TITLE: Informing Mechanistic Toxicology with Computational Molecular Models

Authors: Michael-R. Goldsmith, Shane D. Peterson, Daniel T. Chang, Thomas R. Transue, Rogelio Tornero-Velez, Yu-Mei Tan and Curtis C. Dary

Affiliations:

National Exposure Research laboratory, US Environmental Protection Agency, RTP NC 27711

National Center for Computational Toxicology, US Environmental Protection Agency, RTP NC 27711

Lockheed Martin Information Technology, RTP, NC, 27711

i. Summary/Abstract
Computational molecular models of chemicals interacting with biomolecular targets provides toxicologists a valuable, affordable, and sustainable source of in silico molecular level information that augments, enriches and complements in vitro and in vivo efforts. From a molecular biophysical ansatz we describe how 3D molecular modeling methods used to numerically evaluate the classical pair-wise potential at the chemical/biological interface can inform mechanism of action and the dose-response paradigm of modern toxicology. With an emphasis on molecular docking, 3D-QSAR and pharmacophore/toxicophore approaches, we demonstrate how these methods can be integrated with chemoinformatic and toxicogenomic efforts into a tiered computational toxicology workflow. We describe generalized protocols in which 3D computational molecular modeling is used to enhance our ability to predict and model the most relevant toxicokinetic, metabolic and molecular toxicological endpoints, thereby accelerating the computational toxicology-driven basis of modern risk assessment while providing a starting point for rational sustainable molecular design.

ii. Key Words
Docking, molecular model, virtual ligand screening, virtual screening, enrichment, toxicity, toxicoinformatics, discovery, prediction, 3D QSAR, toxicophore, toxicant, in silico, pharmacophore

1. Introduction

1.1 Overview of Molecular Modeling and its role in Computational Toxicology: Filling the data gaps in Mechanistic Toxicology

Modern computational molecular modeling methods are some of the most well-established, versatile and vital computational chemistry methods that are at the very core of the emerging field of both mechanistic [1] and computational toxicology and sustainable molecular design.[i] The use of molecular modeling coupled to mathematical and chemical–biological inquiry is crucial “to better understand the mechanisms through which a given chemical induces harm and, ultimately, to be able to predict adverse effects of the toxicants on human health and/or the environment” [2].

A first step in considering the use of three-dimensional (3D) computer assisted molecular modeling (CAMM) methods is the awareness of molecular level questions one can address with the various techniques. Molecular modeling can be used in the context of toxicological inquiry to address three

molecular level aspects of both individual small-molecule (ligand) or biological macromolecules (targets), or the resultant interactions of the ligand/target complex , namely: (1) Structure (2) properties and (3) (re)activity. In the context of toxicological and chemical genomic research (or toxicogenomics) one is interested in or requires downstream information that makes use of "optimized" structures or geometries of ligands or biological targets. Of the properties one may be interested in, molecular complementarity is a key objective along with catalytic competence of a chemical and possibly molecular susceptibility (or reactivity of a molecule). Similarly, there are two main research efforts one wishes to inform in mechanistic toxicology, namely:

(1) toxicokinetics or ADME (rate of fate within the body)

(2) toxicodynamics or molecular toxicological interactions that result in a cellular response,

By considering two principal research paths and the biological macromolecular target space to which these coupled processes are related (Table 1) it becomes evident that the subset of molecular modeling tools that will be used by a toxicologist is not much different than the in silico drug discovery workflows [3], with the exception that there is less of an emphasis on lead optimization, and more of an incentive on modeling approaches that possess an ability to both accurately and efficiently prioritize and categorize chemicals to their respective macromolecular targets; in silico methods that are complementary to modern experimental toxicogenomic inquiry.

"Toxico-"	Coupled Processes in Toxicology	Examples of Process relevant Macromolecular protein targets	Molecular Modeling Methods
(I) -KINETICS (biological fate models of chemicals or disposition models)	(A)bsorption (i.e. dermal, oral, inhalation)	-Ion channels (PgP) -molecular transporters - Cell membranes (lipid bilayer considered for passive properties)	A) Geometry Optimization B) Partial charge calculation/assignment Target –Specific endpoint data available C) pharmacophore modeling and D) 3D-QSAR Target Structure available E) Target geometry Optimization and/or homology modeling F) A priori Small-molecule/target interaction evaluation by molecular docking G) Molecular mechanics or empirical pose scoring Structure-based Virtual Ligand Screening (SB-VLS) ...+SB-VLS + SB-VLS + SB-VLS …. =in silico chemical genomics
	(D)istribution (i.e. Target Tissue of target organ)	-extracellular protein binding (e.g. human serum albumin or alpha-fetoprotein or immunoglobulin binding), intracellular solute carrier proteins (eg. SHBG or FABP)	
	(M)etabolism (enzyme mediated chemical transformations typically associated with hepatic clearance mechanisms)	(inhibitors or substrate binding related properties) -Phase I/II enzymes (i.e. CYP450s, oxidoreductases, carboxylesterases, * For kinetic properties, such as rate constants consider QM formalism (see * in methods as well)	
	(E)limination/excretion (Renal or Biliary elimination processes)	-ion channels, organic molecular ionic transporters, globulins, active transporters	
(II) -DYNAMICS (response or effect Models)	Molecular (T)oxicology (ligand /receptor pathways)	Nuclear receptors, G-Protein Coupled Receptors, ion channels (eg. for neuronal impulse propagation or cardiac charge regulation, examples are Sodium-gated Ion channels, or HERG2 channel)	

Table 1 This table shows the overlap between macroscopic and mechanistic toxicology, examples of targets for which pair-wise ligand/target interactions are most often sought after, and the molecular modeling methods used to inform the toxicological questions. Toxicology research streams (toxicokinetics/dynamics), specific toxicology related processes (ADME/T), examples of toxicologically related biological macromolecules implicated in specific processes and several 3-dimensional Computational Molecular Modeling methods (3D-CAMM) are mentioned.

It is estimated that there exist in the order of 7,000,000 chemical leads for small-molecule drug-discovery and ~ 80,000 to 100,000 chemicals under the auspice of environmental chemicals for which the data matrix for risk is sparsely populated (i.e. environmental chemicals), and so there is a need for large-scale screening efforts for prioritization and categorization of these large inventories. [1, 3-6]. Due to the scale of chemical inventories of interest and variety of toxicologically implicated targets of interest, the most appropriate starting point for 3D-CAMM most frequently applied to toxicology (pharmacophores, 3D QSAR and molecular docking) is the use of molecular mechanics force fields to describe or determine the 3D structure of a chemical/biological molecular system of interest. [7-13] In this approach both ligands and biological macromolecular targets are mathematically described and modeled by applying classical Newtonian mechanics to atomic (not electronic) systems which in turn are numerically evaluated using modern computational implementations of the underlying biophysical models. We stress the importance of delineating the fundamental choice of molecular mechanics as opposed to quantum mechanical approaches for answering questions typical of chemical/biological perturbations due to the size domain, and information criteria of the part of mechanistic toxicology one most often wants to inform in the computational toxicology framework; the pair-wise interaction potential between ligand and macromolecular target. To better understand the difference between 2D/3D molecular modeling methods as applied to computational toxicology, we present a symbolic graphic (Figure 1) outlining the three main 3D molecular modeling techniques used in this chapter.

