DEVELOPING ADVANCED RULES AND TOOLS TO IMPROVE *IN VIVO* QSAR MODELS

by

MARLENE T. KIM

A dissertation submitted to the

Graduate School-Camden

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computational and Integrative Biology

Written under the direction of

Hao Zhu

And approved by

_____

Hao Zhu

_____

Menghang Xia

_____

Jongmin Nam

_____

Desmond Lun

Camden, New Jersey

May 2016

ABSTRACT OF THE DISSERTATION

Developing Advanced Rules and Tools to Improve *In Vivo* QSAR Models

by MARLENE T. KIM

Dissertation Director:
Hao Zhu

Together oral bioavailability and hepatotoxicity determine the fate and failure of a new drug in clinical trials. A promising drug candidate that has little to no oral bioavailability could be considered an ineffective treatment. If the drug does have high oral bioavailability, yet exhibits severe hepatotoxicity, the drug will be withdrawn from clinical trials. Thus, oral bioavailability and hepatotoxicity account for a substantial number of drugs eliminated from therapies and withdrawn from the market. Not only is this a severe financial loss for a pharmaceutical company; it is also a loss for patients who will no longer be able to benefit from the therapeutic effects. The most common approach to testing oral bioavailability and hepatotoxicity before clinical trials is through animal testing. The information learned from animal testing is highly valuable, but is expensive, time consuming, and has low throughput. Cell based assays developed to study specific biochemical mechanisms and toxicity are heavily used as an alternative; however, correlations between complex *in vitro* and *in vivo* endpoints are not very clear. Another alternative is Quantitative Structure Activity Relationship (QSAR) approach. A QSAR model could evaluate millions of chemicals without requiring them to be synthesized, which saves money, time, and has high throughput. Many predictive QSAR models have been developed since the 1960's, yet predictive *in vivo* QSAR models are

difficult to build and rare to find. Developing new methods to integrate mechanistic information and improve *in vitro-in vivo* correlation(s) for QSAR modeling purposes have been shown to be beneficial and are the focus of this thesis. The complex *in vivo* endpoints oral bioavailability and hepatotoxicity were used as example endpoints to model. Methods to curate biological information and *in vitro* data, specifically high-throughput screening data, and incorporate it into the QSAR models are discussed in great detail. The performance of the resulting models, as well as their shortcomings is also discussed. Overall, incorporating biological information into the QSAR modeling workflow greatly improved predictions from both the oral bioavailability and hepatotoxicity models. The new techniques can be adapted to model other complex biological endpoints.

ACKNOWLEDGEMENTS

To my advisor and friend, Hao Zhu, thank you for everything. I have learned so much over the years about research and life. Without you, I would not be where I am today.

To my committee members, Jongmin Nam and Menghang Xia, thank you for your generous support and guidance.

To my friend and mentor, Alexander Sedykh, thank you so much to for all your help, inspiration, and guidance. The magnitude of your influence is greater than you can imagine. I cannot thank you enough.

To the CCIB systems administrator, Kevin Abbey, without you none of this would be possible. All your help was greatly appreciated. Thank you.

To my lab members and life-long friends, Abena Boison, Dan Russo, Dan Pinolini, and Wenyi Wang, thank you for making graduate school an enjoyable experience. Your expertise was greatly valued and always improved my work. Together, we accomplished great projects that would not have been realized had we did it ourselves. I will miss spending time with you all.

DEDICATION

To my dad, Paul Kim, you are my hero. You have encouraged me to pursue my passion when know else believed I could. To my mom, Anh Kim, your resilience taught me to never give up and give my all; there is always a solution, no excuses. To my family, our family is huge. Therefore, I cannot acknowledge you all, but thank you for your unconditional love. You all have taught me the importance of sharing, giving, and cooperating; especially when I did not want to. These skills are essential, and I attribute them to my success in graduate school.

To my dogs Lilo and Stitch, you are my inspiration to work. Each day, my love for the both of you grows and increases my passion to advocate for alternatives to animal testing.

To my best friends, Hetal Vasavada and Sojaita "Jenny" Mears, I would not be here without either of you, literally. Hetal, you are the nicest person in the world. Thank you for being my conscience and showing me there is a huge, loving world outside of Camden. By the way, thank you for nagging me to go to grad school. Jenny, you are my "big sister." You devote your free time to community members and bear the burden of others, when you do not have to. Your example encourages me to do the same for others, especially when I am exhausted.

To my high school science teachers, Mr. Grous, Ms. Cook, and Mr. Snyder, thank you for making me fall in love with chemistry. You all said a career in the sciences is not easy, the road will be tough, and despite that, everything will turn out fine. And it did! Mr. Snyder, thank you for teaching me that, "Perfect practice makes perfect."

TABLE OF CONTENTS

Chapter 1. Introduction

Animal toxicity testing is utilized in the pharmaceutical industry, academia and regulatory agencies to evaluate the potential risks and safety of chemical substances. The information learned from animal testing is highly valuable, but comes at a high cost, is time consuming, and has low throughput. Sometimes the data from animal testing could not predict compound toxicity in human due to the species difference. Cell based assays developed to study specific cellular and biochemical mechanisms are commonly used as an alternative; however, correlations between *in vitro* and complex *in vivo* endpoints, such as hepatotoxicity and bioavailability, are not very clear (Kim et al. 2014, 2015). Another alternative is Quantitative Structure Activity Relationship (QSAR) approach, which explores the relationships between the biological activity of a chemical and its sub-structural features and/or physicochemical properties. The unique advantage of using a QSAR model, unlike experimental models which require all chemicals to be evaluated to also be synthesized, a chemical could be evaluated without being synthesized. QSAR saves time, money, and resources. Thus many predictive QSAR models have been developed since its inception in the 1960's; however, not many scientists have been able to develop predictive *in vivo* QSAR models (Cherkasov et al. 2014; Ekins 2014).

In the 1960's, Corwin Hansch evaluated two experimentally based physicochemical properties such as Hammett constant and water partition coefficient and found that these parameters were highly correlated with biological activity in plants (Hansch et al. 1963). His studies, as well many other early QSAR studies, analyzed how the change of functional groups and atoms affect the physicochemical properties and biological activity of a chemical (Fujita et al. 1964; Klopman 1984). Hansch later

discovered that these parameters had general applicability in insects, rats, and bacteria as well (Hansch and Fujita 1964). These studies suggested that using experimental results as descriptors (quantitative variables based on chemical structure) in QSAR modeling is beneficial.

To stress the strong appeal of the traditional QSAR approach, it should be made clear that from a statistical viewpoint, QSAR is a special type of statistical data mining and data modeling approach. Adding the extra parameter, *i.e.,* quantitative chemical information in the form of descriptors to a statistical model transforms it into a QSAR model (Hansch 1969; Tropsha and Golbraikh 2007). The data is then reformatted to represent chemicals described by multiple descriptors, and the robust correlation between descriptors and a target property is sought. Whether the models are publically or commercially available, they all rely on comprehensively curated databases, and the underlying methodologies rely on traditional QSAR modeling workflows (Tropsha and Golbraikh 2007).

Today, most QSAR tools and platforms are relatively streamlined, and thus helpful in risk assessment research and the early stages of drug discovery. Due to these advances, QSAR has gained increased acceptance for use in regulatory agencies and medicinal chemistry (Cherkasov et al. 2014; Judson et al. 2012; Kruhlak et al. 2007; Worth 2010). QSAR models for *in vitro* endpoints such as mutagenicity, skin sensitization, and drug absorption show acceptable predictivity and are widely used in the pharmaceutical industry, academia and regulatory agencies (Chaudhry et al. 2010; Ekins 2014). QSAR models for complex endpoints such as oral bioavailability, hepatotoxicity, and carcinogenicity are also available, but have not been accepted by the general

scientific community as an alternative to animal testing (Ekins 2014; Kruhlak et al. 2012; Schilter et al. 2014). Creating sufficiently predictive models of *in vivo* endpoints has been shown to be most difficult, and new techniques to improve previously established approaches are urgently needed (Thomas et al. 2012; Zhu et al. 2009).

The major issue with modeling *in vivo* endpoints is that *in vivo* responses are the manifestation of many cellular and biochemical events. Since a tradition QSAR model can only predict one chemical or biological event at a time, it is understandable that a QSAR model cannot predict *in vivo* outcomes. One solution to this is to include mechanistic information in the form of *in vitro* data to improve the *in vivo* QSAR model (Zhu et al. 2008). Previous QSAR studies on *in vivo* endpoints have shown that relevant bioassay data can be viewed as additional biological information that characterizes a compound; and incorporating the data into QSAR models have been shown to be beneficial (Low et al. 2011; Sedykh et al. 2011; Zhu et al. 2009).

The remainder of this thesis provides details on how to curate biological information and *in vitro* data, specifically HTS data, and incorporate it into QSAR models. The complex *in vivo* endpoints human oral bioavailability and hepatotoxicity were used as example endpoints in this thesis. Ways to include mechanistic information to improve *in vitro-in vivo* correlations for modeling purposes were also explored to improve the QSAR models. The performance of the resulting models, as well as their shortcomings, is discussed in great detail. Overall, incorporating biological information into the QSAR modeling workflow greatly improved the model predictions, and the new techniques can be adapted to model other complex biological endpoints.

Chapter 2. Curating and Preprocessing High Throughput Screening Data for Quantitative Structure Activity Relationship Modeling

Chapter Overview

Publically available bioassay data often contains errors. Curating the data, especially high throughput screening (HTS) data, for Quantitative Structure-Activity Relationship (QSAR) modeling will require the assistance of automated data curation tools. Using automated data curation tools are beneficial to users, especially ones without prior computer skills, because many platforms have been developed and optimized based on standardized requirements. As a result the users do not need to extensively configure the curation tool prior to the application procedure. In this chapter, a freely available automatic tool to curate and preprocess HTS data for QSAR modeling purposes will be described.

Introduction

A typical high throughput screening (HTS) data set can contain over 10,000 compounds. Although they are potential resources for developing Quantitative Structure-Activity Relationship (QSAR) models, normally these public HTS data sets cannot be used directly for modeling purposes due to the presence of duplicates, artefacts, and other issues. There are public chemical data repositories such as PubChem, ChemSpider, and ChEMBL that contain lots of HTS data available for download, but the original data stored in these resources still need further data curation and processing.

Chemical structure curation and standardization is an integral step in QSAR modeling. This step is essential since it is likely the same compound could be represented differently among different sources. For example, organic compounds could

be drawn with implicit or explicit hydrogens, in aromatized or Kekulé form, as well as in different tautomeric forms. These differences in chemical structure representations could influence the computed chemical descriptor values for the same compound and greatly affect the usefulness and quality of the resulting QSAR models. Furthermore, inorganic compounds and mixtures, which are not suitable for traditional QSAR modeling studies, need to be removed from the data set. More details on chemical structure curation can be found elsewhere (Fourches et al. 2010). The large size of HTS data sets makes it very inefficient, and usually ineffective, to curate all the chemical structures manually. The assistance of an automated curation tool is highly recommended.

Another issue with HTS data is that it is very common for it to have an unbalanced distribution of activities, where there are substantially more inactive than active compounds. This unbalanced distribution of activities (*i.e.,* low active ratio) could result in biased QSAR model predictions. Data sampling, an approach that selects and analyzes a subset of the overall data, can resolve this issue. The specific data sampling method that will be discussed in this chapter is down-sampling, since it is most relevant to HTS data preprocessing. Down-sampling is an approach that ignores most of the data points that are in the largest activity category. This will allow you to select a sample of the inactive compounds from the data set to balance the distribution of activities for modeling. Moreover, smaller data sets are easier to manage and, in most cases, more informative since it captures the most important elements of the data.

In this chapter, automatic data curation workflows to standardize/harmonize chemical structures and down-sample a HTS data set will be described. The approaches to construct the modeling and validations sets, including balancing the HTS activity *via*

down-sampling, were developed using Konstanz Information Miner (KNIME ver. 2.10.1) (www.knime.org) workflows. The workflows incorporate the two most common approaches for selecting a sample size: random and rational selection approaches. These approaches are based on basic statistical approaches (Daniel 2009) and will transform an original public HTS data set into a curated format suitable for QSAR model development and other relevant *in silico* modeling efforts. The quantitative high throughput screening (qHTS) antioxidant response element assay data set obtained from PubChem (PubChem AID 743219) will be used to illustrate this data curation process.

Methods

Materials

Automated procedures to curate chemical structures and down-sample the large data set will be described in this chapter. All of the workflows were developed and executed in the open source platform KNIME. The output files of the workflows are curated data sets with standardized structures that are ready to be processed by QSAR modeling tools. The workflows can be downloaded as a zip file at https://github.com/zhu-lab.

Procedure

1. Prepare an input file for the curation workflow

An input file should be a tab delimited multiple column *txt* file (*FileName.txt*) with a header to each column, where one column must contain the structure information as a SMILES code (Weininger 1988). The input file (a sample file was provided within the zip file) should have at least three columns: *ID*, *SMILES* and *activity*. If needed, other

useful features of compounds (*e.g.,* compound names) could also be included as extra columns.

2. Prepare the curation workflow

Install the KNIME software. It can be downloaded from www.knime.org. Download the curation workflows (https://github.com/zhu-lab) and extract the zip files into a computer directory.

3. Configure the workflow

In the *File* menu bar of KNIME, select *"Import KNIME workflow…"* to import the structure standardizer workflow into KNIME. Now in the pop-up window (Figure 2.1), click on *"Source: Select root directory,"* and find the computer directory that the zip file was extracted to. Select the destination directory, which will be where the output files will be saved to. In *"Workflows:"* select the *"Structure Standardizer"* workflow and click *"Finish"* (Figure 2.1).

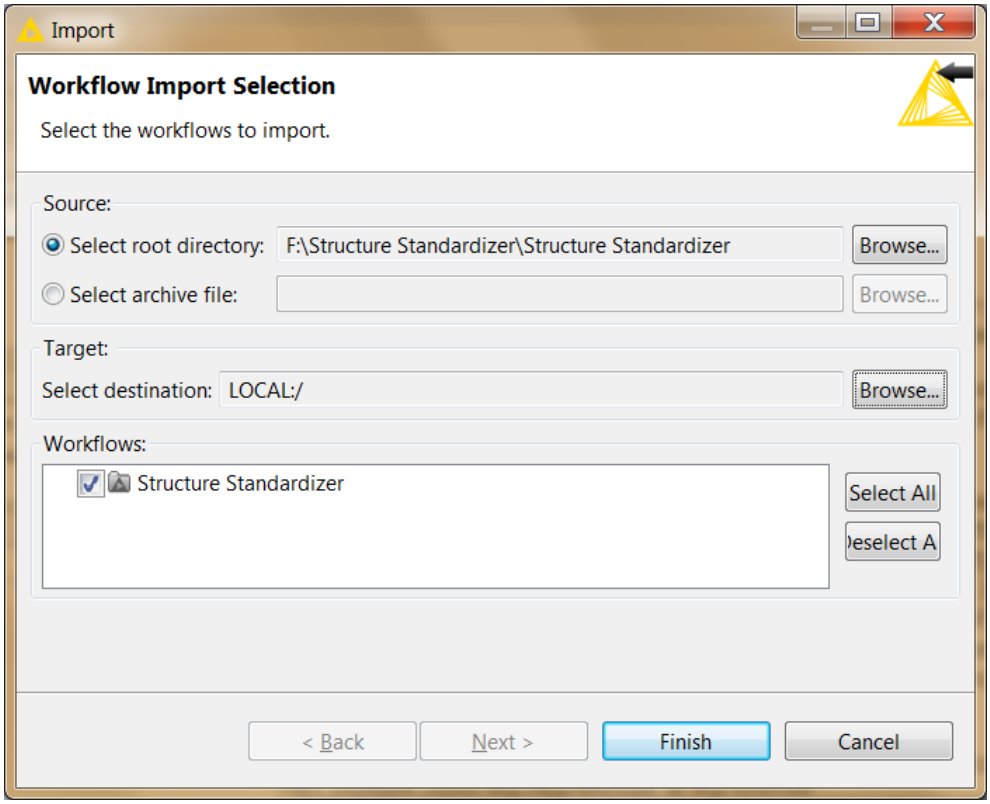


Figure 2.1. The KNIME "Workflow Import Selection" window

4. Set up parameters and run the workflow

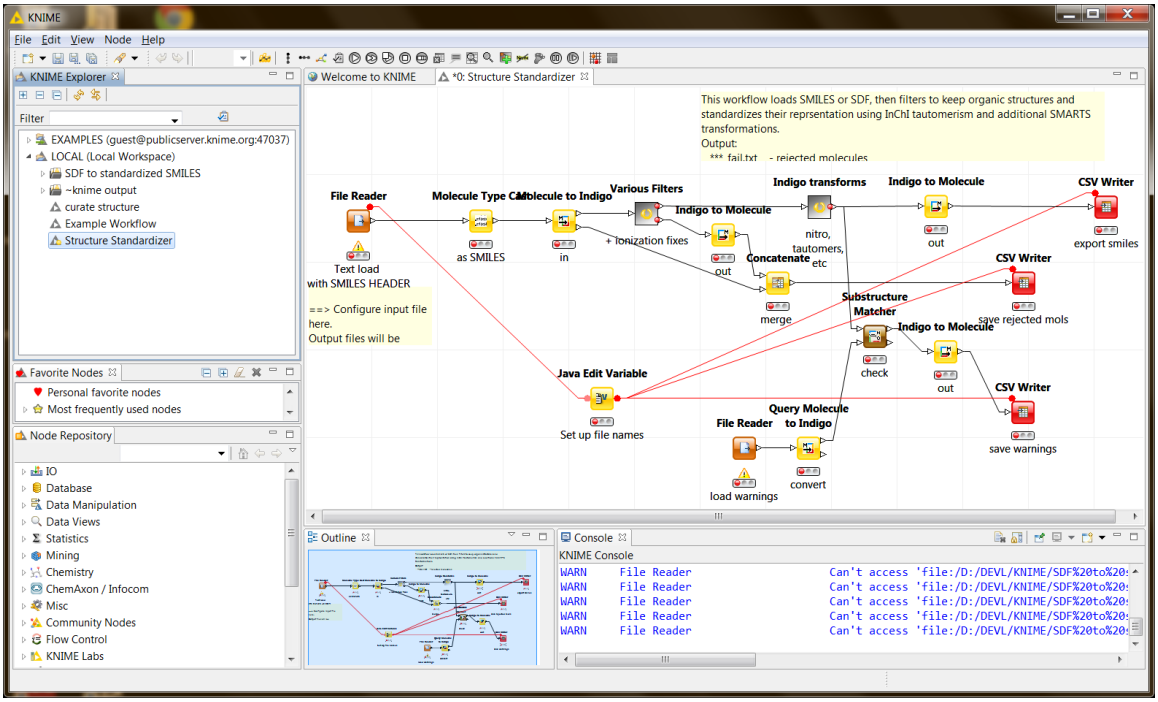


Figure 2.2. A window of the KNIME "Structure Standardizer" workflow

To open the workflow, double click on the Structure Standardizer in the *"KNIME Explorer"* window under *"LOCAL (Local Workspace)"* located in the top left side bar. At this time, the workflow will show up in the main space, which is called the workflow editor (Figure 2.2). If these windows are not present, go to *"View"* in the top menu bar and select *"Reset Perspective…"* Right click the "*File Reader"* node and select "*Configure*." In the pop-up windows, input the valid file location of the input file that has been prepared in the previous step. Make sure the headers of the input file are read correctly. Click *"OK"* to save the changes and close the configuration window. Next, right click the "*Java Edit Variable"* node in the bottom left and change the variable *v_dir* to the directory of the folder where all the files are extracted in the second step. Then, configure sub-workflows individually by double clicking on each node. Within each sub-

workflow, configure the "*Java Edit Variable"* node the same as described above. After closing the sub-workflow windows, the yellow lights on all the nodes should be on, indicating that the workflow is ready to be used. Click on the green *"double-arrow button"* located in the top menu bar to execute the whole workflow and the green lights on all nodes should be on.

Three output files should have been generated in the same folder as the input file (*FileName_fail.txt, FileName_std.txt,* and *FileName_warn.txt)*. (Or the files will be in a folder directory substituting all spaces with %20 if spaces are in the directory (*e.g.,* if input file is in F:\Structure Standardizer\output, then output file would be in F:\Structure%20Standardizer\output).

Description of the three output files:

1. *FileName_fail.txt* contains compounds that could neither be standardized nor be used in QSAR modeling (*e.g.,* mixtures, inorganics, large molecules, and polypeptides).
2. *FileName_std.txt* contains the remaining structurally standardized compounds in which the SMILES are curated as the canonical format.
3. *FileName_warn.txt* contains compounds with potential problems that require further review. For example, compounds with positive/negative charges need to be compared to their original structures to decide the correct structure information. These compounds with warnings will not be removed from the data set and are included in the *FileName_std.txt file*.

The standardized compounds will be in canonical SMILES format. Compounds in this file are curated, standardized, and represented in canonical form. Metals were removed, tautomers were de-isomerized, salts and charges were neutralized, and aromatic rings were de-aromatized. For more information please look into the commented .smk files. The file *FileName_std.txt* is the data set curated for modeling purposes.

5. Preparing the chemical descriptor file

With the chemical structures curated, the chemical descriptors can be calculated by using various descriptor generators, such as RDKit (http://www.rdkit.org/), Molecular Operating Environment® (MOE) (https://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm), Dragon® (http://www.talete.mi.it/products/dragon_description.htm), and etc.

The descriptor values of the whole data set need to be normalized between 0 and 1 before QSAR model development. If there are too many descriptors (e.g., the number of resulting Dragon descriptors is normally over 1,000), it is necessary to reduce the number of descriptors to reduce the computing time for model development. Performing a pairwise comparison between any two descriptor values is one way to find correlated and redundant descriptors. This can be done by constructing a scatter plot for every pair of descriptors and determining the Pearson's product-moment coefficient for every pair (Daniel 2009).

6. Preparing the modeling and validation set files

To develop a predictive QSAR model, the compound classifications in the modeling set need to be balanced. For example, the ARE data set contains 958 active and

4,526 inactive compounds (Figure 2.3). To this end, the inactive compounds of HTS data need to be down-sampled to be similar to the number of actives in the modeling set.

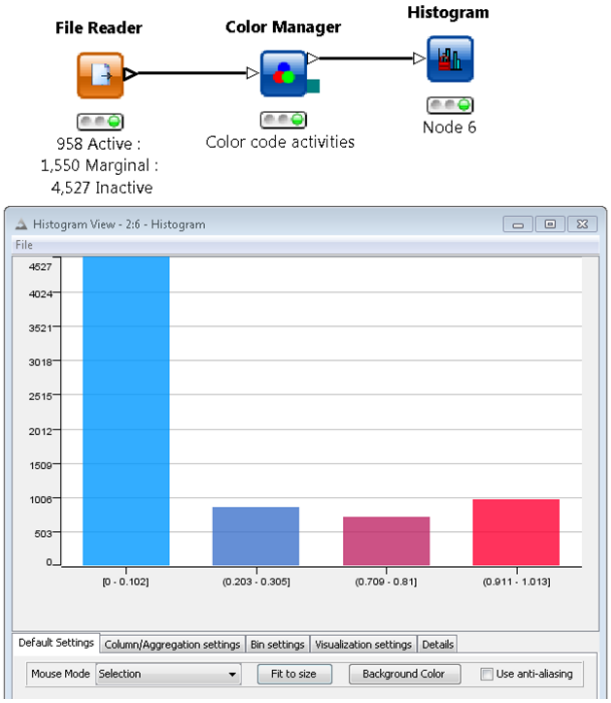


Figure 2.3. Example of KNIME histogram plot workflow and the resulting histogram plot showing the frequency of activity values 0 (inactive, blue), 0.25 to 0.50 (marginal, gray), and 1 (active, red)

There are two methods that can be applied for this purpose: random and rational selection. It has been reported that there is little difference in the QSAR model performance resulting from either random or rational selection approaches (Martin et al. 2012). The random selection approach will randomly select an equal number of inactive compounds compared to the actives. It ensures that the relationships between each compound selected for the model development and validation purposes were not explicitly selected. The rational selection approach uses a quantitatively defined threshold of similarity to select inactive to active compounds. Rational selection ensures

that the test set will have structurally similar analogs in the modeling set, but this cannot be guaranteed for external set compounds. However, the rational selection approach may be advantageous when the applicability domain of the QSAR model needs to be clearly defined (Golbraikh et al. 2003).

