

In Silico Investigation on Potential Off-target Binding and Repurposable Drugs in the Treatment of Spinal Muscular Atrophy

Xiaolong Qiu

University of California, San Diego. La Jolla, California, U.S.A.

Monash University, Caulfield Campus. Melbourne, Victoria, Australia

Abstract

To effectively avoid unexpected adverse side-effect of commercial drugs, it is crucial to identify possible off-target bindings in their preclinical design processes. In this case study, identification of potential off-target bindings is focused on drugs that can potentially be repurposed to target DcpS, which is a valuable therapeutic target in the treatment of spinal muscular atrophy (SMA). By carrying out protein functional site similarity search, small molecule screening, and protein-ligand binding affinity profile analysis, possible off-target bindings in the treatment of SMA, as well as drugs that can potentially be repurposed to target DcpS are identified.

Introduction

Spinal muscular atrophy (SMA) is a group of inherited neuromuscular diseases that causes progressive muscle degeneration and weakness. The cause of SMA is the deletion or mutational inactivation of survival motor neuron 1 (SMN1) gene. Study has shown that SMA affects 1 in 6000 live births and is the leading cause of hereditary infant death (Singh et al.,

2008). Nevertheless, SMA has been identified as a curable disease by the National Institute of Health (NIH).

Possible treatment for SMA has recently been presented (Singh et al., 2008). Proposed treatment involves increasing the amount of survival motor neuron (SMN) protein expressed and presented through gene replacement, increasing SMN transcription, correct splicing, increased translation, and stabilizing the protein. More specifically, one of the treatments proposed is to be carried out by activating SMN2, a duplicate copy of the SMN1 gene in humans that is located immediately centromeric to the functional gene. C5-substituted quinazolines have been proven to be able to increase SMN mRNA levels in SMA patient-derived fibroblasts as well as SMN protein levels. Further researches on C5-substituted quinazolines have shown that these quinazolines are potent inhibitors to DcpS, which is a scavenger decapping enzyme. C5-substituted quinazolines are shown to be highly effective at inhibiting recombinant DcpS decapping activities, and the potency of inhibition correlates with potency for SMN2 promoter induction. Binding of C5-substituted quinazolines to DcpS holds the enzyme in an open, catalytically incompetent conformation. Protein microarray identification of DcpS has suggested an array of C5-substituted quinazolines as potent DcpS binders.

Potential multiple binding mechanisms have raised the concern of off-target binding. Research has shown that off-target binding appears to be the norm rather than the exception in preclinical drug design. Off-target binding implies the binding of a small molecule of therapeutic interest to a protein target other than the primary target for which it was intended, which may lead to detrimental side-effects (Xie et al., 2011). Since proteins with binding pockets similar to DcpS are likely to exist, off-target binding is likely to be the case in the treatment of SMA.

Methods

The Protein Data Bank (PDB) IDs for drug target DcpS are 3BL7, 3BL9, and 3BLA. Available structures of existing drug targets are to be compared with DcpS using SMAP, which is a software package designed for the comparison of shape description of protein structures and applications in predicting ligand binding sites.

The existing druggable human proteome is determined by mapping drug targets against human protein sequences using Protein Specific Interactive-Basic Local Alignment Search Tool (PSI-BLAST). All drug targets are downloaded from DrugBank, an online bioinformatics and cheminformatics database; while all protein sequences are downloaded from Protein Data Bank, an online information portal to biological macromolecular structures. 4001 targets are mapped to 194743 protein sequences. A reliable homology model is recognized in the PSI-BLAST results if both the expect value and the identity of a protein sequence are less than 0.001 and greater than 30 percent, respectively.

After getting a list of PDB IDs that are mapped to the drug targets, SMAP is used to find potential off-targets by searching for proteins that have similar binding sites compared to DcpS. Using the structures of DcpS and drug targets as the templates and queries, respectively, 31935 pair-wise sequence comparisons are executed in SMAP installed on the Nimrod portal (Abramson et al., 2000). Nimrod is a parameter sweeping and cloud execution tool that is capable of running a large amount of jobs simultaneously utilizing multiple compute resources. The queries are then prioritized based on their raw scores and p-values. Raw score is a profile-profile alignment score between the binding pockets of two proteins. It evaluates the evolutionary and geometric similarities of the two binding pockets. P-value estimates the

statistical significance of the raw score by considering the background probability distribution of the binding site alignment scores. In addition, the ligand-binding information of the queries are also parsed and collected.

