ENVIRONMENTAL BIOTECHNOLOGY 9 (2) 2013, 39-51; doi:10.14799/ebms198

Molecular modelling techniques in environmental research

Rafał D. Urniaż[1], Ewelina Rutkowska[1], Jan P. Jastrzębski[2], Paweł Książek[3], Katarzyna Rudnicka[4]

[1]Laboratory of Medicinal Chemistry and Neuroengineering, Medical University of Lublin, Chodzki 4a, 20-093 Lublin, Poland
[2]Department of Plant Physiology, Genetics and Biotechnology, University of Warmia and Mazury in Olsztyn, Oczapowskiego 1a, 10-719 Olsztyn, Poland
[3]Chair of Physics and Biophysics, University of Warmia and Mazury in Olsztyn, Oczapowskiego 4, 10-719 Olsztyn, Poland
[4]Department of Environmental Biotechnology, University of Warmia and Mazury in Olsztyn, Sloneczna 45G, 10-718 Olsztyn, Poland

Corresponding author: Rafał D. Urniaż, Laboratory of Medicinal Chemistry and Neuroengineering, Medical University of Lublin, Chodzki 4a, 20-093 Lublin, Poland; Phone +48 728 815 151, E-mail: Rafal.Urniaz@gmail.com

Key words: molecular dynamics, molecular modelling, pharmacokinetics, pharmacophore modelling, QSAR

Received in March 2013. Published in January 2014.

ABSTRACT

Over the last few decades significant increase in computational methods (*in silico*) was annotated. Novel methods have been developed and applied for hypothesis improvement and testing in regions of industrial, pharmaceutical and environmental research. The term *in silico* methods include variety of approaches. Considerable attention has been attracted to databases, data analysis tools, quantitative structure-activity relationships (QSAR), pharmacophore models, molecular docking and dynamics, pharmacokinetics and other molecular modelling techniques. *In silico* methods are often accompanied by experimental data, both to create the model and to test it. Such models are frequently used in the discovery and optimization of novel molecules with expected affinity to a target, the estimation of absorption, distribution, metabolism, excretion and toxicity properties as well as physicochemical characterization. The review summarizes briefly the applications of most common molecular modelling techniques and evaluates their application in environmental research. Additionally, this study considers computer aided methods as potential and complex tools that may serve as valuable partnership with wet-lab experiments and may provide a rational aid to minimize the cost and time of research.

INTRODUCTION

At the beginning of 1980`s the significant increase of computer technology was annotated. Structural biology primarily focused on obtaining macromolecules and relating their structures to biological function was enriched by new field called 'computational biology'. In the beginning it was associated with finding similarities among nucleotide strings of known genetic sequences and relating them to evolutionary commonalities. Furthermore, the increase of computing power allowed applications of virtual modelling to more complex molecular biology aspects, such as prediction of the three-dimensional molecules arrangement or prediction of their interactions with endogenous ligands. Nowadays, computers have become more and more powerful and have established the development of new modelling techniques called as 'molecular modelling'. The approach contains wide range of tools for visualisation and manipulation of molecules in *in silico* environment.

Molecular modelling is concerned with the description of the atomic and molecular interactions that fulfil microscopic and macroscopic behaviour of physical systems. The essence of molecular modelling resides in the connection between the microscopic and the macroscopic world provided by the theory of statistical mechanics. The fundamental concept of molecular modelling is a "model". It is defined as mathematical or physical description that contains analogous properties and acts in analogous way as modelled object. Therefore,

model is a tool that can be used to describe the system and its behaviour under the different modelled conditions. To illustrate the concept a model of perfectly elastic collision (fundamental model in physics) can be considered as an example (Figure 1).

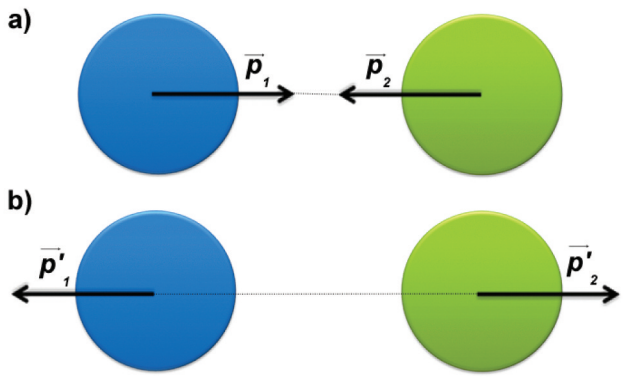

Figure 1. Illustration of perfectly elastic collision model; a) balls before impact; b) balls after impact. More details in the text.

The model illustrates collision of two objects, e.g. billiard balls. In contrast to reality, in the above model, objects do not deform themselves and energy of the system is not lost (for example by heating up). Due to this simplification, the assumption of the law of conservation of energy and momentum are fulfilled. Although, the mathematical calculation of the collision phenomena is not fully precise, in some cases the approximation may be accurate enough to be a good model to describe the phenomena. Especially, when the impact should be calculated for billions of objects and extra precision is not obligatory. Concept of using models to simplify the reality did not come with the advent of molecular modelling. It was known since the creation of science in wide range of scientific areas such as *in situ*, *in vivo* and *in vitro*. The novel concept was to accommodate the traditional areas of modelling and carry them into a new level. It has opened new possibilities to work with new research areas. Nowadays, *in silico* environment has become a valuable partnership for wet-lab experiments, what provides a great aid to minimize the cost and time of research. This review briefly summarizes the most current and popular techniques considering their application in environmental research.

Biological databases

Biological databases are a collection of biological data organized according to few models: flat file model, network model, hierarchical model, relational model, object model or object-relational model. The most effective and applied in the largest biological databases is the object-relational model. Depending on the source of collected data, biological databases are divided into two types: primary and secondary databases. Primary databases are the libraries of experimentally derived data such as nucleotide sequences derived from sequencing (e.g. EMBL (Kanz et al. 2005; Kulikova et al. 2004) currently ENA (Hoopen et al. 2010)). Secondary databases are the collection of computer-processed primary information such as amino acidic sequence translated *in silico* from RNA or DNA sequence (e.g. TrEMBL; nowadays TrEMBL is the part of UniProt). Three primary nucleotide sequence databases (DDBJ, ENA(EMBL) and GenBank) collaborate to acquire, maintain, share complex nucleotide sequence information in the project named International Nucleotide Sequence Database Collaboration (INSDC) (Cochrane et al. 2011; Karsch-Mizrachi et al. 2011). Moreover, the special types of biological databases are metadatabases, which collect data from diverse sources and make them available in unified form. Thus, metadatabases are the database of databases (Hsu et al. 1991). Most of the biological databases are part of biological, biomedical, biotechnological services, which are the collections of databases, tools, document allowed by scientific institutes (e.g. GenBank being a part of NCBI – National Center for Biotechnology Information, ENA being a part of EBI – European Bioinformatics Institute) (Table 1).

Once a year, the journal of Nucleic Acids Research publishes the list of all available biological databases with brief descriptions and direct links to them. In 2013 the published list consists of 1512 items grouped in categories: Nucleotide Sequence Databases, RNA sequence databases, Protein sequence databases, Structure Databases, Genomics Databases (non-vertebrate), Metabolic and Signaling Pathways (Human and other Vertebrate Genomes) Human Genes and Diseases, Microarray Data and other Gene Expression Databases, Proteomics Resources, Other Molecular Biology Databases, Organelle databases, Plant databases, Immunological databases, Cell biology (Fernández-Suárez and Galperin 2013). In the succeeding paragraphs the most useful databases in molecular modelling have been summarized.