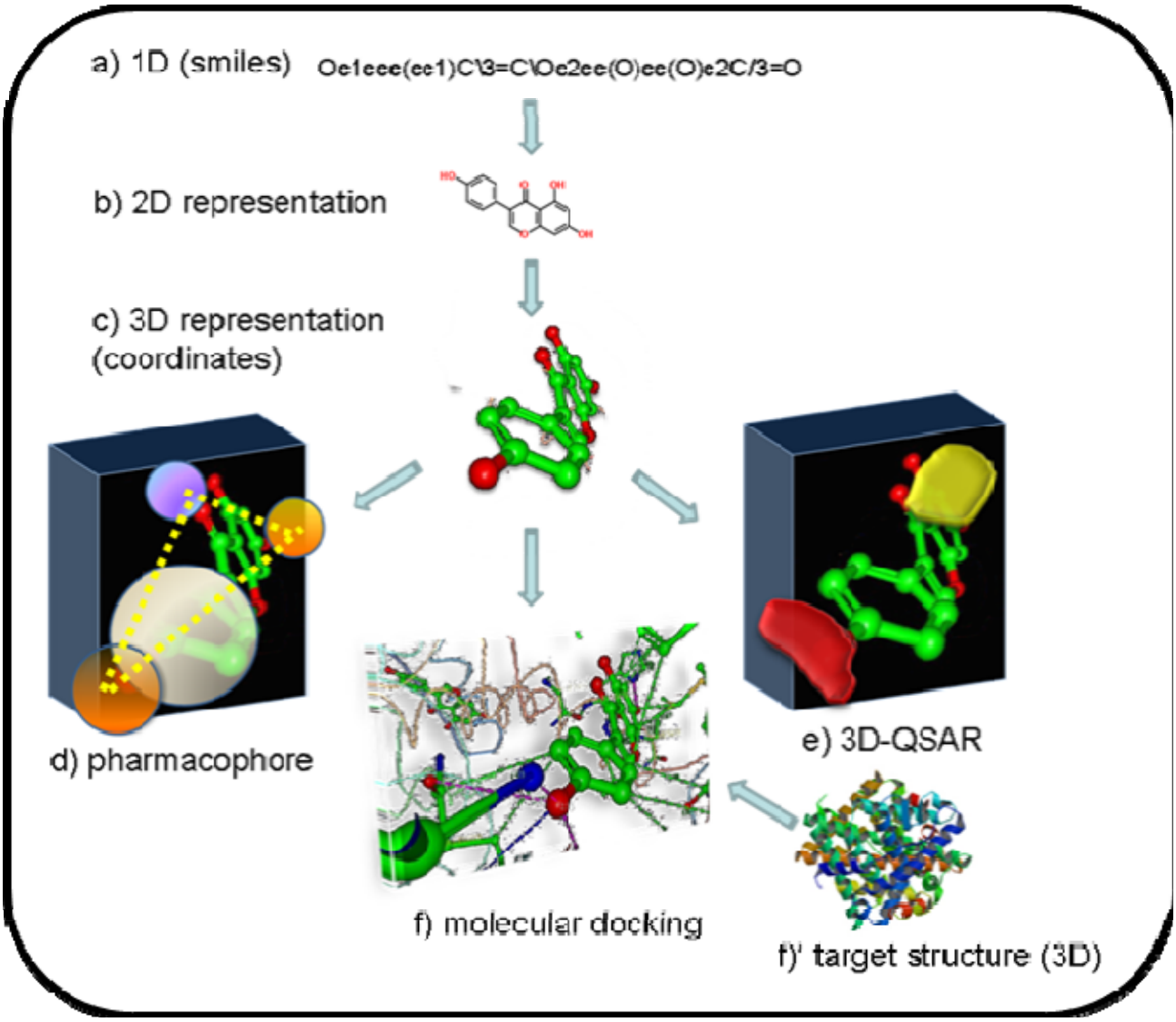


Figure 1. Point of departure from a 1D chemical smiles notation to 3D representation, with atom type and specific coordinates spatially defined. (d-f) the three major classes of molecule modeling methods used to evaluate ligand/target interactions.

Although intrinsic property or functional group chemical filters (leadscope) in addition to classical QSAR approaches [14] and decision tree classifiers [15] are both pragmatic and parsimonious components of the chemoinformatic toolkit of computational toxicologists, they lack the intimate molecular level detail of the biomolecular interaction that could only be resolved by 3D-CAMM. Often chemoinformatics methods alone are unsuitable to address structurally-related questions that require target-specific insight. For instance, for cases that fail to be able to resolve stereoisomerism and its implications in biomolecular interactions, species-related differences in sequences, polymorphism related extrapolations in susceptible populations, and structural bases for mechanistic variability (inhibition versus substrate, agonist versus antagonist), there is little question that primarily 3D modeling methods such as (I) pharmacophore mapping, (II) 3D-QSAR, and (III) molecular docking methods that necessitate detailed structural information (i.e. Cartesian coordinates of atoms and their specific connectivity) and are the only viable alternatives for reliable a priori estimates for risk assessment .

1.2 Exploring ligand:target interactions implicitly: 3D-QSAR and pharmacophores

Unlike specific models of both biological macromolecules and small-molecule ligands, both 3D-QSAR and pharmacophore methods address the fundamental chemical/biological aspects of pair-wise interactions implicitly. Although both deal with the explicit (i.e. full 3D) structure of a chemical of interest, and both require either a training/test set of chemicals with known activities for a given target for a given mode of action (i.e. agonist or antagonist, substrate or inhibitor), neither pharmacophore approaches or 3D QSAR approaches can provide specific molecular level detail between atoms on both macromolecule coupled to those of the ligands that give rise to said activity. Pair-wise interactions between the ligand and the target molecule must be spatially defined. Nonetheless, both methods are a step in the right direction from traditional 2D-QSAR since inherently both 3D-QSAR and pharmacophore models have the ability to discriminate activity based on 3-dimensional topology (i.e. inform stereochemical interactions or regiospecific interactions) without providing residue-specific interactions that could give rise to the specific interaction.

According to IUPAC, a pharmacophore (or in the case of toxicology, a toxicophore) is “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response” [16]. In this sense a pharmacophore model’s objective is to characterize a molecule’s atomic constituents in terms of the primary interaction types that give rise to pair-wise interactions; from multiple atoms to a subset of binding “features.” The features most common for a set of known chemical actors on a known biological target of undefined tertiary structure are: hydrophobic, aromatic, hydrogen bond acceptor, hydrogen bond donor, cationic or anionic or metal interactions. Furthermore there may be exclusion volumes and feature directionality included in the pharmacophore. The pharmacophore features are elucidated by comparing multiple known actors in terms of common overlaid structural features (or alignments). Next, if one is to investigate a series of chemicals and test the pharmacophore one would sample conformational space and any structure that contained a conformation that satisfied the spatial and feature requirements of the pharmacophore model would be considered a complete or partial “hit”.

On the other hand, being able to assess which functional groups or specific spatial features have the ability to modulate the chemical/biological interactions in a quantitative sense is the area of 3D QSAR . Although there are cases of simple QSAR models dating back to the late 1800’s [17], 3D-QSAR is a much more recent approach. While classical QSAR models are useful for rapidly predicting chemically-induced effects based on physicochemical properties, its main weakness is that it does not account for 3-dimensional molecular shape, a critical aspect of intermolecular interaction. Instead of relying on physicochemical properties as molecular descriptors, 3D-QSAR interprets molecular shape using

interaction energies from force field calculations. The huge number of individual interaction energies was historically difficult to correlate with biological activity and it was not until the advent of PLS [18] that 3D-QSAR became technically feasible.

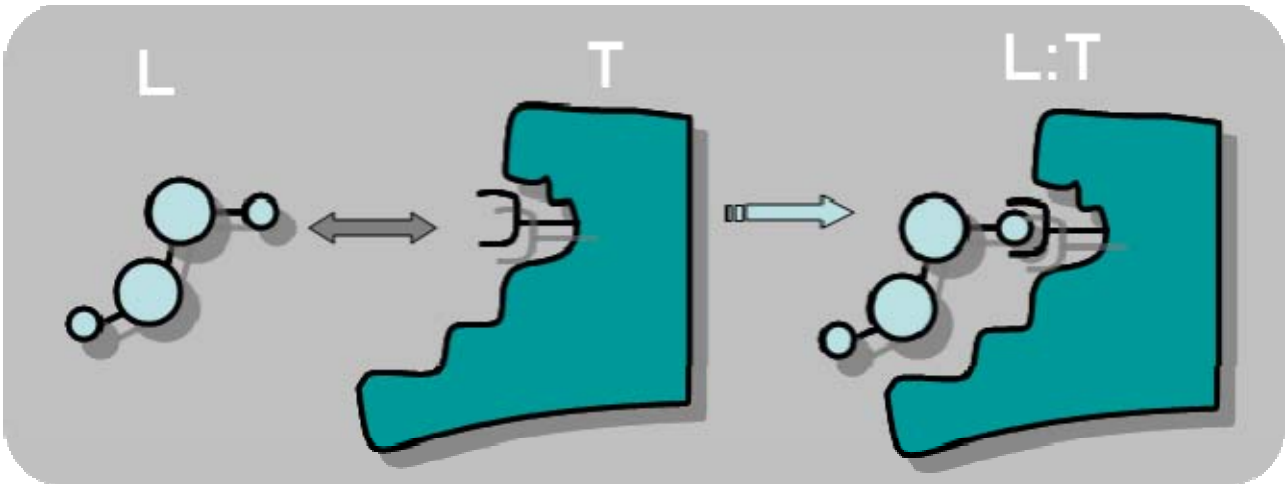
The first, and still most widely used 3D-QSAR method, is known as Comparative Molecular Field Analysis (CoMFA) [19]. Other methods have since emerged, including Comparative Molecular Similarity Indices Analysis (CoMSIA) [20], ALMOND [21], three-dimensional QSAR (TDQ) [22], Catalyst [23] and Phase [23], generally to either improve predictive performance or simplify the model development process. The main drawback of 3D-QSAR is the time requirements and difficulty in preparing the data set for model development.

Further details and steps required for both pharmacophore elucidation/mapping and 3D QSAR as applied to toxicology are elaborated in the Methods section.

1.3 Modeling Explicit Pair-wise interaction potential of ligand-target: Molecular Mechanics, empirical scoring, and the need for structurally informed molecular models

In the specific case of modeling ligand/target interactions for virtual ligand screening as applied to toxicology, certain methods for evaluating pair-wise interaction energy are too computationally expensive/intensive and scape poorly with system size; Quantum Mechanical (QM or sometimes referred to as quantum chemical or electronic structure theory methods) are highly accurate but not ideal (hence not pertinent) due to their computational demand for almost all of the said interaction partners and processes listed in Table 1, with perhaps the exception of bond breaking/making processes inherent in metabolic reactions or irreversible binding. Although the principal focus of in silico methods to estimate metabolic rate constants have been quantum mechanical [24, 25] the majority of pair-wise interactions a toxicologist will require are related to ligand/macromolecular target pair-wise interactions, comprised of both bonded (ligand and target "self energy") and non-bonded interactions between a small molecule and target biological macromolecule (i.e. receptor or enzyme) for which all structural optimization routines are adequate within a classical physics formalism, or more specifically, within a molecular mechanics (MM) framework in which the smallest unit of relevance are atoms (not electrons as in the case of QM approaches).

