Digging Deeper into Deep Data: Molecular Docking as a Hypothesis-driven Biophysical Interrogation System in Computational Toxicology

Michael Goldsmith, Stephen Little, David Reif, James Rabinowitz • U.S. EPA National Center for Computational Toxicology (NCCT)

MOLECULAR MODELS

Background

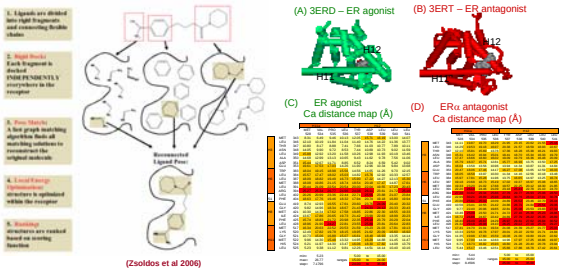
•Developing and evaluating predictive strategies to elucidate the mode of biological activity of environmental chemicals is a major objective of the concerted efforts of the US-EPA's computational toxicology program. Aligning these strategies with the Agency's ongoing chemical-specific risk-assessment needs will provide additional molecular-level insight for decision-making purposes.

•Often, data required for extrapolations inherent in human risk assessment are unavailable. ***In silico*** methods can be used to provide molecular-level information surrogates that are vital for toxicological mechanistic insight.

•Employing a virtual screening approach, a diverse set of chemicals were computationally docked into multiple macromolecular targets (nuclear receptors) using an exhaustive docking algorithm. The individual chemical-target poses, scores, and the chemical-protein contacts generated by this approach afforded a virtual affinity fingerprint matrix that provides mechanistic molecular-level insight. Knowledge gained from quantitative and visual analyses (clustering/heat maps, and linkage networks) of these virtual screens demonstrate the utility of these approaches and their ability to resolve differences in ligand panagonism, receptor promiscuity.

•These virtual affinity fingerprint matrices, coupled to tissue-specific receptor distribution data and inference mapping of downstream signal transduction elements, provide a molecular level of accountability that complements experimental high-throughput screening and toxicogenomic endeavors. ***[This work was reviewed by EPA and approved for publication but does not necessarily reflect official Agency policy.]***

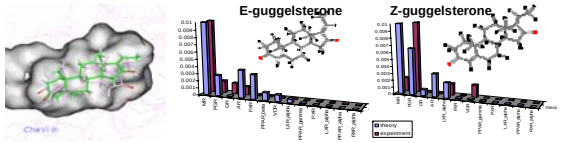
Method: Molecular Docking



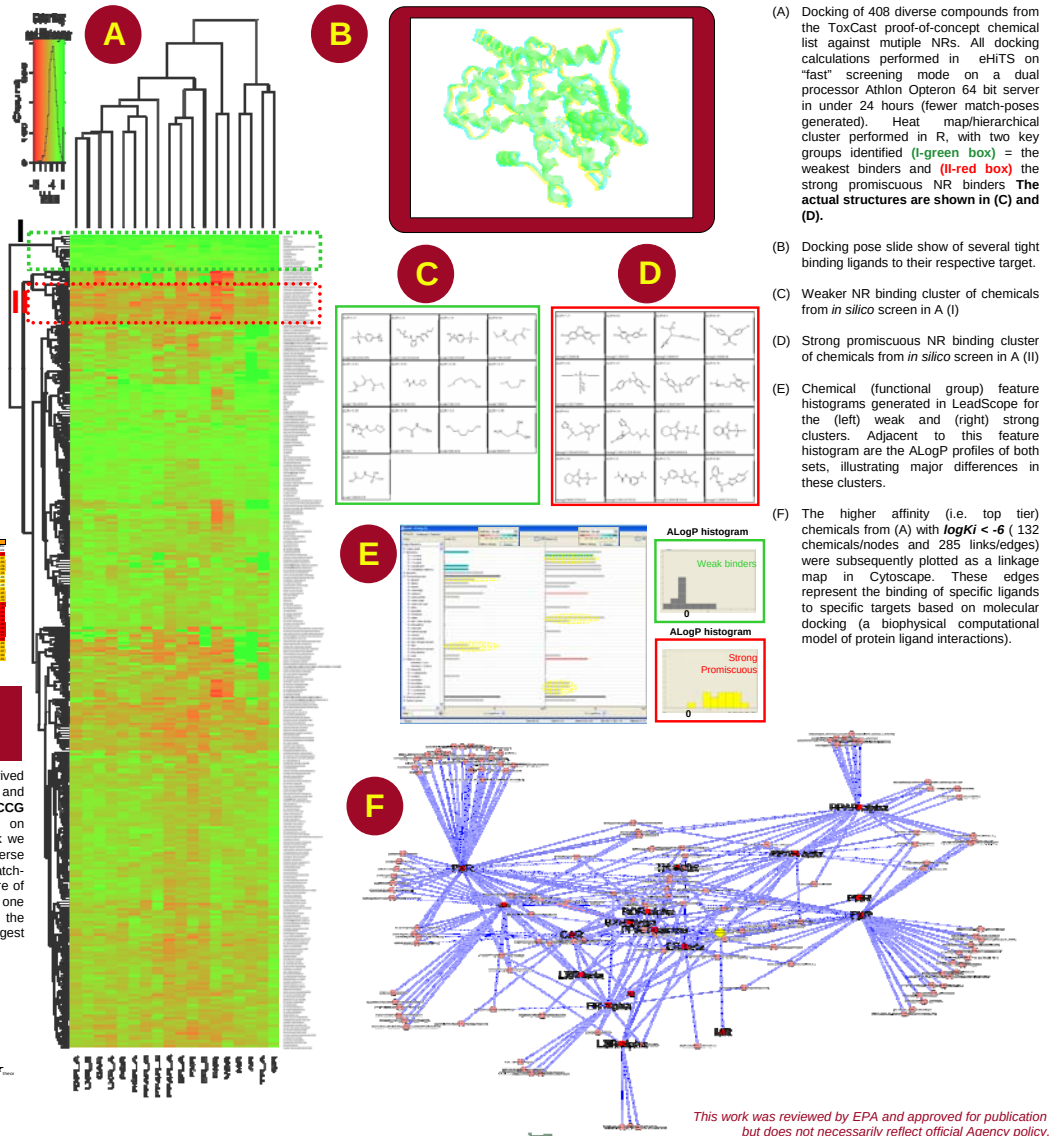
(Zsoldos et al 2006)