The Protein database in NCBI is the metadatabase which collects the amino acidic sequence data collected from a variety of sources, including SwissProt, PIR, PRF, PDB and translations from annotated coding regions in GenBank, RefSeq, TPA, EST.

The Protein Information Resource (PIR) is a complex public resource of proteins for proteomic research studies. PIR manages the Protein Sequence Database (PSD). PIR-PSD contains protein records classified in taxonomic order (Wu et al. 2003).

Table 1. The table contains the name or abbreviation, one sentence summary and web address of described below biological databases.

No.	Name	Short description	WWW
1	INSDC	Collaboration of ENA, DDBJ and GeneBank	http://www.insdc.org/
2	EBI	European Bioinformatic Institute	http://www.ebi.ac.uk/
3	ENA	Primary nucleotide sequence database	http://www.ebi.ac.uk/ena/
4	UniProt	Metadatabase of protein sequence	http://www.ebi.ac.uk/uniprot/
5	DDBJ	Primary nucleotide sequence database	http://www.ddbj.nig.ac.jp/
6	NCBI	The institute of biotechnology in USA	http://www.ncbi.nlm.nih.gov/
7	GenBank	Primary nucleotide sequence database	http://www.ncbi.nlm.nih.gov/genbank/
8	MMDB	Database of 3D structures of macromolecules	http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml
9	PIR	Resource for proteomic research studies	http://pir.georgetown.edu/
10	ExPASy	Bioinformatics resource portal	http://expasy.org/
11	PDB	Database of 3D structures of macromolecules	http://www.rcsb.org/pdb/home/home.do
12	Pfam	Database of protein families and domains	http://pfam.sanger.ac.uk/
13	SCOP	Database of structural classification of proteins	http://scop.mrc-lmb.cam.ac.uk/scop/
14	CATH	Hierarchical classification of protein domain structures	http://www.cathdb.info/
15	KEGG	Database of molecular interaction networks	http://www.genome.jp/kegg/
16	BRENDA	Database of enzyme	http://www.brenda-enzymes.info/
17	PDBsum	Protein database supported by EBI	http://www.ebi.ac.uk/pdbsum/
18	OPM	Orientations of Proteins in Membranes (OPM) database	http://opm.phar.umich.edu/
19	PPM	OPB service for positioning the transmembrane and peripheral proteins in membranes	http://opm.phar.umich.edu/server.php

The Molecular Modelling Database (MMDB) is one of NCBI databases, contains three-dimensional structures of macromolecules such as proteins and polynucleotides. Each record in MMDB is linked to the rest of the NCBI databases, including sequences, bibliographic citations, taxonomic classifications, and sequence and structure neighbours.

The Universal Protein Resource (UniProt) is a metadatabase of protein sequence and annotation data. The UniProt includes the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef) and the UniProt Archive (UniParc). UniProt is a collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics and the Protein Information Resource (PIR). The UniPort databases are accessible by the Bioinformatics Resource Portal – ExPASy. In addition, the ExPASy provides access to bioinformatic tools in molecular modelling. One of these tools is SWISS-MODEL. It is a protein structure homology-modelling server, accessible by WWW, or from the program DeepView (Swiss Pdb-Viewer). The goal of this server is to make protein modelling available and easy to use to all scientists.

The UniProt Knowledgebase (UniProtKB) is the collection of: the UniProtKB/Swiss-Prot, the UniProtKB/TrEMBL, the UniProt Reference Clusters (UniRef), the UniProt Archive (UniParc) and the UniProt Metagenomic and Environmental Sequences (UniMES).

The Swiss-Prot (actual full name - UniProtKB/Swiss-Prot) is the manually annotated and reviewed part of the UniProt Knowledgebase (UniProtKB). It is a non-redundant protein sequence metadatabase, which combines experimental data, computed features and scientific conclusions.

TrEMBL (full name UniProtKB/TrEMBL) is automatically annotated and not reviewed database of protein information, including function, classification, and cross-reference.

The Protein Data Bank (PDB) is a database of experimentally solved data of three dimensional structures of biological macromolecules. PDB provides information of primary structure (sequence), full 3D structure information (secondary, tertiary and quaternary structures), ligand structures, structural classification (CATH, SCOPE), taxonomy, enzyme classification (BRENDA, KEGG), literature, biological specification, chemical properties and domain occurrence (e.g. Pfam). PDB provides numerous tools to analyse the protein structure and links to external databases and tools. PDB is the key database in molecular modelling because it is a collection of real structures, which is an important structural reference for modelling (Berman et al. 2000).

The Pfam is a widely used database of conserved protein families and domains. The goal of Pfam is to make a comprehensive and accurate classification of all known protein sequences (Finn et al. 2010).

The Structural Classification of Proteins (SCOP) database aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in the Protein Data Bank (PDB). Proteins are classified to reflect both structural and evolutionary relatedness. The levels in the classification are: **Family** (evolutionarily relationship), **Superfamily** (probable common evolutionary origin) and Fold (structural similarity) (Murzin et al. 1995; O'Maille et al. 2002).

The Protein Structure Classification (CATH) is the semi-automatic procedure for hierarchical classification of protein domain structures. This procedure classifies proteins in the four main levels: class (C - the secondary structure of the domain), architecture (A - structural similarity but no evidence of homology - Fold in SCOP), topology (T - grouping of topologies sharing structural features) and homologous superfamily (H - evolutionary relationship- Superfamily in SCOP) (Orengo et al. 1997).

The CATH Protein Family Database (CATH-PFDB) contains structural domains, homologous superfamilies, fold groups, grouped sequences and structures (Cuff et al. 2011; Pearl et al. 2001).

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is the reference database that integrates knowledge on molecular interaction networks. The KEGG database consists of three categories of sixteen databases: Systems information (e.g. Pathway/Disease/Drug database), Genomic information (e.g. Orthology/Genome/Genes database), Chemical information (Compound/Reaction/Enzyme database) (Kanehisa et al. 2004).

The BRENDA (BRaunschweig ENzyme DAtabase) is the largest open access system of databases containing a biochemical and molecular information on all classified enzymes, bioinformatic tools for querying the database and calculating molecular properties. It contains classification and nomenclature, reaction and specificity, functional parameters, occurrence, enzyme structure and stability, mutants and enzyme engineering, preparation and isolation, the application of enzymes and ligand-related data. The data in BRENDA are manually elaborated. Records are linked to a literature reference, the origin of organism and, where available, to the protein sequence of the enzyme protein (Chang et al. 2009).

PDBsum is a protein database supported by EBI. PDBsum records contain complex, easy to analyse information about sequence, domains, spatial conformation and structure, interactions of the proteins deposited in the Protein Data Bank (Laskowski 2009). The main advantage of PDBsum is extensive list of references and links from each record to external databases (such as PDBe, RCSB, SRS, MMDB, PDBWiki, Proteopedia, SCOP, CATH, Pfam, UniProt, Gene Ontology, HSSP) and protein structure analysing tools (for instance WHAT_CHECK, PROCHECK, ArchSchema).