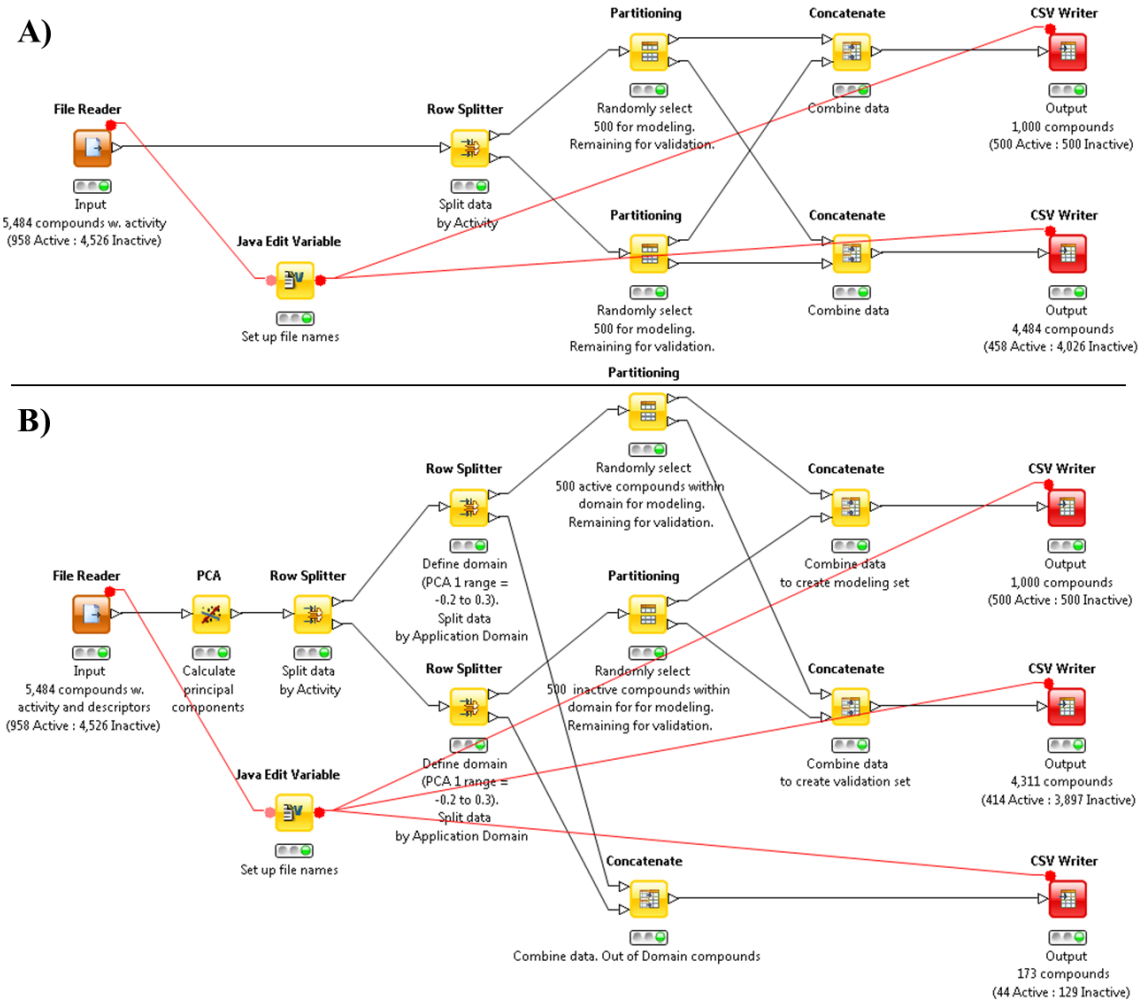


Figure 2.4. Example of KNINE workflow for selecting compounds and partitioning data set into modeling and validation sets using A) random and B) rational selection approaches.

Random Selection. Figure 2.4A shows a KNIME workflow that could be used to randomly select compounds and partition the data set into modeling and validation sets.

(KNIME also has an "Equal Size Sampling" node that automatically down-samples the data set and it can be substituted into the workflow. However, it does not partition the data set into modeling and validation sets.) To run the workflow, first input the curated file (*e.g.,* the file *FileName_std.txt*) from previous step with a minimum of two columns for the *ID* and *activity* in the "*File Reader"* node. Then right click on the *activity* column header to open the "*Column Properties"* and set the "*Type"* as "*String."* The workflow has already been configured to randomly select 500 active and 500 inactive compounds; however, the numbers of active/inactive compounds can be changed. Click on the green *"double-arrow button"* located in the top menu bar to execute the whole workflow. Two files will be generated in the destination directory: *ax_input_modeling.txt* and *ax_input_intValidating.txt*. The *ax_input_modeling.txt* file contains the 500 active and 500 inactive compounds randomly selected to balance the distribution of activities in the modeling set. The *ax_input_intValidating.txt* file contains the remaining compounds (*e.g.,* 458 active and 4,026 inactive compounds from the sample data set) that could be used for validation purposes.

Rational Selection. Figure 2.4B shows a KNIME workflow that could be used to rationally select compounds for QSAR model development, based on the threshold defined using principal component analysis (See Section 7), and partition the data set into modeling and validation sets. In this case, inactive compounds that share the same descriptor space of active compounds will be selected and successively defines the applicability domain in the resulting QSAR models (Tropsha and Golbraikh 2007). The KNIME workflow described here differs slightly from the random selection workflow described above in that it allows one to quantitatively define the similarity threshold

using PCA. To run the workflow, first input the curated file (*e.g., FileName_std.txt*) from the previous step with columns for the *ID*, *activity,* and *descriptors* into the "*File Reader"* node. Then right click on the *activity* column header to open the "*Column Properties"* and set the "*Type"* to "*String."* The workflow has already been configured to select 500 active and 500 inactive compounds and the numbers of active/inactive compounds can be changed. Click on the green *"double-arrow button"* located in the top menu bar to execute the whole workflow. Three files will be generated in the destination directory: *ax_input_ratl_modeling.txt, ax_input_ratl_intValidating.txt,* and *ax_input_ratl_outAD.txt*.

7. Verification: Visualizing the chemical space covered by the data set using principal components

Principal component analysis is a statistical method that reduces the dimensions of descriptors in a data set by finding groups of descriptor combinations (Izenman 2008). It also provides one of the most informative statistics about the data. The first principal component covers the largest amount of variances in the data set. Each consecutive principal component will cover another portion of the variances, but less than the previous one. Therefore, the combination of all principal components represents all the variances in the data set. And the total number of principal components is less than the number of descriptors. All these calculation can be done in software like KNIME and Molecular Operating Environment.

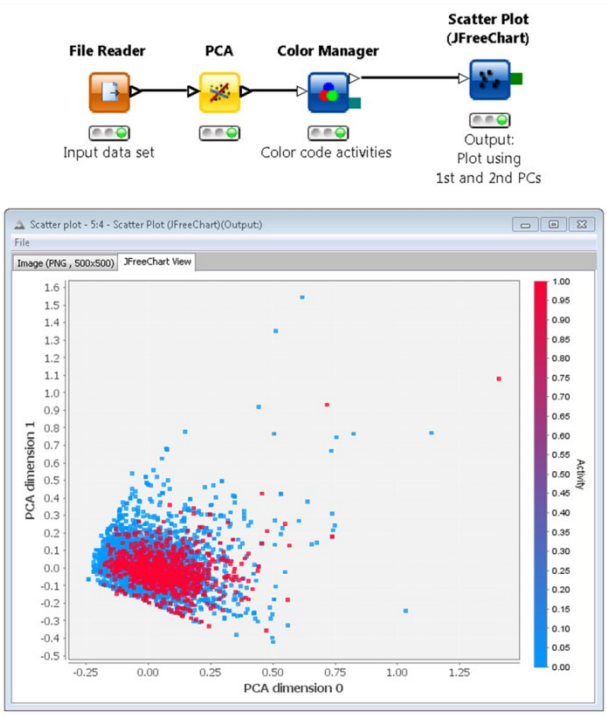


Figure 2.5. Example of KNIME workflow for visualizing the chemical space of all active and inactive compounds

Typically the first 3 principal components can be used to analyze the diversity of the chemical space and the overall relationships in the model. For example, in the sample descriptor file there are 10 descriptors calculated for the whole data set. A principal component analysis was performed to generate 6 principal components. Principal components 1 and 2 are plotted in a scatter plot to show the chemical space. Figure 2.5 shows the KNIME node that can be used generate the principal components and scatter plot of principal components 1 and 2 using all active and inactive compounds (n=5,484). Similar compounds will be clustered together and dissimilar compounds will be dispersed. In this case, the modeling set shows that the active and inactive compounds share the same chemical space. If active and inactive compounds occupy different spaces in the scatter plot, QSAR models will not be able to be developed.

A principal component analysis was performed in KNIME on all the active and inactive compounds in the ARE data set of 5,484 compounds (Figure 2.5). From the scatter plot of principal components 1 versus 2, it was noticeable that most of the compounds clustered at principal component 1 values between −0.2 and 0.3. Therefore, the applicability domain of the resulting model can be defined as any compound that falls within this range. To adjust this applicability domain in the KNIME workflow described in Section 3.6 and depicted in Figure 2.4B, adjust both *"Row Splitter"* nodes by right clicking the node, under *"use range checking,"* adjust the *"lower bound"* and *"upper bound."* Under this condition 500 active and 500 inactive compounds within the range of −0.2 and 0.3, will be selected for the modeling set, while the others will be placed into the validation set. Compounds that were out of domain will be placed into the *ax_input_ratl_outAD.txt* file.

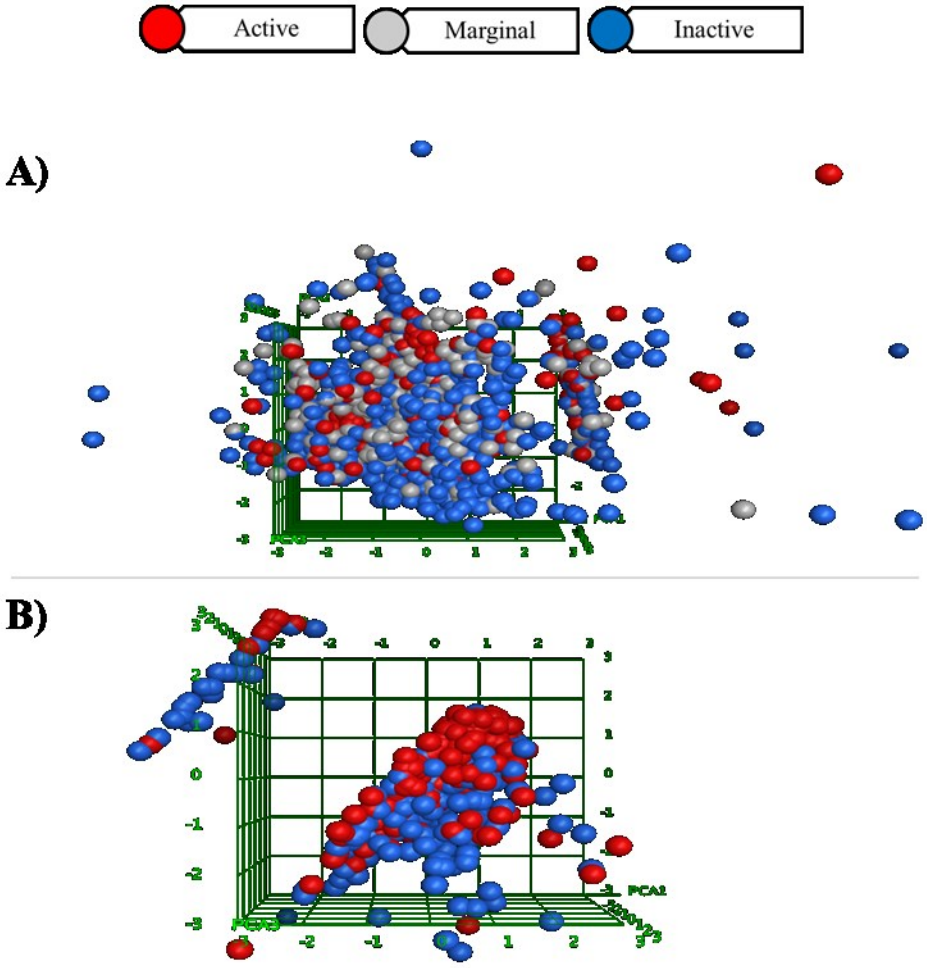


Figure 2.6. 3-D plots of ARE data set using A) all 7,034 data points, and B) modeling set using principal components 1-3 generated using 10 MOE descriptors

After the modeling and validation sets are created, the chemical space can be visualized. The chemical space defines the applicability domain of resulting QSAR models. The chemical space of a data set can be shown in a 3-D plot using the first three principal components (of the descriptor space) generated from MOE chemical descriptors. The MOE descriptors used in this study were *FCharge, PC+, PC-, TPSA, Weight, a_acc, a_don, density, logP(o/w),* and *logS*. In Figure 2.6A, the chemical space using the first three principal components of the entire ARE data set 7,034 compounds

was plotted. Then, 500 active and 500 inactive compounds organized the chemical space for the modeling set, as shown in Figure 2.6B. The inactive compounds were selected based on the similarity to the actives, so the chemical space occupied by the modeling set is clearly different from to the whole data set. Therefore, the predictions of resulting QSAR models should be considered reliable within the chemical space (*i.e.,* the applicability domain) of modeling set.

Summary

Publically available HTS data contains chemical structure errors and unbalanced activity distributions that need to be addressed before the data can be modeled. Due to its size, curating the data for QSAR modeling purpose requires automated computational tools. Furthermore, the activity distribution in HTS data are usually heavily skewed towards inactive compounds, which leads to biased predictions. To avoid biased predictions in the resulting QSAR models, the number of inactive and active compounds selected for modeling needs to be balanced. Down-sampling using either random or rational selection approaches mitigates this issue and results in a sample data set suitable for QSAR modeling. The technology described in this chapter enables one to use automated approaches to curate and prepare the public HTS data for modeling purposes.

Chapter 3. Critical Evaluation of Human Oral Bioavailability for Pharmaceutical Drugs by Using Various Cheminformatics Approaches

Chapter Overview

Oral bioavailability (%F) is a key factor that determines the fate of a new drug in clinical trials. To date, traditional Quantitative Structure-Activity Relationship (QSAR) models have not been able to successfully predict %F. Since %F is complex and relies on drug interactions with biological molecules (*e.g.*, Cytochrome P450 metabolism, liver enzymes), QSAR alone cannot be used to accurately predict %F. Biological information is needed. I hypothesized that incorporating Human Intestinal Transporters (HIT), known for pumping drugs in and out of cells, data could improve the QSAR predictions. For example, if a drug is highly likely to interact with a HIT that removes a drug out of blood circulation, then it is highly likely to have a low %F. Thus, a %F dataset of 995 drugs from public sources was compiled to develop a traditional combinatorial QSAR model, which was validated using five-fold cross-validation. Then, the QSAR models of human %F were integrated with HIT data. To validate my hypothesis, A Bootstrap method (non-parametric permutation N=10,000, α=95%) was used to compare the %F predicted by traditional combinatorial QSAR models versus QSAR models integrated with HIT data. The external predictivity of %F using the traditional combinatorial QSAR model was poor (R^2=0.28, n=995, MAE=24), but was improved (R^2=0.40, n=362, MAE=21) by filtering unreliable predictions that had a high probability of interacting with MDR1 and MRP2 transporters. The best model had a Pearson *p*-value of 0.01, indicating statistically significant improvement. To my knowledge, I created the first computational %F model that attempted to simulate drug interactions using HIT data. In this study, I developed

predictive %F QSAR models that could be used to evaluate new drug compounds, and integrating drug-transporter interactions data greatly benefited the resulting models.

Introduction

Drug oral bioavailability is the fractional extent of the drug dosage that finally reaches the therapeutic site of action and is quantitatively symbolized as %F (Buxton and Benet 2011). In many cases, most of the orally administered drug is metabolized and eliminated before reaching systemic blood circulation (Buxton and Benet 2011). Therefore, poor bioavailability may cause a new drug to fail clinical trials, even if it has high efficacy in previous *in vitro* and/or *in vivo* tests. The traditional process for measuring the %F of a drug is expensive, costly, and time-consuming. Using computational methods as an alternative to calculating the %F of new drug candidates, even before synthesizing the compound, would be advantageous by saving resources and provides a promising alternative to traditional experimental protocols.

Table 3.1. Brief description of previous QSAR oral bioavailability models

Source	Description	Performance	Train/Test Set Sizes
(Moda et al. 2007)	Hologram QSAR, CNT, modeling %F	q^2=0.35-0.70/R_{ext}^2=0.85	250/52 (mostly highly bioavailable drugs)
(Tian et al. 2011)	Combinatorial QSAR, CNT, modeling %F	R^2_{ext}=0.50 (after removing outliers)	916/80(?)
(Andrews et al. 2000)	Stepwise Regression; CNT: modeling %F	R^2_{ext}=0.58	473/118
(Ma et al. 2008)	Combinatorial QSAR, CTG, modeling: positive (%F≥20), negative (%F<20)	CCR_{Train} (5-fold CV) =62%/ CCR_{Test}= 59-71%	690/76 (mostly highly bioavailable drugs)

q^2 - Cross validated correlation coefficient; R^2 - Coefficient of determination;

To date there are many computational oral bioavailability models that are available (Andrews et al. 2000; Hou et al. 2007; Lipinski et al. 2001; Ma et al. 2008; Martin 2005; Moda et al. 2007; Paixão et al. 2012; Tian et al. 2011; Varma et al. 2010; Veber et al. 2002). Some are based on Quantitative Structure-Activity Relationship (QSAR) models that predict the oral bioavailability of new compounds directly from the molecular structure. Table 3.1 lists several major QSAR studies on oral bioavailability. In 2000, Andrews *et al.* developed a computational oral bioavailability model using linear regression. This model was able to predict highly bioavailable compounds accurately, but had poor performance for low bioavailable compounds (Andrews et al. 2000). Moda *et al.* developed hologram QSAR oral bioavailability models that predicted %F using fragment descriptors. However, poorly soluble and non-oral bioavailable drugs were excluded intentionally from the modeling set (Moda et al. 2007). Ma *et al.* used a Combinatorial QSAR (Combi-QSAR) approach to develop an oral bioavailability

classification model. Although the unbalanced accuracy for a five-fold cross-validation of their modeling set was 80%, the specificity (correct predictive rate for inactive compounds) was only 20% due to the high imbalance between actives and inactives (Ma et al. 2008). More recently, Tian *et al.* attempted to create multiple linear regression human oral bioavailability models by combining molecular properties and structural fingerprints with genetic function approximation. The predictivity of the reported model was acceptable (R^2_{ext}=0.50), but the structural fingerprints used to generate the training set does not apply to all drug classes. This limits the applicability of the model for predicting new classes of compounds (Tian et al. 2011).

Table 3.2. List of rules from previous studies on drug oral bioavailability and absorption

Ref.	Rules	Database
(Lipinski et al. 2001)	The 'rule of 5' can be used to estimate drug permeation.	>5,000
(Veber et al. 2002)	Humans - Reducing the overall flexibility and polar surface area of a drug could increase %F. But high molecular rigidity does not guarantee high %F.	277
	Rats - A drug with <10 rotatable bonds and polar surface area equal ≤140 $Å^2$ is likely to be highly bioavailable.	3061
(Hou et al. 2007)	Simple molecular property-based rules are more effective at predicting oral absorption than oral bioavailability.	768
(Martin 2005)	The bioavailability score (probability drug will have a %F>10) from rat studies can be used to identify poorly and highly absorbed compounds for humans.	533
(Varma et al. 2010)	Bioavailability is predominantly limited by drug absorption. Drug absorption is influenced by all the parameters previously discussed. Gut-wall and hepatic elimination are highly influenced by lipophilicity.	309
(Paixão et al. 2012)	Models that incorporated *in vitro* data performed significantly better than pure *in silico* models.	164

In addition to the QSAR models mentioned above, previous research suggests that the rule-based models, such as the rule-of-five (Lipinski et al. 2001), are not sufficient enough for evaluating the oral bioavailability of drugs (Hou et al. 2007; Martin 2005; Veber et al. 2002). Nevertheless such empirical rules are useful for qualitative assessment and are listed in Table 3.2 several *rules* previously developed for assessing drug oral bioavailability and absorption. In 2002, Veber *et al.* studied the molecular properties and *in vivo/in vitro* pharmacokinetic parameters that affect oral bioavailability (Veber et al. 2002). The authors concluded that the molecular properties of the drug, target receptor, cell membrane, and transporter proteins should all be studied during drug development.

Ignoring one factor can result in poor bioavailability (Veber et al. 2002). More recently, property-based rules for bioavailability (Tian et al. 2011) and parameters needed for optimal oral bioavailability classification (Varma et al. 2010) were evaluated. There are certain physical properties that contribute to oral bioavailability, but these parameters are better at predicting intestinal absorption (Tian et al. 2011; Varma et al. 2010; Veber et al. 2002). Recently, Paixão used *in vitro* test results as parameters to develop an oral bioavailability model (Paixão et al. 2012). Incorporating *in vitro* data helped improve the prediction accuracy of the resulting models.

In this study, several novel models of human oral bioavailability of pharmaceutical drugs were developed. After compiling over one thousand drugs and their experimental %F values, the data entry errors were corrected using both automatic tools and manual curation steps. The Combi-QSAR approach was utilized to develop several computational oral bioavailability models. A series of individual category (CTG) and continuous (CNT) models were developed and validated using a five-fold cross-validation. Human Intestinal Transporter (HIT) interactions were integrated into the final predictions to improve the predictivity of the resulting QSAR models. This hybrid approach was able to exclude compounds with considerable prediction errors from the final predictions. Our predictive Combi-QSAR oral bioavailability models can be used to assess and evaluate new drug candidates. Furthermore, similar approaches could be developed and utilized to model other complex biological activities for drug and drug like molecules.

Methods

Human Oral Bioavailability Dataset

The human oral bioavailability dataset was compiled from various public and private sources (Benet et al. 2011; Hou et al. 2007, 2009; Moda et al. 2007; MultiCASE; Thummel et al. 2011; Tian et al. 2011; Zhu et al. 2011). Originally it contained over 1,300 entries. Several tools (CASE Ultra, Chem Axon Standardizer, Chem Axon Structure Checker) were used for chemical structure curation and standardization. For duplicate entries, one was removed. For stereoisomers, the structure of the compound with the highest activity was kept. For salts, the chemical structure was neutralized. Mixtures were separated and the largest component was kept. All metals, metaloorganics, and inorganic entries were removed.

The experimental %F values in the dataset were carefully evaluated. It was common to find different %F values for the same compound among different sources. The %F values reported in, *Goodman & Gilman's The Pharmacological Basis of Therapeutics,* were selected over the %F values reported from other sources, because the bioavailability data in this book was curated and harmonized by experienced medicinal chemists (Thummel et al. 2011). In other cases, the values were harmonized if the range of the %F values were less than 10 for the same compound. If the %F value for the salt and neutral forms were different, the %F value for the neutral form was kept. For compounds with disparate %F values, the experimental studies that reported the values were carefully evaluated. After comparing sources, the %F from the study that clearly defined the method for determining the %F value was selected. A total of 995 unique compounds remained for the following modeling process after the curation.

After harmonizing the %F values, the compounds were classified as low bioavailable (%F<50, n=454) and high bioavailable (%F≥50, n=541). There is no universal criterion to define high and/or low bioavailable compounds. The %F=50% was used as an arbitrary classification threshold in this study since it could also balance the ratio of two classifications in the dataset. Non-oral drugs (%F=0), e.g., compounds commonly administered by intramuscular or intravenous injection, were included in the low bioavailable group. Also, using sigmoid function, the %F values were transformed to logK(%F), a pseudo-equilibrium constant, as it has more balanced distribution of values and could afford improved models.

$$logK(\%F) = log\left(\frac{\%F}{100-\%F}\right) \quad (1)$$

The distribution of all 995 compounds based on the %F values is displayed in Figure 3.1. Table A1 lists all 995 compounds, the oral bioavailability values, and the corresponding references.

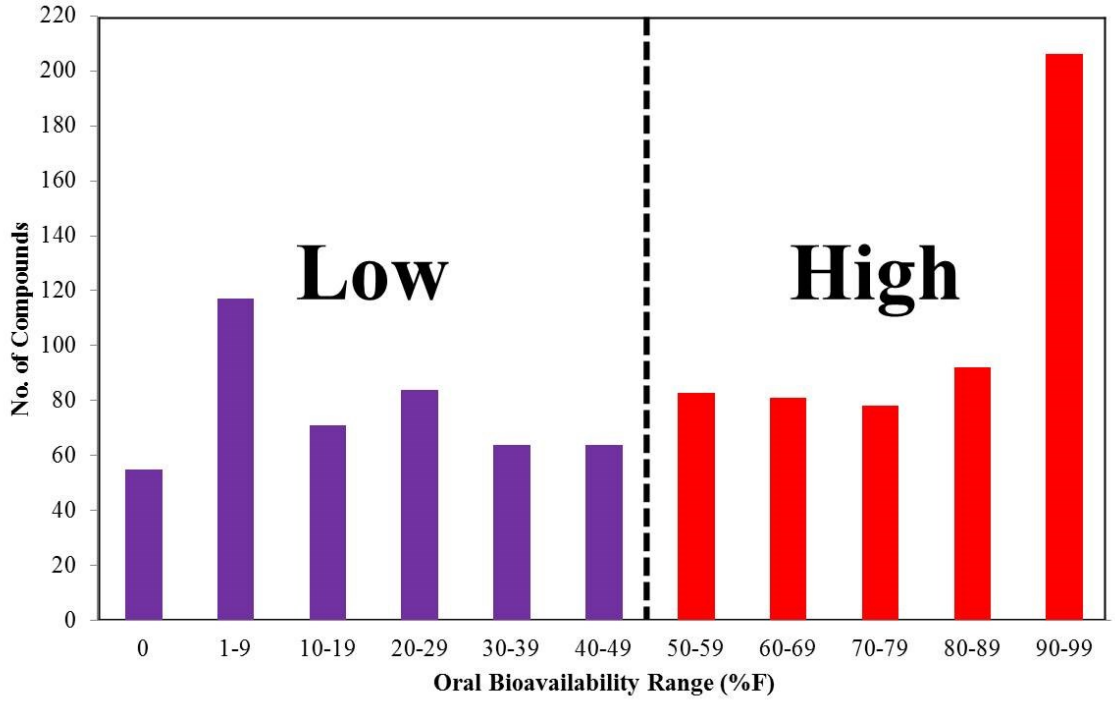


Figure 3.1. Distribution of compounds by various %F ranges

Chemical Descriptors

Chemical descriptors for each compound were generated using 2-D chemical descriptors from Dragon ver. 6.0 (Talete SRL, Milano, Italy) and Molecular Operating Environment (MOE) ver. 2011.10. Dragon descriptors included constitutional indices, ring descriptors, topological indices, walk and path counts, connectivity indices, matrix-based descriptors, autocorrelations, Burden eigenvalues, edge adjacency indices, functional group counts, atom-centered fragments, atom-type, E-state indices, atom pairs, molecular properties, and drug-like indices. MOE descriptors included physical properties, structural keys, E-state indices, topological polar surface area, and topological indices. Initially, the Dragon and MOE software generated 3,753 and 186 descriptors, respectively. Since many Dragon descriptors in this dataset were redundant, the number

of Dragon descriptors was reduced by removing low variance (standard deviation <0.01 or missing values) and highly correlated (r >0.95) descriptors. The remaining 1,597 Dragon and 186 MOE descriptors were range-scaled to [0,1] and used in the modeling process except for CASE Ultra, which has its own built-in fragment descriptors.