The classical physics approach to modeling molecules requires the assumptions of molecular mechanics which makes use of atom-specific functions, or force fields parameters, that have been developed by a variety of experimental or high-level theoretical calculations (i.e. ab initio or semi-empirical QM). These are related to atom-specific terms that describe all bonded, and non-bonded interactions (conformational energy, as a function of dihedral angles, bond angles and bond lengths intramolecular electrostatic interactions and van der waals, or dispersion forces) in Cartesian space that are ultimately integrated over all space of the individual molecule or ligand/biomolecule complex to estimate "intermolecular" interaction energy. The "pair-wise interaction potential" between a ligand and a macromolecular target is provided in a simplified form in Figure 2.



$$\mathbf{1}.\Delta E_{L:T} = \sum_{L}\sum_{T}\left(E_{bonded} + E_{non\text{-}bonded} + E_{solvation}\right)_{L:T} - \sum_{L,T}\left(E_{bonded} + E_{non\text{-}bonded} + E_{solvation}\right)_{L,T}$$

$$\mathbf{2}.\Delta E \approx \Delta G_{association} = \Delta H_{L:T} - T\Delta S_{L:T} = -RT\ln\left(K_{A_{L:T}}\right) = -RT\ln\left(1/K_{D_{L:T}}\right)$$

$$\mathbf{3}.K_{D_{L:T}} \approx K_i$$

Figure 2. Fundamental classical expressions evaluated in pair-wise interaction modeling between ligand (L) and macromolecular target (T), and the resulting affinity of the complex (L:T). (1) The first expression relates the energy of the molecular components as a difference of the complex's bonded/non-bonded atomic potential from the energy of the individual partners (L,T). (2) The approximation that the energy function is related to free energy of a system, the thermodynamic representation in terms of enthalpy (H) and entropy (S), and the thermodynamic interpretation of transition-state theory and molecular driving forces for association (K_a). (3) Finally, the relationship between the Ligand:Target complex affinity constant, K_a or dissociation K_d and the approximate translation to a toxicologists metric of the inhibition constant K_i.

In the fields of both statistical thermodynamics and transition state theory for chemical reaction rates, is the relationship between reaction free energy and thermodynamic variables. As provided in Equation 2 of Figure 2, if one has a method for capturing interaction free energy of complex formation (or association) of a ligand/target complex this thermodynamic variable, dG can be cast in terms of an equilibrium process via the expression in lines 2 and 3, where the association constant of an L:T complex, K_a = [LT]/[L][T] K_d = 1/K_a and K_i, the inhibition constant from competitive inhibition assays that in vitro assays often quantify is directly proportional to K_d (dissociation constant) of the ligand with respect to a reference probe [26].

In theory, it is tempting to believe that the free energy from scoring or force-field functions should directly correlate with the experimentally-determined biological activity (K_d or K_i) of complex formation as evaluated by pair-eise interaction schemes, the problem is significantly more involved. The complexity of the problem and inherent simplifications in molecular docking often result in an ability to enrich a dataset in question in such a way that "actives" (i.e. biologically active molecules, or "hits" for a target) considered above some threshold expectation value for binding are guessed several orders of magnitude better than a random guess. For screening this is a reasonable expectation. Details of the various steps for 3D molecular modeling are addressed briefly in the methods section (3.Methods), with focus on how to use these optimized structures for 3D pharmacophore elucidation, 3D-QSAR, and molecular docking. For more extensive methodological resources for any of the methods provided we refer the reader to Table 3 which contains expansions of the topics covered in this chapter. It is strongly encouraged to familiarize oneself with these tools through practice if one wants to apply these techniques to individual toxicological research efforts.

1.4 The use of molecular modeling in computational toxicology: The integrated modeling workflow to in silico chemicalgenomics

Although we have provided an overview of the most popular and useful aspects of 3D-CAMM that could be used to inform mechanistic toxicology, we need to understand how they fall in to the computational toxicology framework. To know how and when these methods are applied in practice, and by whom, we have devised a workflow (Figure 3) that highlights some of these components and how they may complement experimental High Throughput Screening protocols. The objective is to enrich the understanding of chemical/biological interactions through toxicogenomic inquiry. This is achieved by an in silico (filters -> 2D QSAR > pharmacophore -> docking/3D QSAR) tiered approach that is tightly coupled to experimental in vitro screening efforts (i.e. protein ligand binding assays, transient activation assays, gene expression profiling, cytotoxicity assays, etc) to encode a chemical-specific biological activity fingerprint or signature. This conceivably can also be performed in silico using multiple target screening, and used as a metric for chemical/biological activity comparisons (i.e. similarity based on multi-target virtual affinity fingerprint as opposed to structure alone).

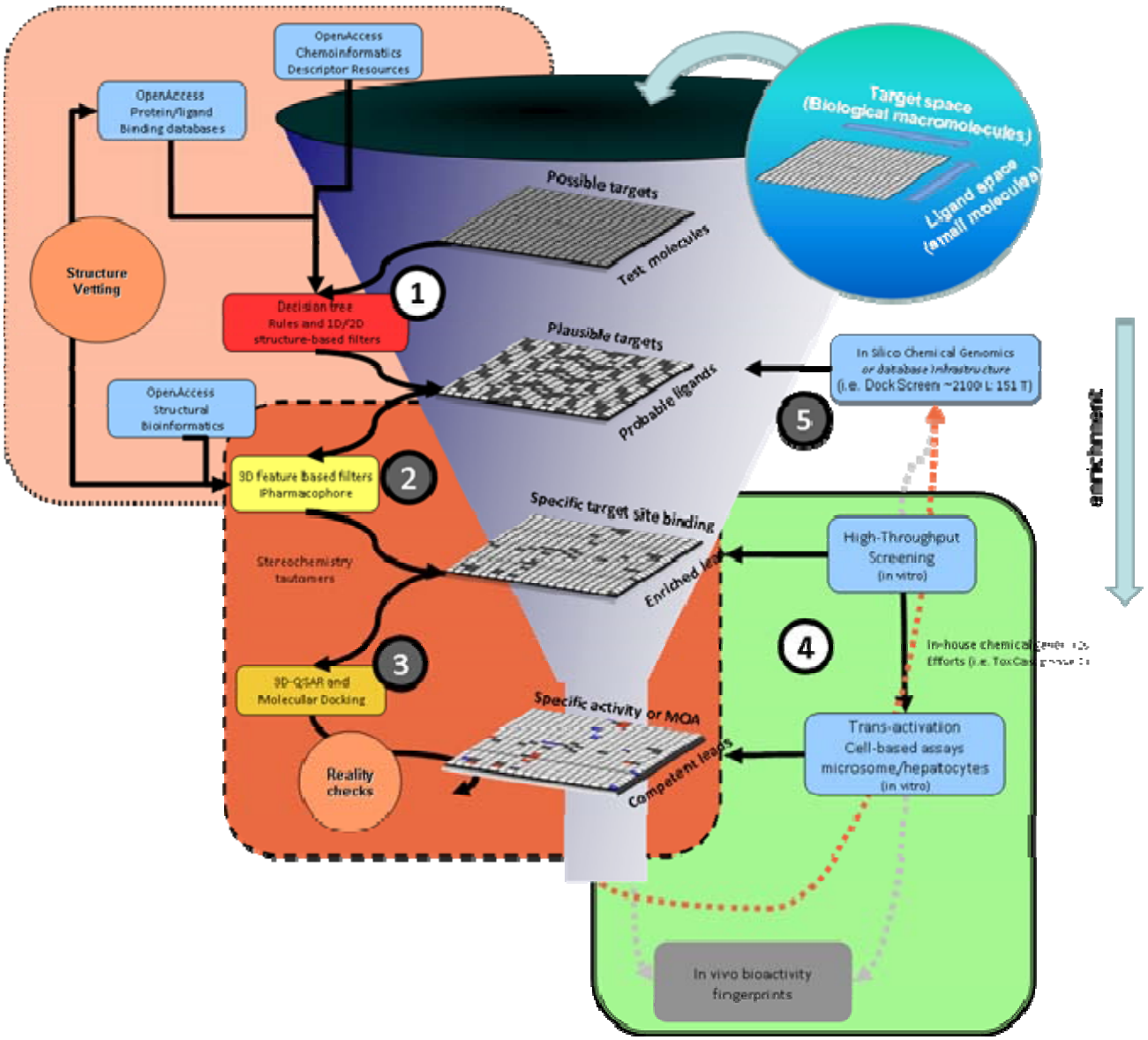


Figure 3. Computational toxicology modeling workflow showcasing the in silico, In vitro and in vivo integration of data and models within an informatics framework.

All of the data from a tiered approach to virtual ligand screening could and should ultimately be encoded or captured within a database framework so that easy recall could be performed to inform molecular level resolution data gaps as they arise. We add that the development of a database infrastructure that can subsequently capture the resultant poses and pair-wise interaction energy (surrogates for affinity) holds value in being able to query molecular-level insight for an experimental chemical genomics screen.