Orientations of Proteins in Membranes (OPM) database accesses the information about spatial orientation of proteins with respect to the surface of lipid membranes (Lomize et al. 2006). Proteins collected in the OPM are taken from the Protein Data Base, these are transmembrane proteins, peripheral proteins and membrane-active peptides. **PPM server** (Lomize et al. 2012) is a tool implemented in OPM webpage, which allows calculating the positions of any transmembrane and peripheral proteins in membranes using PDB input file of analysed protein.

Aforementioned chemical repositories consist of a wealth of biological targets and chemicals data that can be screened and used by *in silico* approaches. Although much of the molecular modelling research to date has been focused on human targets, many of these databases contain reliable data from other species that would shed light on species differences and aid discovery of molecules for animal healthcare as well as assist in understanding the importance of toxicological profile improvement for chemicals released into the environment.

Pharmacophore modelling

Understanding the chemistry behind molecular recognition is of utmost importance to conduct successful research. One of the computational tools that can successfully

facilitate derivation of information from a set of compounds is generation of a pharmacophore. The most widely accepted definition considers a pharmacophore as the largest common denominator shared by a set of active molecules. A pharmacophore model represents molecular features which are indispensable for binding of a ligand to a target macromolecule. The IUPAC defines a pharmacophore as "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or block) its biological response" (Wermuth et al. 1998). Molecular features are defined by pharmacophoric descriptors, e.g. H-bond donor or acceptor properties, hydrophobic and electrostatic interaction sites. A pharmacophore model explains how structurally diverse ligands within a set can bind to a common receptor site.

Pharmacophore model can be constructed via ligand-based or receptor-based approach. The former uses ligands having activity against a target of unknown structures. Ligands are used to search for pharmacophore features and models. The latter approach uses resolved ligand-target complexes to illustrate the complementarity of the receptor to the ligand. These models expend molecular recognition by additional target features and correct ligand geometry (Figure 2).

The pharmacophore pattern allows screening (virtual screening) for novel ligands that share the same features arranged in the same relative orientation thus binding to the same receptor site. Recent study showed application of module Phase as a tool to find matches to a given pharmacophore hypothesis in a database (López-Ramos and Perruccio 2010). The target protein was the iron-containing enzyme, 4-hydroxyphenylpyruvate dioxygenase (HPPD). The protein has proven to be a very successful target for the development of herbicides with bleaching properties, and today HPPD inhibitors are well established in the agrochemical market. The aim of the study was to identify compounds with inhibitory properties using computational tools and to compare the performance of these tools. One of the methods applied was pharmacophore-based search. Researchers sought out a database consisting of compounds active on HPPD and decoy compounds. Subsequently, the search was performed using the conformers stored in the previously prepared Phase database. In this study pharmacophoric features of inhibitor candidate were characterized. Further information can be derived from what is already known about the binding mode of the compounds. In the case of HPPD inhibitors, using this pharmacophore search in the database along with Phase outperformed all the other screening methods. The pharmacophore search performed with the Phase

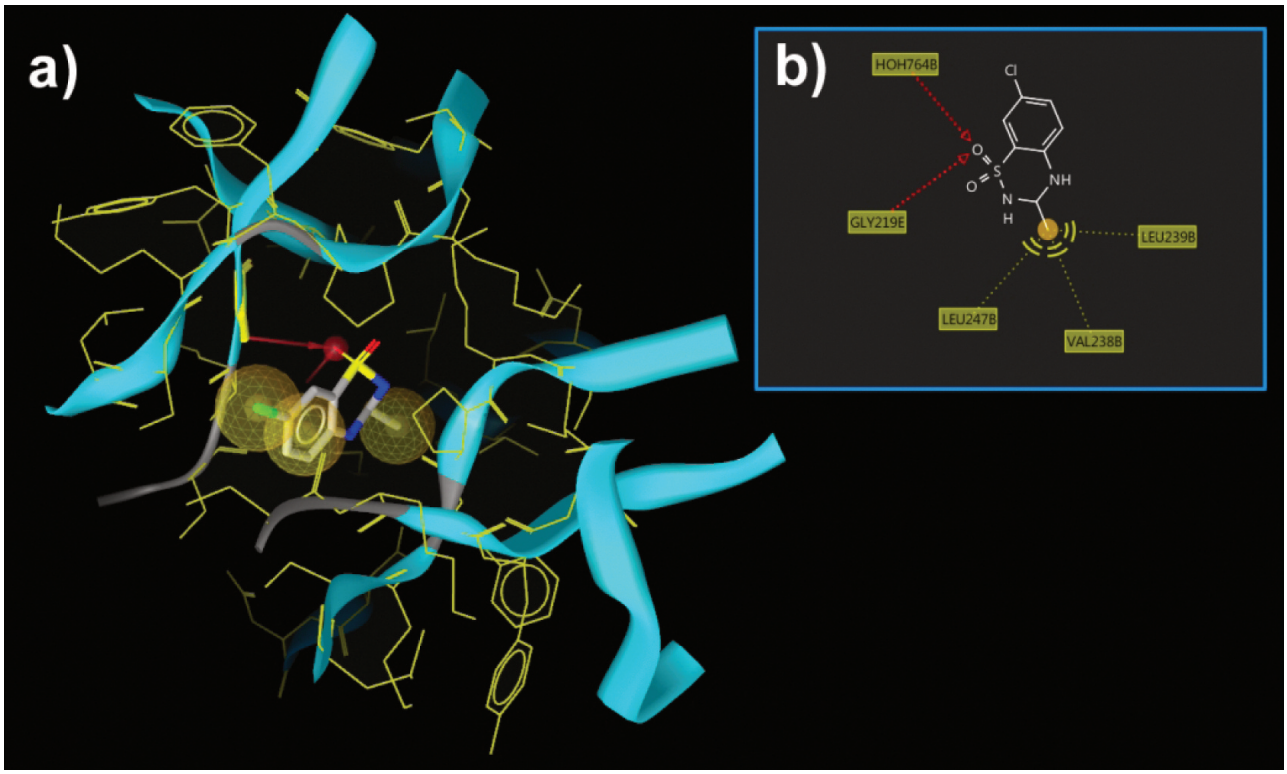

Figure 2. Illustration of receptor-based pharmacophore modelling applied for crystal structure of the AMPA receptor (PDB ID: 3il1) bound to the allosteric modulator IDRA-21 (a) and ligand map of accompanying interactions (b). Arrow – H-bond acceptor (HBA); yellow sphere - steric interactions.

pharmacophore tool gave the best results in the retrieval of active compounds.

Pharmacophore based virtual screening can be done using databases, e.g. ZINC (Irwin et al. 2012) particularly ZINCPharmer dedicated to pharmacophores (Koes and Camacho 2012). Using the pharmacophore approach one can explore the features of macromolecule-ligand interactions through "pharmacophore map". The map simultaneously shows the interactions made by all ligands with their receptors in 3D on the computer graphics (McGregor 2007). It is instructive to view all the ligands overlapped in the common reference frame, with bond connectivity removed and atoms coloured by pharmacophore features. This approach highlights the basis for ligand activity against several targets. Another major task is then to address the question of selectivity between targets and to explore the differences between them.