Modeling Approaches

In this study, the implementation of the Random Forests (RF) (Breiman 2001) and Support Vector Machine (SVM) (Kovatcheva et al. 2005, 2004; Vapnik 2000; Votano et al. 2006) algorithms available in R.2.15.1 (Dalgaard 2008) were used. The *k* Nearest Neighbor (*k*NN) models (Zheng and Tropsha 2000) were built using Chembench (chembench.mml.unc.edu). Below is a brief description each algorithm.

Random Forests (RF)

The algorithm for inducing a RF was developed by Leo Breiman and Adele Cutler (Breiman 2001). A RF is a predictor that consists of many decision trees and outputs a prediction from combined outputs from individual trees. In the random forest modeling procedure, *n* samples are randomly drawn with replacement from the original data set of N compounds. These samples are used to construct *n* training sets and to build *n* trees. For each node of the tree, *m* variables are randomly chosen from the all the available chemical descriptors, and the best data split from the *m* variables is used. Each new variable is pushed down the tree, and the output of the predictions are averaged into a final value. For each compound, each tree "votes" on a classification and the forest selects the classification that has the most votes.

Support Vector Machine (SVM)

SVMs are comprised of supervised learning models that analyze data and recognize patterns for classification and regression analysis (Vapnik 2000) . An SVM model works by representing variables of the training set as points in *n*-dimensional space, and linearly separating the categories by a hyperplane. The regions bounded by the hyperplane are called the margin. The largest marginalization possible is sought and is the optimal hyperplane. Data points that cannot be separated linearly can be transformed using kernel functions. Data points that lie closest to the margins are the support vectors. The test set is then mapped in the same space as the training set, and predicted to belong to a category based on which side of the hyperplane they fall on.

***k*-Nearest Neighbor (*k*NN)**

The *k*-nearest neighbor (kNN) QSAR method (Zheng and Tropsha 2000) employs the *k*NN classification principle and the variable selection procedure. Briefly, a subset of *nvar* (number of selected variables) descriptors is selected randomly at the onset of the calculations. The *nvar* is set to different values and the training set models are developed with leave-one-out cross-validation, where each compound is eliminated from the training set and its activity is predicted as the average activity of *k* most similar molecules where the value of *k* is optimized as well (*k* = 1 to 5). The similarity is characterized by Euclidean distance between compounds in multidimensional descriptor space. A method of simulated annealing with the Metropolis-like acceptance criteria is used to optimize the selection of variables. The objective of this method is to obtain the best leave-one-out cross-validated CCR possible by optimizing the *nvar* and *k*.

CASE Ultra

CASE Ultra is a QSAR expert system and can automatically generate a predictive model from a training set of non-congeneric compounds with associated biological activity data. The training set usually contains examples of both active and inactive chemicals and the algorithm identifies positive and deactivating alerts (structural fragments statistically related to activity and inactivity) after processing them. These alerts form a CASE Ultra model that can be used to predict activity of a test chemical (Chakravarti et al. 2012; Saiakhov et al. 2013).

Combinatorial QSAR Modeling Workflow

The entire Combinatorial QSAR modeling workflow is shown in Figure 3.2. Individual models were developed using Dragon (denoted by the prefix "D") or MOE descriptors and either RF, SVM, or *k*NN modeling methods. CASE Ultra was used to develop a single CTG model. This resulted in seven different CTG, four different CNT-%F, and four different CNT-logK(%F) models. The individual CTG models were D-RF, D-SVM, D-*k*NN, MOE-RF, MOE-SVM, MOE-*k*NN, and CASE Ultra. The individual CNT-%F and CNT-logK(%F) models were D-RF, D-SVM, MOE-RF, and MOE-SVM. The results for each CTG model and CNT model were averaged to generate the corresponding consensus CTG, CNT-%F, and CNT-logK(%F) predictions, which will be further referred to as consensus models (Figure 3.2).

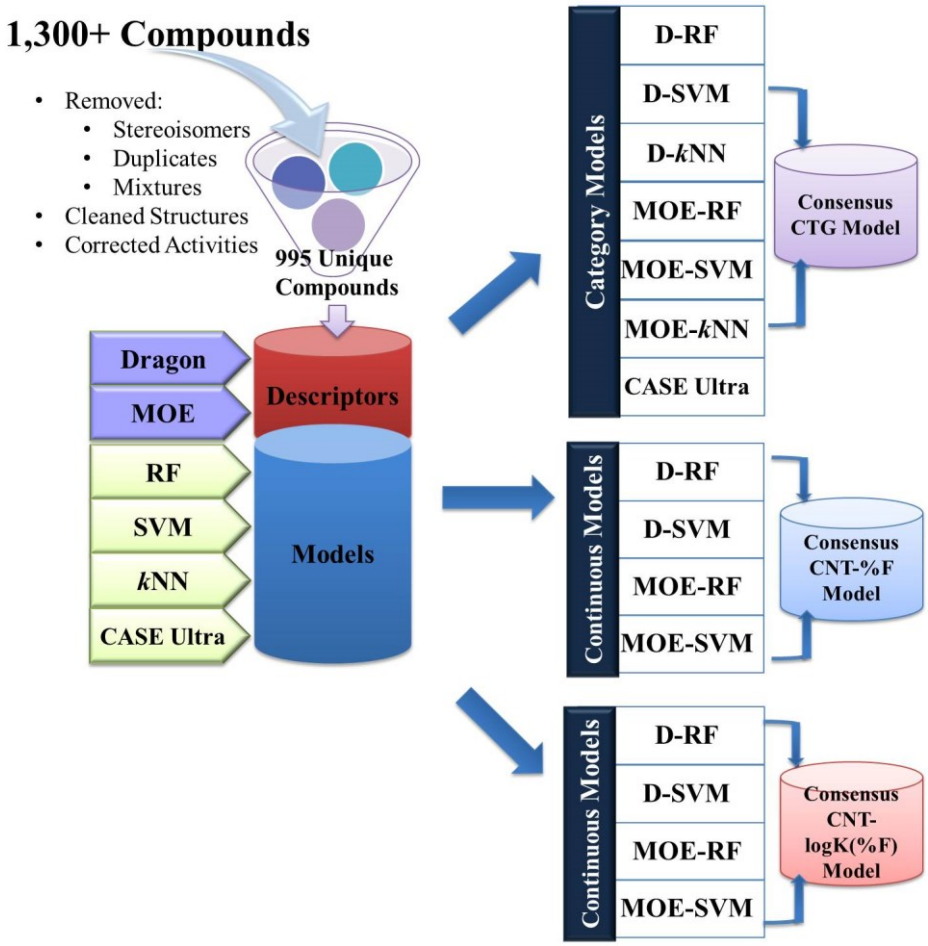


Figure 3.2. Combinatorial QSAR modeling workflow

All models were validated using five-fold external cross-validation. Briefly, the oral bioavailability dataset was randomly divided into five equal subsets. One subset was used as the validation set (20%) and the other four subsets (80%) were used as the training set. The training set was used to develop the models and the models were validated by the left-out validation set. The procedure was repeated five times so that each compound was in a validation set. Additional details about the modeling approaches can be found elsewhere (Golbraikh et al. 2003; Tropsha and Golbraikh 2007).

Universal Statistical Figures of Merit for All Models

Since various modeling approaches and different descriptors were used in the modeling process, universal statistical metrics were needed to evaluate the performance of the models developed individually. The results were harmonized by 1) using sensitivity (percentage of high oral bioavailable drugs predicted correctly), specificity (percentage of low oral bioavailable drugs predicted correctly), and CCR (correct classification rate or balanced accuracy) for CTG models; and 2) Pearson's multiple linear correlation coefficient (R^2) and mean absolute error (MAE) for CNT models. These parameters are defined as followed:

$$\%\ sensitivity = \left(\frac{true\ positives}{true\ positives + false\ negatives}\right) 100 \quad (2)$$

$$\%\ specificity = \left(\frac{true\ negatives}{true\ negatives + false\ positives}\right) 100 \quad (3)$$

$$\%\ CCR = \left(\frac{sensitivity + specificity}{2}\right) 100 \quad (4)$$

$$R^2 = \frac{regression\ sum\ of\ squares}{total\ sum\ of\ squares} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |predicted\ value_i - true\ value_i| \quad (6)$$

Integrating Human Intestinal Transporters Interactions of Compounds into Oral Bioavailability Predictions

Recently a QSAR study for predicting interactions for different HITs (Sedykh et al. 2013) was published. These HIT models were used to generate the transporter

interaction scores for the drugs in our oral bioavailability dataset. Interactions between molecules and HITs depend on the size, shape, charge, and the chemical properties of the molecule (Giacomini and Sugiyama 2011). Most of the compounds in our dataset have aromatic rings, bulky groups, and are ionizable. Compounds with these features are commonly removed from the enterocytes by the efflux transporters Multidrug Resistance Protein 1 (MDR1) and Multidrug Resistance-Associated Protein 2 (MRP2), which could decrease their oral bioavailability (Giacomini and Sugiyama 2011). Therefore, the interaction parameters of MDR1 and MRP2 were used to filter predictions of compounds from our models.

Results

Overview of Dataset

A comprehensive analysis was done on the chemical structures and relevant bioavailability data from the public databases used in this study. This comparison revealed that only 80% of the entries in current oral bioavailability databases are accurate. There were discrepancies between reports from different sources, affecting both molecular structures and %F values. For some compounds, the substituent groups were placed at incorrect positions. Figure 3.3 lists several examples of incorrect chemical structures that were identified from the original sources and corrected.

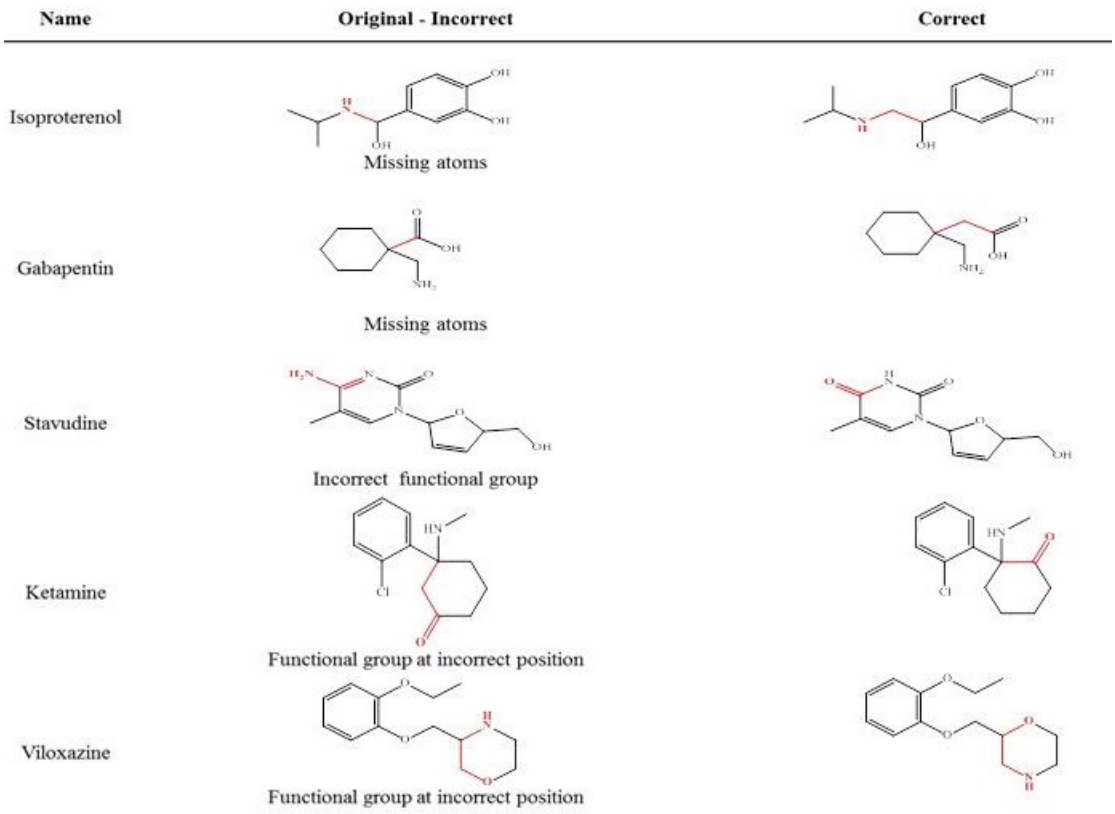


Figure 3.3. Examples of incorrect chemical structures from different sources

Furthermore, Buxton *et al.* indicated that it would be normal for different sources to report different %F values for the same compound (Buxton and Benet 2011). However, the compounds with disparate %F values needed to be harmonized for modeling purpose. Furthermore, the errors from the reported %F values occurred when a source incorrectly used the neutral names and salt forms of a molecule interchangeably. All of the errors were carefully examined and corrected.

The structural similarities between the compounds in the dataset can be analyzed by performing a principal component analysis on the chemical descriptors. After generating the principal components using the 186 MOE descriptors for all of the

compounds in the database, the top three most important components were selected to create a three-dimensional plot (Figure 3.4) for all 995 compounds. These three principal components capture around 50% of the variance in our database. This plot could be viewed as the chemical structure space covered by all the compounds in our oral bioavailability dataset.. According to this analysis, there are about 10 structural outliers that are dissimilar to the majority of the compounds. Most of these compounds represent non-bioavailable or low bioavailable drugs, including antibiotics, neuronal drugs, and intravenous drugs. Some previous studies showed that removing structural outliers before the modeling process was beneficial to the results of the QSAR models (Lipinski et al. 2001; Moda et al. 2007). In this study, these outliers were not removed since they only represent a small portion (~0.1%) of the whole dataset. Furthermore, removing the outliers did not improve the resulting models (data not shown).

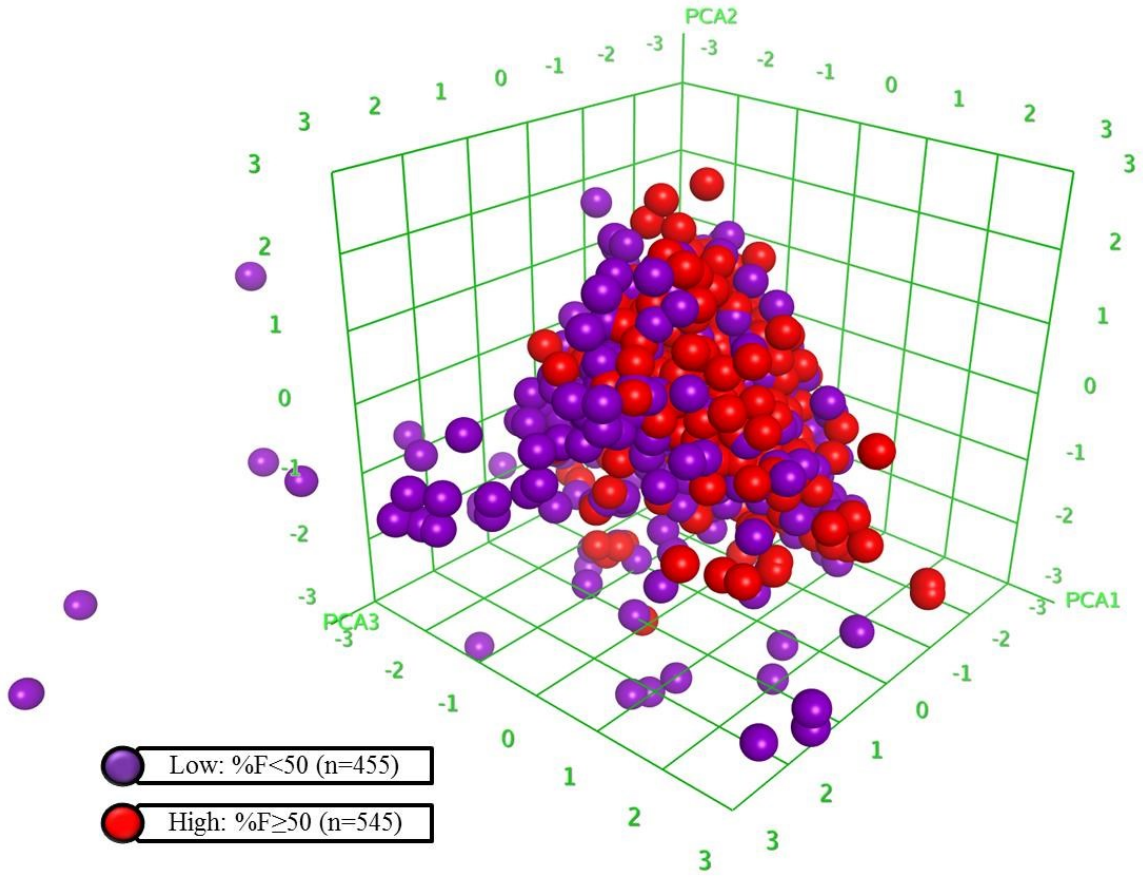


Figure 3.4. Chemical space of human %F database (n=955) using top 3 principal components of MOE descriptors

Category Models

Seven individual and one consensus model was developed by using two bioavailability categories ("low", %F<50% and "high", %F ≥ 50%; see Methods). The five-fold external cross-validation results for all CTG models are shown in Figure 3.5. The sensitivity, specificity, and CCR for the individual models ranged from 59-72%, 61-70%, and 62-70%, respectively. The D-SVM model had the lowest predictivity (CCR=62%). The MOE-RF model had the highest specificity and CCR of 70%. The MOE-*k*NN model had the highest sensitivity of 72%. Compared to the best individual model, the consensus model showed similar statistics, with sensitivity, specificity, and

CCR as 72%, 69%, and 70%, respectively. The model obtained from our commercial modeling software, CASE Ultra, had intermediate results with sensitivity, specificity, and CCR all as 65%.

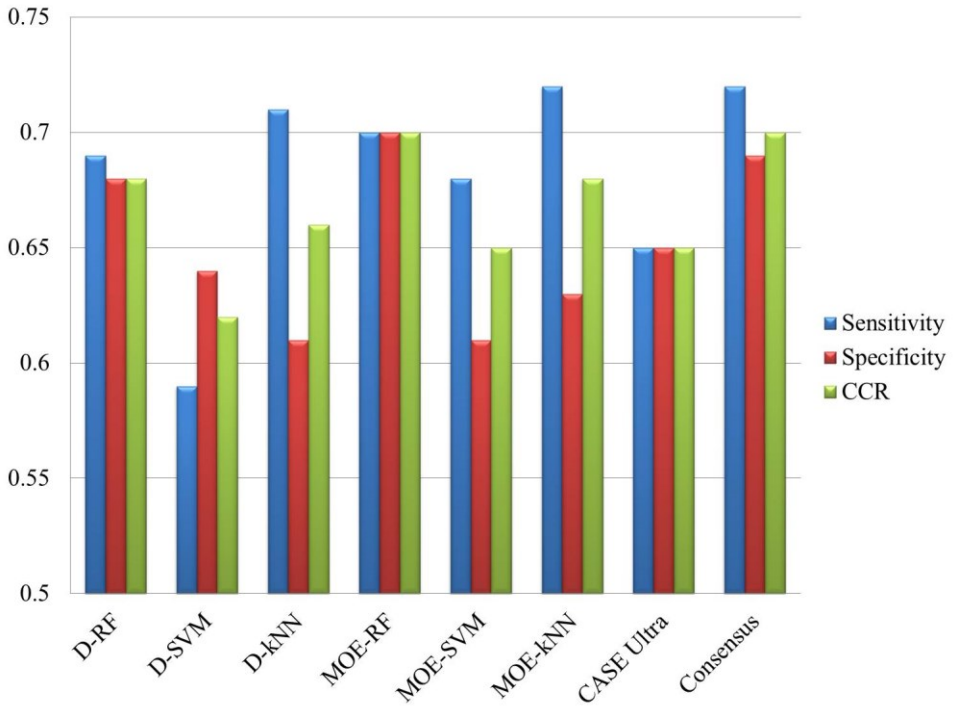


Figure 3.5. Performance of CTG QSAR models using five-fold cross-validation

Furthermore, Consensus Prediction Thresholds (CPT) were implemented (31), to the prediction results by using different low bioavailable and high bioavailable thresholds. The prediction results from each individual model had continuous scores that ranged from 0 to 1. The 0.5 mark was initially used as the single threshold to distinguish compounds predicted as low bioavailable (CPT<0.5) and high bioavailable (CPT≥0.5). Using stricter thresholds, the compounds that were predicted around 0.5 should be considered as "inconclusive." These inconclusive predictions were removed by using

different CPTs to define low bioavailable and high bioavailable compounds. Two CPTs were defined: 1) <0.4 as low bioavailable and >0.6 as high bioavailable (CPT-1 scheme); 2) <0.3 as low bioavailable and >0.7 as high bioavailable (CPT-2 scheme).

Implementing CPT-1 and CPT-2 schemes enhanced the predictivity of the individual and consensus CTG models. For the individual CTG models with CPT-1 and CPT-2, the sensitivity, specificity, and CCR ranges were between 61-87%, 50-82%, and 59-83%, respectively (results not shown here). In the consensus CTG model, the sensitivity, specificity, and CCR were 78%, 74%, 76%, respectively for CPT-1 and 82%, 77%, 79%, respectively for CPT-2 (Figure 3.6). As the tradeoff for excluding compounds with inconclusive predictions, using CPT-1 and CPT-2 decreased the consensus model coverage to 71% and 46%, respectively.

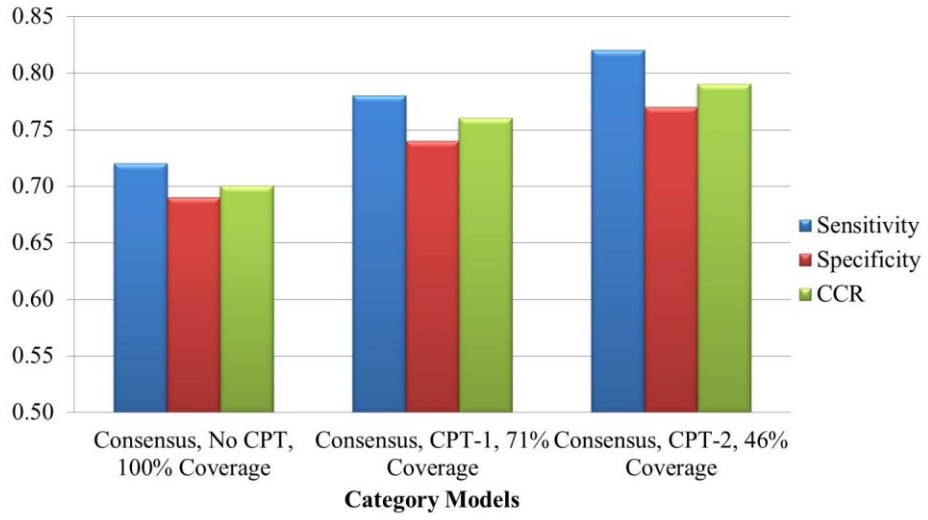


Figure 3.6. Predictivity of consensus CTG model with different consensus prediction thresholds (CPT)

Continuous Models

Four individual and one consensus model was also developed for the CNT-%F and CNT-logK(%F) bioavailability datasets. The results for both types of models are shown in Table 3.3. The statistics for the four individual CNT-%F models were relatively poor (R^2=0.13-0.30 and MAE=~24-53). Using Applicability Domain (AD) to remove unreliable predictions of structurally dissimilar compounds, as described previously (Zheng and Tropsha 2000), did not give significant improvement to our models (results not shown). Therefore AD was not used for the analysis. Compared to the individual models, the consensus CNT-%F model was also close to the upper boundary (R^2=0.28 and MAE=~24). To verify the statistical significance of all the models (in comparison to random chance performance), a two-way ANOVA test with a confidence level of 95% was performed for each model (Fisher 1935). The obtained *p* values were lower than 0.05.