Cross-docking studies are carried out with the aid of AutoDock Vina (Trott and Olson, 2010). Referring back to the PSI-BLAST results, 21 unique top hits generated from SMAP correspond to 790 DB IDs in DrugBank. Drugs that correspond to the top ten hits in the prioritized query lists are extracted from DrugBank in SMILE format, and converted to PDB format by using CORINA, an online three-dimensional structure generator for drug-like molecules. The PDB files are then converted to PDBQT format by AutoDock Vina running on the Nimrod portal. Each of these PDBQT file is run as a ligand against all 3 receptors in AutoDock Vina, and the ones with more negative binding affinities are predicted to be drugs that can potentially be repurposed to target DcpS.

The optimal binding mode is selected by AutoDock Vina, and graphics of the protein-ligand binding sites are predicted and illustrated by PyMol, a molecular visualization software.

Results

The most probable ligand that may bind to receptor 3BL7 is predicted to be DB01721, which is an analogue of indinavir drug. Indinavir is an inhibitor of the human immunodeficiency virus (HIV) protease. Commercial drugs of indinavir sulfate are sold as capsules in various strengths. Side-effects of indinavir drugs include fever, sore throat, headache, pale or yellowed skin, increased urination, and extreme thirst.

The most probable ligand that may bind to receptor 3BL9 is predicted to be DB03467, which is naringenin. Naringenin is the predominant flavanone found naturally in grapefruit.

Tolerance of naringenin varies from person to person. Genetic variation, food interactions, and metabolism-based drugs are some major causes of side-effects related to naringenin.

There are two most probable ligands with the same binding affinities that may bind to receptor 3BLA. They are DB01649, which is 7-methyl-gpppa; and DB03958, which is mRNA cap analog N7-methyl gpppg. Both of these ligands are suggested as DcpS binding enzymes in DrugBank.

The top thirty hits, or the potential off-targets, that are structurally similar to DcpS are shown in Table I, while the top thirty probable ligands, or drugs that can potentially be repurposed to target DcpS, are shown in Table II. The predicted most probable ligand-receptor systems for 3BL7, 3BL9, and 3BLA (two systems with identical binding affinities) are shown in Figure 1, 2, 3, and 4, respectively.

Table I. Off-targets Predicted by SMAP	
3BL7	3DL8, 3FOF, 1YWH, 2GRX, 2PNR, 1HVV, 1QVR, 1C9K, 1YVK, 2BF1
3BL9	2PNR, 3B6H, 1YVK, 1KPA, 1LN2, 1HVV, 3FOF, 2XY9, 2GRX, 1OGV
3BLA	1YVK, 3OFB, 3B6H, 1KPA, 3OAR, 1QVR, 3OE0, 1QVT, 1O57, 1M7W

Table II. Potentially Repurposable Drugs Predicted by AutoDock Vina	
3BL7	DB01721, DB04842, DB02033, DB00246, DB00157, DB00966, DB02668, DB00210, DB00496, DB01267
3BL9	DB03467, DB04216, DB01645, DB03608, DB02162, DB07826, DB00640, DB00817, DB01972, DB00714
3BLA	DB01649, DB03958, DB03461, DB00157, DB00210, DB08446, DB00734, DB07642, DB08460, DB01960

