ACD/Labs

Data sheet

Cytochrome P450 Specificity Module

ACD/Percepta

Overview

ACD/Labs has developed predictive models that calculate how compounds will interact with the five cytochrome P450 (CYP) isoforms: 3A4, 2D6, 2C9, 2C19, and 1A2, that are responsible for the majority of Phase I metabolic reactions. The Cytochrome P450 Specificity prediction package consists of three components:

- P450 Substrates estimates if a compound is metabolized by a particular CYP isoform.
- P450 Inhibitors estimates if a compound is a CYP inhibitor, or efficient CYP inhibitor.
- P450 Regioselectivity predicts sites of metabolism by Human Liver Microsomes or the 3A4, 2D6, 2C9, 2C19 and 1A2 isoforms.

These software tools can assist with the early detection of compounds that may be problematic due to metabolic liabilities or metabolism-based drug interactions, making them a valuable part of a DMPK, comp chem. or med chem workflow.

Features

- Calculates the probability of a compound being a substrate or inhibitor of a particular cytochrome P450 isoform.
- Predicts the probability that individual atoms within a compound will be a site of metabolic reaction catalyzed by cytochrome P450 or other microsomal enzymes.
- View metabolic reactions for each isoform—ranked from most probable, to least likely to occur
- All predictions are supported by Reliability Index values that represent a quantitative evaluation of prediction confidence. A high RI value indicates that the calculated value is likely to be accurate, while a low RI indicates that no similar compounds with consistent data are present in the training set.
- Displays experimental results for up to five most similar compounds from the respective training set.
- Fast batch calculations (hundreds of molecules per minute) without user intervention.
- Training of P450 Inhibitors and Substrates models using ‘in-house’ experimental data is possible, providing the possibility to increase the accuracy of predictions or adapt the model to a particular experimental protocol used in your company.

Technical Information

Predicted endpoints

1. P450 Substrates: Probability that the compound will be metabolized by a certain P450 isoform.

2. P450 Inhibitors: Probability that the compound will inhibit a certain 450 enzyme with IC50 below defined threshold:
    - IC50 < 50 μM – "General inhibition" models
    - IC50 < 10 μM – "Efficient inhibition" models
3. P450 Regioselectivity: Probability to be metabolized in human liver microsomes or by a specific CYP450 enzyme for every atom in the molecule (see below for a definition of atoms that are considered possible sites of metabolism).

Sources of Experimental data

1. **P450 Metabolism** (Substrate specificity and metabolism sites): only experimental data from original scientific publications were used for modeling of cytochrome P450 metabolism sites. The full dataset contained ~700 compounds with >800 possible metabolism sites. The literature dataset was expanded with information about marketed drugs' metabolism and the expanded dataset was used for cytochrome P450 substrate modeling.
2. **P450 Inhibition:** Two types of experimental data were used for cytochrome P450 inhibition modeling, including data from original scientific publications, information about marketed drugs, as well as data from the NCBI PubChem project.

Data set sizes

Table 1. The sizes of the data sets used to develop the predictive models of substrate and inhibitor specificity.

Cytochrome P450 isoform	Substrates	Inhibitors (IC_{50} < 50 μM)	Efficient Inhibitors (IC_{50} < 10 μM)
CYP1A2	935	4867	5815
CYP2C9	867	7666	7677
CYP2C19	794	6899	6833
CYP2D6	1001	7707	7507
CYP3A4	960	6684	7927

Regioselectivity models were based on experimental data for 873 compounds collected from publications dealing with analytical identification of the metabolites observed after the incubation of compound with human liver microsomes or recombinant cytochrome P450 enzymes.

Any carbon atom with at least one hydrogen attached was considered a potential site of metabolism. For each of these carbon atoms, a value of 1 was assigned if a metabolic reaction occurs at this site and a value of 0 was assigned otherwise. Carbon atoms with no hydrogens attached were ignored.

For dealkylation reactions, carbon atoms of the leaving groups were marked in the same manner. Some sites were marked as "inconclusive" and consequently not used in the modeling. Table 2 shows the overall number of marked atoms used for building the models:

Table 2. The sizes of the data sets used to develop the predictive models of HLM Regioselectivity.