The statistics for the four individual CNT-logK(%F) models were similar (R^2=0.11-0.30 and MAE=~23-28). The consensus CNT-logK(%F) model was also close to the upper boundary (R^2=0.25 and MAE=24). The obtained *p* values were lower than 0.05. Nevertheless, the distribution of errors was very different for the CNT-logK(%F) model compared to %F scale (Figure 3.7). Compounds with very low and very high %F values were predicted more accurately by the CNT-logK(%F) model.

Table 3.3. Performance of individual and consensus CNT models using a five-fold cross-validation (n=995)

Models		CNT-%F					CNT-logK(%F)				
		D-RF	D-SVM	MOE-RF	MOE-SVM	Consensus	D-RF	D-SVM	MOE-RF	MOE-SVM	Consensus
R^2		0.30	0.13	0.29	0.19	0.28	0.30	0.11	0.28	0.14	0.25
Error Analysis	Mean Absolute Error (%F)	51.33	52.65	23.88	52.66	24.05	22.93	28.38	23.09	28.28	24.02
	Standard Deviation (%F)	16.51	26.02	17.16	25.37	18.46	22.77	31.66	23.79	32.73	26.4
Two-Way ANOVA Test, α=95%	$F_{significance}$ (p-value)	$1.00x10^{-78}$	$2.33x10^{-31}$	$5.33x10^{-76}$	$2.48x10^{-48}$	$6.53x10^{-71}$	$1.81x10^{-80}$	$3.60x10^{-27}$	$2.48x10^{-74}$	$3.71x10^{-34}$	$3.57x10^{-65}$
	$F_{Calculated}$	423.96	145.48	406.2	238.21	373.65	435.44	123.73	395.44	160.25	337.87

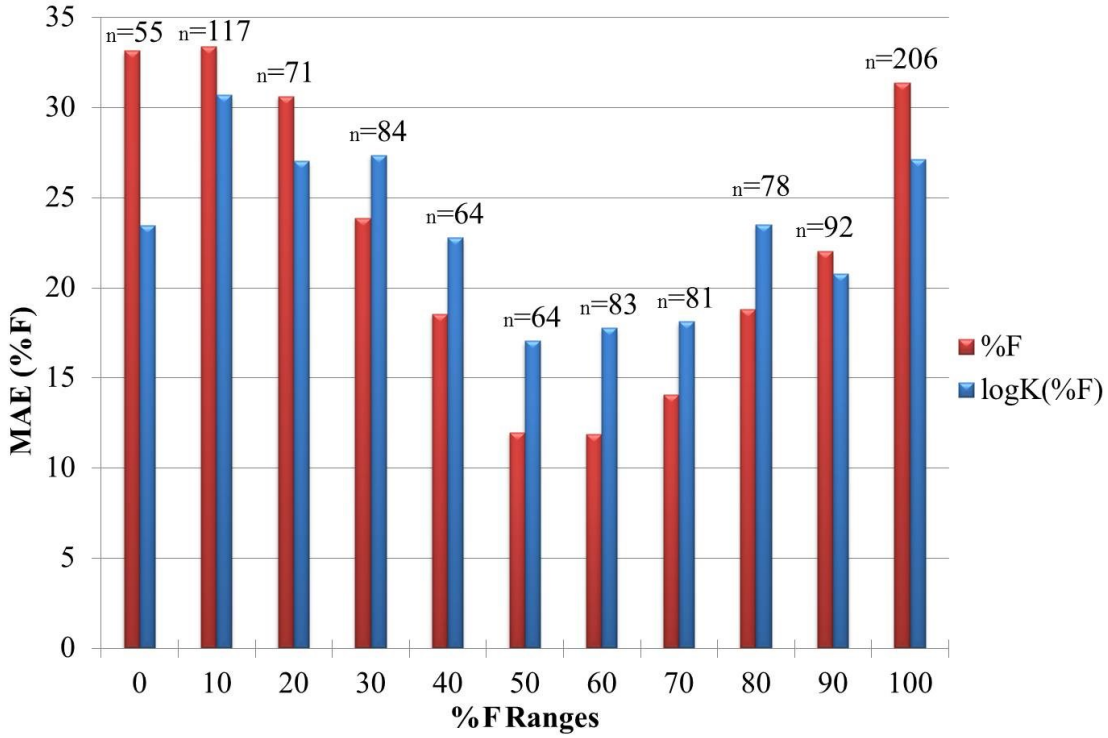


Figure 3.7. Distribution of prediction errors (as MAE) relative to experimental %F. Red and blue bars represent consensus CNT-%F and CNT-logK(%F) models respectively.

Integrating Human Intestinal Transporter Parameters into the CNT-%F Bioavailability Model

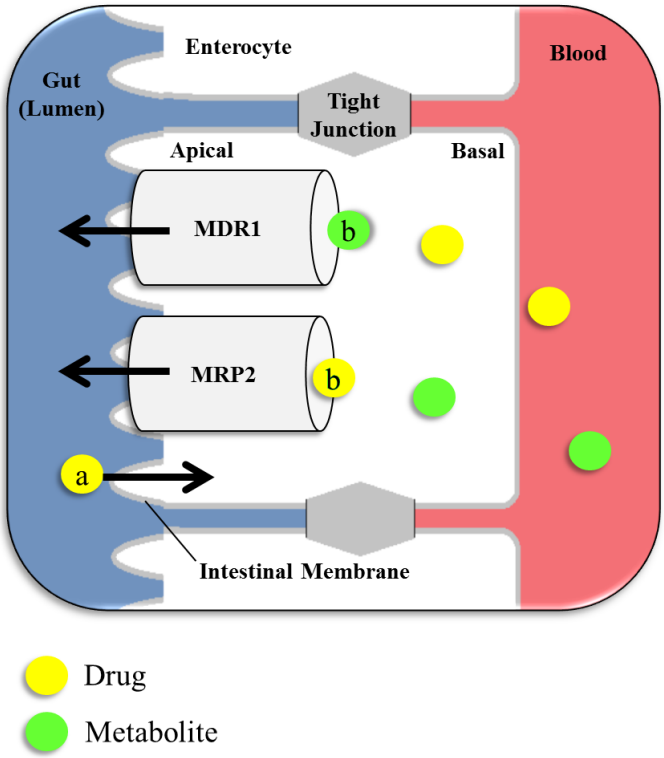


Figure 3.8. Drug efflux by intestinal transporters Multidrug Resistance Protein 1 (MDR1) and Multidrug Resistance-Associated Protein 2 (MRP2) in an enterocyte; a) drug passively diffusing through the intestinal membrane; b) drug and metabolite transported out of the enterocyte.

HITs are an important factor in intestinal absorption, which greatly affects oral bioavailability and other pharmacokinetic properties of their substrate (Shugarts and Benet 2009). Figure 3.8 depicts the transportation of drug molecules by MDR1, MRP2, and by passive diffusion in an enterocyte. It is known that both MDR1 and MRP2 are

responsible for the active efflux of drug molecules from the enterocyte to the lumen (Giacomini and Sugiyama 2011). For this reason, a drug with low passive diffusion, but high substrate affinity to MDR1 and/or MRP2 is not likely to be highly bioavailable. Breast Cancer Resistant Protein, another major efflux transporter, was also considered but its imputed interactions did not enhance the results (data not shown) and subsequently excluded it from further analysis. Four MDR1 and MRP2 model predictions (Sedykh et al. 2013) were used to calculate the probability of interaction (POI) for the compounds in our dataset. Then, the mean probability of interaction (MPOI) for MDR1-s, MDR1-i, MRP2-s, and MRP2-i for drugs in various bioavailability ranges were calculated (Figure 3.9), where *s* and *i* represent substrates and inhibitors, respectively.

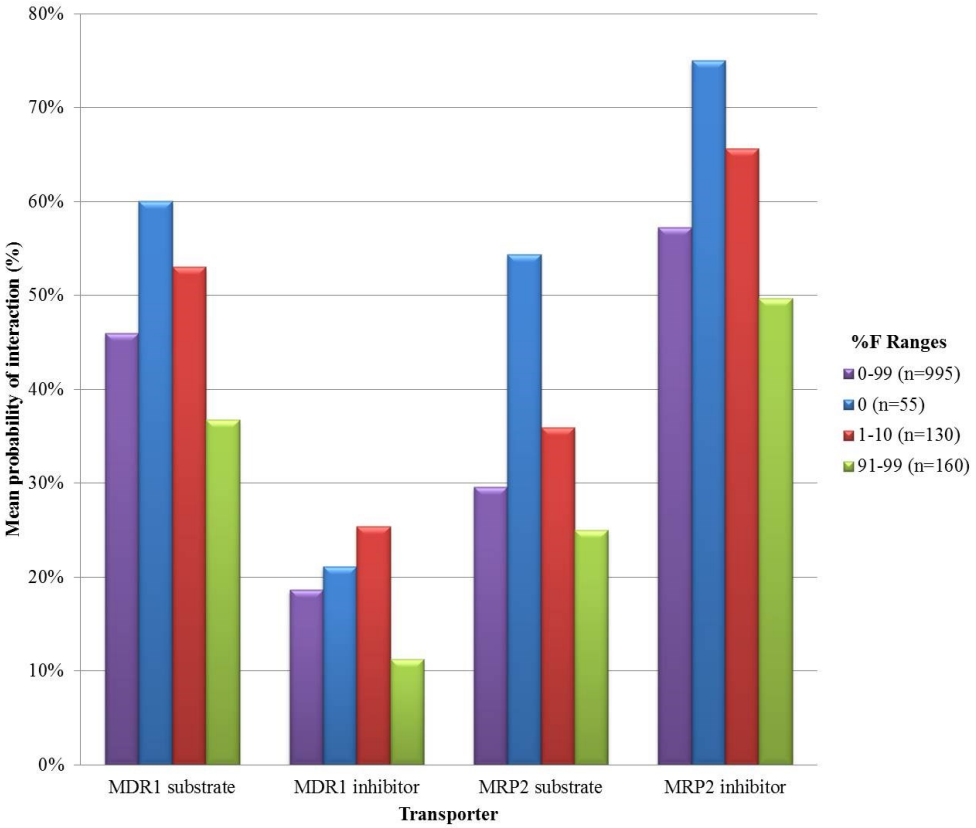
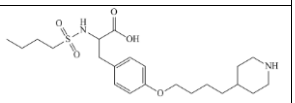
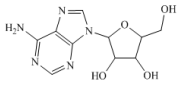
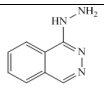
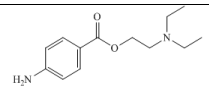
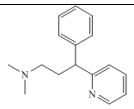
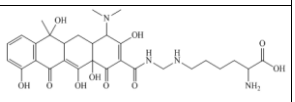


Figure 3.9. Mean probability of interaction (MPOI) for compounds in specified %F ranges

Table 3.4. Examples of compounds with high prediction errors that were successfully (No. 1-3) or not successfully (No. 4-6) removed after combining the interaction properties of MDR1-s, MDR1-i, MRP2-s, and MRP2-i.

No.	Compounds	Transporter Score[a]				%F	Pred. %F	Structure
		MDR1-s	MDR1- i	MRP2-s	MRP2- i			
1	Tirofiban	1	1	1	1	0	54	
2	Vidarabine	1	1	1	1	0	36	
3	Hydralazine	1	1	1	1	23	74	
4	Procaine	0	0	0	0	0	66	
5	Pheniramine	0	0	0	0	99	34	
6	Lymecycline	1	1	1	1	99	7	

[a] Transporter score: 0 (POI<MPOI) and 1 (POI>MPOI)

I hypothesized that the drugs with a POI value greater than the MPOI value of orally non-bioavailable drugs in each transporter models should not have a predicted %F value greater than 10. Table 3.4 lists examples of drugs, with large %F prediction errors, and their HIT classifications. To simplify the discussion, the HIT predictions for each compound were classified as 0 (POI<MPOI) or 1 (POI>MPOI) based on my hypothesis. For example, all the HIT predictions and the predicted %F for Tirofiban were 1 and 54%, respectively. Therefore, Tirofiban was considered an outlier and was subsequently removed from the final model. On the other hand, Procaine had a high prediction error and could not be removed, because all the HIT predictions were classified as 0. In this case, the low bioavailability of this drug may be due to other HITs, metabolism, or other reasons. Table 3.4 lists examples of compounds with high prediction errors that were successfully removed (No. 1-3), and missed (No. 4-6) by our rule. The predictivity of compounds, that are substrates of the two transporters, could not be improved by this rule. These type of compounds with large prediction errors (e.g. compounds 4 and 5) may be due to other factors, such as metabolic stability. For example, Procaine (%F=0, Pred. %F=66) was predicted as a false positive and is metabolized by an esterase in the liver (Inoue et al. 1980). The bioavailability model will be expected to be further improved by integrating metabolism-related parameters, such as CYPs interactions. The compound Lymecycline (%F=99, Pred. %F=3), in Table 3.4 is a specific case. It was predicted to be the substrate of the two transporters, but it is actually a high bioavailable drug. Lymecycline is water-soluble at physiological pH and is readily absorbed through the gastrointestinal tract (Dubertret et al. 2003).

Using the HIT interaction rule described above to remove unreliable predictions improved the prediction accuracy of the current CNT-%F models, especially the consensus model. HIT interactions were not incorporated into the CNT-logK(%F) models. Since the overall results for the two types of models were similar, doing so would have been redundant. The results

for integrating various HIT parameters into the consensus CNT-%F model are listed in Table 3.5. It is noticeable that the use of four HIT parameters affects the predictions differently. However, the best results were obtained from combining all four transporters parameters. The R^2 coefficient enhanced from 0.28 to 0.40 and the MAE reduced from 24 to 21. Subsequently, using HIT parameters reduced the prediction coverage to 30%. A two-way ANOVA (α=95%) (Fisher 1935) and Bootstrap Non-Parametric Permutation (N=10,000; α=95%) (Kornbrot 2005; Spearman 1987) analysis revealed that the observed improvements are statistically significant. Therefore integrating HIT information with the oral bioavailability models was a valid approach. It was also noticed that the relationship between %F and drug interactions with MDR1 and/or MRP2 is non-monotonic. Some HIT combinations were better than others and incremental improvements were not always achieved when integrating another HIT parameter. This is understandable as there is overlap in substrate specificity between different efflux transporters (Sedykh et al. 2013).

Table 3.5. Performance of consensus CTN-%F model with and without integrating HIT POI

Models		Consensus	Combinations of Consensus with HIT POI				
			MDR1-s	MRP2-s	MRP2-s & MRP2-i	MRP2-i & MDR1-s	MDR1-s, MDR1-i, MRP2-s, & MRP2-i
R		0.28	0.31	0.26	0.36	0.40	0.40
n		995	558	758	450	362	304
Coverage		100%	56%	76%	45%	36%	30%
Error Analysis	Mean Absolute Error	24.05	22.97	23.83	21.54	21.61	21.00
	Standard Deviation	16.05	16.35	16.38	15.71	16.13	16.11
Two-Way ANOVA Test α=95%	$F_{significance}$ (p-value)	$6.53x10^{-71}$	$3.42x10^{-48}$	$2.74x10^{-57}$	$6.67x10^{-45}$	$6.66x10^{-41}$	$6.55x10^{-35}$
	$F_{Calculated}$	373.65	259.48	259.50	248.83	232.89	197.85
Bootstrap, Non-Parametric Permutation N=10,000 α=95%	Spearman (p value)	N/A	0.33	0.74	0.08	0.06	0.20
	Pearson (p value)	N/A	0.17	0.17	0.04	0.01	0.02

Discussion

Although the results for the CNT-%F and CNT-logK(%F) models are relatively low, each model has their advantages. The MAE was determined for the various %F ranges for both of the models (Figure 3.7). For predicting compounds with extreme %F values (%F≤20% and %F≥90%), the CNT-logK(%F) models performed better. For the mid %F ranges (%F=20-90%), the CNT-%F model yielded more accurate results. Both types of models can be used to predict oral bioavailability. Using the CNT-logK(%F) model can be advantageous if higher accuracy is needed for very low or very high bioavailability ranges. However, combining the results of the CNT-%F and CNT-logK(%F) models did not result in better statistics (data not shown).

Interpretation of QSAR models

There are many factors that affect oral bioavailability. Some examples are intestinal absorption, water solubility, and lipophilicity (Buxton and Benet 2011). These parameters can be modified to increase or decrease oral bioavailability by slightly changing certain chemical features on a compound. The chemical structures potentially related to oral bioavailability were evaluated by analyzing Dragon descriptors. Dragon descriptors contain more diverse structural descriptors compared to MOE, so they more are practical for the model interpretations. The average values of the most important structural Dragon descriptors for both the 100 least bioavailable drugs (%F=0-10) and 100 most bioavailable drugs (%F=90-99) were caluclated (Figure 3.10).

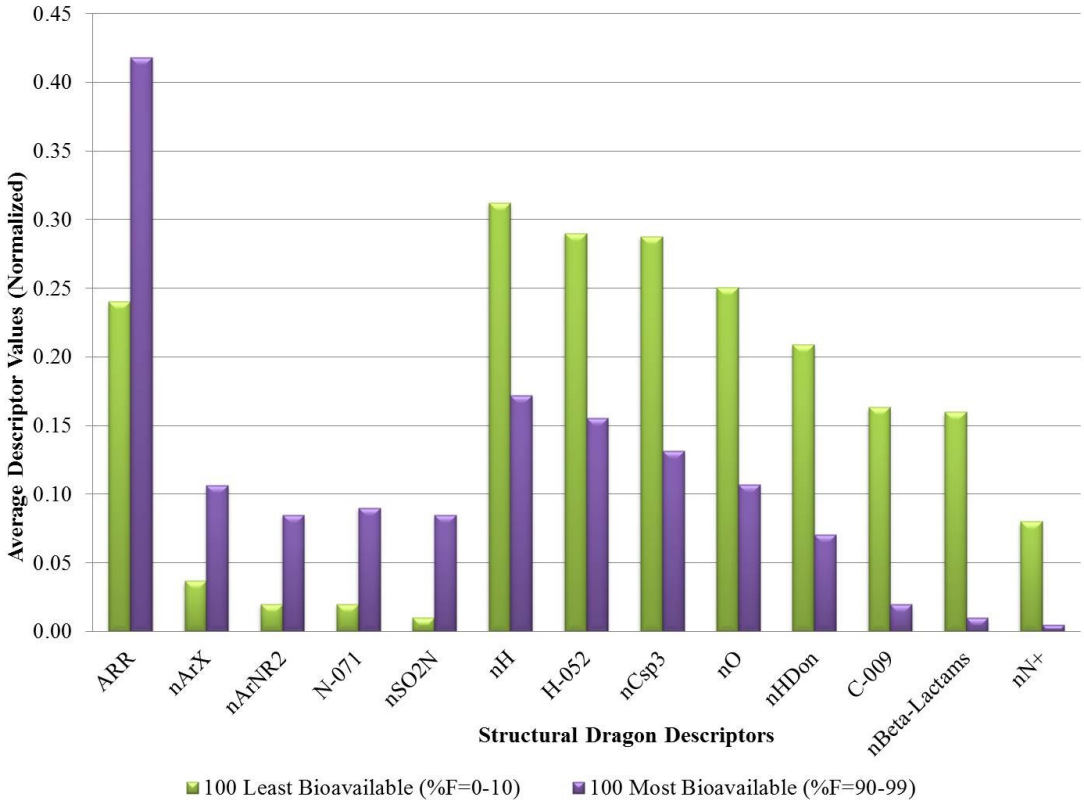


Figure 3.10. Chemical structure analysis for the 100 least and 100 most bioavailable compounds

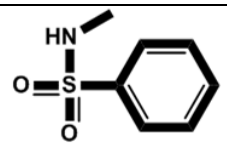
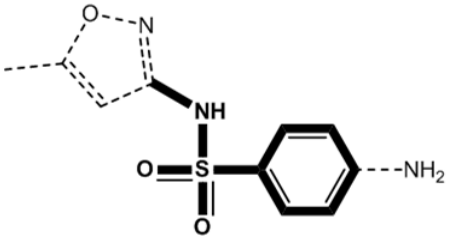
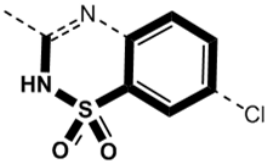
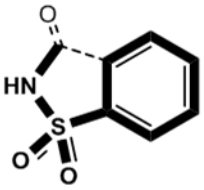
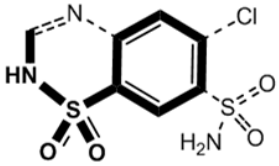
There were more descriptors related to low bioavailable drugs than high bioavailable drugs. For example, compounds with high %F normally have aromatic groups (descriptor ARR). Compounds with multiple aromatic rings like, Anthracene and Naphthalene, can readily pass through biological membranes, which facilitate their absorption and increase their bioavailability (Utvik and Johnsen 1999). Compared to aromatic rings, drugs with aliphatic carbon chains (descriptors nCsp3 and C-009) were likely to have lower %F since these kinds of drugs are poorly soluble in water, which greatly lowers their bioavailability (Hermann 1972). An example of this type of drug is Docosanol (%F=0%). There are several descriptors that refer to the Lipinski rule of five (Lipinski et al. 2001).

According to the Lipinski rule of five, the existence of over five hydrogen bond donors and/or acceptors may cause the decrease of the drug bioavailability (Lipinski et al. 2001). The descriptor nO refers to the number of oxygen atoms (oxygen is a potential hydrogen bond acceptor or donor), and the descriptor nHDon refers to the number of hydrogen donor groups. Our modeling results support this hypothesis since there are greater descriptor values for low bioavailable drugs than high bioavailable drugs. On the other hand, the presence of aromatic halogens (descriptor nARX) was prevalent in highly bioavailable drugs. It was reported that the existence of an appropriate number of aromatic halogens can enhance the lipophilicity and aqueous solubility of a drug, two properties critical for absorption and bioavailability (Birnbaum 1985). Descriptors nArNR2 and N-071 represent tertiary aromatic amines and aromatic amines, respectively. Drugs with aromatic amines can be readily absorbed through the gastrointestinal tract (Holland et al. 2005; Stillwell et al. 1999). Beta-Lactams (descriptor nBeta-Lactams) tend to have low %F due to their low lipophilicity which makes it difficult to passively diffuse across the intestinal membrane. Beta-lactams that have high %F are typically transported by intestinal influx transporters like Peptide Transporter 1 (Saitoh et al. 1996), which increases %F.

Some features, which are considered to be important for bioavailability, represent complex mechanisms. For example, N-alkylation (refer to the descriptor nN+) is a common procedure used to increase the aqueous solubility of drug molecules which have low bioavailability, such as Bupivacaine (%F=0%) (Nielsen et al. 2005). However, this also reduced lipophilicity and the net effect on the oral bioavailability is hard to measure. In our dataset this descriptor was considered to be relevant to low bioavailability since

this feature was found mostly in low bioavailable compounds. The arylsulfonamide moiety (represented by the descriptor nSO2N) was associated with high oral bioavailability. A similar fragment descriptor was also identified by CASE Ultra as the top biophore. There were 23 drugs in our dataset that contained this structural feature and their average %F was 77%. Examples of these drugs are shown in Table 3.6. Methods for improving the oral bioavailability of sulfonamides have been studied for many decades. Previous studies found that the nitrogen atom of this fragment (as shown in Table 3.6) plays an important role in binding to the receptor and is critical to membrane permeability and bioavailability (Sawa et al. 2005; Wu et al. 2001). However, the potential mechanisms that are relevant to the bioavailability of sulfonamides are still not well understood.

Table 3.6. Examples of compounds with the arylsulfonamide structural feature found using the CASE Ultra model

Arylsulfonamide sub-structure Avg. %F=77	
Sulfamethoxazole %F=99	
Diazoxide %F=90	
Saccharin %F=84	
Chlorothiazide %F=13	

Cytochrome P450 (CYP) enzymes have a crucial impact on the metabolic stability of a drug (Meunier et al. 2004). Some descriptors in Figure 3.10, such as the number of hydrogens (nH), hydrogens attached to sp^3 carbon atoms (H-052), and number of sp^3 carbon atoms (nCsp3), were found to be correlated with low bioavailability. It was reported that CYP enzymes hydroxylate the C-H bond on sp^3 carbon atoms (Meunier et al. 2004). Thus, these three descriptors may represent the structural features with low

metabolic stability. Interestingly, halogenated hydrocarbons are also susceptible to oxidative dehalogenation by CYP enzymes (Meunier et al. 2004). However, our descriptor analysis shows that aromatic halogens are related to high bioavailability (likely via enhancing membrane permeability). This relationship could be further explored in the future.

Conclusion

In this study, a database containing 995 unique human oral bioavailable drugs was compiled. The diverse drugs in this data set include molecules with both low and high bioavailability. Then, the %F values were harmonized and all chemical structures were evaluated to ensure that the database is accurate.

The bioavailability database was used to develop both CTG and CNT models by using various modeling approaches. The consensus predictions show better performance than individual models for both CTG and CNT models. Although the results of CNT models are relatively poor, adding HIT parameters indeed improved the model prediction accuracy. Correctly using HIT parameters based on the transport direction allowed us to remove some compounds with high predictions errors. Efflux transporters that transport drugs out of the enterocytes can limit the oral bioavailability of their drug-substrates. In this study the two efflux transporters, MDR1 and MRP2, were found to be important for enhancing the oral bioavailability predictions in the models.