We provide, in brief, a workflow that demonstrates how to pair or couple experimental, in silico, and 3D in silico methods and the various pipes of data that allows one to build a virtual ligand–target complex structural database. This type of strategy had been adopted to build our own in-house resource to support toxicogenomic inquiry (DockScreen) which is explained in the Examples section.

2. MATERIALS

There are well over 350 independent packages (computational codes) available for various aspects of the molecular modeling or Virtual Ligand Screening (or 3D VLS) paradigm that capture the various components required for 3D modeling of ligand / biomolecule interactions: all chemoinformatics and QSAR development, docking, homology modeling, pharmacophore elucidation, chemical structure manipulation, structure building, refinement, optimization, and finally bioinformatics applications.

For the case of computational toxicology the lead optimization procedure/process typically associated with in silico drug discovery or rational drug design and associated methods and coded implementations are essentially dropped (although they may persist for sustainable molecular design). These packages run on many different platforms including but not limited to Windows/PCs, SGI, Mac, Linux (UNIX workstations), and some limited functionality molecular modeling utilities are even available for hand-held devices and smart-phones. For practical purposes, we have typically chosen one of several commercial suites that with the following features:

a) Platform independence (works on heterogeneous network architecture)
b) Token-key license structure (check out by user when required)
c) Many independent molecular modeling methods, bioinformatics, chemoinformatics and data mining methods combined
d) Built-in functionality for scripting, piping data, and automated/macro workflows
e) Is well documented and has good active and passive support networks (technical service, and FAQ/scripting forums)

Public available resources for the “non-expert” or experts are included in Table 2 and provide numerous links for a variety of software packages, both commercial and open-source, in addition to visualization tools and databases relevant for informing ligand/target pair-wise interaction moedeling.

Individual 3D CAMM lists	Uniform Resource Locator (URL)
Directory of In SIlico Drug Design Tools (Swiss Institute of Bioinformatics)	http://www.click2drug.org
Universal Molecular Modeling List (NIH)	http://cmm.info.nih.gov/modeling/universal_software.html
Free computer tools in Structural Bioinformatics and Chemoinformatics	http://www.vls3d.com/links.html
Computational Chemistry List, ltd. (CCL.NET) Software Related Sites (Note, these include "ALL" chemistry related sites above and beyond the scope of this paper	http://www.ccl.net/chemistry/links/software/index.shtml
Virtual Library: Science: Chemistry: Chemistry Software Houses from the University of Liverpool (UK)	http://www.liv.ac.uk/Chemistry/Links/softwarecomp.html

Table 2: A list of several comprehensive software/tool/data resources lists available on the WWW that provide access to various commercial and open-source software packages, in addition to open-access database resources.

From an application stand point the authors have required both bioinformatics and chemoinformatics tools, structural database capabilities, and the ability to perform geometry optimization of structures, molecular docking, homology modeling of target structures, conformational searches, pharmacophore elucidation, and QSAR development. However, we have primarily used Chemical Computing Group's Molecular Operating Environment (MOE) [27] for all database manipulation, QSAR development, library development, structural optimization and descriptor calculation. Similarly, for ADME related parameter estimation via QSAR we use Schrodinger's QikProp [28] which has been vetted against various animal and human drug targets or ADME related endpoints (i.e. LogPBB, LogKhsa, #metabolites, CACO2, or MDCK permeability, etc.)

3. Methods

As mentioned in the introduction, the classical physics approach to modeling molecules requires the assumptions of molecular mechanics which makes use of atom-specific functions, or force field parameters, that have been developed often for specific classes of molecules. Force field calculations have been successfully performed on larger polypeptides and protein structures. Park and Harris [24] utilized AMBER force fields to develop an all-atom model for CYP2E1 which was subsequently used for docking studies. A comprehensive review of AMBER protein force field development can be found elsewhere [7, 29]. Several studies have also assessed the relative performance of CHARMM, MMx, OPLS and AMBER force fields [30, 31]. Gundertofte et al. [32] have assessed the relative accuracies of MMx and AMBER force fields. Jorgenson et al. [33, 34] have also examined the performance of their OPLS force field in the context of proteins and organic liquids. Regardless of the framework details, a molecular mechanics force field is always chosen for structural optimization, and the specific force field selected is usually chosen that best captures the atom-type diversity in the dataset (i.e. chemical space of the training fragments or atoms). A broad overview of all the step-wise modeling procedures is provided in Figure 4.

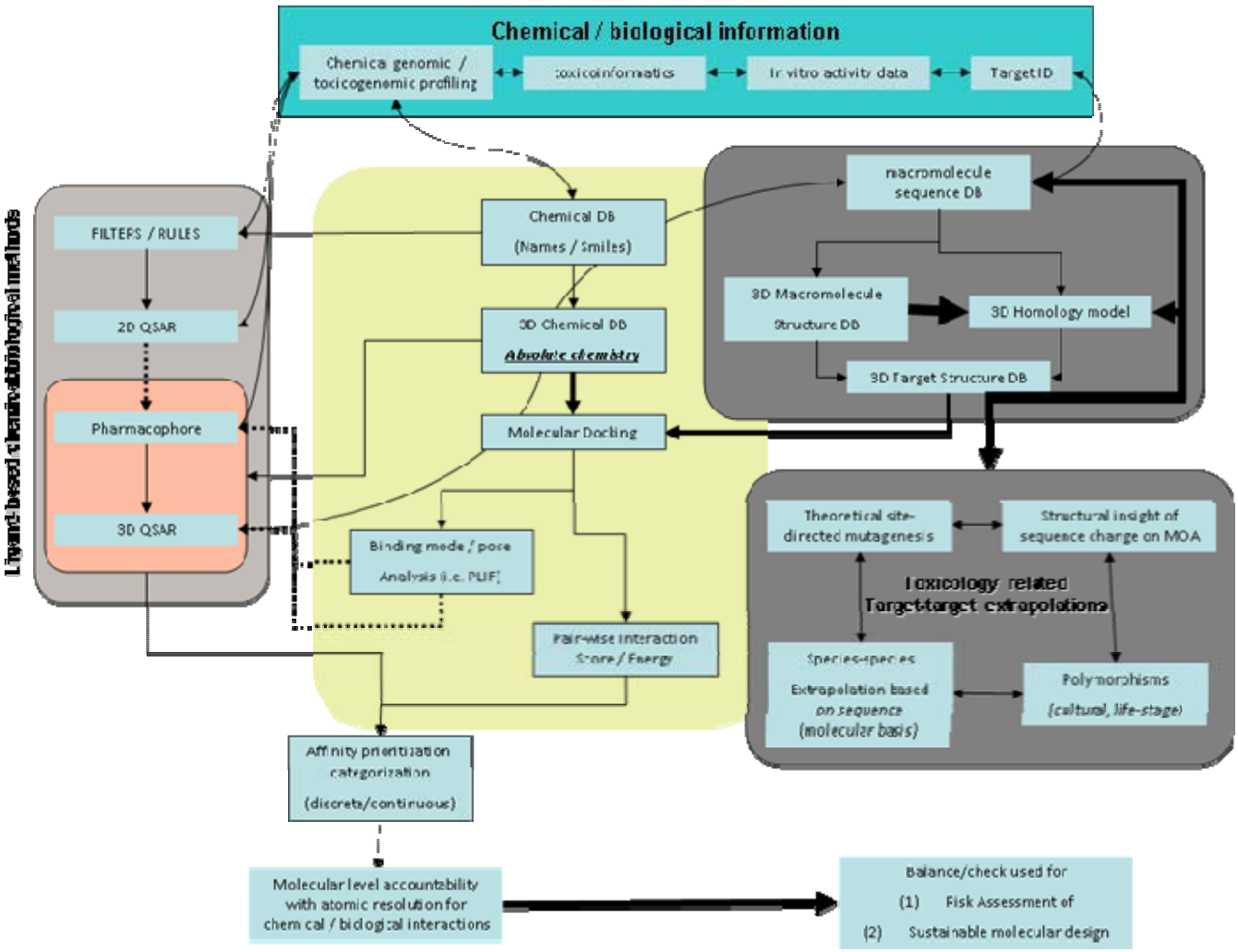


Figure 4. Specific steps in 3D-CAMM (computer assisted molecular modeling) Workflows, chemical/biological knowledge-based boundary conditions (top box), ligand-based approaches (grey, left box), structure-guided methods (yellow central box), structural biological target inventories and types of questions related to toxicologically relevant target-target extrapolations one can inform from structure based approaches (both dark grey boxes).

For our molecular modeling needs, (i.e. in the case of small molecule ligands with environmental chemicals) we have almost exclusively used the MMFFx force field (MMFFx, [35]) for 3D geometry optimizations of entire libraries of chemicals.

Subsequently, partial atomic charges are assigned either using empirical (i.e, Gasteiger) or semi-empirical based charge model representations of the electrostatics of the system (i.e., AM1-BCC) and are stored in a 3D chemical structural database.

These structures could be used directly with target specific activity information as the seed for a conformational search (spanning all rotatable bonds to predict other relevant geometries) and aligned to other known biologically-active chemicals to generate 3D-QSAR or pharmacophore models.

However, if the structure is known and the target protein sequence is known and a crystal structure or near-neighbor homolog exists, it is conceivably simple to optimize hydrogens on the crystal structure obtained from the literature, or perform theoretical site-directed mutagenesis or threading, the basis for homology modeling based off of a known structural template.