Compounds that match the pharmacophore contours of the molecule serve as potential lead compounds for further development. A computer–based similarity search via easily accessible databases can be performed to speed up the process of lead identification. Pharmacophore generation can be regarded as a method by which the binding site on a protein can be located and mapped and the protein–ligand intermolecular interactions can be studied in the context of environmental research or health sciences.

Quantitative Structure–Activity Relationship (QSAR)

One of the general purposes in environmental research is to predict the impact of newly synthesized compounds as well as to explain the mechanism of existing ones. Quantitative structure-activity relationship (QSAR) methods perform an important role in this process. The approach indicates relationship derived from analysis of the properties of previously characterised compounds (Kubinyi 1995). Classical QSAR method correlates biological activities of tested compounds with their physicochemical properties encoded by certain structural features. Moreover, it is a technique for building computational or mathematical models which attempts to find a statistically significant correlation between compounds, structure and function using chemometric techniques. In general, QSAR methods are reduced to solve the "simple" structure – activity correlation equation (Equation 1). It assumes that compound's activity is a direct function of the synergic impact of their physiochemical and/or structural properties.

$$Activity = f(physiological\ and/or\ structural\ properties) \quad (1)$$

Equation 1. Equation illustrates main concept of QSAR methods.

In environmental research QSARs are mainly employed to predict measures of toxicity from physical characteristics of the chemical structure (known also as molecular descriptors). Acute toxicities, such as the concentration that causes half the population to die (LD50), are one example of the toxicity measures that may be predicted from QSAR models. Basic QSAR models calculate the toxicity of chemicals using a Multiple Linear Regression (MLR) function of molecular descriptors, such as molecular weight, number of rotatable bonds, Log P and others (Equation 2).

$$Toxicity = a \cdot x1 + b \cdot x2 + c \ldots \quad (2)$$

Equation 2. The example QSAR model of toxicity. *x1* and *x2* - independent descriptors (e.g. molecular weight or octanol-water partition coefficient); *a*, *b*, and *c* - fitted parameters.

Three-dimensional QSAR (3D-QSAR) analysis (being a spatial variant of QSAR) is commonly used in the environmental studies to establish relationships between three dimensional molecular properties and observed biological effects on a series of congeneric compounds (Cramer et al. 1988). One of the widely used 3D-QSAR method is Comparative Molecular Field Analysis (CoMFA). It adapts statistical procedures to correlate molecular features, such as steric and electrostatic properties with biological activities (Kubinyi 1997a). The fields are calculated by testing the interactions between a virtual probe atom and a collection of aligned molecules.

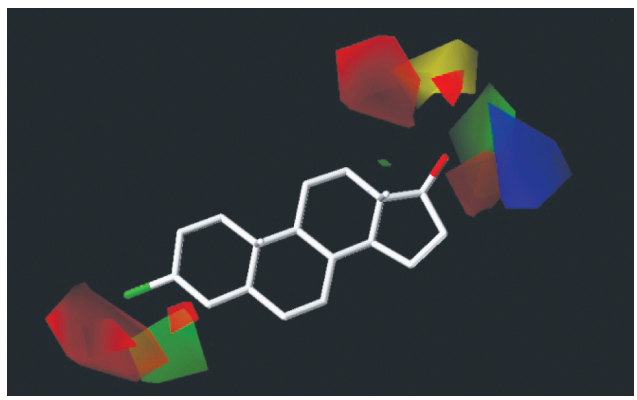

Figure 3. Illustration of CoMFA interaction fields around group of 20 testosterone derivatives. The figure shows 3-chlorodihydrotestosterone, details in the text.

The interactions are evaluated for each molecule at regularly spaced intervals on a virtual grid surrounding the molecules. The result is an interactive contour map with indication of fields which have impact on the activity (Figure 3). Green fields denote desirable steric interactions between ligand and binding site of the molecular target as opposed to yellow fields where steric bulk influence is undesirable. The colour codes indicate regions where electronegative interactions enhance (blue) or reduce (red) the activity. Such approach has essentially two main advantages: firstly, its ability to represent the information in 3D format and highlights regions of the molecules that have positive and negative effects associated with steric and electrostatic

interactions, as shown in Figure 3. Secondly, its ability to predict the unknown biological activity of new compounds based on the calculated model. Over the past few decades, CoMFA methods have become widespread in regions of both industrial and academic research regarding QSAR studies. Such approaches are providing not only the prediction of specific properties of new compounds, but also help to elucidate the possible molecular mechanism of the ligand–target interactions, thus being beneficial especially in cases when experimental NMR or crystal structure of the target molecule is unavailable (Hansch and Fujita 1964).

Apart from the widespread application of QSAR modelling in many regions of science, it is also a well-recognized technique in environmental research, mainly for toxicity prediction (Lo Piparo and Worth 2010; Schultza et al. 2003). One of the priority developments of QSAR toxicity models is mainly employed to identify the adverse effects a chemical may have on sexual function and fertility in adult males and females, developmental toxicity in the offspring, as well as effects on, or mediated via, lactation. Reproductive toxicity refers to a wide range of endpoints relating to mechanisms of action currently unknown or only partially understood at the molecular and cellular level. Along with carcinogenic studies, reprotoxicity studies are among the most costly and time-consuming experimental procedures. Furthermore, reprotoxicity testing requires the highest number of test animals. For all these reasons, the development of alternative (non-animal) methods for toxicity and reprotoxicity assessment is of high political priority (following European Commission recommendation) (Lo Piparo and Worth 2010). Currently there are relatively few models for toxicity endpoints, partly due to the biological complexity of the endpoint, which covers many incompletely understood mechanisms of action, and partly due to the paucity and heterogeneity of high quality data suitable for model development. In contrast, there is an extensive and growing range of software and literature models for predicting activity, e.g. ADMET Predictor, ACD/Tox Suite, CAESAR, Derek, Molcode Toolbox, MultiCASE, OECD QSAR Application Toolbox, PASS, T.E.S.T, TOPKAT, Toxmatch, VirtualToxLab, etc.

Despite the varied applications of QSAR models, its availability for toxicity is limited as a result of the diversity of potential endpoints, and the paucity of data suitable for modelling (Urniaz and Jozwiak 2013). Available models are potentially useful as a means of supporting hazard identification and priority setting, but not yet for the establishment of toxic potencies for use in risk assessment.

Molecular docking

Computational approach that fits small molecule into the structure of macromolecular target and scores and ranks its complementarity was pioneered during the early 1980s (Kuntz et al. 1982). Nowadays docking simulation is a key molecular modelling methodology. It requires structural information for the macromolecule which can be obtained from X-ray crystallography, NMR or homology modelling, thus it is a structure (target)-based approach. Homology modelling is another computational technique used to predict unknown protein structure on the basis of sequence similarity to known protein structure(s). The macromolecule can be a protein (receptor, enzyme), RNA or DNA, while the ligand is usually a small molecule, but can be a protein as well.

The prediction is usually done by searching the ligand's translational and rotational degrees of freedom within the protein cavity, and by searching the conformational states of the ligand itself. The result of docking procedure is a predicted position of ligand's conformer on the receptor binding site (Figure 4). This tool indicates possible conformations of ligand-target complexes and molecular features that are responsible for interactions between them.