Validation: Target Fishing for Promiscuous Ligands

Docking of both E/Z-Guggelsterone geometric isomers against multiple crystal-structure derived human NR targets in their agonist-associated (active) conformation (from **www.pdb.org**) and MMFFx optimized ligand set geometries with AM1-BCC charges assigned from **MOE (CCG Canada)** as found in ***KiBank*** (**Aizawa 2004**), curated from the original publication on guggelsterone polypharmacology (**Burris, 2005**). In the computational toxicology framework we may also pose this question in terms of polytoxicology or *pan-agonism* associated with an adverse rather than therapeutic effect. All performed in **eHiTS** on "fast" screening mode (fewer match-poses generated) (Zsoldos et al 2006) against the diverse set of targets The docked structure of E/Z isomers are shown docked within the binding pocket from MR (mineralcorticoid receptor) one of the top hits for both isomers. The structural formula is shown overlayed on the experiment/theory rank ordered bar graphs (magnitude – normalized binding affinity (K_a) to largest value, so large bars = high affinity).



Results: Docking environmental chemicals into 18 Nuclear Receptors



(A) Docking of 408 diverse compounds from the ToxCast proof-of-concept chemical list against mutiple NRs. All docking calculations performed in eHiTS on "fast" screening mode on a dual processor Athlon Opteron 64 bit server in under 24 hours (fewer match-poses generated). Heat map/hierarchical cluster performed in R, with two key groups identified **(I-green box)** = the weakest binders and **(II-red box)** the strong promiscuous NR binders **The actual structures are shown in (C) and (D).**

(B) Docking pose slide show of several tight binding ligands to their respective target.

(C) Weaker NR binding cluster of chemicals from *in silico* screen in A (I)

(D) Strong promiscuous NR binding cluster of chemicals from *in silico* screen in A (II)

(E) Chemical (functional group) feature histograms generated in LeadScope for the (left) weak and (right) strong clusters. Adjacent to this feature histogram are the ALogP profiles of both sets, illustrating major differences in these clusters.

(F) The higher affinity (i.e. top tier) chemicals from (A) with ***logKi < -6*** (132 chemicals/nodes and 285 links/edges) were subsequently plotted as a linkage map in Cytoscape. These edges represent the binding of specific ligands to specific targets based on molecular docking (a biophysical computational model of protein ligand interactions).

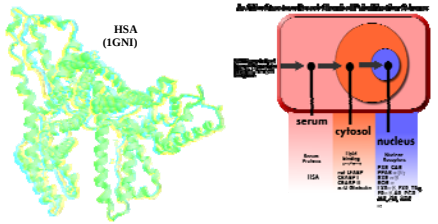
This work was reviewed by EPA and approved for publication but does not necessarily reflect official Agency policy.