Finally, with an optimized target structure database, and an optimized ligand database one could perform molecular docking experiments where the pair-wise ligand:target interactions (bonded and non-bonded terms) are systematically evaluated. The resulting poses from such a docking "run" can each be individually scored based on known binding affinity. There are numerous online resources that provide ligand/target binding affinity data (i.e. www.bindingdb.org).

Using these rank-ordered lists of chemicals based on scored docking poses between a small molecule and a macromolecular target is the starting point for a prioritization or rank-order scheme for screening a specific target: virtual structure (macromolecule) based virtual ligand screening.

A library of structural targets of interest that may have been selected based on their role in a major toxicity pathway a researcher may be studying, has value in being able to fish for targets of any chemical [36]. The next section elaborates on the capabilities of a large-scale ligand/target screening initiative.

For detailed description of external methods we encourage the readers to consult Table 3, which contain more detail for each aspect of the various steps of molecular modeling. Next we provide a step-wise breakdown of various modeling steps required for evaluating ligand/macromolecular target interactions.

Step	Systematic Methods Reference
Molecular Docking	Morris, G. and M. Lim-Wilby, Molecular docking. Methods in Molecular biology (Clifton, NJ), 2008. 443: p. 365.
A general introduction to molecular modelling techniques in the area of protein–ligand interactions	a) Kroemer, R., Molecular modelling probes: docking and scoring. Biochemical Society Transactions, 2003. 31: p. 980-984. and (b) Van Dijk, A., R. Boelens, and A. Bonvin, Data driven docking for the study of biomolecular complexes. Febs Journal, 2005. 272(2): p. 293-312.
Docking Scoring Functions	Pick, D., Novel Scoring Methods in Virtual Ligand Screening. Methods in Molecular biology, 2004. 275: p. 439-448.
Chemical Database Preparation	Bologa, C., M. Olah, and T. Oprea, Chemical database preparation for compound acquisition or virtual screening. Methods in Molecular biology (Clifton, NJ), 2005. 316: p. 375.
Target selection criteria	Wishart, D., Identifying putative drug targets and potential drug leads: starting points for virtual screening and docking. Methods in molecular biology (Clifton, NJ), 2008. 443: p. 333.
Virtual or in silico affinity fingerprints	Briem, H. and U. Lessel, In vitro and in silico affinity fingerprints: Finding similarities beyond structural classes. Perspectives in Drug Discovery and Design, 2000. 20(1): p. 231-244.
3D Structure-Based Virtual Ligand Screening Resources and brief overview	Villoutreix, B., et al., Free resources to assist structure-based virtual ligand screening experiments. Current Protein and Peptide Science, 2007. 8(4): p. 381-411.
In silico chemical genomics – Target and Ligand Preparation	Jongejan, A., et al., The Role and Application of In Silico Docking in Chemical Genomics Research. Methods in Molecular biology (Clifton, NJ), 2005. 310: p. 63.
Analysis of Chemical Space in the context of Domain of Applicability	Jaworska, J., N. Nikolova-Jeliazkova, and T. Aldenberg, QSAR applicability domain estimation by projection of the training set descriptor space: a review. ATLA-NOTTINGHAM-, 2005. 33(5): p. 445.
Analysis of docking data	Bender, A., et al., Chemogenomic data analysis: Prediction of small-molecule targets and the advent of biological fingerprints. Combinatorial Chemistry &# 38; High Throughput Screening, 2007. 10(8): p. 719-731.

Table 3: A comprehensive set of combined reviews and/or methods papers for various aspects of the 3D molecular modeling methods discussed in this chapter. We urge the reader to familiarize themselves with each of the steps associated with their modeling method chosen and the particular toxicological data gaps they may wish to address.

3.1 Ligand Preparation

1) Collect a list of all the chemicals of interest
   a) Dataset augmentation (i.e. adding "simulated metabolites"- Enumerate metabolites using a heuristic (or knowledge base) metabolite enumeration algorithm [37, 38].
2) Curate this list with the smiles representation of the structure of interest
3) Convert this 2D chemical structure dataset to a 3D representation by assigning the chemicals absolute configuration which includes the atom types but their 3D connectivity (bond type and orientation) and selecting an appropriate molecular mechanics force field and charge model of interest (depending on the chemical space of the chemicals of interest in addition to the magnitude of the screening initiative (i.e. hundreds to thousands of chemicals one would be better of going to no more than a classical physics approximation of the molecular geometry)
4) Assign charges to the geometry optimized structures
5) Refine the dataset (see Table of Methods) – [39]
   a) Consider charge state and charge model, force field and domain of applicability dependent on the nature of your chemical
6) Capture all 3D geometries into a database.
   a) Almost all major molecular modeling suites (e.g. Chemical Computing group's Molecular Operating Environment [27], Accelrys Discovery Suite [40], Schrodinger [28] and Tripos [41].) provide database representation of the chemicals of interest, so converting from smiles code to 3D optimized, cleaned, and charge-model applied 3D representation is relatively seamless
   b) STOP

3.2 Target Preparation

1) Coupled to experimental knowledge, searching through chemical genomics databases such as http://stitch.embl.de or the comparative toxicogenomics database (http://ctd.mdibl.org/) often identifies good, relevant targets for a chemical or analog of interest.
2) Finding a suitable target model (typically an X-ray crystal structure from http://www.pdb.org) is the next step. Before assuming that a given target structure will serve as a sound basis for molecular modeling studies, it is critical to understand that "protein structures" are models. Although they are based on experimental data, they are nonetheless prone to bias or ambiguity from several sources. While numeric metrics such as resolution, R-factor, free-R, redundancy, and average I/sigma (signal to noise ratio) are important considerations for the overall reliability of a crystal structure model, at least some local errors or ambiguities are found in nearly all structures. Active sites are often somewhat rigid, especially when bound to ligands, so one can hope that the structure of interest is a sound choice. However, there is usually no substitute for examining electron density (See notes section below and [42]).
3) Selecting the appropriate structure if confronted with several? Perform an RMSD evaluation on a structural superposition. If the geometries are similar they may cluster into most-probable conformation states. Select a representative from each cluster.

4) If the target structure is known and the sequence is known and a crystal structure or near-neighbor homolog exists, it is conceivably simple to optimize hydrogen atoms on the crystal structure obtained from the literature, or perform theoretical site-directed mutagenesis or threading, the basis for homology modeling based off of a known structural template.

5) If the target structure is not known and one wishes to perform structure-based virtual screening or molecular docking, one must build a homology model of the structure of interest using the sequence of the desired target (from www.uniprot.org) and a crystal structure template of the nearest-neighbor homolog (template or crystal structures from www.pdb.org and homology or sequence identify search using BLAST. A protein homology model server is available for integrated web modeling at www.proteinmodelportal.org .

3.3 Molecular Docking

1) With an optimized target structure database, and an optimized ligand database one could perform molecular docking experiments where the pair-wise interactions (bonded and non-bonded terms) between the ligand and the macromolecular target are systematically evaluated. The resulting poses from such a docking "run" can each be individually scored based on known binding affinity data training set of chemicals for a given target.

2) A binding site is identified (co-crystallized ligand site, or rationally selected site) and each ligand is subject to interact with the macromolecular target, where sampling and docking trajectories are subject to the force field approximations. Each individual "pose" is scored or captured for subsequent analysis.

3) Each of the poses are systematically scored using pair-wise interaction potentials that are either derived from classical physics approaches (i.e. force field approximation) or empirical scoring functions that have been optimized to reproduce either experimental in vitro binding affinities (trained scoring function).

4) The results are subsequently validated for their ability to enrich MOA data, or rank-order chemical binding for a known target. Another common validation protocol that has less to do with the binding affinity and more with pose analysis is the ability for the docking algorithm to reproduce the original co-crystallized ligand in the same geometry. Methods that minimize the RMSD between known pose and docked pose are considered optimal. This approach of being able to reproduce experimental crystal structures is termed "pose fidelity" [43, 44] and references within;

5) Docking "experiments" can form the basis of continuous complementarity evaluation of ligand/target complexes (unlike experimental, that rely on binding stronger than a probe chemical, or else result in a "NA" or blank result.). Since one can take the top and bottom rank-ordered chemicals for a target and deduce chemoinformatic filters (i.e. intrinsic functional property or functional group profiling) one could conceivably perform what is known as "progressive docking", where filters from molecular docking simulations are used to create "structure-guided" filters for subsequent chemicals. (Progressive Docking: A Hybrid QSAR/Docking Approach for Accelerating In silico High Throughput Screening [45].

6) Details about assumptions and expectations from structure-based virtual ligand screening models, and the very nature of the target structure used are enumerated in the "Notes section".

It is very important to be aware of the various issues that lead to mismatched expectations (i.e. surprise) when attempting to apply these 3D-CAMM approaches.