All of the models developed in this study can be used to evaluate the bioavailability of new drug candidates. The analysis of the important descriptors in the resulting models showed the relationships between several types of chemical structures and drug oral bioavailability. This type of knowledge could be useful for designing new

drug molecules with suitable oral bioavailability. The use of HIT parameters was beneficial to the model predictions. It is now confirmed that HITs need to be a component in future bioavailability models. Future directions of *in silico* oral bioavailability modeling should also take into consideration interactions with the CYP enzymes. Similar methods could be developed and employed to model other complex bioactivities of drugs and drug-like molecules.

Chapter 4. Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using the Antioxidant Response Element Reporter Gene Assay Models and Big Data

Chapter Overview

Hepatotoxicity accounts for a substantial number of drugs withdrawn from the market. Traditional animal models used to detect hepatotoxicity are expensive and time consuming. Alternative *in vitro* methods, especially cell-based High-Throughput Screening (HTS) studies, have provided the research community with a large data set from cell-based signaling and toxicity assays. Among the various assays used to screen potential toxicants is the Antioxidant Response Element *beta* lactamase reporter gene assay (ARE-*bla*), which identifies chemicals that have the potential to induce oxidative stress and was used to test approximately 10,000 compounds from the Tox21 program. The ARE-*bla* computational model and HTS data from a public data source (PubChem) were used to profile environmental and pharmaceutical compounds with hepatotoxic data. Quantitative Structure-Activity Relationship models were developed based on ARE-*bla* data. The models predicted the potential oxidative stress response for known liver toxicants when there was no ARE-*bla* data available. Liver toxicants were used as probe compounds to search PubChem Bioassay and generate a response profile, which contained thousands of bioassays (> 10 million data points). By ranking the *In Vitro-In Vivo* Correlations (IVIVC), the most relevant bioassay(s) related to hepatotoxicity were identified. The liver toxicants profile contained the ARE-*bla* and relevant PubChem assays. Potential toxicophores for well-known toxicants were created by identifying chemical features that existed only in compounds with high IVIVC. Profiling the chemical IVIVCs created an opportunity to fully explore the source-to-outcome

continuum of modern experimental toxicology using cheminformatics approaches and big data sources.

Introduction

Traditional animal models used to evaluate hepatotoxicity are expensive and time consuming (Hartung 2009). *In vitro* assays are used as an alternative to better understand hepatotoxicity (Adler et al. 2011; Zhu et al. 2014a). However, endeavors to correlate *in vitro* and *in vivo* hepatotoxicity (Moeller 2010) have not successfully replaced *in vivo* hepatotoxicity models (Ekins 2014; MacDonald and Robertson 2009).

There is an unmet need to develop predictive assays for hepatotoxicity (Chen et al. 2014). As an alternative, High-Throughput Screening (HTS) approaches are used to screen large chemical libraries (> 50,000 compounds) to elucidate toxic mechanisms and prioritize candidates for further animal tests (Zhu et al. 2014b). This led to the rapid generation of bioassay data. PubChem, the leading public bioassay data repository, contains > 50 million compounds and > 700,000 assays (Wang et al. 2014). This amount of "big data" is difficult to process and analyze using standard data processing tools.

Another issue with using HTS for toxicological studies is that it tests compounds at one concentration, which may not reveal its toxic effects. This was addressed by the US Tox21 inter-agency collaboration (Attene-Ramos et al. 2013; Collins et al. 2008; Committee on Toxicity Testing and Assessment of Environmental Agents 2007; Dix et al. 2007). Based on their strategy , the National Institutes of Health Chemical Genomics Center (NCGC), now part of the National Center for Advancing Translational Sciences (NCATS), developed Quantitative High-Throughput Screening (qHTS) (Inglese et al. 2006). A qHTS experiment tests approximately 100,000 compounds at 15 different concentrations in three different runs within a week (Attene-Ramos et al. 2013). This approach is more rational than single-dose HTS, because it simulates dose-dependent

animal toxicity effects (Eaton and Gilbert 2010). These results are available online (http://www.ncbi.nlm.nih.gov/pcassay?term=tox21, accessed January 19, 2015).

The Antioxidant Response Element (ARE) pathway plays a major role in regulating and alleviating oxidative stress (Ma 2013), which after long-term exposure causes many pathophysiological conditions, including cancers and hepatotoxicity (Hybertson et al. 2011; Shuhendler et al. 2014). Briefly, ARE pathway is regulated by Kelch-like ECH-associating protein 1 (Keap1) and nuclear factor erythroid 2-related factor 2 (Nrf2). Keap1 contains cysteine residues that interact with reactive oxygen species (ROS) and electrophilic fragments that can trigger the dissociation of the Keap1-Nrf2 complex (Zhang and Hannink 2003). Then, Nrf2 translocates into the nucleus (Kensler et al. 2007), binds to the ARE (Itoh et al. 1997), and regulates the transcription of the antioxidative enzymes (Venugopal and Jaiswal 1998). Hindering transcription can lead to the accumulation of ROS, oxidative stress, and liver toxicity (Shuhendler et al. 2014). The qHTS ARE *beta* lactamase reporter gene assay (ARE-*bla*) can detect compounds that activate the ARE pathway and induce oxidative stress (Attene-Ramos et al. 2013; Shukla et al. 2012; Simmons et al. 2011). However, this assay alone is not sufficient for accessing animal toxicity. The correlations between the ARE pathway and animal toxicity (*i.e.,* hepatotoxicity) are not well understood.

Even with all the data from HTS and/or qHTS studies, the relationship between *in vitro* and *in vivo* toxicity is still unclear (Low et al. 2011; O'Brien et al. 2006) and needs further investigation. In this study, this challenge was addressed by developing chemical *in vitro-in vivo* correlations (IVIVC) between ARE pathway activation and hepatotoxicity (*i.e.,* liver damage). An in-house automated profiling tool and cheminformatics

approaches used qHTS ARE-*bla* and liver toxicity data to retrieve relevant assays, from PubChem, and revealed liver toxicity targets. Analyzing chemical fragments of liver toxicants revealed potential toxicophores (toxic chemical features) with clear IVIVC for a subset of compounds. Our study suggests that the use of assays as an alternative model for toxicity is feasible based on the chemical IVIVC identified from a big data source.

Methods

qHTS ARE-*bla* Dataset.

The initial concentration-response profiles for the Tox21 10K collection tested in the qHTS ARE-*bla* tests were conducted at the NCATS (Attene-Ramos et al. 2013; Shukla et al. 2012). The Tox21 10K chemical library (http://www.epa.gov/ncct/dsstox/sdf_tox21s.html, accessed October 2, 2012) consists of compounds procured from commercial sources by the Environmental Protection Agency (EPA), National Toxicology Program (NTP), and NCGC/NCATS (Huang et al. 2011), for a total of ~10,500 plated compound solutions consisting of 8,311 unique chemical substances including pesticides, industrial, food-use, and drugs. The qHTS ARE-*bla* datasets can also be downloaded from PubChem using Bioassay Accession Identifiers (AID) 743219 and 651741. PubChem is a public repository for chemical structures and their biological properties (Wang et al. 2014). Bioactivity data in PubChem are contributed by hundreds of institutes, research laboratories, and specifically those screening centers under the NIH Molecular Libraries Program (MLP) and the Tox21 program. Descriptions of the individual datasets are listed in Table 4.1.

Table 4.1. Comprehensive toxicity databases compiled from public sources

Names	Types	Description	Number of compounds
Tox21 phase I (NTP and EPA) ARE-*bla* (https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=651741, accessed August 29, 2015)	*In vitro*	Compounds characterized in traditional toxicology tests and/or known to be harmful to humans and the environment	2,617
Tox21 phase II 10K ARE-*bla* (http://www.epa.gov/ncct/dsstox/sdf_tox21s.html, accessed October 2, 2012)	*In vitro*	Diverse compounds (pesticides, industrial, food-use, drugs, etc.) with chemical features that are of interest to toxicologists	8,311
FDA liver damage (Zhu and Kruhlak 2014)	*In vivo*	Drugs known to cause liver damage (*e.g.,* necrosis, lesions, traumatic liver injury)	1,314
PubChem Bioassay (http://www.ncbi.nlm.nih.gov/pcassay/, accessed February 27, 2014)	*In vitro & in vivo*	Compounds that have been validated and screened in different bioassays	48M+

The concentration-responses were normalized, range-scaled to [0, 100], and converted into curve fingerprints (Sedykh et al. 2011) using an in-house program. The source code can be downloaded from GitHub (https://github.com/sedykh/curvep). Each curve fingerprint was summed into one value termed "CurveP." CurveP represents the overall signal of the compound from its qHTS concentration-response curve that was noise filtered (*e.g.,* CurveP = 0 means no significant signals observed). Three criteria were used to classify each compound with regard to activity: 1) CurveP, 2) maximum

concentration-response, and 3) number of concentration-responses ≥ 20. The latter two describe the consistency in the concentration-responses. The scheme is detailed in Table 4.2. For example, a compound was classified as active if CurveP was > 0 and more than one concentration-response ≥ 20. Lastly, since all compounds were tested multiple times and all data were available, activities of each compound were averaged before classification.

Table 4.2. Definition of compound activity categories from concentration-response curves and the CurveP algorithm for the qHTS ARE-bla datasets

Category	Activity	CurveP	Maximum response	Number of responses > 20 units
Active[a]	1	> 0	≥ 20	> 1
Potential active[b]	0.75	> 0	≥ 20	= 1
Inconclusive[c]	0.25	= 0	< 20	= 0
Inactive[d]	0	= 0	< 10	= 0

[a]Strong ARE-*bla* activation signals observed; [b]Weak ARE-*bla* activation signal observed; [c]Inconsistent ARE-*bla* activation signal(s) observed; [d]Negligible or no ARE-*bla* activation signals observed.

In Vivo Hepatotoxicity Dataset

A liver damage dataset compiled by the Food and Drug Administration (FDA) Center for Drug Evaluation and Research (Zhu and Kruhlak 2014) and Multicase Inc., originally contained 1,314 compounds (661 toxic and 653 non-toxic), but six compounds could neither be matched to a CID nor CASRN**.** Therefore, only the remaining 1,308 compounds were used in this study (658 toxic and 650 non-toxic).

Chemical Structure Curation

The structures of all compounds used in this study were curated to remove errors and standardized to a uniform representation. Konstanz Information Miner (KNIME)

version 2.9.2 matched all compound names and PubChem Compound Accession Identifiers (CID) with its appropriate Simplified Molecular-Input Line-Entry System (SMILES) from PubChem. The in-house descriptor generators could not process large molecules (molecular weight > 2000 g/mol) and compounds without chemical structures. These compounds were removed. ChemAxon Standardizer and Structure Checker version 6.2.2 and CASE Ultra version 1.5.0.1 curated, standardized, and converted all the chemical structures into 2-D SMILES. Stereoisomers were considered as one compound. Metalorganics were removed and all salts were neutralized, because the descriptor generator cannot process them. Mixtures were manually evaluated and the major component was kept.

Measures of Quality and Reliability

To systematically evaluate the quality and reliability of the Quantitative Structure-Activity Relationship (QSAR) models and IVIVCs developed in this study, the sensitivity and specificity of each assay relative to *in vivo* animal toxicity data was calculated, and the correct classification rate (CCR) where CCR = [(sensitivity + specificity) / 2] × 100 was derived (Daniel 2009; Kim et al. 2014). In addition, the likelihood parameter (L) as an indication of the likelihood that active responses in a bioassay correlated with *in vivo* toxicity outcomes, where L = sensitivity × [(false positives + true positives)/ (false positives + 1)] (Zhang et al. 2014) was calculated. The statistical significance of the IVIVCs were determined using Chi square (X^2) tests comparing the *in vitro* assay predictions to expectations based on *in vivo* toxicity data, under the null hypothesis of no association between the two data sources (Daniel 2009).

Workflow for Profiling the Mechanisms of Liver Toxicants

The chemical IVIVC between qHTS ARE-*bla* perturbation or relevant PubChem assays and liver damage was evaluated. The profiling workflow has three major stages (Figure 4.1): 1) automated biological response profiling, 2) QSAR modeling of qHTS ARE-*bla* activation, 3) chemical IVIVC evaluation.

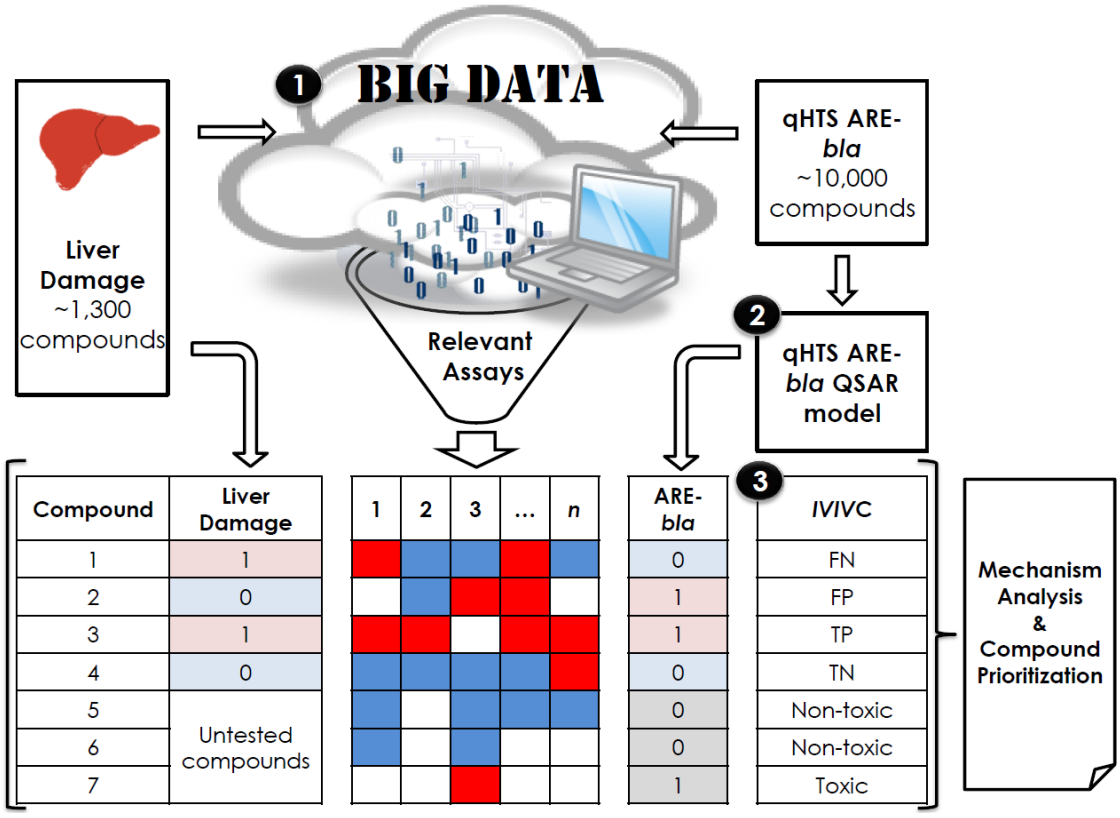


Figure 4.1. The workflow for profiling liver toxicants consists of three major stages: (1) automated biological response profiling, (2) QSAR modeling of qHTS ARE-bla activation, (3) chemical IVIVC evaluation. In the columns [Liver Damage, 1, 2, 3, "…", *n*, ARE-bla], actives are red color and "1;" inactives are blue and "0;" and inconclusive or untested are white and empty.

Automated Biological Response Profiling

The biological response profile was constructed from PubChem Bioassay data (http://www.ncbi.nlm.nih.gov/pcassay/, accessed February 27, 2014) with an in-house automated profiling tool (Zhang et al. 2014), which resulted in two profile groups. One group was related to qHTS ARE-*bla* activation and the second was related to liver damage. The correlations between all bioassays (> 2,000) and ARE-*bla* and liver damage were calculated (sensitivity, specificity, CCR, and L). Only bioassays that fit the following criteria were considered for the final biological response profile: 1) appeared in both profile groups; 2) contained > 10 active responses that matched the inputted data; 3) correlation was better than random (CCR > 0.5 and L ≥ 1); and 4) is an *in vitro* assay. Lastly, bioassays were selected for further analysis if there was literature evidence that showed these assays were used to study oxidative stress and/or liver damage.

It was hypothesized that compounds that were active in multiple assays, but were neither pan assay interference compounds (Baell and Holloway 2010) (*i.e.,* compounds that show false positive results in many assays due to assay technology specific artifacts) nor cytotoxic, were more likely to cause liver damage. Using the responses from the selected assays, the Rate of Actives (RA) was calculated to represent all the bioassay responses for each compound:

$$Rate\ of\ actives = \frac{A}{A+I}, \quad (1)$$

where A is the number of active responses and I is the number of inactive responses for a compound. The RA parameter was designed for this big data research since missing data can occur in the response profiles for target compounds. For example, if four assays were identified and a compound tested in all four assays was active in one assay, and

negative in the other three assays, it would have a RA = 0.25. However, if another compound was active in one assay, negative in two assays, and has no data or an inconclusive result for the fourth assay, it would have a RA = 0.33. Thus, potential bias due to missing assay data was reduced. An arbitrary RA threshold was used to distinguish toxic from non-toxic compounds (RA > 0.25 as toxic, RA ≤ 0.25 as non-toxic). The RA values were used to determine the IVIVC between liver damage and the assays. To measure the quality and reliability, each RA value was classified as true positive (TP), true negative (TN), false positive (FP), or false negative (FN) for a X^2 test ($\alpha = 0.05$).

QSAR Modeling of the ARE-*bla* Pathway

The qHTS ARE-*bla* datasets were used to develop qHTS ARE-*bla* combinatorial QSAR models. 2-D chemical descriptors for each compound were generated using Molecular Operating Environment (MOE) version 2011.10 and Dragon 6 version 6.0. All descriptors were normalized and range scaled to [0, 1]. 186 MOE and 2,629 Dragon descriptors were used to model qHTS ARE-*bla* activation.

The qHTS ARE-*bla* dataset was down-sampled using a chemical similarity search approach to balance the ratio of active and inactive compounds selected for modeling (Sedykh et al. 2011; Willett et al. 1998). This prevents the development of biased models. Active and inactive compounds from the Tox21 phase II dataset were selected to create the modeling set, since it was much larger than the Tox21 phase I dataset (Golbraikh et al. 2003; Tice et al. 2013). Using all 186 MOE descriptors, a principal component analysis was performed. Individual models were developed using the combination of MOE or Dragon descriptors and with either Random Forest (RF)

(Breiman 2001), Support Vector Machine (SVM) (Vapnik 2000), or *k*-Nearest Neighbor (*k*-NN) (Zheng and Tropsha 2000) algorithms. Six different combinations of descriptors and algorithms were used for modeling: MOE-RF, MOE-SVM, MOE-*k*-NN, Dragon-RF, Dragon-SVM, and Dragon-*k*-NN. Modeling results were averaged into a consensus model. Models were validated using 5-fold external cross-validation (80/20% split). Additional details about QSAR modeling and validation approaches can be found elsewhere (Golbraikh et al. 2003; Kim et al. 2014; Tropsha and Golbraikh 2007).

Since prediction values ranged from [0,1], two Consensus Prediction Thresholds (CPT) (Kim et al. 2014) were defined to classify compounds as active or inactive: CPT-1($\geq$0.5 as actives and $<$ 0.5 as inactives) and CPT-2 ($\geq$ 0.8 as actives and $\leq$ 0.3 as inactives). Predictions between CPT-2 thresholds ($<$ 0.8 and $>$ 0.3) were inconclusives. An Applicability Domain (AD) determined whether the external compounds were structurally dissimilar to the modeling set compounds or not (Tropsha and Golbraikh 2007). Predictions of compounds outside the AD were considered unreliable. Therefore, the coverage (fraction of compounds that are within the AD) was calculated when applying AD to the predictions.

Chemical IVIVC Evaluation

Potential toxicophores, chemical fragments with significant IVIVC, were identified by inputting compounds active in the qHTS ARE-*bla* and liver damage datasets into CASE Ultra and ChemoTyper version 1.0. The substructure search tool in KNIME searched the qHTS ARE-*bla* and liver damage datasets for compounds containing the potential toxicophores. The qHTS ARE-*bla* combinatorial QSAR models predicted compounds from the liver damage dataset that have not been tested in the qHTS ARE-*bla* assay. The

predictions were classified as TP, TN, FP, or FN to evaluate the chemical IVIVC for each subset of compounds with the potential toxicophores. The chemical IVIVC results were indicated using sensitivity, specificity, CCR, X^2 ($\alpha = 0.05$) (Daniel 2009).

Results

Overview of qHTS ARE-*bla* Dataset

The original qHTS ARE-bla data contained two datasets (Tox21 phase I and phase II). After combining, curating, and standardizing the chemical structures and activities, 6,767 unique compounds (919 actives, 748 potential actives, 760 inconclusives, and 4,340 inactives) remained. Potentially active and inconclusive compounds were excluded from further analyses. The remaining Phase I dataset consists of 1,474 unique compounds (341 actives and 1,133 inactives) and Phase II dataset consists of 5,134 unique compounds (878 actives and 4,256 inactives).

qHTS ARE-*bla* Combinatorial QSAR Models and Validation Sets

Six individual and one consensus qHTS ARE-*bla* QSAR models were developed for the modeling set (7 models total). The down-sampled modeling set contained 1,550 (750 actives and 800 inactives) unique compounds. Compounds left out of the modeling sets were placed into external validation sets. The chemical space, in a 3-D plot, covered by the modeling set versus its left out compounds and the liver damage dataset are shown in Figure 4.2A and 4.2B, respectively. External validation sets I [from Tox21 phase I] and II [for Tox21 phase II] contained 1,148 (175 active and 973 inactive) and 3,584 (128 active and 3,456 inactive) compounds, respectively. Finally, validation set III consists of compounds from the FDA liver damage data set that have not been experimentally tested in the qHTS ARE-*bla*. The predictions of these QSAR models for new compounds

represent the potential effect of these chemicals (either activation or no effect) in the qHTS ARE-*bla*.

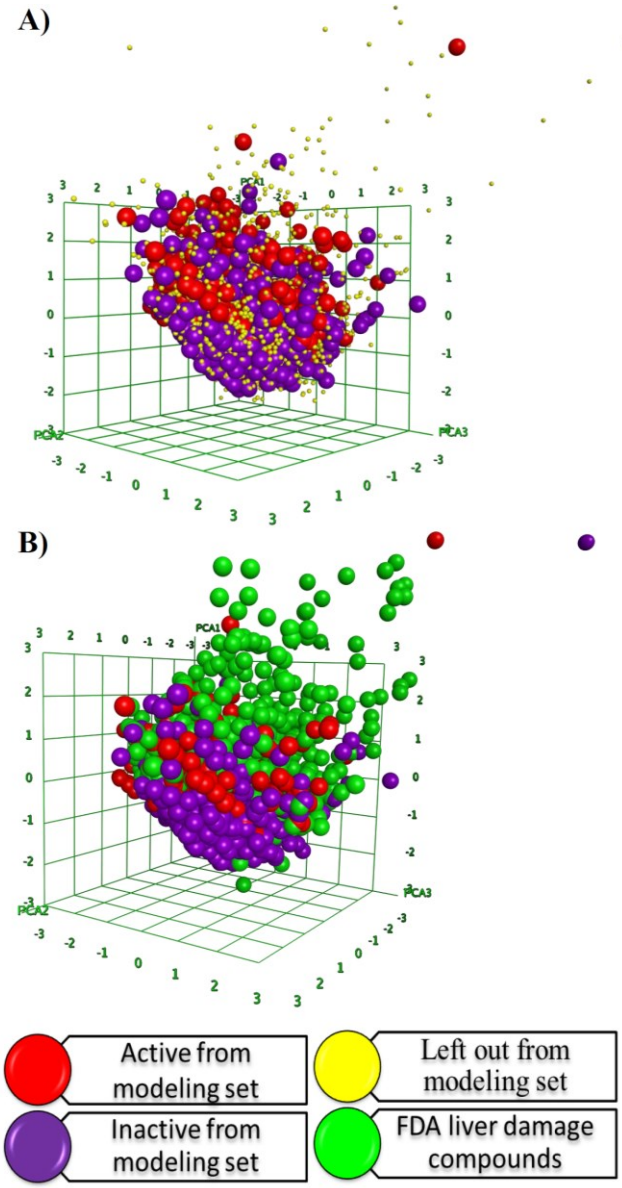


Figure 4.2. Chemical space plot of (A) the modeling set (actives = red, inactives = purple) vs. its left out compounds (yellow) and (B) the modeling set vs the FDA liver damage compounds (green) using the top three principal components generated using 186 MOE 2-D descriptors.