Cytochrome P450 isoform	No. of atoms			
	Positive	Inconclusive	Negative	Total
CYP1A2	383	61	6020	6464
CYP2C19	249	15	5210	5474
CYP2C9	288	43	6314	6645
CYP2D6	354	49	6305	6708
CYP3A4	795	176	6757	7728
Overall (HLM)	1269	340	7182	8791

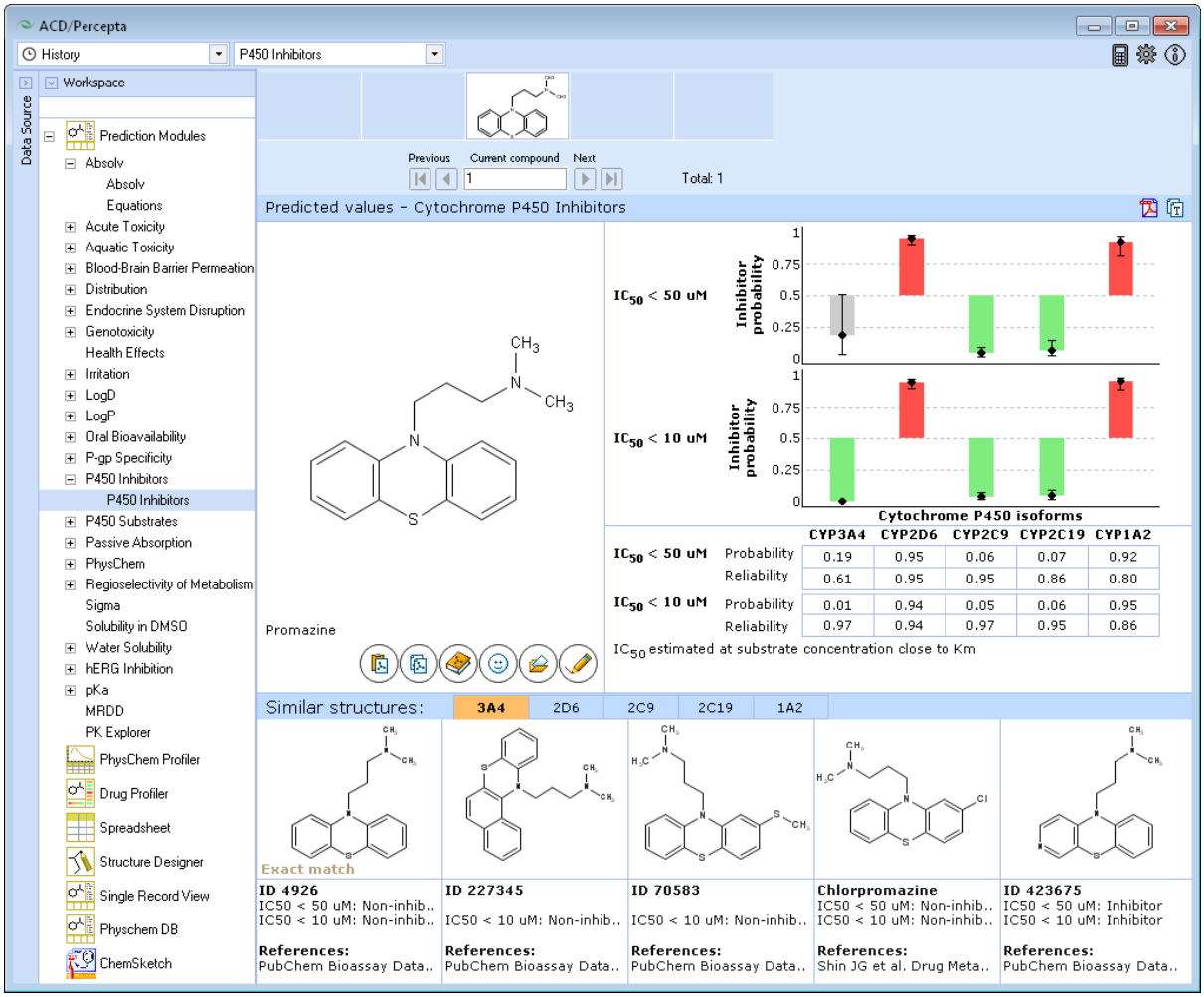

Figure 1. *Screenshot of ACD/Percepta P450 Inhibitors module*

Modeling details & prediction accuracy

Only a short summary of the main technical aspects of model development is given here. For a more detailed description please refer to the following articles:

- Dapkunas J et al. *Chem Biodivers*. **2009** Nov;6(11):2101-6. [1] – P450 Regioselectivity
- Didziapetris R et al. *J Comput Aided Mol Des.* **2010**;24(11):891-906. [2] – CYP3A4 Inhibition

Probabilistic predictive models for both P450 substrate and inhibitor specificity were derived using GALAS (Global, Adjusted Locally According to Similarity) modeling methodology. A GALAS model consists of two parts:

- **Global** (baseline) statistical model based on binomial PLS with multiple bootstrapping, using a predefined set of fragmental descriptors.
- **Local** correction to baseline prediction based on the analysis of model performance for similar compounds from the training set (the so called Self-training Library).

