a specific molecular docking protocol is capable of predicting the crystallographic position for the ligand in the protein structure, a procedure called redocking. The most used criteria to evaluate redocking success are the root-mean-square deviation (RMSD) between the crystallographic position for the ligand and the pose (generated by the computer simulation). In docking simulations, we expect that the best results generate RMSD values less than 2.0 Å compared with crystallographic structures [12–16].

Furthermore, if we have more than one structure complexed with a ligand, we can take the validation process further, applying the molecular docking protocol to an ensemble of complexes structures. In this ensemble, we could have the same protein structure in complex with different ligands. For instance, a search in the PDB for structures containing the name

cyclin-dependent kinases (CDKs) and for which there is inhibition constant (K_i) information returned 31 structures. These structures have water molecules close to the active ligand and without repeated ligands (search carried out on March 20, 2019). This data set is an ensemble of CDK structures, where each entry is a structure complexed with a different ligand. This ensemble of structures can be employed to validate a docking strategy for a specific protein target. Moreover, it could also be used to test scoring functions.

For validation of scoring functions, it is common to investigate the correlation between the experimental binding affinity with scoring functions. Here we evaluated the predictive performance using squared Pearson's (R^2) or Spearman's (ρ) correlation coefficients [21]. Application of machine learning methods can improve the predictive performance of scoring functions trained against data sets composed of experimentally determined structures for which ligand-binding data is available [22–32].

The focus of the present book is on recent developments in docking simulations for target proteins. We have chapters dealing with specific techniques or applications for docking simulations. For instance, we describe the major docking programs. Also, we explain the scoring functions developed for the analysis of docking results and to predict ligand-binding affinity. Due to the importance of docking simulations for the initial stages of drug discovery, we believe that the present volume will appeal to those interested in molecular docking simulation and also in the application of these methodologies for drug discovery.

Finally, I would like to express my gratitude to all authors who accepted the challenge of bringing to a book their scientific knowledge. I want to thank Prof. John M. Walker (series editor for the *Methods in Molecular Biology* series) for his patience and assistance during the editorial process. This book wouldn't be possible without the aid of Anna Rakovsky (Assistant Editor at Springer Science + Business Media, LLC). Many others contributed directly or indirectly to this book. I want to thank all my students who tested the tutorials and protocols described here. They did a great job of helping to improve the quality of the material described in this work. This book is a dream coming true, and it wouldn't be possible without the comprehension and love of my wife Carminha (Maria do Carmo Dantas de Santana Azevedo) who understood my absence and helped me during the months of preparation of this book. To her: "Obrigado minha linda. Este livro é para você. Te amo muito."

Porto Alegre, RS, Brazil *Walter Filgueira de Azevedo Jr.*

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al (2000) The protein data bank. Nucleic Acids Res 28(1):235–242
2. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K et al (2002) The protein data bank. Acta Crystallogr D Biol Crystallogr 58(Pt 6 No 1):899–907
3. Westbrook J, Feng Z, Chen L, Yang H, Berman HM (2003) The protein data bank and structural genomics. Nucleic Acids Res 31(1):489–491
4. de Azevedo WF Jr (2010) Structure-based virtual screening. Curr Drug Targets 11:261–263
5. de Ávila MB, Bitencourt-Ferreira G, de Azevedo Jr. WF (2019) Structural basis for inhibition of enoyl-[acyl carrier protein] reductase (InhA) from mycobacterium tuberculosis. Curr Med Chem doi: 10.2174/0929867326666181203125229

