Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012

Enhancing the effectiveness of virtual screening by using the ChemBioServer: Application to the discovery of PI3Kα inhibitors

Paraskevi Gkeka, Emmanouil Athanasiadis, George Spyrou and Zoe Cournia
Biomedical Research Foundation of the Academy of Athens (BRFAA)
Soranou Efesiou 4, Athens, Greece.
{ pgkeka, mathan, gspyrou, zcournia }@bioacademy.gr

Abstract — The application of an in-house developed web server called ChemBioServer to the filtering and selection of drug candidates for the inhibition of the PI3Kα protein is presented. 1000 candidate molecules were initially selected from a virtual screening experiment. Those molecules were then filtered for steric clashes, physicochemical and toxicity properties and grouped into clusters using the ChemBioServer web application. During this filtering process, 400 compounds were rejected and the remaining 600 were clustered in 20 different groups, allowing for a more efficient visual inspection of the compounds. Representatives of these clusters were then selected for further experimental study. Four out of the seven selected molecules inhibited PI3Kα activity in vitro, indicating that the workflow described herein can be successfully applied in drug discovery. ChemBioServer proved to assist the post-processing application of top-ranked molecules resulting from a docking exercise by increasing the efficiency and the quality of compound selection that passed to the experimental test phase.

Keywords-component; cheminformatics, virtual screening, drug discovery, PI3Kα.

I. INTRODUCTION

The PI3Kα protein is implicated in signaling cascades, which lead to cell proliferation, survival, and cell growth. PI3Kα is one of the most frequently mutated proteins in human cancers and is thus an attractive target for anti-cancer drug discovery [1]. The application of rational, structure-based drug design is proven to be more efficient than the traditional way of drug discovery since it aims to understand the molecular basis of a disease and utilizes the knowledge of the three-dimensional (3D) structure of the biological target in the process. State of the art structure-based drug design methods include virtual screening, which serves as an efficient, alternative approach to high-throughput experimental screening. In virtual screening, large libraries of drug-like compounds that are commercially available are computationally screened against targets of known structure, and those that are predicted to bind well are experimentally tested. Computer-aided drug discovery has recently had important successes: new ligands have been predicted along with their receptor-bound structures and in several cases the achieved hit rates (ligands discovered per molecules tested) have been significantly greater than with high-throughput screening [2]-[4]. Up to now, several chemical compound databases have been developed, including Zinc, PubChem, Chembank and many others [5]-[7]. Nevertheless, online open-access web applications for compound mining are limited in number and, importantly, in pipeline integration level. Another rate-limiting step in computer-aided drug design is often the final selection of compounds to be tested experimentally. The selected compounds should possess favorable computed binding affinities and at the same time be devoid of unwanted characteristics such as intra-ligand steric clashes, undesirable physicochemical properties, and toxic moieties [4].

To overcome these limitations during the virtual screening process, we have developed the ChemBioServer [8] in order to enhance the effectiveness of the virtual screening process. Herein, we describe a virtual screening workflow to identify novel inhibitors of the H1047R mutant form of PI3Kα frequently found in human cancers. The workflow incorporates the use of ChemBioServer, a free web-based application, which enables the efficient final selection of compounds to be tested experimentally.

II. MATERIAL AND METHODS

The crystal structure of the mutant H1047R PI3Kα (PDB ID: 3HIZ) was complemented for missing parts using a combination of homology and loop modeling in order to create the full atomistic model of the full-length H1047R mutant. The Modeller software 9v8 was used for homology and ab initio loop modeling [9]. For the parts of the protein that required homology modeling, the human WT protein (PDB ID: 2RD0 and 2ENQ) was used as a template. The resulting model was solvated in water and employed in Molecular Dynamics (MD) simulations using the NAMD package [10]. The CHARMM22 force field [11] and [12] was used to model all protein interactions and the TIP3P model [13] was used for water. The solvated protein system was energy-minimized and gradually heated from 0 to 310 K with constraints of 1 kcal mol^{-1} $Å^{-2}$ applied on the backbone protein atoms under constant volume. An equilibration run was then performed under constant pressure and constant temperature. Non-bonded forces were calculated with a 2-fs time step and a 12 Å cut-off using the CHARMM switch potential between 10-12 Å. Bonds involving

hydrogen were kept rigid by using the SHAKE algorithm [14] for the protein and the SETTLE algorithm [15] for water. Periodic boundary conditions were applied and the Particle Mesh Ewald method [16] was used to calculate electrostatic interactions every 4 fs. The pressure was maintained at 1 atm with the Langevin piston method [17], while the temperature was maintained at 310 K by means of Langevin dynamics with a damping coefficient of 5 ps^{-1}. Atomic coordinates of the systems were saved every 2 ps. The total simulation time was 70 ns.