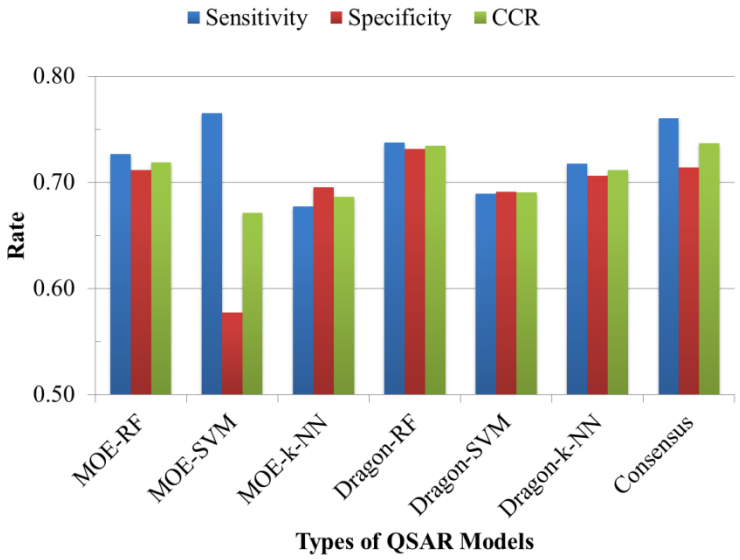


Figure 4.3. Performance of all individual models and the consensus model in the 5-fold cross validation

Table 4.3. qHTS ARE-*bla* combinatorial QSAR consensus model performance in 5-fold cross validation and against external validation sets, with and without Applicability Domain (AD), Consensus Prediction Thresholds (CPT) 1-2

Statistics		5-fold cross validation (80/20% split)	Validation set I	Validation set I + AD	Validation set II	Validation II + AD	Validation set III	Validation set III + AD
n (active/inactive)		750/800	175/973	132/757	128/3,456	59/2,566	174/274	3/9
CPT-1[a]	Sens[c] (%)	76	76	73	83	80	97	67
	Spec[d] (%)	71	83	85	72	78	4	11
	CCR[e] (%)	74	80	79	77	79	50	39
	Coverage[f] (%)	100	100	77	100	73	100	3
CPT-2[b]	Sens. (%)	75	68	62	93	90	100	0
	Spec. (%)	92	99	99	92	95	1	0
	CCR (%)	84	84	80	92	93	50	0
	Coverage (%)	35	40	34	45	37	71	0

[a]CPT-1:QSAR prediction ≥0.5 as actives and QSAR prediction < 0.5 as inactives; [b]CPT-2: QSAR prediction ≥ 0.8 as actives and QSAR prediction ≤ 0.3 as inactives; [c]Sens, sensitivity - percentage of active or toxic compounds predicted correctly; [d]Spec, specificity - percentage of inactive or non-toxic compounds predicted correctly; [e]CCR, correct classification rate; [f]Coverage - fraction of compounds that are within the applicability domain.

The performance of the qHTS ARE-*bla* combinatorial QSAR consensus model in the 5-fold cross validation and against the external validation sets, with an AD for CPTs 1 and 2 are shown in Table 4.3. The consensus modeling set showed good performance in the 5-fold cross validation (sensitivity = 75-76%, specificity = 71-92%, and CCR = 74-84%). The performance of the consensus model against external validation sets I and II without AD was satisfactory (sensitivity = 68-93%, specificity = 72-99%, and CCR = 77-92%). Using an AD, the external validation sets still resulted in acceptable performance (sensitivity = 62-90%, specificity = 78-99%, CCR = 79-93%, coverage = 34-77%). The individual models showed acceptable performance in the 5-fold cross validation (sensitivity = 68-77%, specificity = 58-73%, and CCR = 67-73%) (Figure 4.3). Overall, the consensus prediction results are comparable to the results of the best individual model which is Dragon-RF (sensitivity = 74%, specificity = 73%, CCR = 73%) (Figure 4.3). For validation set III, the consensus model did not perform well and most of the compounds were not within AD (sensitivity = 97-100%, specificity = 1-4%, and CCR = 50%). Using an AD did not improve the performance either (sensitivity = 0-67%, specificity = 0-11%, and CCR = 0-39%).

Liver Toxicants Profile and Its IVIVCs.

The goal of the automatic data mining and extraction tool used in this study is to reduce the big data pool to a much smaller size, which can be curated manually by experts. The profiling tool identified 2,978 assays (available upon request from the corresponding author) relevant to qHTS ARE-*bla* activation and/or liver damage, 958 of which existed in both profiles. Automated data extraction identified 20 PubChem assays based on the first three criteria for assay selection (appeared in both profile groups,

contained > 10 active responses that matched the inputted data, CCR > 0.5 and L ≥ 1). The assays are listed in the Appendix, Table A3. However, automatic methods cannot detect the detailed characteristics of an assay and distinguish the difference between *in vitro* and *in vivo* assays. The 20 assays identified by the initial automated screening procedure were manually reviewed to confirm that they met the *in vitro* selection criterion. For example, AID 1199, was identified as an *in vivo* assay. It did not fit the "*in vitro* assay" criterion and was removed. A total of eight non-*in vitro* assays were removed in this step and there were 12 *in vitro* assays left. Through the literature search, there is no information to support the relevance of six assays (AIDs 121, 123, 589, 590, 2330, and 720532) to either liver damage or oxidative stress. Six assays remained and two of them had redundant activities. For example, AIDs 686978 and 686979 refer to the qHTS human tyrosyl-DNA phosphodiesterase 1 (TDP1) assay tested under two different conditions, and the activities for most of the compounds were the same. AID 686978 was selected since the condition was performed in absence of the topoisomerase I poison camptothecin, which was more suitable for this study. AIDs 743065 and 743067 refer to the qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway. AID 743067 was selected because it was a summary assay (included both primary and cell viability counter screen results). After removing the redundant assays and evaluating the remaining assays by their mechanisms, four PubChem assays remained: AID 686978 qHTS for inhibitors of TDP1, AID 743067 qHTS assay to identify small molecule antagonists of the TR signaling pathway, AID 743140 qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway, and AID 743202 which was the qHTS

ARE-*bla* assay used in the QSAR models above. These assays are relevant to ARE perturbation and liver damage according to literature (Fielden et al. 2007; Königer et al. 2014; Malik and Hodgson 2002; Mantena et al. 2008) and were combined to create the biological response profile (Figure 4.3A).

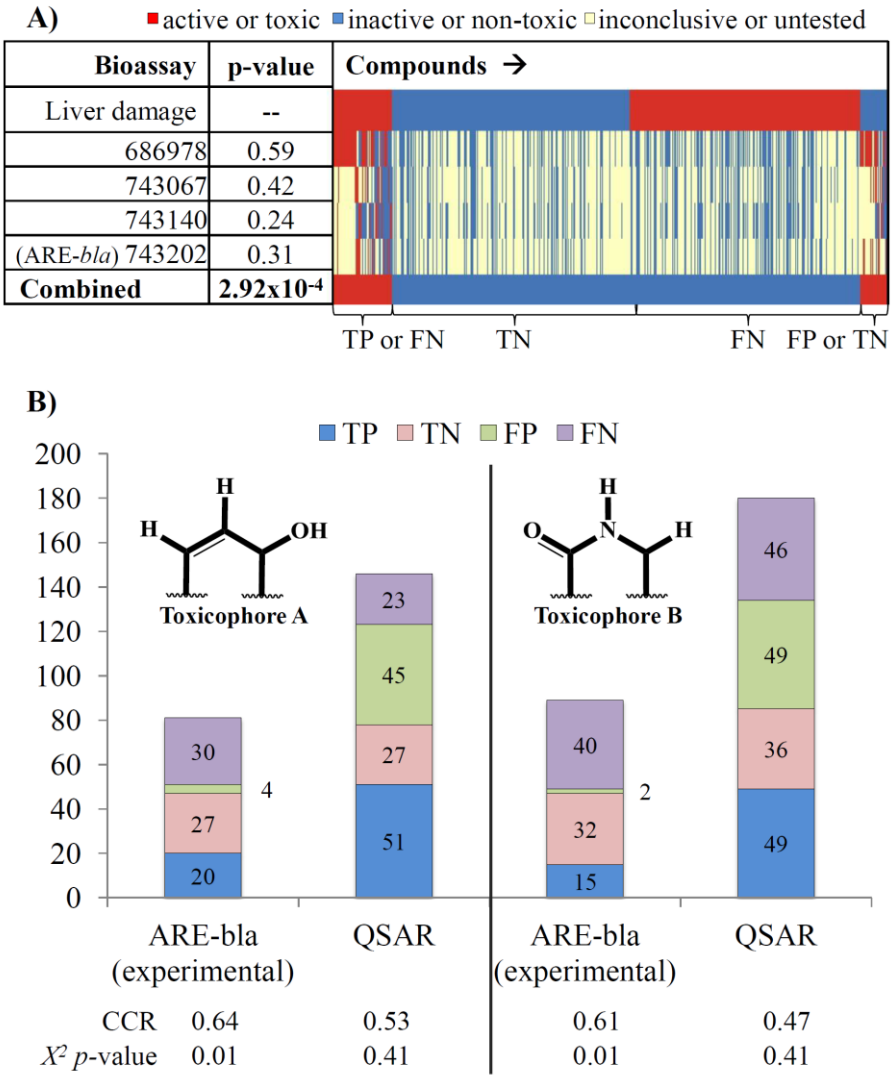


Figure 4.4. The IVIVC between selected assays and liver damage was evaluated by classifying responses as true positive (TP), true negative (TN), false positive (FP), or false negative (FN) for a X^2 (α = 0.05) or CCR test. (A) The biological response profile (red = active or toxic, blue = inactive or non-toxic, yellow = inconclusive or untested) of liver damage compounds represented in the heat map using the top four assays (AIDs 686978, 743067, 743140, and 743202). Individual assays show weak IVIVC, but the combined responses of the assays using threshold RA > 0.25 as active resulted in a statistically significant IVIVC (X^2 p-value = 2.92×10^{-4}). (B) The IVIVC between experimental qHTS ARE-*bla* activation and liver damage and the QSAR predictions for each liver damage compound, for subsets of overlapping compounds with potential toxicophores A (left) and B (right).

Although these top four assays met the selection criteria, the individual assay predictions were not significantly associated with *in vivo* liver damage (X^2 *p*-values for the independence of assays and *in vivo* liver damage = 0.24-0.59). However, combining these four assays and defining toxicity as RA > 0.25 resulted in a statistically significant association (X^2 *p*-value = 2.92×10^{-4}). The biological profile shows the responses for 953 compounds from the liver damage dataset against the top four assays and their combined responses, using threshold RA > 0.25 (Figure 4.4A). 361 liver damage compounds are not shown, because there was no bioassay data available for them.

The qHTS ARE-*bla* dataset used in this study contains > 6,000 compounds, but does not cover all the compounds in the liver damage dataset (860 tested, 448 untested). Therefore, qHTS ARE-*bla* combinatorial QSAR model was used to predict the activity of compounds that were not tested in the qHTS ARE-*bla* study. Table A2 lists the qHTS ARE-*bla* QSAR model consensus predictions and the experimental testing results, if available, of liver damage data set. The experimental calls represent the aggregated activity values derived from the concentration-response curves and CurveP values. It is important to mention that the liver damage dataset consists of mostly drug-like compounds that were outside of the AD of the QSAR models. In previous studies, QSAR models normally cannot predict the compounds out of AD as accurately as the compounds within AD (Tropsha and Golbraikh 2007). As shown in the principal component analysis (Figure 4.2B) and according to the AD analysis, most of the liver damage dataset compounds either share the same chemical space as the actives in the modeling set or are out of AD, meaning they are likely to be predicted as active by the QSAR models. This resulted in the increase of false positives in the later IVIVC

analysis, which provides a hint that extra experimental ARE data are still needed for the drug-like compounds of interest in the future study.

Using CASE Ultra and ChemoTyper, two subsets of compounds were identified. Subsets contained a chemical fragment that showed a statistically significant IVIVC between ARE-*bla* activation and liver damage in the X^2 test with *p*-values of 0.01 and are referred to as potential toxicophores A and B (Figure 4.4B), respectively. There are more true positives than false positives. Therefore, the active responses in this assay are potential signals of liver damage for the compounds that contain the potential toxicophores.

Furthermore, the qHTS ARE-*bla* combinatorial QSAR models were used to predict liver damage dataset compounds without experimental qHTS ARE-*bla* perturbation results. Figure 4.4B shows the IVIVC (TP, TN, FP, and FN) between the qHTS ARE-*bla* activation and liver damage, for compounds with potential toxicophores A and B, using experimental ARE-*bla* data and QSAR predictions. When using only QSAR results, the IVIVC was not statistically significant (X^2 *p*-value = 0.41) for both potential toxicophores. This is due to structural differences between the drugs in the liver damage dataset and the compounds in the Tox21 dataset, used to develop the qHTS ARE-*bla* combinatorial QSAR model, as described above. The result suggests the limitation of applying QSAR models to predict new compounds that are out of AD.

Discussion

ARE pathway perturbation is an important mechanism for alleviating and preventing oxidative stress (Ma 2013). In this study, qHTS ARE-*bla* data and the resulting QSAR models were used to study the relationship between oxidative stress and liver damage. When qHTS ARE-*bla* data for a compound was not available, the

combinatorial QSAR models were used to fill-in the empty entries. This technique can be adapted to populate response profiles for other assays.

The workflow created in this study used data from PubChem, a publicly available big data source, to create and populate a bioassay response profile and revealed the relationship between oxidative stress and liver damage (Figure 4.1). Furthermore, the workflow in this study can be adapted to develop adverse outcome pathways (AOP) (Ankley et al. 2010). Our study identified a combination of molecular initiating events (MIE) (Allen et al. 2014) between some drugs and biomolecules that could cause the adverse outcome resulting in liver damage. The combination of drugs or compounds (*i.e.,* lipids) carrying fragments susceptible to free radical oxidation and fragments causing the inhibition of signaling pathways meant to alleviate or prevent oxidative stress can all lead to liver damage. These MIEs and their adverse outcome(s) are described in the following paragraphs and are illustrated in Figure 4.5.

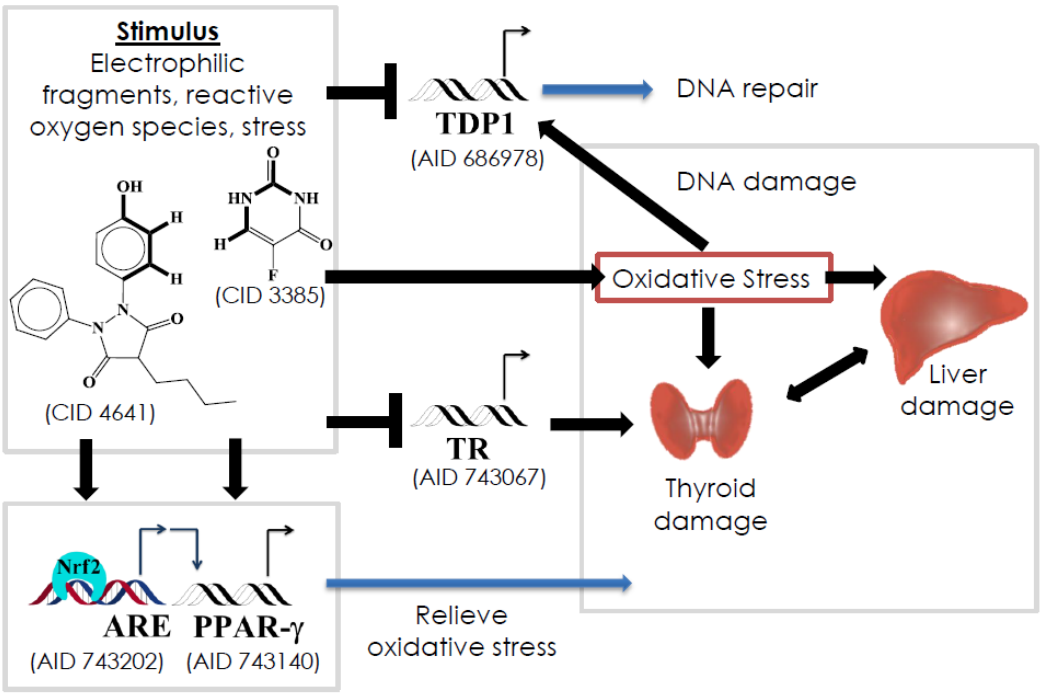


Figure 4.5. The potential liver toxicity mechanism of the compounds, like oxyphenbutazone (CID 4641) and 5-fluorouracil (CID 3385), that contain either of the proposed toxicophores A or B can generate reactive oxygen species. These types of stimuli activate the Antioxidant Response Element signaling pathway (ARE) (AID 743202) and peroxisome proliferator-activated receptor gamma signaling pathway (PPARg) (AID 743140), inhibit human tyrosyl-DNA phosphodiesterase 1 signaling pathway (TDP1) (686978), or disrupt the thyroid receptor signaling pathway (TR) (AID 743067).

The assay AID 686978 identifies inhibitors of human TDP1. TDP1 is an enzyme that repairs single-stranded DNA breaks covalently linked to topoisomerase I. It is known that mutations in TDP1 impair the ability of a cell to repair DNA damaged by oxidation or drugs (Ben Hassine and Arcangioli 2009). When DNA is damaged and TDP1 is inhibited, topoisomerase I stays covalently linked to the DNA during replication and the cell dies (Pouliot et al. 1999). Since the ARE pathway contains a considerable

number of detoxifying genes, it acts as the first line of defense to prevent DNA damage from oxidation or drugs (Kwak et al. 2003).

For AID 743067, active compounds in this assay act as TR antagonist and can disrupt metabolic homeostasis by inhibiting the binding of the thyroid hormone (Jameson and Weetman 2012). The liver plays a major role in thyroid hormone metabolism and liver damage is often associated with thyroid diseases (Huang and Liaw 1995). Furthermore, the liver metabolizes lipids and thyroid hormones regulate hepatic lipid homeostasis (Malik and Hodgson 2002). Lipids autoxidize in the presence of molecular oxygen, a process known as lipid peroxidation (Porter et al. 1995), which forms free radicals and ROS. Normally the ARE pathway will participate in the process of inactivation of ROS (Shukla et al. 2012). Failure to terminate ROS results in oxidative stress (Sies 1997), especially when a TR antagonist has disrupted liver lipid metabolism.

The assay AID 743140 identifies PPARg agonists that activate the PPAR response elements and in this specific case it regulates adipogenesis (Tontonoz et al. 1994). Adipose tissue, especially visceral adipose tissue, releases fatty acids directly into the liver *via* the hepatic portal vein (Lafontan and Girard 2008). Fatty acids are susceptible to lipid peroxidation. Disrupting PPARg and adipogenesis could put the liver at risk for oxidative stress when fatty acids are in excess.

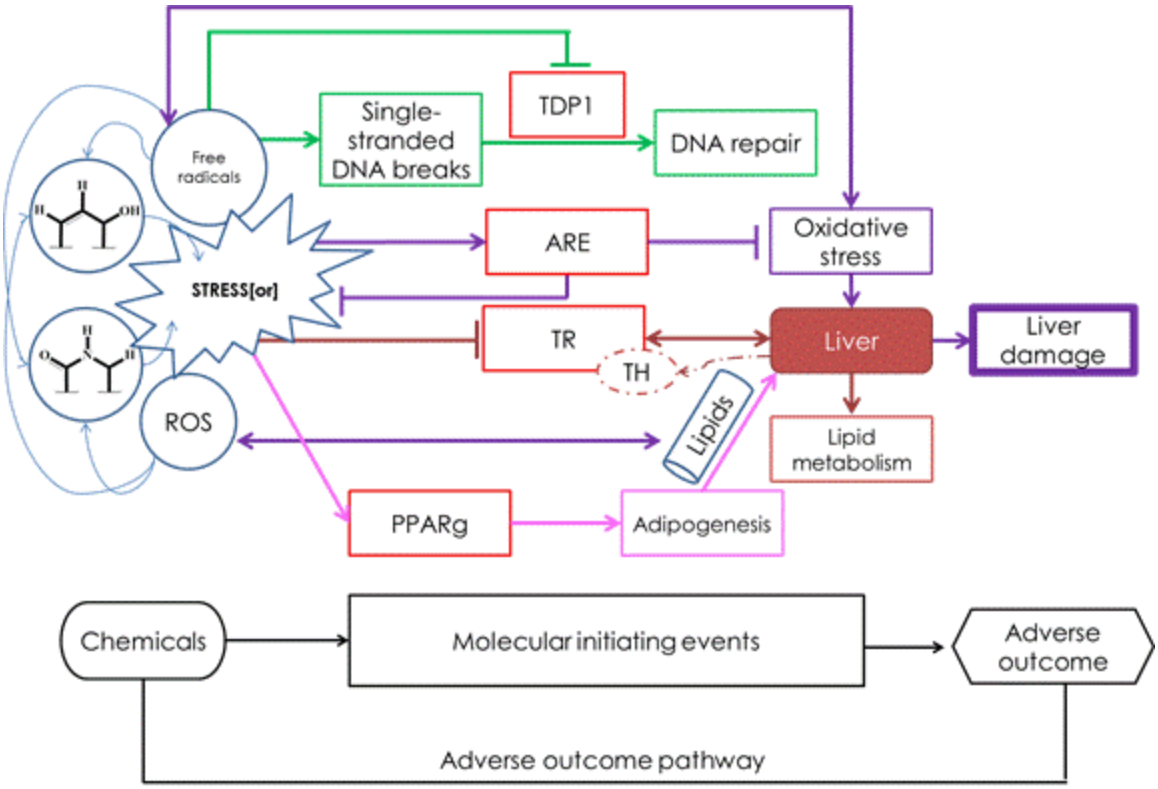


Figure 4.6. The proposed AOP of liver damage. Chemicals with potential toxicophores A and B can generate reactive oxygen species and activate the antioxidant response element signaling pathway (ARE) and peroxisome proliferator-activated receptor gamma signaling pathway (PPARg), inhibit human tyrosyl-DNA phosphodiesterase 1 signaling pathway (TDP1) (686978), or disrupt the thyroid receptor signaling pathway (TR). These interactions represent the molecular initiating events that eventually lead to the adverse outcome liver damage.

The AOP concept was presented as a logical sequence of biological responses that is useful for understanding complex toxicity phenomena (Allen et al. 2014; Ankley et al. 2010). Based on the AOP concept, Allen et al. discussed a unified MIE definition for the AOP framework for risk assessment purposes (Allen et al. 2014). This kind of research classifies compounds by mode of action using *in vitro* methods. Therefore, the chemical *in vitro-in vivo* relationships identified in this study can also be integrated into the AOP

framework of liver damage (Figure 4.6). Potential toxicophore A is an electrophilic fragment highly susceptible to free radical oxidation, due to its allylic hydrogen (Porter et al. 1995). It represents a key chemical property of potential toxicants in an AOP framework. For example, oxyphenbutazone (CID 4641) is known for causing liver damage (Gaisford 1962). It contains potential toxicophore A and is active in AIDs 686978 and 743202 as a TDP1 inhibitor and ARE agonist, respectively. The bioassay results can be viewed as the macro-molecular interactions and the RA value can be considered as a specific cellular response pathway perturbation score (*i.e.,* ARE signaling pathway perturbation and TDP1 inhibition) of AOP for this compound. The molecular mechanism by which oxyphenbutazone causes liver damage is still not clear (Gaisford 1962; Tai 2012). However, it is well established that it is a lipid soluble drug metabolized by liver microsomal enzymes and requires molecular oxygen to metabolize (Davies and Thorgeirsson 1971). Similarly, potential toxicophore B is known as N-methylformamide, a well-known liver toxicant susceptible to free radical oxidation by C-H abstraction from alkyl group(s) adjacent to the nitrogen atom (Borduas et al. 2015). This reaction produces methyl isocyanate, which is highly toxic (Varma 1987). For example, 5-fluorouracil (CID 3385) contains toxicophore B. 5-fluorouracil was shown to be active in both AIDs 686978 and 743067, TDP1 inhibitor and TR antagonist, respectively. If administered orally, 5-fluorouracil is metabolically degraded predominantly in the liver by dihydropyrimidine dehydrogenase (DPD) (Omura 2003). Patients that lack DPD are highly likely to experience liver damage (Chabner et al. 2011). In our current study, it is noticeable that the four major components of an AOP (as defined by Ankley et al. 2010) are included: chemical properties of toxicants, macro-

molecular interactions, cellular responses, and organ responses. Our future study will focus on the AOP framework of liver damage by differentiating the hepatotoxicity mechanisms of liver damage (*e.g.,* acute hepatic failure, cytolytic hepatitis, hepatic necrosis) (Zhu and Kruhlak 2014).

Our findings suggest that the four assays (686978, 743067, 743140, and 743202) could be used to screen for compounds that cause oxidative stress and induce liver damage. When specific chemical features (*e.g.,* potential toxicophores A and B) are present, the active responses obtained from these bioassays suggest potential hepatotoxicity. Although the four assays have covered several important mechanisms of oxidative stress, the negative results from all four assays would not be sufficient to indicate that a chemical is not hepatotoxic. Future work on this project includes the validation of these assays for their predictivity of liver damage, which will be used to optimize predictive liver toxicity models.

Conclusion

A workflow that identified potential assays from a public big data source for the evaluation of liver damage caused by oxidative stress was developed. Although using four assays will not be enough to cover all the relevant toxicity mechanisms of liver damage, this work clearly indicates the benefits of searching for useful toxicity data in the public big data domain for the compounds of interest. The increase in false positives in the IVIVC analysis indicates that the bioassay data is still needed for the compounds out of AD (*e.g.,* drug-like compounds). This issue could be resolved by rational design of the HTS chemical library that covers all the chemical space. New compounds

containing the potential toxicophores can be tested using these four assays to assess the potential liver damage caused by oxidative stress prior to animal testing.