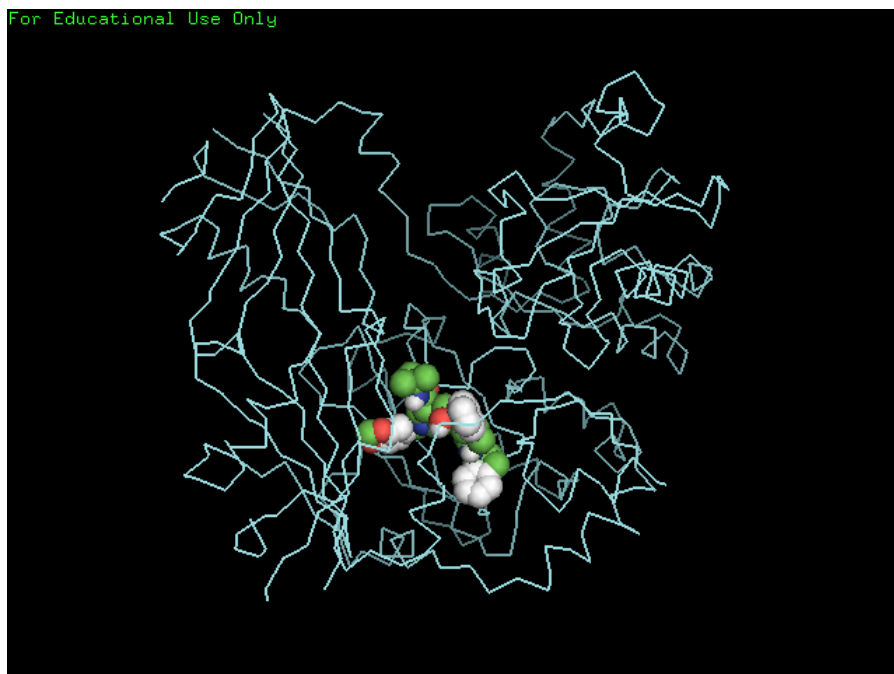


Figure 1. Predicted DB01721-3BL7 Binding Site by PyMol

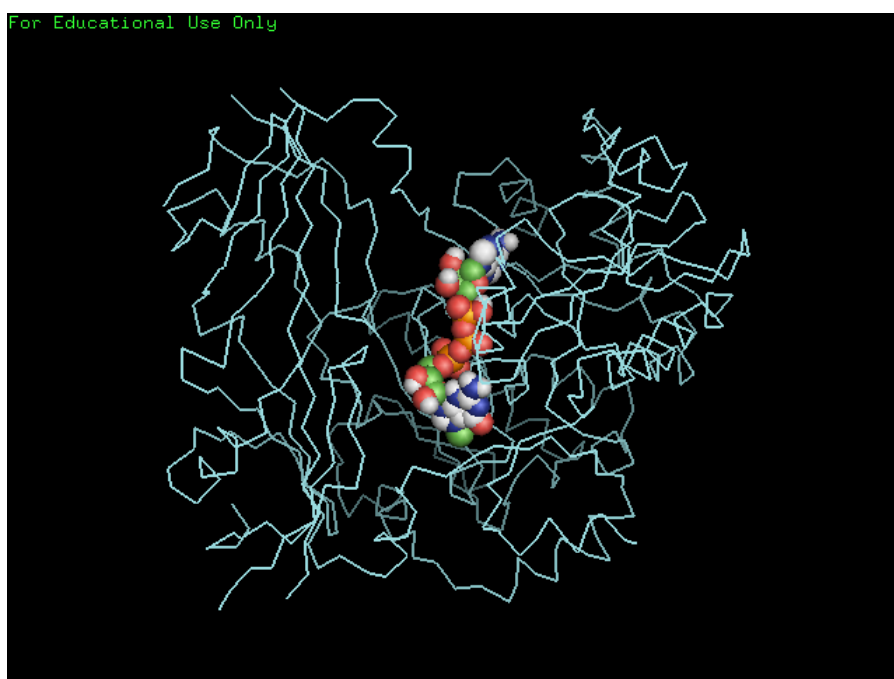


Figure 2. Predicted DB03467-3BL9 Binding Site by PyMol

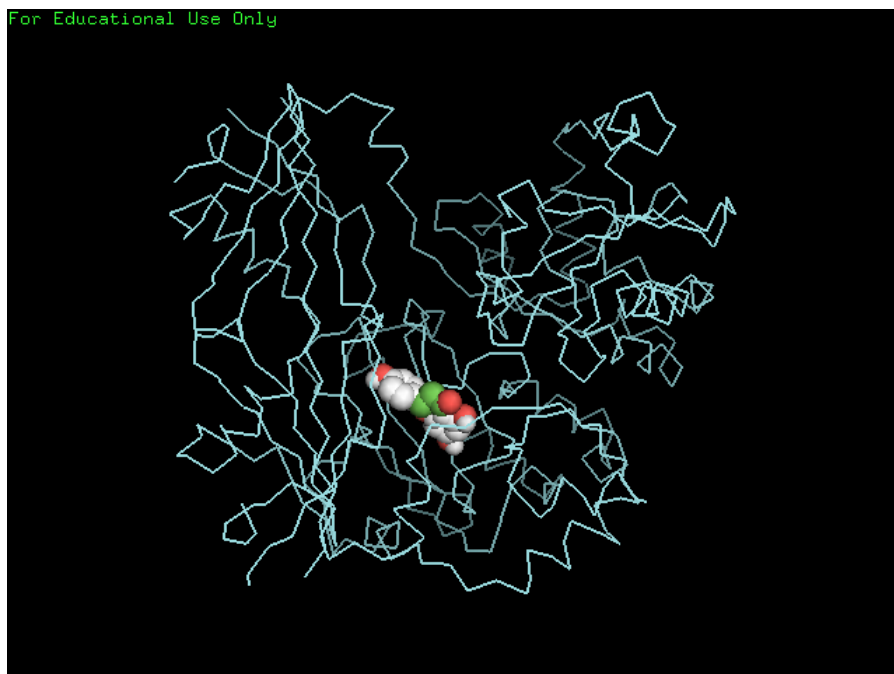


Figure 3. Predicted DB01649-3BLA Binding Site by PyMol

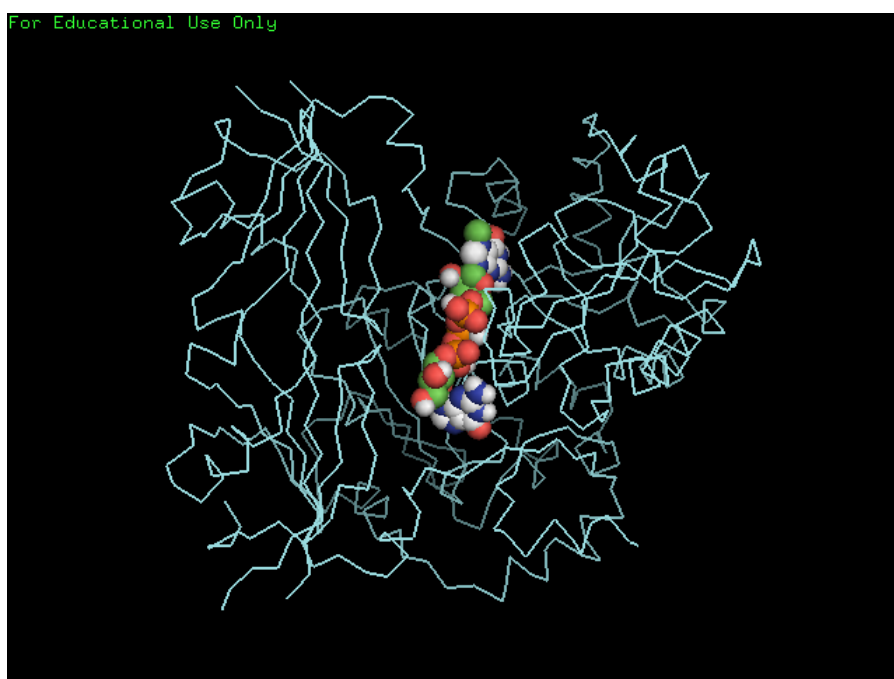


Figure 4. Predicted DB03958-3BLA Binding Site by PyMol

Discussion

Given the fact that computational screening on a proteome-wide scale is not feasible, identification of structurally similar ligand binding sites will greatly reduce the amount of work that docking tools have to perform (Xie et al., 2007). The combination of functional site similarity search on a structural proteome-wide scale, small molecule screening, and protein-ligand docking has made it possible to identify potential off-target bindings and side-effects might be caused by commercial drugs.

In the virtual docking experiment, a system consisting of the receptor and ligand in water is considered. If the potential binding energy of the ligand-receptor system is plotted as a function of its conformation, there is a corresponding potential energy value for each of the conformation. The global minimum corresponds to its native conformation. However, a deeper and narrower potential energy well is often less preferable due to the increase of the system's entropy during its binding process, as governed by the second law of thermodynamics. Thus, the system's chemical potential is a more accurate indicator of the likelihood of predicted bindings. Nevertheless, due to feasibility issues involved in computational calculations of a system's chemical potential, a screen function, a function that is tuned to recognize the promising conformations and predict the affinities, is used instead to predict the likelihood of binding (Trott and Olson, 2010).