6. Volkart PA, Bitencourt-Ferreira G, Souto AA, de Azevedo WF (2019) Cyclin-dependent kinase 2 in cellular senescence and cancer. A structural and functional review. Curr Drug Targets doi: 10.2174/1389450120666181204165344
7. Canduri F, Fadel V, Basso LA, Palma MS, Santos DS, de Azevedo WF Jr (2005) New catalytic mechanism for human purine nucleoside phosphorylase. Biochem Biophys Res Commun. 327 (3):646–649
8. Canduri F, Teodoro LG, Fadel V, Lorenzi CC, Hial V, Gomes RA et al (2001) Structure of human uropepsin at 2.45 A resolution. Acta Crystallogr D Biol Crystallogr 57(Pt 11): 1560–1570
9. de Azevedo WF Jr, Dias R (2008) Experimental approaches to evaluate the thermodynamics of protein-drug interactions. Curr Drug Targets 9(12):1071–1076
10. Delatorre P, Rocha BA, Souza EP, Oliveira TM, Bezerra GA, Moreno FB et al (2007) Structure of a lectin from Canavalia gladiata seeds: new structural insights for old molecules. BMC Struct Biol 7:52
11. Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. J Mol Recognit 9:175–186
12. Azevedo LS, Moraes FP, Xavier MM, Pantoja EO, Villavicencio B, Finck JA et al (2012) Recent progress of molecular docking simulations applied to development of drugs. Curr Bioinform 7:352–365
13. DesJarlais RL, Dixon JS (1994) A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. J Comput Aided Mol Des 8:231–242
14. de Azevedo WF Jr (2010) Structure-based virtual screening. Curr Drug Targets 11:261–263
15. de Azevedo WF Jr (2010) MolDock applied to structure-based virtual screening. Curr Drug Targets 11:327–334
16. Dias R, de Azevedo WF Jr (2008) Molecular docking algorithms. Curr Drug Targets 9:1040–1047
17. Fischer E (1890) Ueber die optischen Isomeren des Traubezuckers, der Gluconsäure und der Zuckersäure. Ber Dtsch Chem Ges 23:2611–2624
18. Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges 27:2985–2993
19. Heberlé G, de Azevedo WF Jr (2011) Bio-inspired algorithms applied to molecular docking simulations. Curr Med Chem 18:1339–1352
20. de Azevedo WF Jr, Dias R (2008) Computational methods for calculation of ligand-binding affinity. Curr Drug Targets 9:1031–1039
21. Zar JH (1972) Significance testing of the spearman rank correlation coefficient. J Am Stat Assoc 67:578–580
22. Bitencourt-Ferreira G, de Azevedo Jr WF (2018) Development of a machine-learning model to predict Gibbs free energy of binding for protein-ligand complexes. Biophys Chem 240:63–69
23. de Ávila MB, de Azevedo WF Jr (2018) Development of machine learning models to predict inhibition of 3-dehydroquinate dehydratase. Chem Biol Drug Des 92:1468–1474
24. Russo S, de Azevedo WF (2019) Advances in the understanding of the Cannabinoid Receptor 1—focusing on the inverse agonists interactions. Curr Med Chem doi: 10.2174/0929867325666180417165247
25. Amaral MEA, Nery LR, Leite CE, de Azevedo Junior WF, Campos MM (2018) Pre-clinical effects of metformin and aspirin on the cell lines of different breast cancer subtypes. Invest New Drugs 36:782–796
26. Levin NMB, Pintro VO, Bitencourt-Ferreira G, Mattos BB, Silvério AC, de Azevedo Jr WF (2018) Development of CDK-targeted scoring functions for prediction of binding affinity. Biophys Chem 235:1–8
27. Freitas PG, Elias TC, Pinto IA, Costa LT, de Carvalho PVSD, Omote DQ et al (2018) Computational approach to the discovery of phytochemical molecules with therapeutic potential targets to the PKCZ protein. Lett Drug Des Discov 15:488–499
28. Pintro VO, Azevedo WF (2017) Optimized virtual screening workflow. Towards target-based polynomial scoring functions for HIV-1 protease. Comb Chem High Throughput Screen 20:820–827
29. de Ávila MB, Xavier MM, Pintro VO, de Azevedo WF (2017) Supervised machine learning techniques to predict binding affinity. A study for cyclin-dependent kinase 2. Biochem Biophys Res Commun 494:305–310
30. Heck GS, Pintro VO, Pereira RR, de Ávila MB, Levin NMB, de Azevedo WF (2017) Supervised machine learning methods applied to predict ligand-binding affinity. Curr Med Chem 24:2459–2470

31. Levin NM, Pintro VO, de Ávila MB, de Mattos BB, De Azevedo WF Jr (2017) Understanding the structural basis for inhibition of cyclin-dependent kinases. New pieces in the molecular puzzle. Curr Drug Targets 18:1104–1111
32. Xavier MM, Heck GS, de Avila MB, Levin NM, Pintro VO, Carvalho NL et al (2016) SAnDReS a computational tool for statistical analysis of docking results and development of scoring functions. Comb Chem High Throughput Screen 19:801–812

Acknowledgments

This work was supported by grants from CNPq (Brazil) (308883/2014-4). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior—Brasil (CAPES)—Finance Code 001. WFA is a researcher for CNPq (Brazil) (Process Numbers: 308883/2014-4 and 309029/2018-0).

Contents

Dedication . . . . . *v*
Preface . . . . . *vii*
Acknowledgments . . . . . *xi*
Contributors . . . . . *xv*
About the Editor . . . . . *xvii*

1 Building Machine-Learning Scoring Functions for Structure-Based Prediction of Intermolecular Binding Affinity 1
Maciej Wójcikowski, Pawel Siedlecki, and Pedro J. Ballester

2 Integrating Molecular Docking and Molecular Dynamics Simulations 13
Lucianna H. S. Santos, Rafaela S. Ferreira, and Ernesto R. Caffarena

3 How Docking Programs Work 35
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

4 SAnDReS: A Computational Tool for Docking 51
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

5 Electrostatic Energy in Protein–Ligand Complexes 67
Gabriela Bitencourt-Ferreira, Martina Veit-Acosta, and Walter Filgueira de Azevedo Jr.

6 Van der Waals Potential in Protein Complexes 79
Gabriela Bitencourt-Ferreira, Martina Veit-Acosta, and Walter Filgueira de Azevedo Jr.

