pkCSM: predicting small-molecule pharmacokinetic properties using graph-based signatures

Douglas E. V. Pires*, Tom L. Blundell and David B. Ascher*

*Correspondence: douglas.pires@cpqrr.fiocruz.br, dascher@svi.edu.au

Theory - How to interpret pkCSM results

Absorption

Caco-2 Permeability

The Caco-2 cell line is composed of human epithelial colorectal adenocarcinoma cells. The Caco-2 monolayer of cells is widely used as an *in vitro* model of the human intestinal mucosa to predict the absorption of orally administered drugs. This model is based on 674 drug like molecules with Caco-2 permeability values and predicts the logarithm of the apparent permeability coefficient (log Papp; log cm/s).

How to interpret the results:

A compound is considered to have a high Caco-2 permeability if it has a Papp > 8 x 10^{-6} cm/s. For the pkCSM predictive model, **high Caco-2 permeability** would translate in predicted values **> 0.90**.

Water Solubility

The water solubility of a compound (logS) reflects the solubility of the molecule in water at 25°C. Lipid-soluble drugs are less well absorbed than water-soluble ones, especially when they are enteral. This model is built using experimental water solubility measurements of 1708 molecules.

How to interpret the results:

The predicted water solubility of a compound is given as the logarithm of the molar concentration (log mol/L).

Intestinal Absorption (Human)

The Intestine is normally the primary site for absorption of a drug from an orally administered solution. This method is built to predict the proportion of compounds that were absorbed through the human small intestine.

How to interpret the results:

For a given compound it predicts the percentage that will be absorbed through the human intestine. A molecule with an absorbance of **less than 30%** is considered to be **poorly absorbed**.

P-glycoprotein substrate

The P-glycoprotein is an ATP-binding cassette (ABC) transporter. It functions as a biological barrier by extruding toxins and xenobiotics out of cells. P-glycoprotein transport screening is performed using transgenic mdr knockout mice and *in vitro* cell systems. This model was built using 332 compounds that have been characterised for their ability to be transported by Pgp.

How to interpret the results:

The model predicts whether a given compound is likely to be a substrate of Pgp or not.

Theory - How to interpret pkCSM results

Absorption

P-glycoprotein I and II inhibitors

Modulation of P-glycoprotein mediated transport has significant pharmacokinetic implications for Pgp substrates, which may either be exploited for specific therapeutic advantages or result in contraindications. This predictive models were build using 1273 and 1275 compounds that have been characterised for their ability to inhibit P-glycoprotein I and P-glycoprotein I transport, respectively.

How to interpret the results:

The predictor will determine is a given compound is likely to be a P-glycoprotein I/II inhibitor.

Skin Permeability

Skin permeability is a significant consideration for many consumer products efficacy, and of interest for the development of transdermal drug delivery. This predictor was built using 211 compounds whose in vitro human skin permeability has been measured.

How to interpret the results:

It predicts whether if given compound is likely to be skin permeable, expressed as the skin permeability constant logKp (cm/h). A compound is considered to have a relatively low skin permeability if it has a **logKp > -2.5**.

Theory - How to interpret pkCSM results

Distribution

VDss (Human)

The volume of distribution (VDss) is the theoretical volume that the total dose of a drug would need to be uniformly distributed to give the same concentration as in blood plasma. The higher the VD is, the more of a drug is distributed in tissue rather than plasma. It can be affected by renal failure and dehydration. This predictive model was built using the calculated steady state volume of disdribution (VDss) in humans from 670 drugs. The predicted logarithm of VDss of a given compound is given as the log L/kg.

How to interpret the results:

VDss is considered **low** if below 0.71 L/kg **(log VDss < -0.15)** and **high** if above 2.81 L/kg **(log VDss > 0.45)**.

Fraction Unbound (Human)

Most drugs in plasma will exist in equilibrium between either an unbound state or bound to serum proteins. Efficacy of a given drug may be affect by the degree to which it binds proteins within blood, as the more that is bound the less efficiently it can traverse cellular membranes or diffuse. This predictive model was built using the measured free proportion of 552 compounds in human blood (Fu).