The workflow developed in this study can be easily adapted to study the relationship between any bioassay and other *in vivo* exposure data to evaluate complex *in vitro-in vivo* relationships and reveal toxicity mechanisms. Future directions of *in silico* modeling of animal toxicity induced by drugs and oxidative stress could include pharmacology studies.

Appendix

Table A1. Curated oral bioavailability (%F) data set

No	Name	%F	logK(%F)	Category	Updated SMILES
1	3-Ketodesogestrel	76	0.5006	1	OC1(CCC2C3C(C4C(=CC(=O)CC4)CC3)C(CC12CC)=C)C#C
2	Abacavir	83	0.6886	1	OCC1CC(n2c3nc(nc(NC4CC4)c3nc2)N)C=C1
3	Abecarnil	92	1.0607	1	O(Cc1ccccc1)C=1C=CC2=NC=3C(=C2C=1)C(COC)=C(NC=3)C(OC(C)C)=O
4	Acadesine	10	-0.9542	0	NC(=O)c1ncn(C2OC(CO)C(O)C2O)c1N
5	Acamprosate	11	-0.9080	0	S(O)(=O)(=O)CCCNC(=O)C
6	Acarbose	2	-1.6902	0	O1C(C)C(NC2C=C(CO)C(O)C(O)C2O)C(O)C(O)C1OC1C(O)C(O)C(OC1CO)OC1C(O)C(O)C(OC1CO)O
7	Acebutolol	37	-0.2311	0	O(CC(O)CNC(C)C)c1ccc(NC(=O)CCC)cc1C(=O)C
8	Acenocoumarol	60	0.1761	1	O1c2c(cccc2)C(O)=C(C(CC(=O)C)c2ccc([N+](=O)[O-])cc2)C1=O
9	Acepromazine	55	0.0872	1	S1c2c(N(c3c1cccc3)CCCN(C)C)cc(cc2)C(=O)C
10	Acetaminophen	88	0.8653	1	Oc1ccc(NC(=O)C)cc1
11	Acetazolamide	99	1.9956	1	s1c(nnc1S(=O)(=O)N)NC(=O)C
12	Acetohydroxamic Acid	55	0.0872	1	O=C(NO)C
13	Acetylcarnitine	10	-0.9542	0	CC(=O)OC(CC(=O)[O-])C[N+](C)(C)C
14	Acetylcysteine	5	-1.2788	0	SCC(NC(=O)C)C(O)=O
15	Acetyldigitoxin	70	0.3680	1	O1C(C)C(OC2OC(C)C(O)C(OC(=O)C)C2)C(O)CC1OC1C(OC(OC2CC3CCC4C(CCC5(C)C(CCC45O)C4=CC(OC4)=O)C3(CC2)C)CC1O)C
16	Aciclovir	22.5	-0.537	0	O=C1NC(=Nc2n(cnc12)COCCO)N

			1		
17	Acipimox	90	0.9542	1	OC(=O)c1ncc([n+]([O-])c1)C
18	Acitretin	59	0.1581	1	O(C)c1cc(C)c(C=CC(=CC=CC(=CC(O)=O)C)C)c(C)c1C
19	Acrivastine	18	-0.6585	0	OC(=O)C=Cc1nc(ccc1)C(=CCN1CCCC1)c1ccc(cc1)C
20	Adefovir	12	-0.8653	0	P(O)(O)(=O)COCCn1c2ncnc(N)c2nc1
21	Adefovir Dipivoxil	12	-0.8653	0	CC(C)(C)C(=O)OCOP(=O)(COCCn1cnc2c(N)ncnc12)OCOC(=O)C(C)(C)C
22	Adinazolam	39	-0.1943	0	Clc1cc2c(-n3c(nnc3CN(C)C)CN=C2c2ccccc2)cc1
23	Alafosfalin	50	0.0000	1	P(O)(O)(=O)C(NC(=O)C(N)C)C
24	Albendazole	5	-1.2788	0	S(CCC)c1cc2nc([nH]c2cc1)NC(OC)=O
25	Alcuronium	0	-2.0000	0	OCC=C1C2C=3C4N(C=C5C6CC7[N+](CC6=CCO)(CCC67C5N(C=3)c3c6cccc3)CC=C)c3c(C45C([N+](C1)(CC5)CC=C)C2)cccc3
26	Alendronate	1	-1.9956	0	P(O)(O)(=O)C(P(O)(O)=O)(O)CCCN
27	Alfacalcidol	71	0.3889	1	OC1CC(O)CC(=CC=C2C3CCC(C(CCCC(C)C)C)C3(CCC2)C)C1=C
28	Alfentanil	1	-1.9956	0	CCN1N=NN(CCN2CCC(CC2)(COC)N(C(=O)CC)c3ccccc3)C1=O
29	Alfuzosin	55	0.0872	1	O1CCCC1C(=O)NCCCN(C)c1nc(N)c2cc(OC)c(OC)cc2n1
30	Aliskiren	2.5	-1.5911	0	O(CCCOC)c1cc(ccc1OC)CC(C(C)C)CC(N)C(O)CC(C(C)C)C(=O)NCC(C(=O)N)(C)C
31	Alizapride	84	0.7202	1	O(C)c1cc2[nH]nnc2cc1C(=O)NCC1N(CCC1)CC=C
32	Allopurinol	53	0.0522	1	O=C1NC=Nc2[nH]ncc12
33	Almitrine	63	0.2311	1	Fc1ccc(cc1)C(N1CCN(CC1)c1nc(nc(n1)NCC=C)NCC=C)c1ccc(F)cc1
34	Almotriptan	70	0.3680	1	S(=O)(=O)(N1CCCC1)Cc1cc2c([nH]cc2CCN(C)C)cc1

35	Alosetron	55	0.0872	1	O=C1N(CCc2n(c3c(c12)cccc3)C)Cc1nc[nH]c1C
36	Alprazolam	88	0.8653	1	Clc1cc2c(-n3c(nnc3C)CN=C2c2ccccc2)cc1
37	Alprenolol	8	-1.0607	0	O(CC(O)CNC(C)C)c1ccccc1CC=C
38	Alprostadil	15	-0.7533	0	OC1CC(=O)C(CCCCCCC(O)=O)C1C=CC(O)CCCCC
39	Alvimopan	6	-1.1950	0	Oc1cc(ccc1)C1(CCN(CC1C)CC(Cc1ccccc1)C(=O)NCC(O)=O)C
40	Amantadine	90	0.9542	1	NC12CC3CC(C1)CC(C2)C3
41	Ambroxol	77	0.5248	1	Brc1cc(Br)cc(CNC2CCC(O)CC2)c1N
42	Amdinocillin	5	-1.2788	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2N=CN1CCCCCC1
43	Amdinocillin_pivoxil	0	-2.0000	0	S1C2N(C(C(OCOC(=O)C(C)(C)C)=O)C1(C)C)C(=O)C2N=CN1CCCCCC1
44	Amezinium	58	0.1402	1	O(C)c1[n+](ncc(N)c1)-c1ccccc1
45	Amifostine	99	1.9956	1	S(P(O)(O)=O)CCNCCCN
46	Amikacin	0	-2.0000	0	O1C(CN)C(O)C(O)C(O)C1OC1C(O)C(OC2OC(CO)C(O)C(N)C2O)C(NC(=O)C(O)CCN)CC1N
47	Amiloride	50	0.0000	1	Clc1nc(C(=O)N=C(N)N)c(nc1N)N
48	Aminocamptothecin	49	-0.0174	0	OC1(C2=C(CCC1=O)C(=O)N1C(=C2)c2nc3c(cc2C1)c(N)ccc3)CC
49	Aminoglutethimide	95	1.2788	1	O=C1NC(=O)CCC1(CC)c1ccc(N)cc1
50	Aminolevulinic_acid	60	0.1761	1	OC(=O)CCC(=O)CN
51	Amiprilose	63	0.2311	1	O1C2OC(OC2C(OCCCN(C)C)C1C(O)CO)(C)C
52	Amisulpride	48	-0.034	0	S(=O)(=O)(CC)c1cc(C(=O)NCC2N(CCC2)CC)c(OC)cc1N

			8		
53	Amitriptyline	48	-0.0348	0	N(CCC=C1c2c(CCc3c1cccc3)cccc2)(C)C
54	Amlodipine	74	0.4543	1	Clc1ccccc1C1C(C(OCC)=O)=C(NC(C)=C1C(OC)=O)COCCN
55	Amobarbital	95	1.2788	1	CCC1(CCC(C)C)C(=O)NC(=O)NC1=O
56	Amodiaquine	75	0.4771	1	Clc1cc2nccc(Nc3cc(CN(CC)CC)c(O)cc3)c2cc1
57	Amosulalol	99	1.9956	1	S(=O)(=O)(N)c1cc(ccc1C(O)CNCCOc1ccccc1OC)C
58	Amoxapine	36	-0.2499	0	Clc1cc2c(Oc3c(N=C2N2CCNCC2)cccc3)cc1
59	Amoxicillin	93	1.1234	1	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(N)c1ccc(O)cc1
60	Amphotericin_B	3	-1.5097	0	CC1C=CC=CC=CC=CC=CC=CC(CC2C(C(CC(O2)(CC(CC(C(CCC(CC(CC(=O)OC(C(C1O)C)C)O)O)O)O)O)O)O)C(=O)O)OC3C(C(C(C(O3)C)O)N)O
61	Ampicillin	62	0.2126	1	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(N)c1ccccc1
62	Ampiroxicam	0	-2.0000	0	S1(=O)(=O)N(C)C(C(=O)Nc2ncccc2)=C(OC(OC(OCC)=O)C)c2c1cccc2
63	Amrinone	93	1.1234	1	O=C1NC=C(C=C1N)c1ccncc1
64	Amsalog	34	-0.2881	0	S(=O)(=O)(Nc1cc(OC)c(Nc2c3c(nc4c2cccc4C)c(ccc3)C(=O)NC)cc1)C
65	Anagrelide	75	0.4771	1	Clc1c2CN3CC(=O)NC3=Nc2ccc1Cl
66	Anastrozole	80	0.6021	1	n1cn(nc1)Cc1cc(cc(c1)C(C#N)(C)C)C(C#N)(C)C
67	Anisotropine Methylbromide	17.5	-0.6734	0	O(C(=O)C(CCC)CCC)C1CC2[N+](C(C1)CC2)(C)C
68	Antipyrine	99	1.9956	1	O=C1N(N(C)C(=C1)C)c1ccccc1
69	Antrafenine	99	1.9956	1	FC(F)(F)c1cc(N2CCN(CC2)CCOC(=O)c2ccccc2Nc2c3c(ncc2)cc(cc3)C(F)(F)F)ccc1

70	Apomorphine	10	-0.9542	0	Oc1c-2c(CC3N(CCc4c3c-2ccc4)C)ccc1O
71	Aprepitant	62.5	0.2218	1	Fc1ccc(cc1)C1N(CCOC1OC(C)c1cc(cc(c1)C(F)(F)F)C(F)(F)F)CC1=NC(=O)NN1
72	Aprindine	85	0.7533	1	N(CCCN(CC)CC)(C1Cc2c(C1)cccc2)c1ccccc1
73	Arbekacin	0	-2.0000	0	O1C(CO)C(O)C(N)C(O)C1OC1C(O)C(OC2OC(CCC2N)CN)C(N)CC1NC(=O)C(O)CCN
74	Aripiprazole	87	0.8256	1	Clc1c(N2CCN(CC2)CCCCOc2cc3NC(=O)CCc3cc2)cccc1Cl
75	Artemether	58	0.1402	1	O1C2OC3(OOC24C(CCC(C4CC3)C)C(C)C1OC)C
76	Astemizole	2	-1.6902	0	COc1ccc(CCN2CCC(CC2)Nc3nc4ccccc4n3Cc5ccc(F)cc5)cc1
77	Atazanavir	64	0.2499	1	O(C(=O)NC(C(C)(C)C)C(=O)NC(Cc1ccccc1)C(O)CN(NC(=O)C(NC(OC)=O)C(C)(C)C)Cc1ccc(cc1)-c1ncccc1)C
78	Atenolol	56	0.1047	1	O(CC(O)CNC(C)C)c1ccc(cc1)CC(=O)N
79	Atomoxetine	94	1.1950	1	O(C(CCNC)c1ccccc1)c1ccccc1C
80	Atovaquone	23	-0.5248	0	Clc1ccc(cc1)C1CCC(CC1)C1=C(O)C(=O)c2c(cccc2)C1=O
81	Atropine (DL)	50	0.0000	1	O(C(=O)C(CO)c1ccccc1)C1CC2N(C(C1)CC2)C
82	Avitriptan	17.2	-0.6825	0	S(=O)(=O)(NC)Cc1cc2c([nH]cc2CCCN2CCN(CC2)c2nccncc2OC)cc1
83	Azathioprine	60	0.1761	1	S(c1n(cnc1[N+](=O)[O-])C)c1ncnc2nc[nH]c12
84	Azelastine	99	1.9956	1	Clc1ccc(cc1)CC1=NN(C2CCCN(CC2)C)C(=O)c2c1cccc2
85	azidocillin	60	0.1761	1	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)N=[N+]=[N-])C(=O)O)C
86	Azithromycin	34	-0.2881	0	O1C(CC)C(O)(C)C(O)C(N(CC(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C)C)C
87	Azlocillin	0	-2.0000	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(NC(=O)N1CCNC1=O)c1ccccc1
88	Azosemide	10	-0.954	0	Clc1cc(NCc2sccc2)c(cc1S(=O)(=O)N)-c1[nH]nnn1

			2		
89	Aztreonam	1	-1.9956	0	s1cc(nc1N)C(=NOC(C(O)=O)(C)C)C(=O)NC1C(N(S(O)(=O)=O)C1=O)C
90	Baclofen	90	0.9542	1	Clc1ccc(cc1)C(CC(O)=O)CN
91	Balsalazide	1	-1.9956	0	O=C1C=CC(=NNc2ccc(cc2)C(=O)NCCC(O)=O)C=C1C(O)=O
92	Bambuterol	15	-0.7533	0	O(C(=O)N(C)C)c1cc(cc(OC(=O)N(C)C)c1)C(O)CNC(C)(C)C
93	Barnidipine	1	-1.9956	0	CC1=C(C(C(=C(N1)C)C(=O)OC2CCN(C2)CC3=CC=CC=C3)C4=CC(=CC=C4)[N+](=O)[O-])C(=O)OC
94	Benazepril	18	-0.6585	0	O=C1N(c2c(CCC1NC(CCc1ccccc1)C(OCC)=O)cccc2)CC(O)=O
95	Bencyclane	51	0.0174	1	O(CCCN(C)C)C1(CCCCCC1)Cc1ccccc1
96	Bendroflumethiazide	90	0.9542	1	S(=O)(=O)(N)c1cc2S(=O)(=O)NC(Nc2cc1C(F)(F)F)Cc1ccccc1
97	Benperidol	40	-0.1761	0	Fc1ccc(cc1)C(=O)CCCN1CCC(N2c3c(NC2=O)cccc3)CC1
98	Benzquinamide	36	-0.2499	0	O(C(=O)C)C1CC2N(CC1C(=O)N(CC)CC)CCc1cc(OC)c(OC)cc12
99	Benzthiazide	25	-0.4771	0	Clc1cc2N=CN(S(=O)(=O)c2cc1S(=O)(=O)N)CSCc1ccccc1
100	Benztropine	29	-0.3889	0	O(C(c1ccccc1)c1ccccc1)C1CC2N(C(C1)CC2)C
101	Benzydamine	87	0.8256	1	O(CCCN(C)C)c1nn(c2c1cccc2)Cc1ccccc1
102	Bepridil	61	0.1943	1	O(CC(C)C)CC(N(Cc1ccccc1)c1ccccc1)CN1CCCC1
103	Beta-carotene	25	-0.4771	0	C1CCC(C)=C(C=CC(=CC=CC(=CC=CC=C(C=CC=C(C=CC=2C(CCCC=2C)(C)C)C)C)C)C)C1(C)C
104	Betaxolol	85	0.7533	1	O(CC(O)CNC(C)C)c1ccc(cc1)CCOCC1CC1
105	Bevantolol	57	0.1224	1	O(C)c1cc(ccc1OC)CCNCC(O)COc1cc(ccc1)C

106	Bezafibrate	99	1.9956	1	Clc1ccc(cc1)C(=O)NCCc1ccc(OC(C(O)=O)(C)C)cc1
107	Bifemelane	13	-0.8256	0	O(CCCCNC)c1ccccc1Cc1ccccc1
108	Biotin	99	1.9956	1	S1CC2NC(=O)NC2C1CCCCC(O)=O
109	Biperiden	30	-0.3680	0	OC(CCN1CCCCC1)(C1C2CC(C1)C=C2)c1ccccc1
110	Bisacodyl	15	-0.7533	0	O(C(=O)C)c1ccc(cc1)C(c1ccc(OC(=O)C)cc1)c1ncccc1
111	Bisoprolol	90	0.9542	1	O(CC(O)CNC(C)C)c1ccc(cc1)COCCOC(C)C
112	Bosentan	50	0.0000	1	S(=O)(=O)(Nc1nc(nc(OCCO)c1Oc1ccccc1OC)-c1ncccn1)c1ccc(cc1)C(C)(C)C
113	BRETYLIUM	23	-0.5248	0	Brc1ccccc1C[N+](CC)(C)C
114	Bromazepam	84	0.7202	1	Brc1cc2c(NC(=O)CN=C2c2ncccc2)cc1
115	Bromfenac	67	0.3076	1	Brc1ccc(cc1)C(=O)c1cccc(CC(O)=O)c1N
116	Bromhexine	77.5	0.5371	1	Brc1cc(Br)cc(CN(C)C2CCCCC2)c1N
117	Bromocriptine	28	-0.4102	0	CC(C)CC1C(=O)N2CCCC2C3(N1C(=O)C(O3)(C(C)C)NC(=O)C4CN(C5CC6=C(NC7=CC=CC(=C67)C5=C4)Br)C)O
118	Bromopride	67.5	0.3174	1	Brc1cc(C(=O)NCCN(CC)CC)c(OC)cc1N
119	Brompheniramine	89	0.9080	1	Brc1ccc(cc1)C(CCN(C)C)c1ncccc1
120	Brotizolam	70	0.3680	1	Brc1sc2-n3c(nnc3C)CN=C(c2c1)c1ccccc1Cl
121	Bucindolol	30	-0.3680	0	O(CC(O)CNC(Cc1c2c([nH]c1)cccc2)(C)C)c1ccccc1C#N
122	Budesonide	37	-0.2311	0	O1C2(C(OC1CCC)CC1C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)C(=O)CO
123	Budipine	47	-0.0522	0	N1(CCC(CC1)(c1ccccc1)c1ccccc1)C(C)(C)C
124	Buflomedil	65	0.2688	1	O(C)c1cc(OC)cc(OC)c1C(=O)CCCN1CCCC1

125	Bufuralol	46	-0.0696	0	CCC1=CC=CC2=C1OC(=C2)C(CNC(C)(C)C)O
126	Bumetanide	90	0.9542	1	S(=O)(=O)(N)c1cc(cc(NCCCC)c1Oc1ccccc1)C(O)=O
127	Bupivacaine	0	-2.0000	0	O=C(Nc1c(cccc1C)C)C1N(CCCC1)CCCC
128	Buprenorphine	28	-0.4102	0	O1C2C34C5(CC(C(O)(C(C)(C)C)C)C2(OC)CC5)C(N(CC3)CC2CC2)Cc2c4c1c(O)cc2
129	Buspirone	3.9	-1.3917	0	O=C1N(CCCCN2CCN(CC2)c2ncccn2)C(=O)CC2(C1)CCCC2
130	Busulfan	70	0.3680	1	S(OCCCCOS(=O)(=O)C)(=O)(=O)C
131	Butalbital	85	0.7533	1	O=C1NC(=O)NC(=O)C1(CC(C)C)CC=C
132	Butorphanol	17	-0.6886	0	OC12C3(CCCC1)CCN(C2Cc1c3cc(O)cc1)CC1CCC1
133	Butylscopolamine	1	-1.9956	0	O1C2C3[N+](C(CC(OC(=O)C(CO)c4ccccc4)C3)C12)(CCCC)C
134	Cabergoline	65	0.2688	1	O=C(N(CCCN(C)C)C(=O)NCC)C1CC2C(N(C1)CC=C)Cc1c3c2cccc3[nH]c1
135	Caffeine	98	1.6902	1	O=C1N(C)C(=O)N(c2ncn(c12)C)C
136	Calcipotriol	5.5	-1.2351	0	OC1CC(O)CC(=CC=C2C3CCC(C(C=CC(O)C4CC4)C)C3(CCC2)C)C1=C
137	Calcitriol	61	0.1943	1	OC1CC(O)CC(=CC=C2C3CCC(C(CCCC(O)(C)C)C)C3(CCC2)C)C1=C
138	Capecitabine	99	1.9956	1	FC1=CN(C2OC(C)C(O)C2O)C(=O)N=C1NC(OCCCCC)=O
139	Captopril	69	0.3475	1	SCC(C(=O)N1CCCC1C(O)=O)C
140	Carbamazepine	70	0.3680	1	O=C(N)N1c2c(C=Cc3c1cccc3)cccc2
141	Carbenicillin	35	-0.2688	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(C(O)=O)c1ccccc1
142	Carbimazole	0	-2.0000	0	S=C1N(C=CN1C)C(OCC)=O
143	Carmustine	11.5	-0.8862	0	ClCCN(N=O)C(=O)NCCCl

144	Carteolol	88	0.8653	1	O(CC(O)CNC(C)(C)C)c1c2CCC(=O)Nc2ccc1
145	Carvedilol (racemic)	25	-0.4771	0	O(CCNCC(OCc1c2c3c([nH]c2ccc1)cccc3)O)c1ccccc1OC
146	Cefacetrile	0	-2.0000	0	S1C2N(C(=O)C2NC(=O)CC#N)C(C(O)=O)=C(C1)COC(=O)C
147	Cefaclor	90	0.9542	1	ClC=1CSC2N(C(=O)C2NC(=O)C(N)c2ccccc2)C=1C(O)=O
148	Cefadroxil	95	1.2788	1	S1C2N(C(=O)C2NC(=O)C(N)c2ccc(O)cc2)C(C(O)=O)=C(C1)C
149	Cefamandole	96	1.3802	1	Cn1nnnc1SCC2=C(N3C(SC2)C(NC(=O)C(O)c4ccccc4)C3=O)C(=O)O
150	Cefatrizine	66	0.2881	1	S1C2N(C(=O)C2NC(=O)C(N)c2ccc(O)cc2)C(C(O)=O)=C(C1)CSc1nn[nH]c1
151	cefazolin	0	-2.0000	0	s1c(nnc1SCC=1CSC2N(C(=O)C2NC(=O)Cn2nnnc2)C=1C(O)=O)C
152	cefcanel daloxate	40	-0.1761	0	s1c(nnc1SCC=1CSC2N(C(=O)C2NC(=O)C(OC(=O)C(N)C)c2ccccc2)C=1C(OCC=1OC(OC=1C)=O)=O)C
153	Cefdinir	18.5	-0.6440	0	C=CC1=C(N2C(C(C2=O)NC(=O)C(=NO)C3=CSC(=N3)N)SC1)C(=O)O
154	Cefditoren	14	-0.7884	0	s1cnc(C)c1C=CC=1CSC2N(C(=O)C2NC(=O)C(=NOC)c2nc(sc2)N)C=1C(O)=O
155	Cefepime	99	1.9956	1	s1cc(nc1N)C(=NOC)C(=O)NC1C2SCC(C[N+]3(CCCC3)C)=C(N2C1=O)C(O)=O
156	Cefetamet	47	-0.0522	0	s1cc(nc1N)C(=NOC)C(=O)NC1C2SCC(C)=C(N2C1=O)C(O)=O
157	Cefixime	47	-0.0522	0	s1cc(nc1N)C(=NOCC(O)=O)C(=O)NC1C2SCC(C=C)=C(N2C1=O)C(O)=O
158	Cefodizime	0	-2.0000	0	s1c(CC(O)=O)c(nc1SCC=1CSC2N(C(=O)C2NC(=O)C(=NOC)c2nc(sc2)N)C=1C(O)=O)C
159	Cefoperazone	10	-0.9542	0	S1C2N(C(=O)C2NC(=O)C(NC(=O)N2CCN(CC)C(=O)C2=O)c2ccc(O)cc2)C(C(O)=O)=C(C1)CSc1nnnn1C
160	Cefoxitin	0	-2.0000	0	COC1(C2N(C1=O)C(=C(CS2)COC(=O)N)C(=O)O)NC(=O)CC3=CC=CS3

