Toxicology and Applied Pharmacology 234 (2009) 47–57

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

In silico identification of anthropogenic chemicals as ligands of zebrafish sex hormone binding globulin

Nels Thorsteinson [a], Fuqiang Ban [a], Osvaldo Santos-Filho [a], Seyed M.H. Tabaei [a], Solange Miguel-Queralt [b], Caroline Underhill [b], Artem Cherkasov [a,*], Geoffrey L. Hammond [b]

[a] Prostate Centre at the Vancouver General Hospital, University of British Columbia, 2660 Oak Street, Vancouver, BC, Canada V6H 3Z6
[b] Department of Obstetrics and Gynecology, University of British Columbia, Child & Family Research Institute, 950 West 28th Avenue, Vancouver, BC, Canada V5Z 4H4

ARTICLE INFO

Article history:
Received 30 April 2008
Revised 27 June 2008
Accepted 7 July 2008
Available online 25 July 2008

Keywords:
Zebrafish SHBG
Computational toxicology
Virtual screening
Docking
QSAR

ABSTRACT

Anthropogenic compounds with the capacity to interact with the steroid-binding site of sex hormone binding globulin (SHBG) pose health risks to humans and other vertebrates including fish. Building on studies of human SHBG, we have applied in silico drug discovery methods to identify potential binders for SHBG in zebrafish (*Danio rerio*) as a model aquatic organism. Computational methods, including; homology modeling, molecular dynamics simulations, virtual screening, and 3D QSAR analysis, successfully identified 6 non-steroidal substances from the ZINC chemical database that bind to zebrafish SHBG (zfSHBG) with low-micromolar to nanomolar affinities, as determined by a competitive ligand-binding assay. We also screened 80,000 commercial substances listed by the European Chemicals Bureau and Environment Canada, and 6 non-steroidal hits from this in silico screen were tested experimentally for zfSHBG binding. All 6 of these compounds displaced the [^{3}H]5α-dihydrotestosterone used as labeled ligand in the zfSHBG screening assay when tested at a 33 μM concentration, and 3 of them (hexestrol, 4-*tert*-octylcatechol, and dihydrobenzo(a) pyren-7(8H)-one) bind to zfSHBG in the micromolar range. The study demonstrates the feasibility of large-scale in silico screening of anthropogenic compounds that may disrupt or highjack functionally important protein:ligand interactions. Such studies could increase the awareness of hazards posed by existing commercial chemicals at relatively low cost.

Introduction

As in other vertebrates, sex hormone binding globulin (SHBG) transports sex steroids in the blood of fish and regulates their access to tissues (Siiteri et al., 1982). In addition to binding its natural steroid ligands, SHBG in humans bind xenobiotics including potential endocrine disruptors present in waste water systems (Hodgert-Jury et al., 2000). Since aquatic species are in intimate contact with these anthropogenic compounds, fish SHBGs represent interesting targets for computational toxicology studies of potentially harmful environmental contaminants.

For more than two decades, the steroid benchmark set of ligands has been used to study the interactions between steroids and their plasma transport proteins, and as a model system for in silico drug design research (Tuppurainen et al., 2004; Asikainen et al., 2004; Korhonen et al., 2003; Liu et al., 2002; Tuppurainen et al., 2002; Liu et al., 2001; Cherkasov et al., 2005; Cherkasov et al., 2008; Cramer et al., 1988; Klebe et al., 1994). In the early applications, the steroid benchmark set was used to develop popular molecular modeling tools such as comparative molecular field analysis (CoMFA) (Cramer et al., 1988) and comparative molecular similarity index analysis (CoMSiA) (Klebe et al., 1994). Most recently, novel non-steroidal nanomolar ligands of human SHBG (hSHBG) have been identified by applying such methods to an extended and alignment-corrected 76 steroid version of the benchmark steroids (Cherkasov et al., 2008).