The four drugs mentioned in the results section, according to the docking experiments, are the most probable ligands that target each of the three receptors. In other words, the four drugs that are most likely to inhibit DcpS are identified as potentially repurposable drugs. The

structures of the top three drugs that are predicted to bind 3BL7, 3BL9, and 3BLA are shown in Figure 5, 6, and 7, respectively.

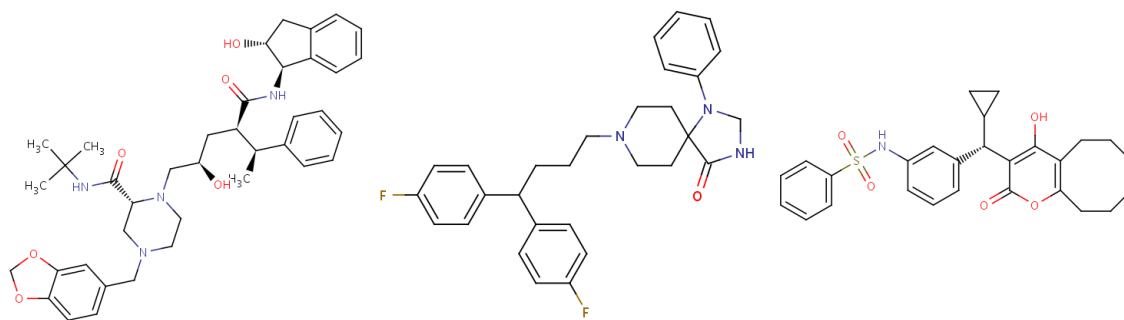


Figure 5. Predicted 3BL7 Binding Drugs

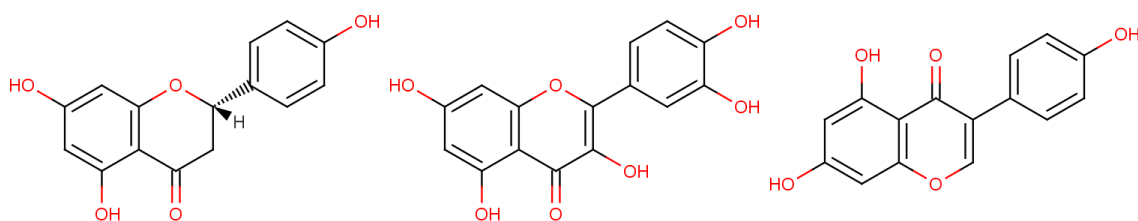


Figure 6. Predicted 3BL9 Binding Drugs

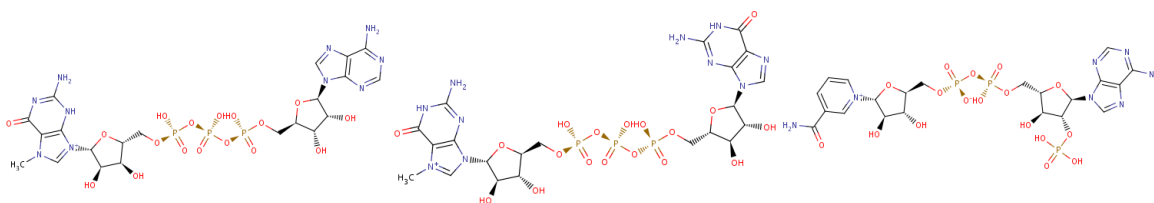


Figure 7. Predicted 3BLA Binding Drugs

The structure of the top three drugs that inhibits 3BL9 are very similar in the sense that they are all consisted of 3 aromatic rings with multiple OH groups attached to the two outer rings and a ketone group attached to the ring in the middle. The binding pocket of 3BL9 may also be hydrophilic since it readily binds with the OH groups of the ligand, and the binding pocket is likely to be positively charged and is ready for any nucleophilic attacks from the ligands. Compared with drugs that are likely to inhibit 3BL9, the top hits of 3BL7 have generally longer carbon chains, more aromatic rings, and are less hydrophilic. There isn't really an obvious pattern that can be observed. The ligands that bind 3BLA have even longer carbon chains, more hydrophilic groups, and the ligands themselves are structurally symmetrical. The phosphate groups that link the two symmetrical parts are likely to break upon binding with the receptor. In addition, both of the most probable ligands that target 3BLA, 7-methyl-gpppa and mRNA Cap analog N7-methyl gpppg, are suggested as DcpS binders in DrugBank.

