iSafeRat® SESAME-3D: Prediction of Protein-Ligand Interactions at the Molecular Level to Detect Endocrine Disrupting Potential of Chemicals

Zlatomir Todorov[1], Franklin Bauer[1], Paul Thomas[1]

[1] KREATiS, 38080 L'Isle d'Abeau – France

Contact us : www.kreatis.eu / contact@kreatis.eu

INTRODUCTION

There is a need for new approach methodologies (NAMs) to identify substances with an endocrine disrupting (ED) modality. Our goal is to provide an accurate, high-throughput, expandable and customizable tool for molecular modeling that can be used to detect potential ED modality. The interaction of chemicals with proteins is a key to investigate the disruption of endocrine pathways. Among the computational methods that could investigate such interactions, molecular docking is a technique that has a reasonable tradeoff between computational investment and accuracy of predictions. Arguably, there is currently no publicly available "gold standard" computational framework employing molecular docking entirely fit for purpose to handle this model; hence, our team is building our own computational framework.

The aim of this study was to show how SESAME-3D performs in the worst-case scenario when faced with decoy molecules which are highly similar to known ligands of EATS targets.

METHODS

Fundamental parts of the framework consist of third-party databases[2] and software[3,4], while the linking of software and some analysis strategies were developed entirely by our team.

Gather experimentally derived 3D models of protein-ligand complexes involving proteins of interest and native or synthetic ligands

↓

Protein 3D models curation and clustering of similar protein conformations

↓

Prepare 3D protein models for molecular docking

Ligand 3D models for molecular docking are prepared *de novo for test and reference molecules*

EATS biological target implemented in the workflow	Conformational clusters
Thyroid hormone receptor α (TRα)	2
Thyroid hormone receptor β (TRβ)	5
Estrogen receptor α (ERα)	18
Estrogen receptor β (ERβ)	10
Progesterone receptor (PR)	11
Androgen receptor (AR)	15

↓

Flexible docking: ligand & protein *N* replicas/ prot. conf. → **Analysis: Calc. Binding Profiles & Classification of test mol**

Small molecules (ligands and decoys) used in this study were retrieved at random from lists specified for each EATS target in NRLiSt BDB[5].

To address a target endpoint, a set of protein conformations was assembled, one conformation from each conformational cluster. We ran 52 replicas of docking searches within each binding site identified before head in each protein structure.

We are actively developing a scoring function termed Binding Profile (**BP**), which accounts for thermodynamic and conformational properties of the docking experiment results.

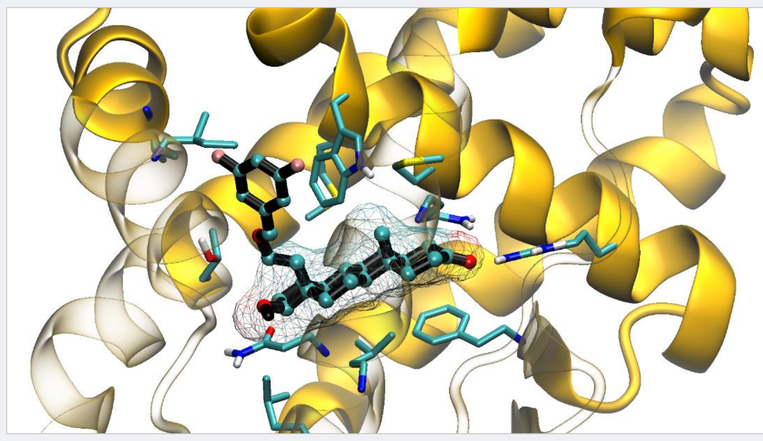
In order to classify a test molecule, its distribution of BP scores obtained from the evaluation of all docking searches addressing the same protein binding site was compared with the corresponding distribution of BP scores computed for a reference molecule (native hormone).

Performance statistics were computed using in-house scripts manipulating external libraries.

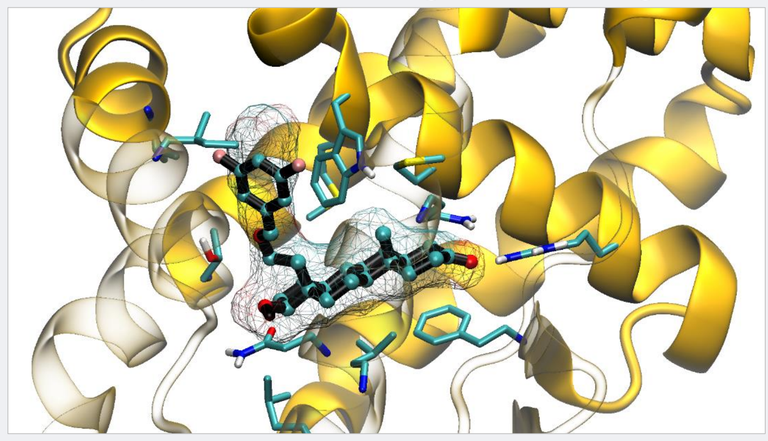
RESULTS OF EXTERNAL VALIDATION

Reliable predictions of protein-ligand binding and accurate quantification of interactions (<1 log unit)

Example of docking experiment results with compound ENM5744[6] (black sticks) in binding site BF-1 of target AR:



Comparison of binding modes of ENM5744 with Testosterone (wireframe molecular surface):
experimental RBA*$_{Tes.**/ENM5477}$ = ~8750;
predicted RBA$_{Tes,/ENM5477}$ = 2677;
*RBA= Relative Binding Affinity; **Test. = testosterone



Comparison of experimental (wireframe molecular surface) and predicted binding modes of compound ENM5477: **experimental binding affinity ~0.0383 nM; predicted binding affinity 0.0294 nM;**
RBA$_{Experimental/Predicted}$ = 1.30

Adaptable highly automated workflow

SESAME-3D proved to be capable of testing any organic compound. It can also be easily extended to investigate any protein target if experimental 3D conformations or trusted predicted 3D models are available.