The biological importance of SHBG in fish is not as well studied as in mammals. It has been demonstrated that the protein is expressed primarily in the liver and intestine of zebrafish (*Danio rerio*) (Miguel-Queralt et al., 2004), and the uptake of steroids from their aquatic environment appears to be influenced in some way by their affinity for SHBG (Scott et al., 2005). Given the role SHBG may play in

Abbreviations: CAS, chemical abstract service; CEPA, Canadian environmental protection act; CoMFA, comparative molecular field analysis; CoMSiA, molecular similarity indices in a comparative analysis; DBP, 9,10-dihydrobenzo(a)pyren-7(8H)-one; DHT, 5α-dihydrotestosterone; DSL, domestic substances list; EE, ethinylestradiol; EINECS, European inventory of existing chemical substances; hSHBG, human sex hormone binding globulin; LIE, linear interaction energy; LOO, leave one out; MD, molecular dynamics; MMF, Merck molecular forcefield; MOE, Molecular Operating Environment; NDSL, non-domestic substances list; OC, 4-*tert*-octylcatechol; OP, 4-*tert*-octylphenol; PAH, polycyclic aromatic hydrocarbon; PDB, protein data bank; PLS, partial least squares; RBA, relative binding affinity; SD format, structure data format; SHBG, sex hormone binding globulin; zfSHBG, zebrafish sex hormone binding globulin; SMILES, simplified molecular input line entry specification; SVL, scientific vector language.

* Corresponding author.
E-mail address: artc@interchange.ubc.ca (A. Cherkasov).

0041-008X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.taap.2008.07.014

regulating the bioavailability of androgens and estrogens during sexual differentiation and the reproductive cycle in fish (Miguel-Queralt et al., 2007), environmental compounds that bind to fish SHBG could adversely influence their reproductive performance. Ethinylestradiol (EE) was of particular interest in this context because it is a well-known endocrine disruptor in some fish and has an unusually high affinity for SHBG in zebrafish (Miguel-Queralt and Hammond, 2008). Moreover, synthetic ligands of fish SHBG sequestered from water may accumulate in the bodies of the fish and subsequently harm predators, including humans. Public concern about the potential toxicity of such xenobiotic substances in the oceans, lakes, and rivers has prompted several government funded environmental agencies, such as the European Chemicals Bureau and Environment Canada, to identify commercial substances that represent health or environmental risks.

The sequence of zebrafish SHBG (zfSHBG) has been reported together with values of its relative binding affinity (RBA) for 19 steroids from the steroid benchmark set (Miguel-Queralt et al., 2004). In the current work, a zfSHBG homology model was built from an hSHBG crystal structure template (Grishkovskaya et al., 2002b). Molecular dynamics (MD) simulations were performed to refine the model; to explain previously observed zfSHBG steroid-binding characteristics; to identify amino acid substitutions responsible for the unique ligand-binding properties of hSHBG and zfSHBG, and to assess the accuracy of the model.

The zfSHBG model was then used in a multi-method virtual screening pipeline which involved large-scale docking as well as CoMFA and CoMSiA modeling to identify zfSHBG ligands from the following lists of existing commercially produced chemicals: the European inventory of existing commercial substances (EINECS) (European Chemicals Bureau, 2002) and Environment Canada's domestic and non-domestic substances lists (DSL and NDSL) (Environment Canada, 2006). A small set of the top virtual screening hits from these lists were validated experimentally in a zfSHBG ligand-binding assay. Virtual screening was also applied to identify potential anthropegenic zfSHBG ligands from the ZINC database (Irwin and Shoichet, 2005). ZINC is a large, freely available database of compounds that have passed criteria making them suitable for virtual screening in drug discovery applications. The compounds in ZINC are commercially available but are not necessarily produced in significant quantity by industry.

Materials and methods

ZINC chemical database preparation. Almost four million structures in SD format from the ZINC database were imported into a database using Molecular Operating Environment (MOE) version 2006.08 (Chemical Computing Group, Inc., 2006). These structures were washed —i.e. all inorganic components were removed, and all ionizable groups were coordinated with pH=7.0 conditions. Next, the database was energy minimized using the MMFF94x forcefield (Halgren, 1996) and exported in SD format for use by the Glide (Schrödinger Inc, 2006) docking program.