3.4 3D QSAR

This section is intended to provide a brief overview of 3D-QSAR, outlining its advantages and disadvantages and describing the basic steps required to derive and validate a model. For greater detail on the topic, the reader is referred to external references. [46-48]. Figure 5 outlines the basic steps involved with developing, validating and using a 3D-QSAR model. Although not listed on the figure, the first step in deriving a 3D-QSAR model is really defining the applicability domain or the chemical space comprising the set of compounds for which the model is to be used. A good way of rapidly assessing the chemical space of two chemical lists is using ChemGPS-NP [49, 50]. When that has been defined, a representative sample of that compound list with known biological activity must be chosen for further development. 3D-QSAR models have been developed using data sets ranging from as little as 20-30 compounds or as much as 200-300. A large data set of compounds will likely cover a larger chemical space, but considerations of the cost of biological testing usually limit this size. Another consideration in the initial stages of data set design is the overall span in biological activity values: a span in activity values of 5 log units is generally considered to be the minimum requirement.

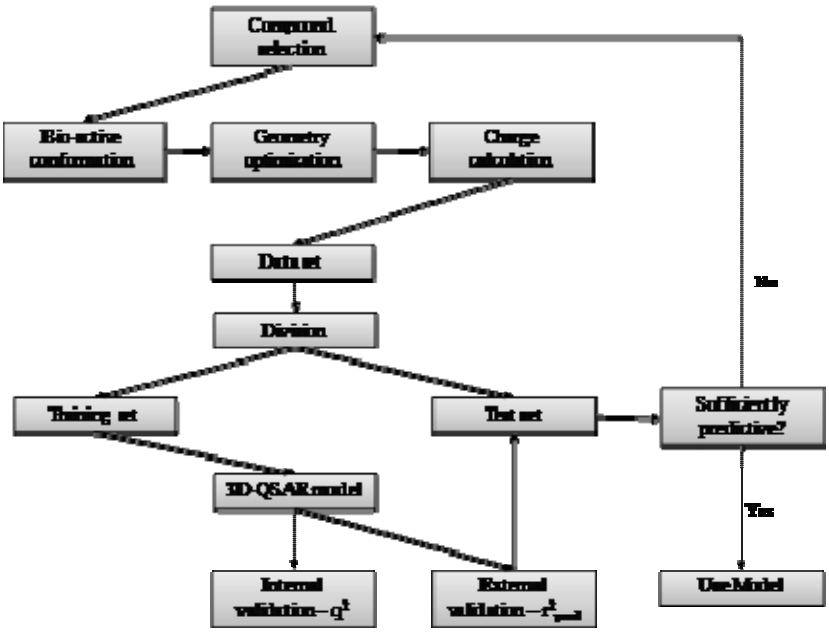


Figure 5. Workflow for deriving, validating and using a 3D-QSAR model.

1) After data set compounds have been carefully chosen, several steps of molecular modeling are necessary to ensure

   a) they are in their biologically active conformation,

   b) geometry optimization has been performed and

   c) accurate partial atomic charges have been assigned.

2) Of these steps, identification of the biologically active conformation is the most important and most difficult. Geometry optimization and charge calculation methods generally have a much lesser effect on model predictive performance.

   a) data set is carefully divided into training and test sets. The training set is a subset of the data set used for deriving the 3D-QSAR model and usually comprises roughly two-thirds of the original data set. The test set is then comprised of the remaining one-third of compounds and is used for evaluation of the predictive performance of the model. Care must be taken to ensure the training and test sets have similar coverage of chemical space as well as a similar span in activity.

   b) Following data set division, the final software-dependent steps in 3D-QSAR model derivation may be taken. Since 3D-QSAR correlates biological activity with differences in structural features, these final steps generally involve aligning important pharmacophore groups or features of the chemical scaffold to bring to light those structural features that impact biological activity.

3) After the 3D-QSAR model has been derived it must be validated. The first step of the validation process normally includes internal validation by cross-validation, which gives an indication of the strength of correlation within the training set compounds. Although this is a useful metric, it gives little indication of how well the model can predict activity data for compounds not included in the training set. To understand this, the model is used to predict the activities of test set compounds. These predicted values are compared with the previously-known activity values to calculate r^2_{pred}, a measure of external predictive performance.

4) If the model is found to be sufficiently predictive following these tests, it may be used to perform predictions on similar compounds for which biological activity is unknown. If it is not, the user must repeat the previous steps leading up to model derivation until the model is sufficiently predictive.

3.5 Pharmacophore / toxicophore elucidation

1) Taking the optimized ligand geometries and knowledge of the specific target for which these chemicals interact with it is possible to superimpose the various ligands, or conformations of the

various ligands, to re-recreate or infer the optimal features for a specific experimental mode of action.

2) Negative data is especially useful in pharmacophore models as it allows one to rule out "impossible" superpositions (hence better molecular level boundary conditions) and increase the predictive accuracy of models to predict "hits" or "non-hits."

3) Flexible alignment or ligand superposition is preceded by geometry optimization and conformation enumeration of each of the 3D geometries that can be made by performing either stochastic or deterministic molecular simulations within a molecular mechanics framework to systematically alter the dihedral angles of the chemicals of interest and localize multiple low energy conformers or rotamers.

4) Once alignment procedures between chemicals have been completed, one often finds common molecular features that one can reduce the ligand structure into (i.e. hydrogen bond donor, hydrogen bond acceptor, hydrophobic contact, aromatic contacts, metal interactions, cationic interactions, anionic interactions). These features can include exclusion volumes or cavities that "wrap" the outer volume of a set of known ligands. For any given chemical, if a conformation falls within the cavity, and the spatial relationship between features are either completely or partially satisfied one would identify "potential hits".

4. Examples

In order to familiarize toxicologists with pertinent examples for further exploration (I) we briefly outline some of the key papers in table format (Table 4) as they address mechanistic toxicology questions, the modeling approaches taken, the software used, and the literature reference of the research effort. Finally (II) we provide a brief description of an in-house in silico chemical genomics program and the key actions taken to bring an in silico molecular modeling results database to fruition to complement screening and toxicogenomic efforts.

Example(s) I: 3D CAMM to inform toxicology, single target research.

Tox data gap	Modeling method chosen / software used	Chemical/target space	citation
1. Absorption	Catalyst (Accelrys) 3D QSAR Pharmacophore Homology modeling,	TARGET: P-gp (P-glycoprotein efflux transporter) LIGANDS : 27 digoxin inhibitors, 21+17 vinblastine inhibitors,	[51]
2. Distribution	a) Homology modeling, 3D QSAR, molecular dynamics, and molecular docking (various - Glide – Schrodinger, and Chemical Computing and Tripos suites) b) Molecular docking (Autodock) molecular modeling (Macromodel/Schrodinger), theoretical site-directed mutagenesis (Sybyl, Tripos) c) 3D-QSAR and molecular docking	TARGET: sex hormone binding globulin / LIGAND: 80,000 ligands TARGET: Human Serum Albumin binding (plasma binding), LIGAND: < 10 structurally-related chemicals to naturally occurring ochratoxin TARGET: Human Serum Albumin binding (plasma binding), LIGAND: 37 structurally related putative interleukin 8 inhbitors	[52-54]
3. Metabolism	Pharmacophore and 3D QSAR	a) TARGET: CYPs 1A2, 2B6, 2C9, 2D6, 3A4 LIGANDS: various drug-like / leads b) TARGET: PXR LIGANDS: various drug-like / leads	[55-57]
4. Elimination	Pharmacophore and 3D QSAR (various in-house packages such as CAMDA, as well as commercial suites such as Sybyl (TRIPOS)	TARGET: rat multi-drug resistance-associated protein 2 and 1 / LIGAND: >4000 conformers of >18 metabolism-like leads.	[58, 59]
5. Molecular	A) homology modeling and molecular docking	TARGET: Estrogen receptor alpha (i.e. nuclear receptor) or	[60, 61]

toxicology (receptors)	(OpenEye's FRED and eHITS)	5-HT6 receptor based off the rhodopsin crystal structure (GPCR modeling)	

Table 4: Selected research papers that exemplify 3D CAMM to inform mechanistic toxicology.

Example II : Building an in silico chemical genomics framework using 3D CAMM to inform toxicology on multiple targets

This next example focuses primarily on the components required to build a multiple target large-scale docking inventory to support chemical genomic research. In an effort to dig deeper into toxicogenomics data we had built an in-house in silico chemical genomics infrastructure. The idea was to complement HTS and chemical genomics programs and support a fully integrated in silico, in vitro, in vivo computational toxicology framework.

Particular examples of building a multiple target, multiple chemical docking database are relatively rare in the literature, and for the most part support drug discovery research. We wanted to demonstrate to the reader how this was done at an overview level of detail to show key considerations, infrastructure and coding requirements needed to be assessed. In this case, we were building a database to inform the data matrix required for risk assessment of many environmental chemicals. One of the efforts from the US-Environmental Protection Agency's National Center for Computational Toxicology is the ToxCast program [5]. In this case, for phase I a total of 320 chemicals (DSSTox dataset) were selected for thorough in vitro work-up. What was not provided by any of these assays, however, was the type of information at molecular resolution as obtained by structure-based virtual ligand screening or in silico chemical genomics. In this case an in silico chemical genomic intiative, DockScreen, was started [62-64].