The workflow has very high sensitivity

The results of this and a previous study (using only a positive validation set of molecules whose binding modes are experimentally known from crystallographic data) show that the framework is extremely good at correctly classifying known ligands.

However, the performance statistics also indicates that the current version of our BP scoring function is not capable of sufficiently distinguishing binding modes of decoy molecules from those of reference molecules, (i.e., scoring the decoy molecules as false positives).

Robust statistical classification

As an example, the comparison of both Binding Profiles (for the docking experiments depicted above) yielded a *p-value = 0.51*, which is not significant at alpha-level 5% or 10%, thus the null hypothesis that compound ENM5477 is a ligand to target AR cannot be rejected.

Metrics	TRα	TRβ	ERα	ERβ	PR	AR
#Positives (ligand dataset)	32	40	40	43	38	45
#Negatifs (decoy dataset)	31	32	32	29	31	30
#TP	27	30	40	43	38	45
#TN	7	3	1	0	2	0
#FP	24	24	31	29	29	30
#FN	5	3	0	0	0	0
Accuracy	0.54	0.55	0.57	0.60	0.58	0.60
Precision	0.53	0.56	0.56	0.60	0.57	0.60
Sensitivity	0.84	0.91	1.00	1.00	1.00	1.00
Specificity	0.23	0.11	0.03	0.00	0.06	0.00
F1_score	0.65	0.69	0.72	0.75	0.72	0.75
ROC(AUC)	0.57	0.54	0.56	0.58	0.62	0.51
Enrichment factor	0.56	0.60	0.73	0.67	0.61	0.64
#top(TP)	18	18	24	29	23	29

Table: Performance statistics at alpha-level 10% for docking experiments within the hormone binding site of EATS targets

Elements of discussion

Although, the performance statistics indicate poor specificity of the framework, it is important to note that the assumed negative dataset consisted of decoys, computationally prepared to be highly similar to the positive ligands. However, the observed enrichment factors are high, leading to the conclusion that the method should be able to correctly classify negative molecules that are dissimilar to the investigated known ligands (NB: future work using molecules negatives *in vitro* for activation of EATS targets).

CONCLUSION & PERSPECTIVES

We present a novel automatic workflow employing cutting-edge drug design modeling techniques, its high-throughput makes it suitable to investigate the binding of chemicals to protein targets as the key initiating event in the scope of toxicological studies. Despite the high computational cost, our framework can provide its prediction results in matter of hours while covering numerous biological targets. Moreover, soon we will proceed to expand the applicability domain of biological targets investigated within the EATS classes and beyond, taking advantage of the already available experimental 3D data of protein conformations from public consortia.

As demonstrated by the validation study using *in vitro* identified ligands and decoy molecules, we are very confident about the high sensitivity of our methodology. We are currently working on assembling a reliable negative testing set in order to assess the specificity of the framework. The BP scoring function will be soon amended with a detailed investigation of the interaction energies. We also intend to expand the conformational space the investigated proteins by simulation of molecular dynamics, allowing our framework to better evaluate the possibility of protein-ligand interactions in the scope of *in-silico* toxicological studies.

REFERENCES

1) Zhaobin Zhang, Chengxia Jia, Ying Hu et. al. (2012). The estrogenic potential of salicylate esters and their possible risks in foods and Cosmetics. Toxicology Letters. 209, 146– 153. DOI: 10.1016/j.toxlet.2011.12.004, 2) The Protein Data Bank H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) Nucleic Acids Research, 28: 235-242. doi:10.1093/nar/28.1.235. 3) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Computational Chemistry 2009, 16: 2785-91. 4) O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. Journal of Cheminformatics, 3(1), 33. https://doi.org/10.1186/1758-2946-3-33. 5) Lagarde, Nathalie, Nesrine Ben Nasr, Aurore Jérémie, Hélène Guillemain, Vincent Laville, Taoufik Labib, Jean-François Zagury, et Matthieu Montes. « NRLiSt BDB, the Manually Curated Nuclear Receptors Ligands and Structures Benchmarking Database ». Journal of Medicinal Chemistry 57, n° 7 (10 avril 2014): 3117-25. 6) Cantin, L., Faucher, F., Couture, J.-F., de Jésus-Tran, K. P., Legrand, P., Ciobanu, L. C., Fréchette, Y., Labrecque, R., Singh, S. M., Labrie, F., & Breton, R. (2007). Structural Characterization of the Human Androgen Receptor Ligand-binding Domain Complexed with EM5744, a Rationally Designed Steroidal Ligand Bearing a Bulky Chain Directed toward Helix 12. Journal of Biological Chemistry, 282(42), 30910–30919.

Poster ID P04-18