How to interpret the results:

For a given compound the predicted fraction that would be unbound in plasma will be calculated.

BBB permeability

The brain is protected from exogenous compounds by the blood-brain barrier (BBB). The ability of a drug to cross into the brain is an important parameter to consider to help reduce side effects and toxicities or to improve the efficacy of drugs whose pharmacological activity is within the brain. Blood-brain permeability is measured *in vivo* in animals models as logBB, the logarithmic ratio of brain to plasma drug concentrations. This predictive model was built using 320 compounds whose logBB has been experimentally measured.

How to interpret the results:

For a given compound, a **logBB > 0.3** considered to **readily cross the blood-brain barrier** while molecules with **logBB < –1** are **poorly distributed to the brain**.

CNS permeability

Measuring blood brain permeability can difficult with confounding factors. The blood-brain permeability-surface area product (logPS) is a more direct measurement. It is obtained from *in situ* brain perfusions with the compound directly injected into the carotid artery. This lacks the systemic distribution effects which may distort brain penetration. This predictive model was built using 153 compounds whose logPS has been experimentally measured.

How to interpret the results:

Compounds with a **logPS > -2** are considered to **penetrate the Central Nervous System (CNS)**, while those with **logPS < -3** are considered as **unable to penetrate the CNS**.

Theory - How to interpret pkCSM results

Metabolism

Cytochrome P450 inhibitors

Cytochrome P450 is an important detoxification enzyme in the body, mainly found in the liver. It oxidises xenobiotics to facilitate their excretion. Many drugs are deactivated by the cytochrome P450's, and some can be activated by it. Inhibitors of this enzyme, such as grapefruit juice, can affect drug metabolism and are contraindicated. It is therefore important to assess a compounds ability to inhibit the cytochrome P450. Models for different isoforms were built (CYP1A2/CYP2C19/CYP2C9/CYP2D6 /CYP3A4) using from over 14000 to 18000 compounds whose ability to inhibit the cytochrome P450 has been determined. A compound is considered to be a cytochrome P450 inhibitor if the concentration required to lead to 50% inhibition is less than 10 uM.

How to interpret the results:

The predictors will assess a given molecule to determine whether it is likely going to be a cytochrome P450 inhibitor, for a given isoform.

CYP2D6/CYP3A4 substrate

The cytochrome P450's are responsible for metabolism of many drugs. However inhibitors of the P450's can dramatically alter the pharmacokinetics of these drugs. It is therefore important to assess whether a given compound is likely to be a cytochrome P450 substrate. The two main isoforms responsible for drug metabolism are 2D6 and 3A4. These models were built using 671 compounds whose metabolism by each cytochrome P450 isoform has been measured.

How to interpret the results:

The predictor will assess whether a given molecule is likely to be metabolised by either P450.

Theory - How to interpret pkCSM results

Excretion

Renal OCT2 substrate

Organic Cation Transporter 2 is a renal uptake transporter that plays an important role in disposition and renal clearance of drugs and endogenous compounds. OCT2 substrates also have the potential for adverse interactions with coadministered OCT2 inhibitors. Assessing a candidates potential to be transported by OCT2 provides useful information regarding not only its clearance but potential contraindications. This model was built using 906 compounds whose transport by OCT2 has been experimentally measured.

How to interpret the results:

The predictor will assess whether a given molecule is likely to be an OCT2 substrate.

Total Clearance

Drug clearance is measured by the proportionality constant CLtot, and occurs primarily as a combination of hepatic clearance (metabolism in the liver and biliary clearance) and renal clearance (excretion via the kidneys). It is related to bioavailability, and is important for determining dosing rates to achieve steady-state concentrations. This predictor was built using the total clearance data for 398 compounds.

How to interpret the results:

The predicted total clearance log(CLtot) of a given compound is given in log(ml/min/kg).