The Dockscreen data is the result of about 2,500 ligands docked into about 150 protein targets using the eHiTS (Simbiosys Ltd, Canada) software package. The result was over 350,000 docking runs resulting in over 9 million ligand poses. These calculations were performed over a period of 2 months on 20 servers and collected a total of ~ 250 Gb of coordinate specific pose data for each of the 2500 chemicals on each of the 151 targets. To store and manage queries to access this data, a MySQL relational database schema was designed with separate tables for ligands, proteins, docking runs, and poses as well as some computational statistics. A custom interface to this database was built in a Linux OS environment with a PHP enabled Apache web server. The acronym "LAMP" is often used to refer to such a combination of Linux, Apache, MySQL, and PHP which have been used in combination to provide web access to many databases. For Dockscreen, only a dozen or so PHP scripts were needed to let users to view, query and select groups of ligands, groups of targets, and statistical calculations on the distribution of docking scores for the runs including such ligands and targets. In addition to numeric statistics, histogram graphics were produced on the fly. An applet allowing users to draw chemicals and search against ligands is also built in. We believe in house tools such as these that provide scientists relatively fast access to "molecular-level" target binding properties and poses is critical for those wishing to focus on chemical risk assessment at a molecular level of accountability.

5.Notes

The good modeling practices as discussed more broadly throughout this book in other chapters still apply to molecular modeling: that is in order to keep an audit trail of the steps and methods applied to virtual screening one must keep track of details used to obtain the numerical results:

- Keep track of crystal structure PDB accession number
- Comment on species type and co-crystallized ligands
- Consider any information with regards to MOA to be pertinent and capture (eg. IS inhibitor or IS substrate, or IS agonist, or IS antagonist)
- Consider keeping the crystal structure of the co-crystallized ligands as methods to test pose fidelity and "accuracy" of your modeling experiment
- When selecting a crystal structure it is good practice to inspect the atoms in the vicinity of the putative binding site for which you will perform docking. If the B or thermal factor is relatively low, then this is a good sign that the active site is relatively rigid and not an "ensemble of conformations." This information is explicitly found in the PDB file (and can be downloaded from http://www.pdb.org). Other derived information about the model geometry can be analyzed with free tools such as MolProbity [65]. This free software helps identify model inconsistencies which may suggest not using a given structure. Similarly, it is vital to consider only targets for which the original X-ray data have been deposited. Using this data, electron density maps can be calculated or downloaded from places such as the Uppsala Electron Density server [66]. The maps and models can be viewed using free programs such as Coot [67], Python Molecular Viewer (PMV) [68, 69], or SwissPDB's DeepView [70]. Even with help from an experience X-ray crystallographer, one can confirm a) that density clearly follows the shape of the model and that b) there is not substantial "difference electron density" to indicate that model atoms are incorrectly placed.
- It is good practice to use crystal structures with relevant co-crystallized ligands as opposed to only resolution criteria.

Every detail counts: knowledge of the pH, solvent medium, ionic strength and buffers used can have implications on the model, the charge state of the model, and the type of charge model you would select to estimate atomic charges.

- Considering the sub-cellular localization can often help in determining charge state for a chemical. For instance, the pH of the cytosol is ~7, the mitochondrial and ER pH is ~5, whereas the pH of the nucleus is ~7.5-8. This may affect the charge models you wish to capture.

For modeling protein binding consider solvation/desolvation description as implemented in molecular mechanics frameworks or molecular docking algorithms as being inadequately captured or addressed.

For modeling more complex cellular activity phenomena (such as receptor mediated transient activation assays) consider cellular transport processes as surrogates to modeling a molecular MOA. For instance, estimating cellular membrane permeability, non-specific target binding and specific target binding may assist in these efforts.

When validation of pose fidelity is not optimal, consider the reasons for failure. “Reasons for Pose fidelity failure - Many of these pose fidelity failures could be attributed to one of four common causes: (a) insufficient conformational sampling of the ligand, particularly of aliphatic ring puckering, (b) symmetric or pseudosymmetric molecule, (c) critical missing water molecules, and (d) ligand pose dominated by electronic (orbital) effects. These issues are common to all docking methods and protocols.” [71]

Putting the pieces together – data and models and different software packages: One may want to consider either the purchase of a workflow manager such as Pipeline Pilot [40] or use of public domain versions such as KNIME [28], Bioclipse [72] or Taverna [73]. Many of these packages contain the necessary elements to address step 1 and 2 of Figure 1 in the in silico in vitro workflow. Then, the selection of a docking package is the final step.

6. Acknowledgements

MRG would like to thank James Rabinowitz and Stephen Little (both of the US-EPA’s National Center for Computational Toxicology) for providing mentorship and assistance during his post-doctoral research, and providing the environment to explore molecular docking in the context of toxicology while providing insight and valuable discussion in the development of the in-house DockScreen project (briefly discussed in the Examples (Example 2) section).

7. References

Footnote: Historically molecular modeling methods, stemming from roots in theoretical and computational chemistry, are comprised of an ensemble of developed and thoroughly vetted computational approaches used to investigate molecular-level processes and phenomena including but not limited to molecular structure, chemical catalysis, geochemistry, interfacial chemistry, nanotechnology, conformational analysis, stereoselectivity, enzyme biochemistry, chemical reaction dynamics, solvation, molecular aggregation and molecular design.

Bibliography

[1] Voutchkova, A; Osimitz, T; Anastas, P. Toward a comprehensive molecular design framework for reduced hazard. Chem Rev. 2010. 110, 5845-5882.

[2] Rusyn, I; Daston, G. Computational toxicology: Realizing the promise of the toxicity testing in the 21st Century. Environ Health Persp. 2010. 118, 1047-1050.

[3] Rabinowitz, J; Goldsmith, M; Little, S; Pasquinelli, M. Computational molecular modeling for evaluating the toxicity of environmental chemicals: Prioritizing bioassay requirements. Environ Health Persp. 2008. 116, 573-577.

[4] Allinger, N; Burkert, U. Molecular Mechanics. An American Chemical Society Publication. 1982.

[5] Dix, D; Houck, K. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci. 2007. 95, 5-12.

[6] Villoutreix, B; Renault, N; Lagorce, D; Sperandio, O; Montes, M; Miteva, M. Free resources to assist structure-based virtual ligand screening experiments. Curr Protein Pept Sc. 2007. 8, 381-411.

[7] Ponder, J; Case, D. Force fields for protein simulations. Adv Protein Chem. 2003. 66, 27-85.

[8] Pearlman, D; Case, D; Caldwell, J; Ross, W; Cheathham, T; DeBolt, S; Ferguson, D; Seibel, G; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate hte structural and energetic properties of molecules. Comp Phys Commun. 1995. 91, 1-41.

[9] MacKerell, A; Brooks, B; Brooks, C; Nilsson, L; Roux, B; Won, Y; Kaplus, M. CHARMM: The energy function and its parameterization with an overview of the program. PvR Scheyer EA (Ed.). John Wiley & Sons: Chichester, 1998.

[10] Case, D; Cheatham, T; Darden, T; Gohlke, H; Luo, R; Merz, K; Onufriev, A; Simmerling, C; Wang, B; Woods, R. The AMBER biomolecular simulation programs. J Comput Chem. 2005. 26, 1668-1688.

[11] Brooks, B; Brooks, C; Mackerell, A; Nilsson, L; Petrella, R; Roux, B; Won, Y; Archontis, C; Bartels, S; Caflish, B; Caves, L; Cui, Q; Dinner, A; Feig, M; Fischer, S; Gao, J; Hodoscek, M; Im, W; Kuczera, K; Lazaridi, T; Ma, J; Ovchinnikov, V; Paci, E; Pastor, R; Post, C; Pu, J; Schaefer, M; Tidor, B; Venable, T; Woodcock, H; Wu, X; Yah, W; York, D; Karplus, M. CHARMM: The biomolecular simulation program. J Comp Chem. 2009. 30, 1545-1615.

[12] Brooks, B; Bruccoleri, R; Olafson, B; States, D; Swaminathan, S; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem. 1983. 4, 187-217.

[13] Allinger, N; Yuh, Y; Lii, J. Molecular mechanics: The MM3 force field for hydrocarbons. J Am Chem Soc. 1989. 111, 8551-8566.

[14] Leo, A; Hansch, C; Elkins, D. Partition coefficients and their uses. Chem Rev. 1971. 71, 525-616.

[15] Lipinski, C; Lombardo, F; Dominy, B; Feeney, P. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Del Rev. 1997. 23, 3-25.

[16] Wermuth, C; Ganellin, C; Lindberg, P; Mitscher, L. Glossary of terms used in medicinal chemistry. Pure Appl Chem. 1998. 70, 1129-1143.

[17] Kubinyi, H; . From narcosis to hyperspace: The history of QSAR. Quant Struct-Act Relat. 2002. 21, 348-356.

[18] Wold, S; Ruhe, A; Wold, H; Dunn, W. The collinearity problem in linear regression - the partial least squares (PLS) approach to generalized inverses. SIAM J Sci Stat Comp. 1984. 5, 735-743.

[19] Cramer, R; Patterson, D; Bunce, J. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc. 1988. 110, 5959-5967.

[20] Klebe, G; Abraham, U; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem. 1994. 37, 4130-4146.

[21] Pastor, M; Cruciani, G; McLay, I; Pickett, S; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. J Med Chem. 2000. 43, 3233-3243.

[22] Norinder, U; . 3D-QSAR investigation of the Tripos benchmark steroids and some protein-tyrosine kinase inhibitors of styrene type using the TDQ approach. J Chemom. 1996. 10, 533-545.