The same general principles were applied for cytochrome P450 regioselectivity modeling with the only exception that here different atoms (sites of metabolism) within the same molecule were considered distinct data points.

Reliability Index (*RI*)

Local part of the model provides the basis for estimating reliability of prediction by the means of calculated Reliability Index (*RI*) values. *RI* is a number ranging from 0 to 1 (0 – unreliable prediction, 1 – idealistic, fully reliable prediction). The following two criteria are applied for reliability estimation:

- Similarity of the analyzed molecule to compounds in the Self-training Library (a reliable prediction cannot be made if no similar compounds have been found in the Library).
- Consistency of model predictions with experimental data for similar compounds (i.e. alternating P-gp inhibitors and non-inhibitors among similar molecules lead to lower *RI* values).

RI can serve as a valuable tool for interpreting prediction results. If a compound obtains *RI* lower than a certain cut-off value (typically, set at 0.3), it means that this compound falls outside of the Model Applicability Domain, and the respective prediction should be discarded from further analysis regardless of calculated probabilities.

Improving Prediction Accuracy via Training

The ACD/P450 Substrates and ACD/P450 Inhibitor modules also implement the Trainability feature. It addresses the issue of the chemical space of ‘in-house’ libraries being considerably wider than that of publicly available data which results in limited applicability of most third-party QSARs for analysis of ‘in-house’ data. The ‘Training engine‘ makes appropriate corrections for systematic deviations produced by the baseline QSAR model based on analysis of similar compounds from the experimental data library.

Addition of user-defined experimental data for new compounds to the model Self-training Library leads to an instant improvement of prediction accuracy for the respective compound classes. Moreover, training procedure also allows adapting the existing model to the particular experimental protocol used in your company thus reducing potential issues related to discrepancies between different experimental methods used for determination of drug interactions with cytochrome P450 enzymes.