Theory - How to interpret pkCSM results

Toxicity

Rat LD50

It is important to consider the toxic potency of a potential compound. The lethal dosage values (LD50) are a standard measurement of acute toxicity used to assess the relative toxicity of different molecules. The LD50 is the amount of a compound given all at once that causes the death of 50% of a group of test animals.

How to interpret the results:

The model was built on over 10000 compounds tested in rats and predicts the LD50 (in mol/kg).

AMES toxicity

The Ames test is a widely employed method to assess a compounds mutageneic potential using bacteria. A positive test indicates that the compound is mutagenic and therefore may act as a carcinogen. This predictive model was built on the results of over 8000 compounds Ames tests.

How to interpret the results:

It predicts whether a given compound is likely to be Ames positive and hence mutagenic.

T. Pyriformis toxicity

T. Pyriformis is a protozoa bacteria, with its toxicity often used as a toxic endpoint. This method was build using the concentration of 1571 compounds required to inhibit 50% of growth (IGC50).

How to interpret the results:

For a given compound, the pIGC50 (negative logarithm of the concentration required to inhibit 50% growth in log ug/L) is predicted, with a value **> -0.5 log ug/L** is considered **toxic**.

Minnow toxicity

The lethal concentration values (LC50) represent the concentration of a molecule necessary to cause the death of 50% of the Flathead Minnows. This predictive model was built on LC50 measurements for 554 compounds.

How to interpret the results:

For a given compound, a log LC50 will be predicted. LC50 values below 0.5 mM **(log LC50 < -0.3)** are regarded as **high acute toxicity**.

Theory - How to interpret pkCSM results

Toxicity

Maximum Tolerated Dose

The maximum recommended tolerated dose (MRTD) provides an estimate of the toxic dose threshold of chemicals in humans. The model is trained using 1222 experimental data points from human clinical trials and predicts the logarithm of the MRTD (log mg/kg/day). This will help guide the maximum recommended starting dose for pharmaceuticals in phase I clinical trials, which are currently based on extrapolations from animal data.

How to interpret the results:

For a given compound, a MRTD of **less than or equal to 0.477 log(mg/kg/day)** is considered **low**, and **high** if **greater than 0.477 log(mg/kg/day)**.

Hepatotoxicity

Drug-induced liver injury is a major safety concern for drug development and a significant cause of drug attrition. This predictor was built using the liver associated side effects of 531 compounds observed in humans. A compound was classed as hepatotoxic if it had at least one pathological or physiological liver event which is strongly associated with disrupted normal function of the liver.

How to interpret the results:

It predicts whether a given compound is likely to be associated with disrupted normal function of the liver.

Oral Rat Chronic Toxicity

Exposure to low-moderate doses of chemicals over long periods of time is of significant concern in many treatment strategies. Chronic studies aim to identify the lowest dose of a compound that results in an observed adverse effect (LOAEL), and the highest dose at which no adverse effects are observed (NOAEL). This predictor was built using the LOAEL results from 445 compounds.

How to interpret the results:

For a given compound, the predicted log Lowest Observed Adverse Effect (LOAEL) in **log(mg/kg_bw/day)** will be generated. The LOAEL results need to be interpreted relative to the bioactive concentration and treatment lengths required.

Skin Sensitisation

Skin sensitisation is a potential adverse effect for dermally applied products. The evaluation of whether a compound, that may encountered the skin, can induce allergic contact dermatitis is an important safety concern. This predictor was built using 254 compounds which have been evaluated for their ability to induce skin sensitisation.

How to interpret the results:

It predicts whether a given compound is likely to be associated with skin sensitisation.

Theory - How to interpret pkCSM results

Toxicity

hERG I and II Inhibitors

Inhibition of the potassium channels encoded by hERG (human ether-a-go-go gene) are the principal causes for the development of acquire long QT syndrome - leading to fatal ventricular arrhythmia. Inhibition of hERG channels has resulted in the withdrawal of many substances from the pharmaceutical market. These predictors were built using hERG I and II inhibition information for 368 and 806 compounds, respectively.

How to interpret the results:

The predictor will determine if a given compound is likely to be a hERG I/II inhibitor.