Following the simulation, binding site analysis was performed using the SiteMap module of Schrodinger v2.4 [18] on the protein conformation corresponding to the last frame of the trajectory (70 ns). An allosteric binding site close to the H1047R mutation was found by SiteMap to be among the top-ranked potential receptor binding sites and was used in the present study. After binding site identification, we performed virtual screening using the docking program Glide 5.7 (Schrodinger, LLC) [19] - [20]. In the process of virtual screening, initially, the all-atom protein model was submitted to a series of restrained, partial minimizations using the OPLS-AA force field within the "Protein Preparation" module of Glide. A benzene molecule was placed in the predicted binding cavity and was used for the "Grid Generation" module of Glide, which prepares a grid for ligand docking. For the protein preparation, grid generation, and ligand docking procedures, the default Glide settings were used. The van der Waals (vdW) radii for nonpolar ligand atoms were scaled by a factor of 0.8, thereby decreasing penalties for close contacts. Receptor atoms were not scaled. The drug-like subset of the HitFinder collection from the Maybridge database (www.maybridge .com) was used for the virtual screening [4]. All structures were docked and scored using the Glide standard precision (SP) mode [20]. The 10,000 top-ranked structures from the SP filter were redocked and rescored using the Glide extra precision (XP) mode [21]. The complexes for the top-ranked 1000 compounds resulting from the XP processing were submitted to further postprocessing with the ChemBioServer.

The ChemBioServer web application is divided into six main sections: (i) basic search, (ii) filtering, (iii) advanced filtering, (iv) clustering, (v) customized pipeline and (vi) visualization of compounds' properties. The application back-end is developed in R programming language (http://cran.r-project.org/), while the front-end is implemented with PHP (http://www.php.net/). 2D and 3D display of compounds is accomplished by means of the open-source Java viewer for chemical structures JChemPaint (http://jchempaint.github. com/) and Jmol (http://jmol.sourceforge.net/), respectively. Compound Fingerprints are generated with Open Babel (http:// openbabel.org/wiki/Main_Page). Briefly, the ChemBioServer provides the following functionalities. The 'Basic Search' section enables the researcher to browse the contents of a compound file that is uploaded to the server. In the 'Filtering' section, compound mining can be performed based on a variety of chemical properties, such as the Lipinski Rule of Five or custom-made filters. In the "Advanced Filtering" section, compounds with steric clashes can be discarded using the vdW filtering by means of energy and radii tolerance. Also, compounds with toxic or unwanted chemical groups can be filtered out. The 'Clustering' section includes a classical (hierarchical) as well as a modern clustering (affinity propagation) algorithm. Visualization of clusters is also available as a dendrogram plot in PDF format. Graphical representations of molecular properties (i.e. PSA, logP, etc.) can be implemented by means of the Raphaël javascript library (http://raphaeljs.com/) using principal components analysis. Finally, a pipeline workflow that combines all or part of the previously described filtering services is provided by the ChemBioServer. ChemBioServer is running on a quad-core Intel Xeon Processor E5420 server with 18GB RAM. Indicative computational time, using an sdf file containing 557 compounds (Test Set 4 available in Example Data webpage section), for vdW filtering was 1.5min, for toxicity filtering was 2min and for AP was 1 min.

To post-process docking results in order to enhance the experimental hit rate, several functionalities of the ChemBioServer were used. Initially, the vdW filtering was applied to remove compounds with steric clashes. Poses that are far from the energy minimum are unlikely to be adopted in nature and hence should be discarded. In this docking exercise with Glide we observed that the post-docking poses often suffered from vdW clashes; even after Glide post-docking minimization, approximately 20% of the generated poses should be discarded due to unrealistic vdW interactions. The compounds that passed vdW filtering were then subjected to physicochemical property filtering based on the Jorgensen rule of 3 [22] as well as toxicity filtering based on a database available in ChemBioServer, which contains known toxic moieties. These filters were not applied in the initial database as compounds with very high docking scores and a few liabilities, e.g. a toxic moiety, may be considered for optimization and/or as a starting point for de novo design. Subsequently, a hierarchical clustering was performed for the remaining compounds using the Tanimoto coefficient and the Ward Clustering Linkage. Finally, the resulting clusters were visually inspected and the most promising compounds were purchased and submitted to in vitro assay testing. The process is described in Fig 1.