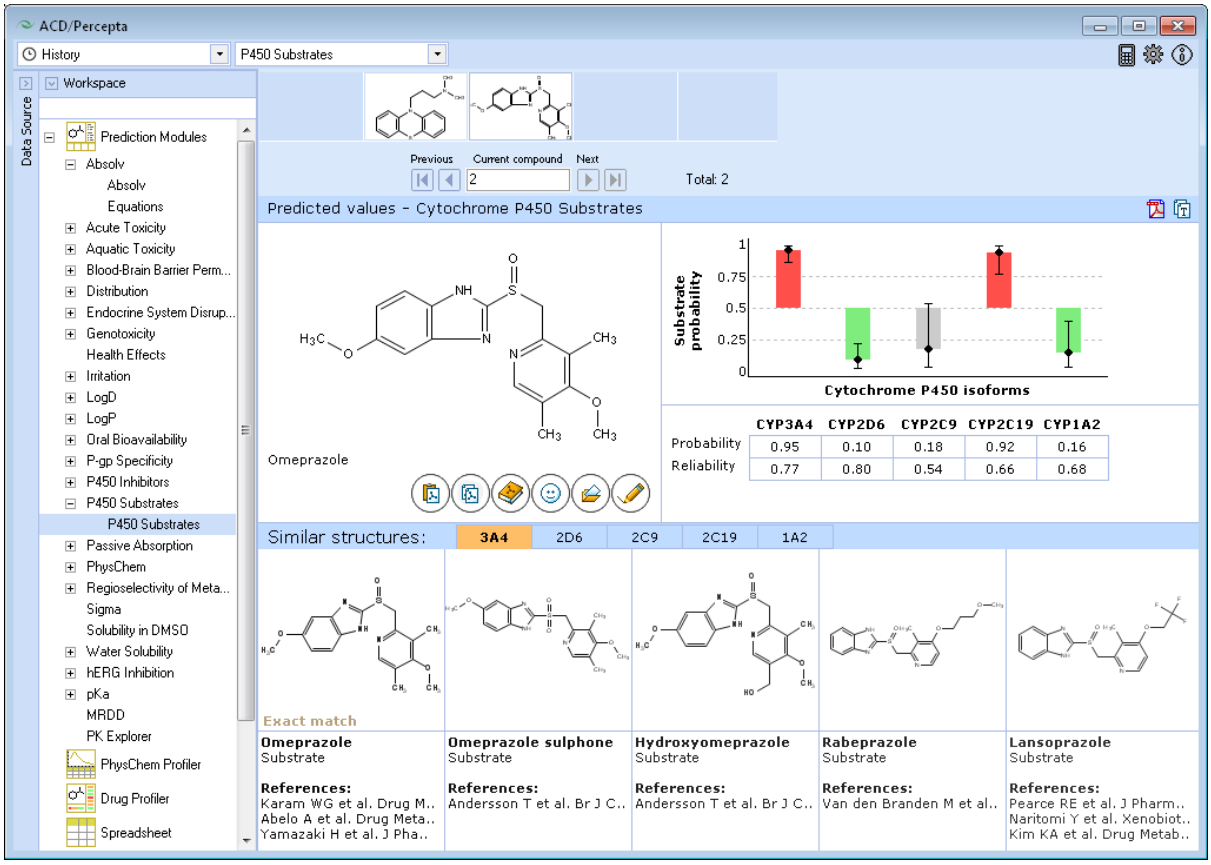

Figure 2. *Screenshot of ACD/Percepta P450 substrates module*

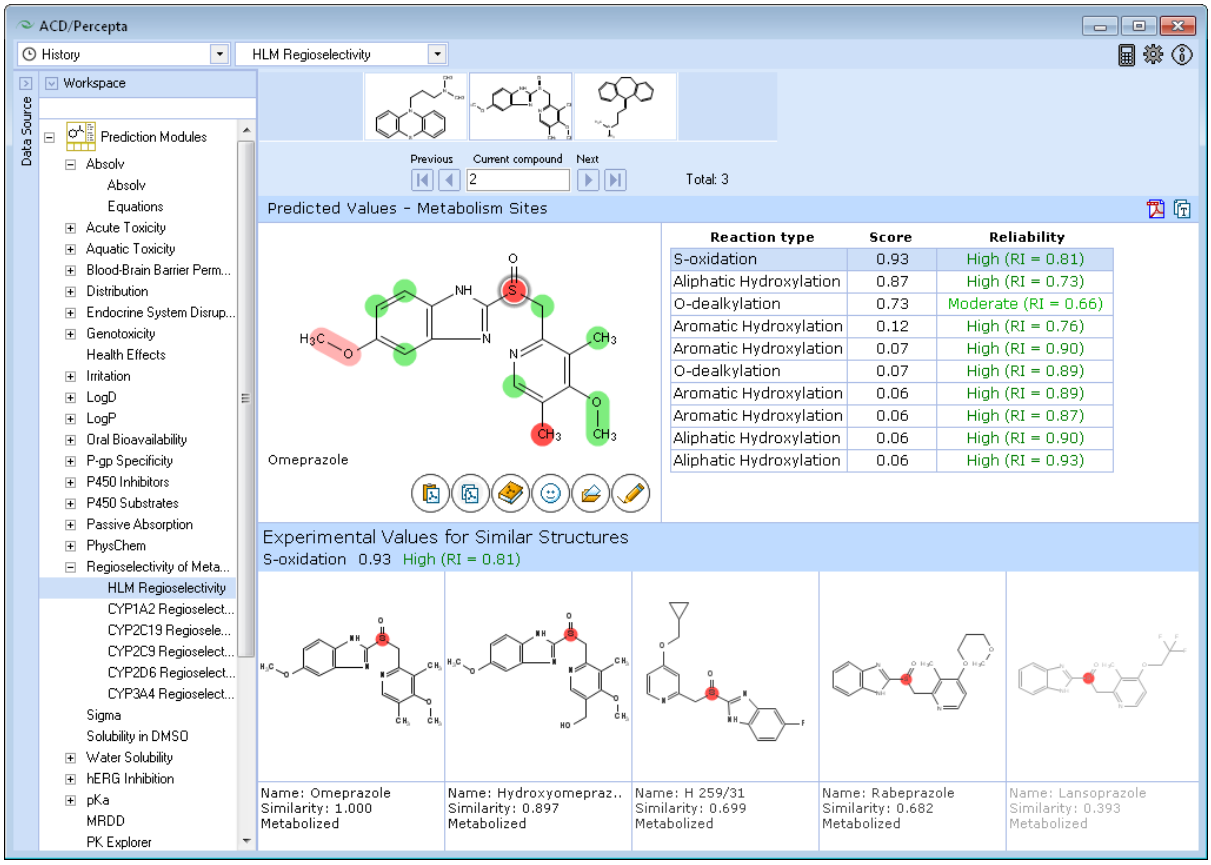

Figure 3. *Screenshot of ACD/Percepta Regioselectivity of Metabolism module*