Commercial chemical database preparation. Structures for over 80,000 compounds from various environmental compounds lists were obtained. The lists include the EINECS (European Chemicals Bureau, 2002) and the Canadian Environmental Protection Act (CEPA) environmental registry's DSL and NDSL (Environment Canada, 2006). The EINECS contains substances manufactured or imported into the European Union and has a high overlap with CEPA's DSL and NDSL. The 80,000 compounds obtained were those within the grasp of our resources and they represent the majority of the substances in the lists.

We obtained 68,970 substances in simplified molecular input line entry specification (SMILES) format from the European Chemicals Bureau (http://ecb.jrc.it/qsar/information-sources/). Most of the remaining EINECS and CEPA structures were also obtained from various online resources as SMILES strings, but we were not able to obtain the remaining EINECS and CEPA substances. In the end, SMILES strings for approximately 80,000 compounds, representing about 70% of the EINECS and CEPA lists, were obtained for in silico screening of zfSHBG binding.

The 80,000 SMILES strings were imported into a MOE database. MOE was used to rebuild the SMILES strings into 3D structures. These structures were washed —i.e. all inorganic components were removed, and all ionizable groups were coordinated with pH=7.0 conditions. Next, the database was energy minimized using the MMFF94x forcefield and exported in SD format for use by the Glide docking program.

Homology modeling of zfSHBG. SHBG protein sequences from five different species: zebrafish, rainbow trout, European seabass, mouse, and human were obtained from the NCBI protein database and correspond to accession numbers AAU14174, BAE48779, AAW23033, NP_035497, and AAC18778, respectively. A multiple alignment of these five sequences was performed and the hSHBG to zfSHBG amino acid mapping within this alignment was used for homology modeling. MOE was utilized with default settings to construct the zfSHBG homology model from the hSHBG template structure of the 1KDM entry (Grishkovskaya et al., 2002b) of the protein data bank (PDB).

Molecular dynamics and binding free energy calculations. All MD simulations and binding free energy calculations were performed on zfSHBG using the GROMACS 3.3 simulation package (van der Spoel et al., 2005). Newton's equations of motion were integrated with a time step of 1.5 fs. Short-range and long-range forces were cutoff at 0.9 nm and 1.4 nm respectively and periodic boundary conditions were applied. The simulations were conducted at 300 K.

The linear interaction energy (LIE) method (Hansson et al., 1998) was used to calculate the ΔG_{bind} values. For each ΔG_{bind} calculation, two 750 ps NVT simulations (constant moles, volume, and temperature) were performed; one with the protein bound with the ligand in water, and one with just the ligand in water. PRODRG (Schuttelkopf and van Aalten, 2004) was used to generate the ligand topologies according to the GROMACS87 forcefield but the ligand Gasteiger partial charges (Gasteiger and Marsili, 1980) were calculated using MOE. The Coulomb and Lennard–Jones interaction between ligand and solvent from the 750 ps simulations were used in the LIE formula with the following parameters: β=0.33, γ=0, α=0.18.

Molecular docking. The Maestro suite (Schrödinger Inc., 2004) was used to prepare the 1LHN hSHBG structure and the zfSHBG model for docking. All water and ion molecules were removed from the corresponding PDB files, and hydrogen atoms were added and adjusted where necessary. The steroid-binding sites were defined as 10 Å surrounding the ligands in all cases. Docking was performed using Glide 4.0 parallel suite with default settings.

The MOE estimated pK_i was calculated for each ligand using the scoring.svl script available through the SVL exchange service (Chemical Computing Group, Inc., 2005). For this, hydrogen atoms were added to the zfSHBG model or to 1LHN and the partial charges were calculated with the AMBER99 forcefield (Wang et al., 2000). The Gasteiger partial charges were calculated for the structures that passed the docking cutoff. The estimated pK_i for these structures were calculated by choosing the dock_p*K*i descriptor with default settings for the molecular database.

CoMFA modeling. A set of 19 ligands from the steroid benchmark set with experimental pK_as in zfSHBG were docked using Glide. The resulting docking poses were used to build CoMFA and CoMSiA models. For hSHBG, 87 steroids from (Cherkasov et al., 2008) were