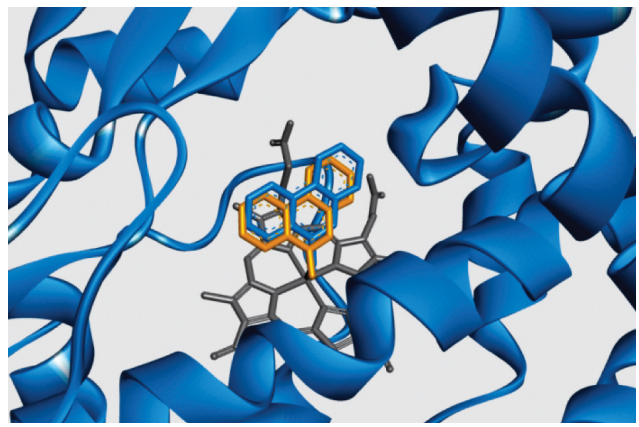

Figure 4. Model of 9-aminophenanthrene (light blue compound) and cofactor (grey) in the active site of mammalian cytochrome P450 monoxygenase (blue protein rendered as secondary structure view) deposited in PDB database (PDB ID: 1EGY). The resulting conformation of docking procedure performed in Molegro Virtual Docker (MVD 2012.5.5.0) is coloured in orange. Docked ligand shows similar pattern of binding and displays the same type of interactions as co-crystallised one.

First step in the docking process is the application of docking algorithms to pose small molecules in the active site. Posing determines whether a ligand conformation fits in the binding site. Algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets (Kitchen et al. 2004). It is a rough measure of the interactions between ligand and target, usually described by using van der Waals term given by a Lennard-Jones potential function and electrostatic energy terms given by a Coulombic interactions. Docking programs use a scoring function which distinguishes among the generated binding modes the best solution. A large number of search algorithms and scoring functions have been developed to model protein-ligand complex.

The scoring function evaluates compound fits on the basis of calculations of approximate shape and electrostatic complementarities. More advanced evaluation of docking results can be performed by ranking the poses. Ranking procedures usually are expended calculations that re-evaluate poses with respect to e.g. entropy or explicit solvation. The results also show that energy minimization and reranking of the top poses can be an effective means to overcome some of the limitations of a given docking function (Perola et al. 2004). It should also be noted that docking method encounters various limitations, e.g. of crystallographic resolution or water influence on binding event. Docking software includes AutoDock, Dock, 3D-Dock, Affinity, LigandFit, FRED, SurFlex, HEX, Glide, GOLD, ICM, and MVD as examples.

The docking study was a valuable tool used by Yang and co-workers to investigate the polybrominated diphenyl ethers (PBDEs) widely used as flame retardants (Schecter et al. 2005; Yang et al. 2010). Unfortunately, PBDEs accumulate in the environment (Birnbaum and Staskal 2004; Darnerud et al. 2001). Furthermore, PBDEs are known to be endocrine-disrupting compounds, altering activity of the human estrogen receptor alpha (hERα). In order to distinguish the ER antagonists among the set of 41 PBDE compounds, the binding features of the target compounds were analysed by molecular docking. Automatic flexible molecular docking program SurFlex-Dock implemented in SYBYL 7.3 generated ligand poses close to the X-ray conformation more often than the other docking programs (Cross et al. 2009). SurFlex-Dock docks ligands automatically into a receptor's ligand binding site using a protomol-based method and an empirically derived scoring function. The protomol is a unique and important factor of the docking algorithm and is a computational representation of the assumed ligands that interact with the binding site. SurFlex-Dock's scoring function contains hydrophobic, polar, repulsive, entropic and solvation terms. Simulations showed that some of the PBDE compounds acting as hERα antagonists extended into the channel of the estrogen receptor (ER), which is usually occupied by the alkylamine side chain of the ER antagonists raloxifene and 4-hydroxytamoxifen, while most PBDE compounds without antiestrogenic activity adopted binding modes similar to that of ER agonist 17β-estradiol, which did not reach into the channel. The study suggested that pose comparison based on docking was useful for discriminating whether or not PBDE compounds have antiestrogenic activity. Knowing the binding modes of compounds in hERα can help to screen out antiestrogenic compounds and further develop descriptive and predictive models in ecotoxicology (Kubinyi 1997b).

Molecular dynamics simulation

Molecular Dynamics (MD) is a computer simulation method that relies mainly on Newton mechanics to simulate the physical interactions and movement of atoms and molecular systems. To obtain the dynamic characteristics and the understanding of interaction mechanism at an atomistic scale, MD simulation packages like AMBER, CHARMM, GROMACS and NAMD are applied widely in the modelling of systems. Generation of a successive series of configurations resulting in a trajectory are very computer time-intensive. Molecular dynamics provides additional tools that complement the static data obtained by crystallization and provides new insights into the molecular mechanisms that govern the binding process. Binding free energy prediction has been regarded as a powerful and valuable tool to explore the binding mechanisms and binding affinity. Simulation consists of the numerical, step-by-step, solution of the classical equations of motion, which for a simple atomic system may be written as Equation 3 and 4.

$$m_i \ddot{r}_i = f_i \tag{3}$$

$$f_i = -\frac{\partial}{\partial r_i} U \tag{4}$$

For this purpose, we need to calculate the forces f_i acting on the atoms, and these are usually derived from a potential energy U (r^N), where $r^N = (r^1, r^2, \ldots r^N)$ represents the complete set of 3N atomic coordinates.

To start simulation, preparation of ligand-target complexes, topology and force field files are needed. The topology file contains all the information about the structure and connectivity of atoms in the system as well as few parameters of the force field. Force field is a function expressing the energy of a system as a sum of diverse molecular mechanic (or other) terms and is governed by a set of predefined parameters. Next step in preparation for MD run is energy minimization (Figure 5). The purpose of this stage is not to find a true global energy minimum, but to adjust the structure to the force field, in particular, distribution of solvent molecules, and to relax possible steric clashes created by assuming coordinates of atoms. Next step is heating the simulation system in a linear manner (e.g. from 0K to 300K) and equilibration. Equilibration stage is used to equilibrate kinetic and potential energies, i.e. to distribute the kinetic energy "pumped" into the system during heating among all degrees of freedom. In other words, the kinetic energy must be transferred to potential energy. As soon as potential energy levels of the equilibration stage are stabilized we can start the simulation. Generated trajectories act as a bridge between theory and experiment. We may test a theory by conducting a simulation. We may test the model by comparing with experimental results. We may also carry out simulations on the computer that are difficult or impossible in the laboratory (for example, working at extreme temperatures or pressures). The evident advantage of MD is that it gives a route to

dynamical properties of the system of interest: transport coefficients, time-dependent responses to perturbations and rheological properties. By introduction of perturbations into the Hamiltonian, or directly into the equations of motion, their effect on the distribution function may be calculated (Lee et al. 2003).

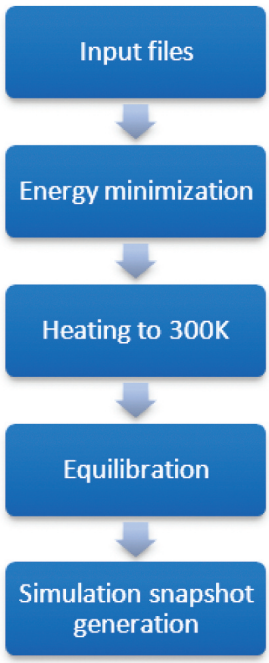

Figure 5. Scheme of basic Molecular Dynamics (MD) workflow.