Conclusion

In conclusion, both potential off-target bindings and drugs that can potentially be repurposed to target DcpS are identified through computational methods in this case study. The results are presented in Table I and II, respectively.

Future Work

In order to show that the ligand-receptor binding systems that predicted by AutoDock Vina are statistically significant, it is suggested to dock a set of random ligands found in Directory of Useful Decoys (DUD) to the same 3 receptors using the exact same methods. After getting the binding affinities for this random set of docking experiment, an unpaired t-test is to be performed to test the validity of the null-hypothesis. In addition, the p-values of the top hits

are to be calculated, and a complete statistical analysis is to be carried out to support the results generated by AutoDock Vina. Furthermore, experimental findings that support the stated computational results presented in the conclusion section is to be compared and analyzed.

Acknowledgements

I am grateful for the financial supports from National Science Foundation, IOSE-0710726 and the UCSD Pacific Rim Undergraduate Experiences (PRIME). I appreciate the guidance and help that I received from Dr. Gabriele Wienhausen, Dr. Peter Arzberger, Teri Simas, and Rob Gray. I would like to say special thanks to Dr. Philip E. Bourne, Dr. David Abramson, Dr. Lei Xie, Dr. Li Xie, and Blair Bethwaite for valuable discussions and technical assistance.

References

1. Abramson, D., Giddy, J. and Kotler, L. (2000) High Performance Parametric Modeling with Nimrod/G: Killer Application for the Global Grid? *International Parallel and Distributed Processing Symposium (IPDPS)*., 520-528.
2. Altschul, Stephen, F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, J. (1997) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.*, **25**, 3389-3402.
3. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A. C., and Wishart, D. S. (2011) DrugBank 3.0: A Comprehensive Resource for 'Omics' Research on Drugs. *Nucleic Acids Res.*, **39**, 1035-1041
4. Huang, N., Shoichet, B. K., and Irwin, J. J. (2006) Benchmarking Sets for Molecular Docking. *J.Med. Chem.*, **49**(23) 6789-6801.
5. Singh, J., Salcius, M., Liu, S., Staker, B. L., Mishra, R., Thurmond, J., Michaud, G., Mattoon, D. R., Printen, J., Christensen, J., Bjornsson, J. M., Pollok, B. A., Kiledjian, M., Stewart, L., Jarecki, J., and Gurney, M. E. (2008) DcpS as a Therapeutic Target for Spinal Muscular Atrophy. *ACS Chemical Biology*., **3**, 711-722.
6. Trott, O., and Olson, A. J. (2010) AutoDock Vina: Improving the Speed and Accuracy of Docking with A New Scoring Function, Efficient Optimization and Multithreading. *Computational Chemistry*., **31**, 455-461.
7. Xie, L., Xie L., and Bourne, P. E. (2011) Structure-Based Systems Biology for Analyzing Off-target Binding. *Current Opinion in Structural Biology*., **21**, 189-199.
8. Xie, L., Xie L., and Bourne, P. E. (2009) A Unified Statistical Model to Support Local Sequence Order Independent Similarity Searching for Ligand Binding Sites and Its Application to Genome-based Drug Discovery. *Bioinformatics*., **25**(12), 305-312.
9. Xie, L., and Bourne, P. E. (2008) Detecting Evolutionary Relationships across Existing Fold Space, Using Sequence Order Independent Profile-profile Alignments. *PNAS*, **105**(14), 5441.
10. Xie, L., and Bourne, P. E. (2007) A Robust and Efficient Algorithm for the Shape Description of Protein Structures and Its Application in Predicting Ligand Binding Sites. *BMC Bioinformatics*., **8**, S9.
11. Xie, L., Wang, J., and Bourne, P. E. (2007) In Silico Elucidation of the Molecular Mechanism Defining the Adverse Effect of Selective Estrogen Receptor Modulators. *PLoS Comput Biol.*, **3**, e217.
12. Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008) DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic Acids Res.*, **36**, 901-906.

13. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2008) DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.*, **34**, 668-672.