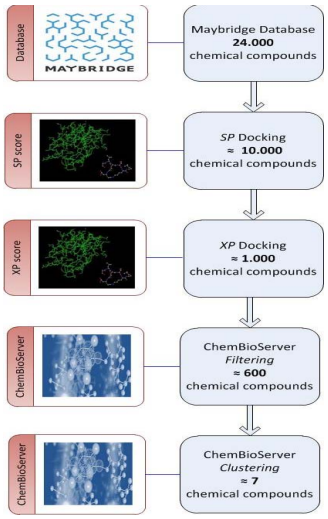

Figure 1 Workflow of Virtual Screening process using the ChemBioServer

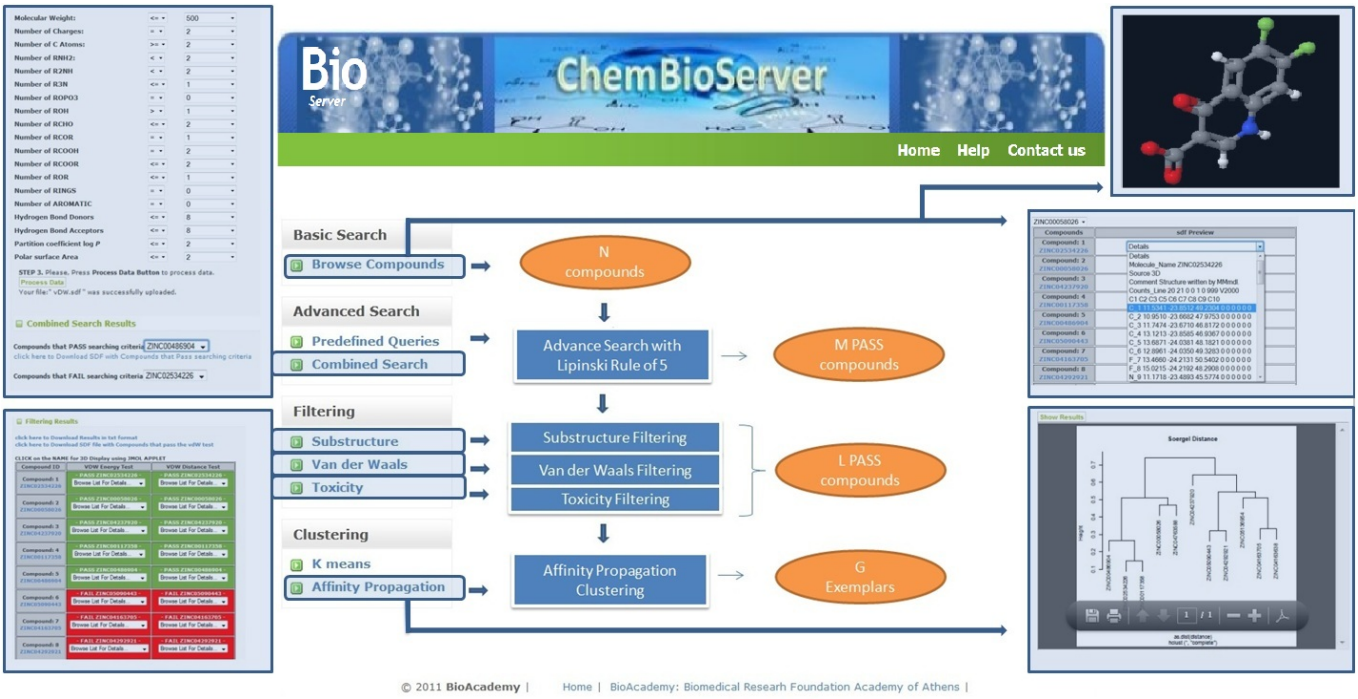

Figure 2 Workflow of the PI3Kα virtual screening process by means of the ChemBioServer

III. Results and Discussion

MD simulations of the PI3Kα protein, a common target in cancer, were performed starting from its crystal structure. Binding site identification on the last frame of the MD trajectory revealed an allosteric binding site, which was further used to perform a docking exercise for the identification of PI3Kα inhibitors. Virtual screening results were then post-processed with physicochemical, toxicity, and structural filters in order to enhance the efficiency and accuracy of the docking exercise. Initially, the 1000 top-scored compounds were filtered for steric clashes, using the vdW filtering available in ChemBioServer using a threshold energy of 50 kcal/mol, which resulted in 250 rejected poses. The remaining 750 drug candidates were subjected to physicochemical/toxicity filtering and the 600 accepted compounds were grouped in clusters via hierarchical clustering using Simple Matching Coefficient (Jaccard / Tanimoto Coefficient), Ward Clustering Linkage and distance 0.8 (see Fig. 2). The clustering resulted in twenty clusters. Maximum one compound per cluster was selected by visual inspection based on a) important ligand interactions with key residues of the binding site and b) promising predicted physicochemical properties. Finally, seven of the most promising compounds were purchased. The compounds were tested with in vitro assays and four inhibited PI3Kα activity in micromolar concentrations, achieving a 57% hit rate and indicating that the workflow described herein can be successfully applied to enhance the hit rate of in silico drug discovery.