A combination of *in silico* approaches was used in a recent study of major products of biotransformation of polychlorinated biphenyls (PCBs), namely hydroxylated polychlorinated biphenyls (HO-PCBs) performed by Li and co-workers (Li et al. 2012). Production and use of PCBs were banned in most of the countries, however, these compounds have still been detected in the environment and have received concerns from environmental and ecological perspectives (Dirtu et al. 2010). In this study, docking runs, molecular dynamics simulation, and structure-based 3D-QSAR models were performed to investigate the detailed binding mode between ERα with HO-PCBs and also to develop a rational estrogenic activity predictive model. Molecular docking and MD simulations optimized the bound ligands into the active site of receptor protein, and investigated protein-ligand interactions. A 3D-QSAR CoMSIA model was then developed. The docked complexes of ERα with two HO-PCBs (highly active compound 4′-OH-CB50 and lowly active compound 2′-OH-CB65) were used as the initial structures for MD simulations. Molecular docking and MD simulations suggested that multiple hydrophobic and hydrogen bond interactions are two predominant factors that affect the binding process. Moreover, the probable binding modes of two compounds with much difference in their activity and ERα were analysed based on the results from molecular docking and MD simulations. Combinational use of QSAR, molecular docking, and MD simulations, was useful in defining the ligand-receptor binding modes and provided possible mechanism interpretations.

Using MD as a tool to support or even to substitute wet laboratory work could assist in focusing the laboratory experiments resulting not only in saving considerable amounts of resources, but also increasing the number of molecules and scenarios investigated. However, this tool requires a lot of compute resources and special technical knowledge, despite these facts molecular dynamic studies are being extended to larger systems and longer time scales. Molecular docking and dynamic studies are of considerable importance in a range of disciplines including molecular biology, drug design, and environmental studies and have attracted much attention over the last two decades.

Pharmacokinetics (Toxicokinetics)

Pharmacokinetics studies the rates of liberation, absorption, distribution, metabolism and excretion of drugs and metabolites in biological systems. Pharmacokinetic modelling is possible due to development of mathematical descriptions for these critical flow rates, commonly referred to as the LADME scheme. Pharmacokinetic models have long been used in the prediction of amounts and concentrations of drugs in the body as functions of time and dosing. This progress gives benefits to toxicology and provides a firm foundation for environmental health research. If applied rationally, pharmacokinetic modelling can be advantageous in the safety evaluation and assessment of the adverse effects of environmental chemicals (Clark and Smith 1984; Young and Holson 1978). Pharmacokinetic models are widely applied to predict the kinetics of chemical residues in the environment, to solve pollution problems, and to help understand and interpret the results of toxicology studies (Krewski et al. 2011). As a consequence, several types of pharmacokinetic models exist, e.g. compartmental and noncompartmental models, physiologically based pharmacokinetic models, and population pharmacokinetic models (Figure 6).

Such models have been successful in risk assessment and safety evaluation, in describing chemical distribution in test organisms under various toxicological conditions and doses, in monitoring chemical transfer rates in ecosystems, and in both optimizing each chemical's utility and minimizing its undesirable side effects (Barron et al. 1990; Karara and Hayton 1984; Landrum et al. 1992). Most importantly, these models are extensively used in establishing governmental guidelines and regulations concerning public health

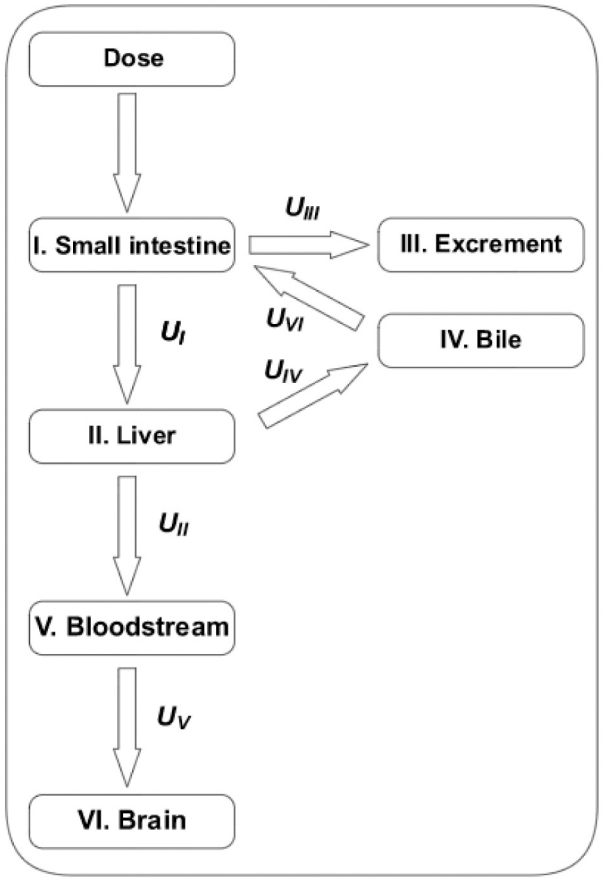

Figure 6. Compartmental model of the pharmacokinetics of drugs excreted by bile or by non-renal mode. Visualization describes the transport of the codeine in modelled organs. Arrows show the direction of codeine flow to succeeding compartments (rectangle). U_I, U_{II}, U_{III}, U_{IV}, U_V, U_{VI} – speed of codeine transfer between compartments. (From Jastrzebski and Urniaz 2011. Adapted with permission from Jagiellonian University, Krakow, Poland).

problems (Oreskes et al. 1994). Pharmacokinetic modelling is currently an area of very intensive research in environmental science. The applications of pharmacokinetic models have been extensive, scientifically precise and accurate. More vigorous utilization of these models is now on the way. Several relevant review articles have focused on the development of models (Barron et al. 1990) and the application of these modelling techniques to toxicological testing (Landrum et al. 1992; Young and Holson 1978). There is also a critical perspective of pharmacokinetic models currently in use (Wen et al. 1999). It is important to remember that pharmacokinetic models are abstractions, subjectively reflecting the defined systems of the organism, and referring to specific scales and parameters. Validation of the predictive capability of pharmacokinetic models is a significant step toward model acceptance. All models must be tested against measurements before their predictions can be regarded as reliable. Quantifying error is an important component of the model description and the first stage in validation. The examination of model sensitivity is of enormous importance to the quality of pharmacokinetic models as well. It gives additional information on the significance of each parameter in the model and offers a powerful tool for model verification. Because modelling is an on-going process, the new models will improve and finally replace the old ones when fresh and more information and technology becomes available.

DISCUSSION

To evaluate the application of molecular modelling techniques in environmental research, we summarized the methods and divided them into four groups. Each group contains methods which may be applied depending on the knowledge available (Table 2). Presented division corresponds to possible stages during the research process. On every step, researchers have to consider different solutions and evaluate them in context to the aim of the study. As it can be observed, the investigators may employ molecular modelling techniques practically on every stage of research advancement. It is a great advantage of molecular modelling methods, although they always stay in close relation to experimental evidence.