7 Hydrogen Bonds in Protein-Ligand Complexes 93
Gabriela Bitencourt-Ferreira, Martina Veit-Acosta, and Walter Filgueira de Azevedo Jr.

8 Molecular Dynamics Simulations with NAMD2 109
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

9 Docking with AutoDock4 125
Gabriela Bitencourt-Ferreira, Val Oliveira Pintro, and Walter Filgueira de Azevedo Jr.

10 Molegro Virtual Docker for Docking 149
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

11 Docking with GemDock 169
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

12 Docking with SwissDock 189
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

13 Molecular Docking Simulations with ArgusLab 203
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

14 Web Services for Molecular Docking Simulations 221
Nelson J. F. da Silveira, Felipe Siconha S. Pereira, Thiago C. Elias, and Tiago Henrique

15 Homology Modeling of Protein Targets with MODELLER 231
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.
16 Machine Learning to Predict Binding Affinity 251
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.
17 Exploring the Scoring Function Space 275
Gabriela Bitencourt-Ferreira and Walter Filgueira de Azevedo Jr.

Index *283*

Contributors

Pedro J. Ballester • *Cancer Research Center of Marseille, INSERM U1068, Marseille, France; Institut Paoli-Calmettes, Marseille, France; Aix-Marseille Université, Marseille, France; CNRS UMR7258, Marseille, France*

Gabriela Bitencourt-Ferreira • *Escola de Ciências da Saúde, Pontifícia Universidade Católica do Rio Grande do Sul—PUCRS, Porto Alegre, RS, Brazil*

Ernesto R. Caffarena • *Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil*

Nelson J. F. da Silveira • *Laboratory of Molecular Modeling and Computer Simulation/MolMod-CS, Institut of Exact Science/ICEx, Federal University of Alfenas/UNIFAL-MG, Alfenas, MG, Brazil*

Walter Filgueira de Azevedo Jr. • *Escola de Ciências da Saúde, Pontifícia Universidade Católica do Rio Grande do Sul—PUCRS, Porto Alegre, RS, Brazil*

Thiago C. Elias • *Laboratory of Molecular Modeling and Computer Simulation/MolMod-CS, Institut of Exact Science/ICEx, Federal University of Alfenas/UNIFAL-MG, Alfenas, MG, Brazil*

Rafaela S. Ferreira • *Laboratório de Modelagem Molecular e Planejamento de Fármacos, Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil*

Tiago Henrique • *Departament of Molecular Biology, Medical School of São José do Rio Preto/FAMERP, São José do Rio Preto, SP, Brazil*

Felipe Siconha S. Pereira • *Laboratory of Computacional Modeling, National Laboratory of Scientific Computing (LNCC), Petrópolis, RJ, Brazil*

Val Oliveira Pintro • *Escola de Ciências da Saúde, Pontifícia Universidade Católica do Rio Grande do Sul—PUCRS, Porto Alegre, RS, Brazil*

Lucianna H. S. Santos • *Laboratório de Modelagem Molecular e Planejamento de Fármacos, Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil*

Pawel Siedlecki • *Institute of Biochemistry and Biophysics PAS, Warsaw, Poland; Department of Systems Biology, Institute of Experimental Plant Biology and Biotechnology, University of Warsaw, Warsaw, Poland*

Martina Veit-Acosta • *Escola de Ciências da Saúde, Pontifícia Universidade Católica do Rio Grande do Sul—PUCRS, Porto Alegre, RS, Brazil*

Maciej Wójcikowski • *Institute of Biochemistry and Biophysics PAS, Warsaw, Poland*

About the Editor

Walter Filgueira de Azevedo Jr. is Frontiers Section Editor (Bioinformatics and Biophysics) for the *Current Drug Targets*, member of the editorial board of *Current Bioinformatics*, and section editor (Bioinformatics in Drug Design and Discovery) for the *Current Medicinal Chemistry*. Prof. Azevedo graduated in physics (BSc in physics) from the University of São Paulo (USP) in 1990. He completed a Master's Degree in Applied Physics also from the USP (1992), working under the supervision of Prof. Yvonne P. Mascarenhas, the founder of crystallography in Brazil. His dissertation was about X-ray crystallography applied to organometallic compounds. During his Ph.D., he worked under the supervision of Prof. Sung-Hou Kim (University of California, Berkeley), on a split Ph.D. program with a fellowship from Brazilian Research Council (CNPq) (1993–1996). His Ph.D. was about the crystallographic structure of CDK2. At present, he is the coordinator of the Structural Biochemistry Laboratory at Pontifical Catholic University of Rio Grande do Sul (PUCRS). His research interests are interdisciplinary with two major emphases: molecular simulations and protein-ligand interactions. He published over 160 scientific papers about protein structures and computer simulation methods applied to the study of biological systems (H-index: 40, RG Index > 41.0). These publications have over 5000 citations.