[23] Kurogi, Y; Guner, O. Pharmacophore modelling and three-dimensional database searching for drug design using Catalyst. Curr Med Chem. 2001. 8, 1035-1055.

[24] Park, J; Harris, D. Construction and assessment of models of CYP2E1: Predictions of metabolism from docking, molecular dynamics and density functional theoretical calculations. J Med Chem. 2003. 46, 1645-1660.

[25] Jones, J; Mysinger, M; Korzekwa, K. Computational models for cytochrome P450: A predictive electronic model for aromatic oxidation and hydrogen atom abstraction. Drug Metab Dispos. 2002. 30, 7-12.

[26] Cheng, Y; Prusoff, W. Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol. 1973. 22, 3099-3108.

[27] MOE. Chemical Computing Group. Montreal, Quebec, Canada.

[28] Schrodinger, Inc. New York, NY.

[29] Cheatham, T; Young, M. Molecular dynamics simulation of nucleic acids: Successes, limitations and promise. Biopolymers. 2001. 56, 232-256.

[30] Roterman, I; Lambert, M; Gibson, K; Scheraga, H. A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. 2. Phi-Psi maps for n-acetyl alanine N'-methyl amide - comparisons, contrasts and simple experimental tests. J Biomol Struct Dyn. 1989. 7, 421-453.

[31] Roterman, I; Gibson, K; Scheraga, H. A comparison of the CHARMM, AMBER and ECEPP potential for peptides. 1. Conformational predictions for the tandemly repeated peptide (Asn-Ala-Asn-Pro)9. J Biomol Struct Dyn. 1989. 7, 391-419.

[32] Gundertofte, K; Liljefors, T; Norrby, P; Petterson, I. A comparison of conformational energies calculated by several molecular mechanics methods. J Comp Chem. 1996. 17, 429-449.

[33] Jorgensen, W; Maxwell, D; Tirado-Rives, J; . Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc. 1996. 118, 11225-11236.

[34] Jorgensen, W; Tirado-Rives, J. The OPLS potential functions for proteins - energy minimizations for crystals of cyclic-peptides and crambin. J Am Chem Soc. 1988. 110, 1657-1666.

[35] Halgren, T; . Merck molecular force field. I. Basis, form, scope parameterization and performance of MMFF94. J Comp Chem. 1996. 17, 490-519.

[36] Chen, Y; Zhi, D. Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins: Struct., Funct., Bioinf. 2001. 43, 217-226.

[37] Ellis, L; Hou, B; Kang, W; Wackett, L. The University of Minnesota Biocatalysis/Biodegradation Database: post-genomic data mining. Nucleic Acids Res. 2003. 31, 262-265.

[38] MetaPrint2d http://www-metaprint2d.ch.cam.ac.uk/metaprint2d.

[39] Bologa, C; Olah, M; Oprea, T. Chemical database preparation for compound acquisition or virtual screening. Methods Mol Biol. 2005. 316, 375.

[40] Accelrys Discovery Suite, Accelrys, Inc. San Diego, CA.

[41] Sybyl. Tripos, Inc. St. Louis, MO.

[42] Schwede, T; Sali, A; Honig, B; Levitt, M; Berman, H; Jones, D; Brenner, S; Burley, S; Das, R; Dokholyan, N; Dunbrack, R; Fidelis, K; Fiser, A; Godzik, A; Huang, Y; Humblet, C; Jacobsen, M; Joachimiak, A; Krystek, S; Kortemme, T; Kryshtafovych, A; Montelione, G; Moult, J; Murray, D; Sanchez, R; Sosinick, T; Standley, D; Stouch, T; Vajda, S; Vasquez, M; Westbrook, J; Wilson, I. Outcome of a workshop on applications of protein models in biomedical research. Structure. 2009. 17, 151-159.

[43] Irwin, J; . Community benchmarks for virtual screening. J Comput Aid Mol Des. 2008. 22, 193-199.

[44] Cross, J; Thompson, D; Rai, B; Baber, J; Fan, K; Hu, Y; Humblet, C. Comparison of several molecuclar docking programs: Pose prediction and virtual screening accuracy. J Chem Inf Model. 2009. 49, 1455-1474.

[45] Cherkasov, A; Fuqiang, B; Li, Y; Fallahi, M; Hammond, G. Progressive docking: A hybrid QSAR/Docking approach for accelerating in silico High Throughput Screening. J Med Chem. 2006. 49, 7466-7478.

[46] Peterson, S; . Improved CoMFA modeling by optimization of settings: Toward the design of inhibitors of the HCV NS3 protease. Uppsala, Sweden:Uppsala University.2007.

[47] Norinder, U; . Recent progress in CoMFA methodology and related techniques. Perspect Drug Discovery Des. 1998. 12/13/14, 25-39.

[48] Kim, K; Grecco, G; Novellino, E. A critical review of recent CoMFA applications. Perspect Drug Discovery Des. 1998. 12/13/14, 257-315.

[49] Rosen, J; Lovgren, A; Kogej, T; Muresan, S; Gottfries, J; Backlund, A. ChemGPS-NPWeb: chemical space navigation tool. J Comput Aid Mol Des. 2009. 23, 253-259.

[50] Larsson, J; Gottfries, J; Muresan, S; Backlund, A. ChemGPS-NP: Tuned for navigation in biologically relevant chemical space. J Nat Prod. 2007. 70, 789-794.

[51] Eakins, S; et al. Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein. Mol Pharmacol. 2002. 61, 964.

[52] Thorsteinson, N; Ban, F; Santos-Filho, O; Tabaei, S; Miguel-Queralt, S; Underhill, C; Cherkasov, A; Hammond, G. In silico identification of anthropogenic chemicals as ligands of zebrafish sex hormone binding globulin. Toxicol Appl Pharm. 2009. 234, 47-57.

[53] Perry, J; Goldsmith, M; Peterson, M; Beratan, D; Wozniak, G; Ruker, F; Simon, J. Structure of the ochratoxin A binding site within human serum albumin. J Phys Chem B. 2004. 108, 16960-16964.

[54] Aureli, L; Cruciani, G; Cesta, M; Anacardio, R; De Simone, L; Moriconi, A. Predicting human serum albumin affinity of interleukin-8 (CXCL8) inhibitors by 3D-QSPR approach. J Med Chem. 2005. 48, 2469-2479.

[55] Ekins, S; de Groot, M; Jones, J. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. Drug Metab Dispos. 2001. 29, 936-944.

[56] Ekins, S; Erickson, J. A pharmacophore for human pregnane X receptor ligands. Drug Metab Dispos. 2002. 30, 96-99.

[57] Lewis, D; . Molecular modeling of human cytochrome P450-substrate interactions. Drug Metab Rev. 2002. 34, 55-67.

[58] Hirono, S; Nakagome, L; Imai, R; Maeda, K; Kusuhara, H; Sugiyama, Y. Estimation of the Three-Dimensional Pharmacophore of Ligands for Rat Multidrug-Resistance–Associated protein 2 using ligand-based drug design techniques. Pharm Res. 2005. 22, 260-269.

[59] DeGorter, M; Conseil, G; Deeley, R; Campbell, R; Cole, S. Molecular modeling of the human multidrug resistance protein 1 (MRP1/ABCC1). Biochem Bioph Res Co. 2008. 365, 29-34.

[60] Rabinowitz, J; Little, S; Laws, S; Goldsmith, M. Molecular modeling for screening environmental chemicals for estrogenicity: Use of the toxicant-target approach. Chem Res Toxicol. 2009. 22, 1594-1602.

[61] Hirst, W; Abrahamsen, B; Blaney, F; Calver, A; Aloj, L; Price, G; Medhurst, A. Differences in the central nervous system distribution and pharmacology of the mouse 5-hydroxytryptamine-6 receptor compared with rat and human receptors investigated by radioligand binding, site-directed mutagenesis, and molecular modeling. Mol Pharmacol. 2003. 64, 1295-1308.

[62] Available online at: http://oaspub.epa.gov/eims/eimscomm.getfile?p_download_id=466705.

[63] Availabel Online at: http://www.epa.gov/ncct/bosc_review/2009/posters/2-06_Rabinowitz_CompTox_BOSC09.pdf.

[64] Goldsmith, M; Little, S; Reif, D; Rabinowitz, J. Digging deeper into deep data: Molecular docking as a hypothesis-driven biophysical interrogation system in computational toxicology. . . ,.

[65] Available Online at: http://molprobity.biochem.duke.edu.

[66] Available online at: http://xray.bmc.uu.se/valid/density/form1.html.

[67] Available Online at: http://www.biop.ox.ac.uk/coot.

[68] Available Online at: http://pmvbase.blogspot.com/2009/04/electron-density-map.html.

[69] Available Online at: http://mgltools.scrips.edu/documentation/tutorial/python-molecular-viewer.

[70] Availabe Online at: http://spdbv.vital.it.ch.

[71] Irwin, J; Shoichet, B; Mysinger, M; Huang, N; Colizzi, F; Wassam, P; Cao, Y. Automated docking screens: A feasibility study. J Med Chem. 2009. 52, 5712-5720.

[72] Bioclipse. Proteometric Group, Department of Pharmaceutical Biosciences, Uppsala University, Sweden & Cheminformatics and Metabolism Team, European Bioinformatics Institute (EMBI).

[73] Taverna. School of Computer Science, University of Manchester, UK.