From the range of applications reported here, it should be clear that molecular modelling is a very versatile technique and can be applied to many areas of molecular studies. Unequivocal and pure model prediction in which no direct experimental data are used is still an area that must be approached cautiously. Difficulties and pitfalls associated with modelling are still problems for unwary users. Successful prediction methods require careful development being a highly realistic and also still computationally tractable. Due to the increasing trend for creating prediction more valuable is desirable. It allows numerous commotional methods to be applied on a common set of problems and for them to be evaluated in a common way. It is often difficult to evaluate the relative merits of methods from different laboratories or variations in methods from the same laboratory. The development of parameters for molecular mechanic simulations requires a high degree of skill and care. Knowledge of limitations of a particular package is important for its effective use. There are still cases in the literature where molecular modelling methods have been applied to a system for which it was not parameterised. In the case, the quality of results will then be unpredictable. Modelling type simulations are utilised with experimental determination method, where the experimental data are incorporated into model. Simulation schemes can incorporate data from a number of experiments. As computer systems become more powerful with time the utility of methods like modelling and simulation can only increase. This may happen in two ways

Table 2. Table illustrates the applicability of molecular modelling techniques in different research stages.

Number of compounds	Structure of molecular target	
	Unknown	Known
Low	Create the database of known structures and examine their differentiation	Find molecular fragments or ligands filling (grant) into active site
	Apply the automatic synthesis approach	Databases high-throughput screening
		Compounds *de novo* building
High	Build a model of the active site – pharmacophore	Structure based pharmacophore
	Search databases for suitable compounds, fulfil the pharmacophore model assumption	Static and dynamic models of ligand – molecular target complex - molecular docking, molecular dynamics
	Multidimensional QSAR (e.g. 3D-QSAR)	Pharmacokinetics (Toxicokinetics)

- firstly, existing types of simulations will be able to be run for longer time periods, thus allowing better sampling of conformational space and properties. Secondly, more realistic (but more expensive) computation schemes will be accessible in a reasonable time. The growth of structure databases such as PDB will make the application of methods much powerful in context to protein homology building and molecular docking applicable to a wider range of proteins. Molecular modelling methods have much to contribute to our understanding of structural biology and common knowledge. Although molecular modelling base on the models some being simplification of real system, its application may facilitate or even explain the potential interacting mechanisms. Additionally, molecular modelling is commonly used to describe and correlate the experimental results with occurring phenomena. As it was described previously, it is a noticeable fact that the approach increases the speed and efficiency in the discovering process. Moreover, collecting the pieces of information from different experiments may help to coordinate the information and make the experiments more rational. The utility of such analyses is even more evident if in silico precisions stay in close relation with traditional experimental techniques. Molecular modelling approaches are aimed to increase the speed and efficiency in the research process. However, it is not (in some cases) an independent source of knowledge (due to its dependence of high quality data) it provides a more detailed map to the goal.

CONCLUSIONS

Presented brief overview does not exhaust impressive range of molecular modelling techniques, however evaluates current and popular techniques considering their application. Although there are some limitations and imperfections of molecular modelling, the approach can be successfully applied in situations when current knowledge is limited. *In silico* modelling helps summarize current stage of research or propose new developmental directions, becoming a valuable partnership with experiments.

ACKNOWLEDGEMENTS

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Authors are grateful to Dr. Naomi D'souza for editorial help and language support.

REFERENCES

Barron, M., G. Stehly, W. Hayton. 1990. Pharmacokinetic modelling in aquatic animals. 1. Models and concepts. Aquatic Toxicology 18: 61-86.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. 2000. The Protein Data Bank. Nucleic Acids Research 28: 235-242.

Birnbaum, L.S., D.F. Staskal. 2004. Brominated flame retardants: Cause for concern? Environmental Health Perspectives 112: 9–17.

Chang, A., M. Scheer, A. Grote, I. Schomburg, D. Schomburg. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Research 37: D588-D592.

Clark, B., D. Smith. 1984. Pharmacokinetics and toxicity testing. CRC Critical Reviews in Toxicology 12: 343-385.

Clark, R.D., E. Abrahamian. 2009. Using a staged multi-objective optimization approach to find selective pharmacophore models. Journal of Computer-Aided Molecular Design 23: 765–771.

Clark, R.D., U. Norinder. 2012. Two personal perspectives on a key issue in contemporary 3D QSAR. WIREs Computational Molecular Science 2: 108–113.

Cochrane, G., I. Karsch-Mizrachi, Y. Nakamura. 2011. The international nucleotide sequence database collaboration. Nucleic Acids Research 39: D15–D18.

Cramer, R.D., D.E. Patterson, J.D. Bunce. 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society 110: 5959-5967.

Cross, J.B., D.C. Thompson, B.K. Rai, J.C. Baber, K.Y. Fan, Y.B. Hu, C. Humblet. 2009. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. Journal of Chemical Information and Modeling 49: 1455–1474.

Cuff, A.L., I. Sillitoe, T. Lewis, A.B. Clegg, R. Rentzsch, N. Furnham, M. Pellegrini-Calace, D. Jones, J. Thornton, C.A. Orengo. 2011. Extending CATH: increasing coverage of the protein structure universe and linking structure with function. Nucleic Acids Research 39: D420-D426.

Darnerud, P.O., G.S. Eriksen, T. Jóhannesson, P.B. Larsen, M. Viluksela. 2001. Polybrominated diphenyl ethers occurrence, dietary exposure, and toxicology. Environmental Health Perspectives 109: 49–68.

Dirtu, A.C., V.L.B. Jaspers, R. Cernat, H. Neels, A. Covaci. 2010. Distribution of PCBs, their hydroxylated metabolites, and other phenolic contaminants in human serum from two European countries. Environmental Science and Technology 44: 2876-2883.

Fernández-Suárez, X.M., M.Y. Galperin. 2013. The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. Nucleic Acids Research 41: 1-7.

Finn, R.D., J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman. 2010. The Pfam protein families database. Nucleic Acids Research 38: 211-222.

Hansch, C., T. Fujita. 1964. Rho-Sigma-Pi-Analysis. A method for the correlation of biological activity and chemical structure. Journal of American Chemical Society 86: 1616–1626.

Hoopen, P. 2010. ENA as an Information Hub Received 25 October 2010 12:35 UTC; Posted 25 October 2010 (http://precedings.nature.com/documents/5108/version/1) (accessed September 2013).

Hsu, C., M. Bouziane, L. Rattner, L. Yee. 1991. Information resources management in heterogeneous, distributed environments: A metadatabase approach. IEEE Transactions on Software Engineering 6: 604-624.

Irwin, J.J., T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman. 2012. ZINC: a free tool to discover chemistry for biology. Journal of Chemical Information and Modeling 52: 1757–1768.

Jastrzebski, J.P., R. Urniaz. 2011. Compartmental model of the pharmacokinetics of drugs excreted by bile or by non-renal mode. Bio-algorithms and Med-systems 7: 31-37.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, M. Hattori. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Research 32: 277-280.

Kanz, C., P. Aldebert, N. Althorpe, W. Baker, A. Baldwin, K. Bates, P. Browne, A. van den Broek, M. Castro, G. Cochrane, K. Duggan, R. Eberhardt, N. Faruque, J. Gamble, F. Garcia Diez, N. Harte, T. Kulikova, Q. Lin, V. Lombard, R. Lopez, R. Mancuso, M. McHale, F. Nardone, V. Silventoinen, S. Sobhany, P. Stoehr, M.A. Tuli, K. Tzouvara, R. Vaughan, D. Wu, W. Zhu, R. Apweiler. 2005. The EMBL nucleotide sequence database. Nucleic Acids Research 33: D29-D33.