IV. Conclusion

A virtual screening experiment was performed on the PI3Kα protein, a known cancer target. The resulting top-scored compounds were subjected to structural, physicochemical/ toxicity, and clustering filters within the ChemBioServer web application. Compounds were finally visually inspected and the most promising ones were purchased. Four out of the seven purchased molecules inhibited the PI3Kα protein activity in vitro in micromolar concentrations, indicating that the workflow described herein can be successfully applied in drug discovery.

Acknowledgments

This work was funded by the NSRF 2007 – 2013, co-funded by the European Regional Development Fund and national resources, under grant “Cooperation” [No. 09ΣΥΝ-11-675].

References

[1] Liu, P.; Cheng, H.; Roberts, T. M.; Zhao, J. J. “Targeting the phosphoinositide 3-kinase pathway in cancer”. Nature Rev. 8:627-644., 2009.

[2] Shoichet BK. “Virtual screening of chemical libraries”. Nature, 16;432(7019):862-5, 2004.

[3] Andricopulo AD, Salum LB, Abraham DJ. “Structure-based drug design strategies in medicinal chemistry.”, Curr Top Med Chem., 9(9):771-90, 2009.

[4] Cournia, Z., Leng, L., Gandavadi, S., Du, X., Bucala, R., Jorgensen, W.L. “Discovery of human macrophage migration inhibitory factor (MIF)-CD74 antagonists via virtual screening”. Journal of medicinal chemistry, 52, 416-424, 2009.

[5] Irwin, J.J. and Shoichet, B.K. “ZINC--a free database of commercially available compounds for virtual screening”. Journal of chemical information and modeling, 45, 177-182, 2005.

[6] Li, Q. Cheng, T., Wang, Y., Bryant, S.H. “PubChem as a public resource for drug discovery”. Drug discovery today, 15, 1052-1057, 2010.

[7] Seiler, K.P., George, G.A., Happ, M.P., Bodycombe, N.E., Carrinski, H.A., Norton, S., Brudz, S., Sullivan, J.P., Muhlich, J., Serrano, M., Ferraiolo, P., Tolliday, N.J., Schreiber, S.L.,Clemons, P.A. "ChemBank: a small-molecule screening and cheminformatics resource database". Nucleic acids research, 36, D351-359, 2008.

[8] Athanasiadis, E., Cournia, Z., Spyrou, G. "ChemBioServer: A web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery", Bioinformatics doi: 10.1093/ bioinformatics/bts551, 2012.

[9] Sali, A., Blundell, T.L. "Comparative protein modelling by satisfaction of spatial restraints", Journal of Molecular Biology, 234(3):779-815, 1993.

[10] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 26:1781-1802, 2005.

[11] MacKerell J., A.D., Bashford, D., Bellott, M., Dunbrack J., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., H. S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M. "All-atom empirical potential for molecular modeling and dynamics studies of proteins". J. Phys. Chem. B, 102: 3586-3616, 1998.

[12] Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD Jr. "Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme" , Biophys J.15;90(4):L36-8, 2006.

[13] Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W. and Klein M. L. "Comparison of simple potential functions for simulating liquid water". J. Chem. Phys., 79:926-935, 1983.

[14] Ryckaert JP, Ciccotti G, Berendsen HJC "Numerical-integration of cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". J. Comput. Phys., 23: 327-341, 1977.

[15] Miyamoto S, Kollman PA "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". J. Comput. Chem., 13: 952-962, 1992.

[16] Darden T, York D, Pedersen L "Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems". J. Chem. Phys., 98: 10089-10092, 1993.

[17] Feller SE, Zhang YH, Pastor RW, Brooks BR "Constant pressure molecular dynamics Simulation: The Langevin piston method". J Chem. Phys., 103: 4613-4621, 1995.

[18] SiteMap, version 2.4, Schrödinger, LLC, New York, NY, 2011.

[19] Glide, version 5.7, Schrödinger, LLC, New York, NY, 2011.

[20] Friesner, R.A, Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy", J. Med. Chem., 47, 1739–1749. 2004.

[21] Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C., Mainz, D.T. "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes", J. Med. Chem., 49, 6177–6196, 2006.

[22] Kerns, E., Di, L. "Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization", 1st Ed., Elsevier, 2008.