Highlights & Future Direction

•The guggelsterone screen qualitatively and quantitatively agree with experiment, a useful tool for ***screening or "fishing" for putative targets***.

•Docking studies demonstrates the practicality of an approach that identifies and clusters both (a) chemicals in a given target-space, and (b) targets in a given chemical space: this experiment has identified compounds that show (I) high NR promiscuity and affinity as well as (II) high NR specificity with varying degrees of affinity. The top structure in this class is strikingly similar (structurally analogous) to tributyltin, a known environmental obesogen. Target-space clustering in the context of these ligands suggests weaker binders are smaller than tighter promiscuous binders, have a higher heteroatom (O,N,S) count (tight have higher halogen count) and have greater degrees of freedom.

•Will consider additional targets **(see schema below)** such as human serum albumin **(shown below in 3D)** and lipid binding proteins required to translocate chemicals from the cytosol to the nucleus.



•Will also perform analogous screen on rodent targets (mouse and rat) for which sufficient in vitro and in vivo data exists, although this may require homology modeling sparse target sets (most crystallized forms of targets shown in this study are protein sequences from humans expressed in a secondary system).

•More efficient identification and enumeration of biologically/environmentally relevant permutations and progeny of the chemical structures in question would be highly desirable. These include stereoisomers, tautomers, protonation states, metabolites and degradation products)

•vHTS studies of parent compounds provide valuable molecular-level detail in the toxicant-target paradigm. These details, along with additional experimental information, may be used for hypothesis generation and are complementary to hypothesis-driven toxicogenomic inquiry.

References

1.(a) eHiTS v. 6.1, Simbiosys Inc. Toronto Canada. (b) Zsolt Zsoldos et al.: "Software tools for structure based rational drug design", Journal of Molecular Structure: THEOCHEM, Volumes 666-667, 29 December 2003, Pages 659-665 (c) Zsolt Zsoldos, Darryl Reid, Aniko Simon, Bashir S. Sadjad and A. Peter Johnson: "eHiTS: An Innovative Approach to the Docking and Scoring Function Problems", Current Protein and Peptide Science, 2006, 7, 000-000 (in press)

2. Burris, Thomas P., Montrose, Chahrzad, Houck, Keith A., Osborne, Harold E., Bocchinfuso, Wayne P., Yaden, Benjamin C., Cheng, Christine C., Zink, Richard W., Barr, Robert J., Hepler, Christoper D., Krishnan, Venkatesh, Bullock, Heather A., Burris, Lorri L., Galvin, Rachelle J., Bramlett, Kelli, Stayrook, Keith R. The Hypolipidemic Natural Product Guggulsterone Is a Promiscuous Steroid Receptor Ligand.; Mol Pharmacol 2005 67: 948-954

3. Molecular Operating Environment (MOE) (Chemical Computing Group Inc., Quebec, Canada)

4. R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

5. (a) Aizawa, M., Onodera, K., Zhang, J.-W., Amari, S., Iwasawa, Y., Nakano, T. and Nakata. K., KiBank: A Database for Computer-Aided Drug Design Based on Protein-Chemical Interaction Analysis. Yakugaku Zasshi, 124(9), 613-619 (2004). (b) Zhang, J.-W., Aizawa, M., Amari, S., Iwasawa, Y., Nakano T. and Nakata, K.: Development of KiBank, a database supporting structure-based drug design. Computational Biology and Chemistry, 28 (5-6), 401-407 (2004).

6. Toxicol Sci. 2007 Jan;95(1):5-12. Epub 2006 Sep 8. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ.

7. M.A. Wolf and A.M. Richard (2006) DSSTox Master Structure-Index File: SDF File and Documentation, Launch version DSSToxMaster_v1a_8804_10Apr2006, www.epa.gov/ncct/dsstox/

(PDB IDs: 1GNI, 1LFO, 1CBR, 1CBS, 2A2G, 1SKX, 1I7G, 1DB1, 2H77, 1KNU, 2AWH, 1A28, 2A3I, 1PQ9, 2ACL, 1P93, 1OSH, 1U3R, 1XPC, 2AMA, 1MV9, 1CBR).

Acknowledgements

MRG was Supported in part by a NHEERL-DESE Cooperative Training in Environmental Research, EPA CR83323601. Special thanks to Dr. T. Transue and Dr. M. Wolf (both Lockheed Martin) for assistance.