Karara, A., W. Hayton. 1984. Pharmacokinetic model for the uptake and disposition of di-2-etylhexylphtalane in sheepshead minnow *Cyprinodon variegatus*. Aquatic Toxicology 5: 181-195.

Karsch-Mizrachi, I., Y. Nakamura, G. Cochrane. 2011. International nucleotide sequence database collaboration. "The International Nucleotide Sequence Database Collaboration". Nucleic Acids Research 40: D33–D37.

Kitchen, D.B., H. Decornez, J.R. Furr, J. Bajorath. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery 3: 935-949.

Koes, D.R., C.J. Camacho. 2012. ZINCPharmer: pharmacophore search of the ZINC database. Nucleic Acids Research 40: W409–W414.

Krewski, D., M. Westphal, M. Al-Zoughool, M.C. Croteau, M.E. Andersen. 2011. New directions in toxicity testing. Annual Review of Public Health 32: 161-78.

Kubinyi, H. 1995. Strategies and recent technologies in drug discovery. Pharmazie 50: 647-662.

Kubinyi, H. 1997a. QSAR and 3D QSAR in drug design. I. Methodology. Drug Discovery Today 2: 457-467.

Kubinyi, H. 1997b. QSAR and 3D QSAR in drug design. II. Applications and problems. Drug Discovery Today 2: 538-546.

Kulikova, T., P. Aldebert, N. Althorpe, W. Baker, K. Bates, P. Browne, A. van den Broek, G. Cochrane, K. Duggan, R. Eberhardt, N. Faruque, M. Garcia-Pastor, N. Harte, C. Kanz, R. Leinonen, Q. Lin, V. Lombard, R. Lopez, R. Mancuso, M. McHale, F. Nardone, V. Silventoinen, P. Stoehr, G. Stoesser, M.A. Tuli, K. Tzouvara, R. Vaughan, D. Wu, W. Zhu, R. Apweiler. 2004. The EMBL nucleotide sequence database. Nucleic Acids Research 32: D27-D30.

Kuntz, I.D., J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin. 1982. A geometric approach to macromolecule–ligand interaction. Journal of Molecular Biology 162: 269-288.

Landrum, P., H. Lee, M. Lydy. 1992. Toxicokinetics in aquatic systems - model comparison and use in hazard assessment. Environmental Toxicology and Chemistry 11: 1709-1725.

Laskowski, R.A. 2009. PDBsum new things. Nucleic Acids Research 37: 355-359.

Lee, S.H., D.K. Park, D.B. Kang. 2003. Molecular dynamics simulations for transport coefficients of liquid argon: New approaches. The Bulletin of the Korean Chemical Society 24: 178-182.

Li, X., L. Ye, X. Wang, X. Wang, H. Liu, X. Qian, Y. Zhu, H. Yu. 2012. Molecular docking, molecular dynamics simulation, and structure-based 3D-QSAR studies on estrogenic activity of hydroxylated polychlorinated biphenyls. Science of the Total Environment 441: 230–238.

Lomize, M.A., A.L. Lomize, I.D. Pogozheva, H.I. Mosberg. 2006. OPM: orientations of proteins in membranes database. Bioinformatics 22: 623-625.

Lomize, M.A., I.D. Pogozheva, H. Joo, H.I. Mosberg, A.L. Lomize. 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Research 40: 370-376.

Lo Piparo, E., A. Worth. 2010. Review of QSAR models and software tools for predicting developmental and reproductive toxicity (http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/doc/EUR_24522_EN.pdf) (accessed September 2013).

López-Ramos, M., F. Perruccio. 2010. HPPD: Ligand- and target-based virtual screening on a herbicide target. Journal of Chemical Information and Modeling 50: 801–814.

Martin, Y.C., J.L. Kofron, L.M. Traphagen. 2002. Do structurally similar molecules have similar biological activity. Journal of Medicinal Chemistry 45: 4350–4358.

McGregor, M.J. 2007. A pharmacophore map of small molecule protein kinase inhibitors. Journal of Chemical Information and Modeling 47: 2374-2382.

Murzin, A.G., S.E. Brenner, T. Hubbard, C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247: 536-540.

O'Maille, P.E., M. Bakhtina, M.D. Tsai. 2002. Structure-based combinatorial protein engineering (SCOPE). Journal of Molecular Biology 321: 677–691.

Orengo, C.A., A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton. 1997. CATH – a hierarchic classification of protein domain structures. Structure 5: 1093-1109.

Oreskes, N., K. Shrader-Frechette, K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the Earth sciences. Science 263: 641-646.

Pearl, F.M.G., N. Martin, J.E. Bray, D.W.A. Buchan, A.P. Harrison, D. Lee, G.A. Reeves, A.J. Shepherd, I. Sillitoe, A.E. Todd, J.M. Thornton, C.A. Orengo. 2001. A rapid

classification protocol for the CATH Domain Database to support structural genomics. Nucleic Acids Research 29: 223-227.

Perola, E., W.P. Walters, P.S. Charifson. 2004. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. PROTEINS: Structure, Function, and Bioinformatics 56: 235–249.

Schecter, A., O. Papke, K.C. Tung, J. Joseph, T.R. Harris, J. Dahlgreen. 2005. Polybrominated diphenyl ether flame retardants in the U.S. population: Current levels, temporal trends, and comparison with dioxins, dibenzofurans, and polychlorinated biphenyls. Journal of Occupational and Environmental Medicine 47: 199–211.

Schultza, T.W., M.T.D. Croninb, T.I. Netzevab. 2003. The present status of QSAR in toxicology. Journal of Molecular Structure (Theochem) 622: 23–38.

Urniaz, R.D., K. Jozwiak. 2013. X-ray crystallographic structures as a source of ligand alignment in 3D-QSAR. Journal of Chemical Information and Modeling 53: 1406-1414.

Verma, J., V.M. Khedkar, E.C. Coutinho. 2010. 3D-QSAR in drug design - a review. Current Topics in Medicinal Chemistry 10: 95-115.

Wen, Y.H., J. Kalff, R.H. Peters. 1999. Pharmacokinetic modeling in toxicology: a critical perspective. Environmental Reviews 7: 1–18.

Wermuth, C.G., C.R. Ganellin, P. Lindberg, L.A. Mitscher. 1998. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure and Applied Chemistry 70: 1129–1143.

Wu, C.H., L.S. Yeh, H. Huang, L. Arminski, J. Castro-Alvear, Y. Chen, Z. Hu, P. Kourtesis, R.S. Ledley, B.E. Suzek, C.R. Vinayaka, J. Zhang, W.C. Barker. 2003. The protein information resource. Nucleic Acids Research 31: 345-347.

Yang, W.H., Z.Y. Wang, H.L. Liu, H.X. Yu. 2010. Exploring the binding features of polybrominated diphenyl ethers as estrogen receptor antagonists: docking studies. SAR and QSAR in Environmental Research 21: 351–367.

Young, J., J. Holson. 1978. Utility of pharmacokinetics in designing toxicological protocols and improving interspecies explorations. Journal of Environmental Pathology and Toxicology 2: 169-186.