161	Cefpodoxime Proxetil	52	0.0348	1	s1cc(nc1N)C(=NOC)C(=O)NC1C2SCC(COC)=C(N2C1=O)C(OC(OC(OC(C)C)=O)C)=O
162	Cefprozil	90	0.9542	1	S1C2N(C(=O)C2NC(=O)C(N)c2ccc(O)cc2)C(C(O)=O)=C(C1)C=CC
163	Ceftazidime	0	-2.0000	0	s1cc(nc1N)C(=NOC(C(O)=O)(C)C)C(=O)NC1C2SCC(C[n+]3ccccc3)=C(N2C1=O)C(O)=O
164	Ceftibuten	82.5	0.6734	1	s1cc(nc1N)C(=CCC(O)=O)C(=O)NC1C2SCC=C(N2C1=O)C(O)=O
165	Ceftizoxime	0	-2.0000	0	s1cc(nc1N)C(=NOC)C(=O)NC1C2SCC=C(N2C1=O)C(O)=O
166	Ceftriaxone	0	-2.0000	0	CON=C(C(=O)NC1C2SCC(=C(N2C1=O)C(=O)O)CSC3=NC(=O)C(=O)NN3C)c4csc(N)n4
167	Cefuroxime	32	-0.3274	0	S1C2N(C(=O)C2NC(=O)C(=NOC)c2occc2)C(C(O)=O)=C(C1)COC(=O)N
168	Celiprolol	40	-0.1761	0	O(CC(O)CNC(C)(C)C)c1ccc(NC(=O)N(CC)CC)cc1C(=O)C
169	Cephalexin	90	0.9542	1	S1C2N(C(=O)C2NC(=O)C(N)c2ccccc2)C(C(O)=O)=C(C1)C
170	Cephalothin	0	-2.0000	0	s1cccc1CC(=O)NC1C2SCC(COC(=O)C)=C(N2C1=O)C(O)=O
171	Cephapirin	0	-2.0000	0	S1C2N(C(=O)C2NC(=O)CSc2ccncc2)C(C(O)=O)=C(C1)COC(=O)C
172	Cephradine	94	1.1950	1	S1C2N(C(=O)C2NC(=O)C(N)C=2CC=CCC=2)C(C(O)=O)=C(C1)C
173	Cerivastatin	60	0.1761	1	Fc1ccc(cc1)-c1c(COC)c(nc(C(C)C)c1C=CC(O)CC(O)CC(O)=O)C(C)C
174	Cetirizine	90	0.9542	1	Clc1ccc(cc1)C(N1CCN(CC1)CCOCC(O)=O)c1ccccc1
175	Chlorambucil	82	0.6585	1	ClCCN(CCCl)c1ccc(cc1)CCCC(O)=O
176	Chloramphenicol	82	0.6585	1	ClC(Cl)C(=O)NC(C(O)c1ccc([N+](=O)[O-])cc1)CO
177	Chloramphenicol_palmitate	80	0.6021	1	ClC(Cl)C(=O)NC(C(O)c1ccc([N+](=O)[O-])cc1)COC(=O)CCCCCCCCCCCCCCC

178	Chlordiazepoxide	99	1.9956	1	Clc1cc2c(N=C(NC)C[N+]([O-])=C2c2ccccc2)cc1
179	Chlorhexidine	1	-1.9956	0	Clc1ccc(NC(=NC(=NCCCCCCCN=C(N=C(Nc2ccc(Cl)cc2)N)N)N)N)cc1
180	Chlormezanone	99	1.9956	1	Clc1ccc(cc1)C1S(=O)(=O)CCC(=O)N1C
181	Chloroquine	80	0.6021	1	Clc1cc2nccc(N(CCCCN(CC)CC)C)c2cc1
182	Chlorothiazide	13	-0.8256	0	Clc1cc2N=CNS(=O)(=O)c2cc1S(=O)(=O)N
183	Chloroxine	89	0.9080	1	Clc1cc(Cl)c2c(nccc2)c1O
184	Chlorpheniramine	41	-0.1581	0	Clc1ccc(N(CCN(C)C)c2ncccc2)cc1
185	Chlorpromazine	25	-0.4771	0	Clc1cc2N(c3c(Sc2cc1)cccc3)CCCN(C)C
186	CHLORPROPAMIDE	95	1.2788	1	Clc1ccc(S(=O)(=O)NC(=O)NCCC)cc1
187	Chlorprothixene	41	-0.1581	0	Clc1cc2c(Sc3c(cccc3)C2=CCCN(C)C)cc1
188	Chlortetracycline	30	-0.3680	0	CC1(C2CC3C(C(=O)C(=C(C3(C(=O)C2=C(C4=C(C=CC(=C41)Cl)O)O)O)O)C(=O)N)N(C)C)O
189	Cibenzoline	86	0.7884	1	N1CCN=C1C1CC1(c1ccccc1)c1ccccc1
190	Cicaprost	99	1.9956	1	OC1CC2C(CC(C2)=CCOCC(O)=O)C1C#CC(O)C(CC#CCC)C
191	Ciclesonide	1	-1.9956	0	O1C2(C(OC1C1CCCCC1)CC1C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)C(=O)COC(=O)C(C)C
192	Cicloprolol	99	1.9956	1	CC(C)NCC(COC1=CC=C(C=C1)OCCOCC2CC2)O
193	Cidofovir	5	-1.2788	0	P(O)(O)(=O)COC(CN1C=CC(=NC1=O)N)CO
194	CILAZAPRIL	60	0.1761	1	O=C1N2N(CCCC1NC(CCc1ccccc1)C(OCC)=O)CCCC2C(O)=O
195	Cilazaprilat	26	-0.4543	0	O=C1N2N(CCCC1NC(CCc1ccccc1)C(O)=O)CCCC2C(O)=O

196	Cilomilast	99	1.9956	1	O(c1cc(ccc1OC)C1(CCC(CC1)C(O)=O)C#N)C1CCCC1
197	Cimetidine	60	0.1761	1	S(Cc1nc[nH]c1C)CCNC(=NC)NC#N
198	Cimetropium	2	-1.6902	0	O1C2C3[N+](C(CC(OC(=O)C(CO)c4ccccc4)C3)C12)(CC1CC1)C
199	Cinacalcet	20	-0.6021	0	FC(F)(F)c1cc(ccc1)CCCNC(C)c1c2c(ccc1)cccc2
200	Cinolazepam	95	1.2788	1	Clc1cc2c(N(CCC#N)C(=O)C(O)N=C2c2ccccc2F)cc1
201	cinoxacin	72.5	0.4210	1	O1c2c(OC1)cc1N(N=C(C(O)=O)C(=O)c1c2)CC
202	Ciprofibrate	99	1.9956	1	ClC1(Cl)CC1c1ccc(OC(C(O)=O)(C)C)cc1
203	Ciprofloxacin	60	0.1761	1	Fc1cc2c(nc1N1CCCCC1)N(C=C(C(O)=O)C2=O)C1CC1
204	Cisapride	40	-0.1761	0	Clc1cc(C(=O)NC2CCN(CC2OC)CCCOc2ccc(F)cc2)c(OC)cc1N
205	Cladribine	45	-0.0872	0	Clc1nc(N)c2ncn(c2n1)C1OC(CO)C(O)C1
206	Clarithromycin	55	0.0872	1	O1C(CC)C(O)(C)C(O)C(C)C(=O)C(CC(OC)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C
207	Clavulanic Acid (Clavulanate)	75	0.4771	1	O1C2N(C(C(O)=O)C1=CCO)C(=O)C2
208	Clemastine	37	-0.2311	0	Clc1ccc(cc1)C(OCCC1N(CCC1)C)(C)c1ccccc1
209	Clenbuterol	99	1.9956	1	Clc1cc(cc(Cl)c1N)C(O)CNC(C)(C)C
210	Clinafloxacin	90	0.9542	1	Clc1c2N(C=C(C(O)=O)C(=O)c2cc(F)c1N1CC(N)CC1)C1CC1
211	Clindamycin	87	0.8256	1	ClC(C(NC(=O)C1N(CC(C1)CCC)C)C1OC(SC)C(O)C(O)C1O)C
212	Clobazam	90	0.9542	1	Clc1cc2N(C(=O)CC(=O)N(c2cc1)C)c1ccccc1
213	Clodronate	1	-1.9956	0	ClC(Cl)(P(O)(O)=O)P(O)(O)=O

214	Clofarabine	50	0.0000	1	Clc1nc(N)c2ncn(c2n1)C1OC(CO)C(O)C1F
215	Clofazimine	57	0.1224	1	Clc1ccc(N2C3=CC(=NC(C)C)C(Nc4ccc(Cl)cc4)=CC3=Nc3c2cccc3)cc1
216	Clofibrate	95	1.2788	1	Clc1ccc(OC(C(OCC)=O)(C)C)cc1
217	Clomethiazole	25	-0.4771	0	ClCCc1scnc1C
218	Clomifene	90	0.9542	1	ClC(=C(c1ccc(OCCN(CC)CC)cc1)c1ccccc1)c1ccccc1
219	Clomipramine	51	0.0174	1	Clc1cc2N(c3c(CCc2cc1)cccc3)CCCN(C)C
220	Clonazepam	98	1.6902	1	Clc1ccccc1C1=NCC(=O)Nc2c1cc([N+](=O)[O-])cc2
221	Clonidine	95	1.2788	1	Clc1cccc(Cl)c1N=C1NCCN1
222	Clopidogrel	50	0.0000	1	Clc1ccccc1C(N1CCc2sccc2C1)C(OC)=O
223	Clorazepate	91	1.0048	1	Clc1cc2c(NC(=O)C(N=C2c2ccccc2)C(O)=O)cc1
224	Clotrimazole	0	-2.0000	0	Clc1ccccc1C(n1ccnc1)(c1ccccc1)c1ccccc1
225	Cloxacillin	40	-0.1761	0	Clc1ccccc1-c1noc(C)c1C(=O)NC1C2SC(C)(C)C(N2C1=O)C(O)=O
226	Clozapine	55	0.0872	1	Clc1cc2N=C(N3CCN(CC3)C)c3c(Nc2cc1)cccc3
227	COCAINE	57	0.1224	1	O(C(=O)c1ccccc1)C1CC2N(C(CC2)C1C(OC)=O)C
228	Codeine	50	0.0000	1	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC)cc3)C=CC2O
229	Colchicine	50	0.0000	1	O(C)C1=CC=C2C(=CC1=O)C(NC(=O)C)CCc1c2c(OC)c(OC)c(OC)c1
230	Colistin	0	-2.0000	0	S(O)(=O)(=O)CNCCC1NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)CCCCC(CC)C)CCNCS(O)(=O)=O)C(O)C)CCNCS(O)(=O)=O)CCNC(=O)C(NC1=O)C(O)C)CCNCS(O)(=O)=O)CC(C)C)CC(C)C)CCNCS(O)(=O)=O
231	Conivaptan	42	-0.1402	0	O=C(N1CCc2nc([nH]c2-c2c1cccc2)C)c1ccc(NC(=O)c2ccccc2-c2ccccc2)cc1
232	Cotinine	97	1.5097	1	O=C1N(C)C(CC1)c1cccnc1

233	Coumarin	3	-1.5097	0	O1c2c(C=CC1=O)cccc2
234	Cyclobenzaprine	55	0.0872	1	N(CCC=C1c2c(C=Cc3c1cccc3)cccc2)(C)C
235	Cyclopenthiazide	99	1.9956	1	Clc1cc2NC(NS(=O)(=O)c2cc1S(=O)(=O)N)CC1CCCC1
236	Cyclophosphamide	74	0.4543	1	ClCCN(P1(OCCCN1)=O)CCCl
237	Cycloserine	80	0.6021	1	O1NC(=O)C(N)C1
238	Cyclosporine	30	-0.3680	0	O=C1N(C)C(C(O)C(CC=CC)C)C(=O)NC(CC)C(=O)N(CC(=O)N(C)C(CC(C)C)C(=O)NC(C(C)C)C(=O)N(C)C(CC(C)C)C(=O)NC(C)C(=O)NC(C)C(=O)N(C)C(CC(C)C)C(=O)N(C)C(CC(C)C)C(=O)N(C)C1C(C)C)C
239	Cyproterone	99	1.9956	1	ClC=1C2=CC(=O)C3C(C3)C2(C2C(C3CCC(O)(C(=O)C)C3(CC2)C)C=1)C
240	Cytarabine	20	-0.6021	0	O1C(CO)C(O)C(O)C1N1C=CC(=NC1=O)N
241	Dantrolene	70	0.3680	1	o1c(ccc1C=NN1CC(=O)NC1=O)-c1ccc([N+](=O)[O-])cc1
242	Dapsone	93	1.1234	1	S(=O)(=O)(c1ccc(N)cc1)c1ccc(N)cc1
243	Daptomycin	0	-2.0000	0	O1C(C)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)CCCCCCCCCC)Cc2c3c([nH]c2)cccc3)CC(=O)N)CC(O)=O)C(=O)NCC(=O)NC(CCCN)C(=O)NC(CC(O)=O)C(=O)NC(C)C(=O)NC(CC(O)=O)C(=O)NCC(=O)NC(CO)C(=O)NC(C(CC(O)=O)C)C(=O)NC(CC(=O)c2ccccc2N)C1=O
244	Darifenacin	17	-0.6886	0	O1CCc2cc(ccc12)CCN1CC(CC1)C(C(=O)N)(c1ccccc1)c1ccccc1
245	Darunavir	37	-0.2311	0	S(=O)(=O)(N(CC(C)C)CC(O)C(NC(OC1C2CCOC2OC1)=O)Cc1ccccc1)c1ccc(N)cc1
246	Dasatinib	15	-0.7533	0	Clc1cccc(C)c1NC(=O)c1sc(nc1)Nc1nc(nc(N2CCN(CC2)CCO)c1)C
247	Deferasirox	70	0.3680	1	O=C1C=CC=CC1=C1NC(NN1c1ccc(cc1)C(O)=O)=C1C=CC=CC1=O
248	Deferoxamine	5	-1.2788	0	O=C(N(O)CCCCCNC(=O)CCC(=O)N(O)CCCCCN)CCC(=O)NCCCCCN(O)C(=O)C
249	Deflazacort	68	0.3274	1	O1C2CC3C4C(C5(C(=CC(=O)C=C5)CC4)C)C(O)CC3(C)C2(N=C1C)C(=O)COC(=O)C

250	Delapril	55	0.0872	1	O(C(=O)C(NC(C(=O)N(CC(O)=O)C1Cc2c(C1)cccc2)C)CCc1ccccc1)CC
251	Delavirdine	85	0.7533	1	S(=O)(=O)(Nc1cc2cc([nH]c2cc1)C(=O)N1CCN(CC1)c1ncccc1NC(C)C)C
252	Demeclocycline	66	0.2881	1	Clc1c2c(C(=O)C=3C(CC4C(O)(C(=O)C(C(=O)N)=C(O)C4N(C)C)C=3O)C2O)c(O)cc1
253	Depamide	68	0.3274	1	O=C(N)C(CCC)CCC
254	Desipramine	82.5	0.6734	1	N(CCCN1c2c(CCc3c1cccc3)cccc2)C
255	Desmopressin	15	-0.7533	0	S1SCCC(=O)NC(Cc2ccc(O)cc2)C(=O)NC(Cc2ccccc2)C(=O)NC(CCC(=O)N)C(=O)NC(CC(=O)N)C(=O)NC(C1)C(=O)N1CCCC1C(=O)NC(CCCN=C(N)N)C(=O)NCC(=O)N
256	Desogestrel	5	-1.2788	0	OC1(CCC2C3C(C4C(CC3)=CCCC4)C(CC12CC)=C)C#C
257	Dexfenfluramine	89	0.9080	1	FC(F)(F)c1cc(ccc1)CC(NCC)C
258	Dexloxiglumide	48	-0.0348	0	Clc1cc(ccc1Cl)C(=O)NC(CCC(O)=O)C(=O)N(CCCCC)CCCOC
259	Dextroamphetamine	75	0.4771	1	NC(Cc1ccccc1)C
260	Dextromethorphan	11	-0.9080	0	O(C)c1cc2C34C(C(N(CC3)C)Cc2cc1)CCCC4
261	Dextromoramide	99	1.9956	1	O1CCN(CC1)CC(C(C(=O)N1CCCC1)(c1ccccc1)c1ccccc1)C
262	Dextrose	99	1.9956	1	O1C(CO)C(O)C(O)C(O)C1O
263	Diacetolol	35	-0.2688	0	O(CC(O)CNC(C)C)c1ccc(NC(=O)C)cc1C(=O)C
264	Diacetylmorphine	1	-1.9956	0	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC(=O)C)cc3)C=CC2OC(=O)C
265	Diatrizoate	5	-1.2788	0	Ic1c(C(O)=O)c(I)c(NC(=O)C)c(I)c1NC(=O)C
266	Diazepam	99	1.9956	1	Clc1cc2c(N(C)C(=O)CN=C2c2ccccc2)cc1
267	Diazoxide	90	0.9542	1	Clc1cc2S(=O)(=O)NC(=Nc2cc1)C

268	Diclofenac	54	0.0696	1	Clc1cc(Cl)ccc1Nc1ccccc1CC(O)=O
269	Dicloxacillin	67.5	0.3174	1	Clc1cccc(Cl)c1-c1noc(C)c1C(=O)NC1C2SC(C)(C)C(N2C1=O)C(O)=O
270	Didanosine	38	-0.2126	0	O1C(CCC1n1c2N=CNC(=O)c2nc1)CO
271	Dienogest	94	1.1950	1	OC1(CCC2C3C(=C4C(=CC(=O)CC4)CC3)CCC12C)CC#N
272	Diflunisal	85	0.7533	1	Fc1cc(F)ccc1-c1cc(C(O)=O)c(O)cc1
273	Digitoxin	95	1.2788	1	O1C(C)C(OC2OC(C)C(O)C(O)C2)C(O)CC1OC1C(OC(OC2CC3CCC4C(CCC5(C)C(CCC45O)C4=CC(OC4)=O)C3(CC2)C)CC1O)C
274	Digoxin	70	0.3680	1	O1C(C)C(OC2OC(C)C(O)C(O)C2)C(O)CC1OC1C(OC(OC2CC3CCC4C(CC(O)C5(C)C(CCC45O)C4=CC(OC4)=O)C3(CC2)C)CC1O)C
275	Dihydrocodeine	20	-0.6021	0	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC)cc3)CCC2O
276	Dihydroergosine	10	-0.9542	0	O1C(NC(=O)C2CC3C(N(C2)C)Cc2c4c3cccc4[nH]c2)(C)C(=O)N2C(CC(C)C)C(=O)N3C(CCC3)C12O
277	Dihydroergotamine	1	-1.9956	0	O1C(NC(=O)C2CC3C(N(C2)C)Cc2c4c3cccc4[nH]c2)(C)C(=O)N2C(Cc3ccccc3)C(=O)N3C(CCC3)C12O
278	Dilevalol	12	-0.8653	0	Oc1ccc(cc1C(=O)C)C(O)CNC(CCc1ccccc1)C
279	Diloxanide	90	0.9542	1	ClC(Cl)C(=O)N(C)c1ccc(O)cc1
280	Diltiazem	38	-0.2126	0	S1c2c(N(CCN(C)C)C(=O)C(OC(=O)C)C1c1ccc(OC)cc1)cccc2
281	Dimercaprol	0	-2.0000	0	SC(CO)CS
282	Diphenhydramine	72	0.4102	1	O(C(c1ccccc1)c1ccccc1)CCN(C)C
283	Diprophylline	90	0.9542	1	O=C1N(C)C(=O)N(c2ncn(c12)CC(O)CO)C
284	Dipyridamole	50	0.0000	1	OCCN(CCO)c1nc(N2CCCCC2)c2nc(nc(N3CCCCC3)c2n1)N(CCO)CCO

285	Dirithromycin	10	-0.9542	0	O1C(CC)C(O)(C2OC(NC(C2C)C(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C)COCCOC)C
286	Disopyramide	83	0.6886	1	O=C(N)C(CCN(C(C)C)C(C)C)(c1ccccc1)c1ncccc1
287	Distigmine	8	-1.0607	0	O(C(=O)N(CCCCCCN(C(Oc1ccc[n+](c1)C)=O)C)C)c1ccc[n+](c1)C
288	Disulfiram	80	0.6021	1	S(SC(=S)N(CC)CC)C(=S)N(CC)CC
289	Docetaxel	0	-2.0000	0	O1C2CC(O)C3(C(C(OC(=O)c4ccccc4)C4(O)CC(OC(=O)C(O)C(NC(OC(C)(C)C)=O)c5ccccc5)C(=C(C4(C)C)C(O)C3=O)C)C2(OC(=O)C)C1)C
290	Docosanol	0	-2.0000	0	OCCCCCCCCCCCCCCCCCCCCCC
291	Dofetilide	96	1.3802	1	S(=O)(=O)(Nc1ccc(cc1)CCN(CCOc1ccc(NS(=O)(=O)C)cc1)C)C
292	Dolasetron	75	0.4771	1	O(C(=O)c1c2c([nH]c1)cccc2)C1CC2N3CC(=O)C(C2)CC3C1
293	Domperidone	15	-0.7533	0	Clc1cc2NC(=O)N(c2cc1)C1CCN(CC1)CCCN1c2c(NC1=O)cccc2
294	Donepezil	99	1.9956	1	O(C)c1cc2c(CC(CC3CCN(CC3)Cc3ccccc3)C2=O)cc1OC
295	Dosulepin	30	-0.3680	0	S1Cc2c(cccc2)C(c2c1cccc2)=CCCN(C)C
296	dOTC	80	0.6021	1	S1C(OCC1N1C=CC(=NC1=O)N)CO
297	Doxapram	64	0.2499	1	O1CCN(CC1)CCC1CN(CC)C(=O)C1(c1ccccc1)c1ccccc1
298	Doxepin	30	-0.3680	0	O1Cc2c(cccc2)C(c2c1cccc2)=CCCN(C)C
299	Doxifluridine	40	-0.1761	0	FC1=CN(C2OC(C)C(O)C2O)C(=O)NC1=O
300	DOXORUBICIN	5	-1.2788	0	O1C(C)C(O)C(N)CC1OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(C1=O)c(OC)ccc3)c2O)C(=O)CO
301	Doxycycline	93	1.1234	1	OC12C(C(N(C)C)C(O)=C(C(=O)N)C1=O)C(O)C1C(=C2O)C(=O)c2c(cccc2O)C1C
302	Doxylamine	24.7	-0.4841	0	O(C(C)(c1ccccc1)c1ncccc1)CCN(C)C

303	Droperidol	75	0.4771	1	Fc1ccc(cc1)C(=O)CCCN1CCC(N2c3c(NC2=O)cccc3)=CC1
304	Drospirenone	76	0.5006	1	O1C2(C3C(C4C5C(CCC24C)C2(C(=CC(=O)CC2)C2C5C2)C)C3)CCC1=O
305	Drotaverine	58	0.1402	1	O(CC)c1cc(ccc1OCC)C=C1NCCc2cc(OCC)c(OCC)cc12
306	Droxidopa	90	0.9542	1	Oc1cc(ccc1O)C(O)C(N)C(O)=O
307	Duloxetine	50	0.0000	1	s1cccc1C(Oc1c2c(ccc1)cccc2)CCNC
308	Dutasteride	60	0.1761	1	FC(F)(F)c1ccc(cc1NC(=O)C1CCC2C3C(CCC12C)C1(C(NC(=O)C=C1)CC3)C)C(F)(F)F
309	Dydrogesterone	28	-0.4102	0	O=C1CCC2(C3C(C4CCC(C(=O)C)C4(CC3)C)C=CC2=C1)C
310	Edrophonium	0	-2.0000	0	Oc1cc([N+](CC)(C)C)ccc1
311	Efavirenz	42.5	-0.1313	0	Clc1cc2c(NC(OC2(C#CC2CC2)C(F)(F)F)=O)cc1
312	Eflornithine	55	0.0872	1	FC(F)C(N)(CCCN)C(O)=O
313	Emedastine	50	0.0000	1	O(CCn1c2c(nc1N1CCCN(CC1)C)cccc2)CC
314	Empronium	1	-1.9956	0	CC[N+](C)(C)C(C)CC(C1=CC=CC=C1)C2=CC=CC=C2
315	Enalapril	63.5	0.2405	1	O(C(=O)C(NC(C(=O)N1CCCC1C(O)=O)C)CCc1ccccc1)CC
316	Enalaprilat	40	-0.1761	0	OC(=O)C1N(CCC1)C(=O)C(NC(CCc1ccccc1)C(O)=O)C
317	Encainide	27	-0.4320	0	O(C)c1ccc(cc1)C(=O)Nc1ccccc1CCC1N(CCCC1)C
318	Endralazine	75	0.4771	1	O=C(N1CC2=CC(NN=C2CC1)=NN)c1ccccc1
319	Enoxacin	87	0.8256	1	Fc1cc2c(nc1N1CCNCC1)N(C=C(C(O)=O)C2=O)CC
320	Entacapone	42	-0.1402	0	Oc1c([N+](=O)[O-])cc(cc1O)C=C(C(=O)N(CC)CC)C#N
321	Entecavir	99	1.9956	1	OC1CC(n2c3NC(=NC(=O)c3nc2)N)C(=C)C1CO
322	Epanolol	8	-1.060	0	O(CC(O)CNCCNC(=O)Cc1ccc(O)cc1)c1ccccc1C#N

			7		
323	Eplerenone	69	0.3475	1	O1C2(CCC1=O)CCC1C3C4(OC4CC12C)C1(C(CC3C(OC)=O)=CC(=O)CC1)C
324	Epristeride	93	1.1234	1	OC(=O)C=1CCC2(C3C(C4CCC(C(=O)NC(C)(C)C)C4(CC3)C)CC=C2C=1)C
325	Eprosartan	13	-0.8256	0	s1cccc1CC(=Cc1n(Cc2ccc(cc2)C(O)=O)c(nc1)CCCC)C(O)=O
326	Eproxindine	70	0.3680	1	O(C)c1c2c(n(c1C(=O)NCC(O)CN(CC)CC)-c1ccccc1)cccc2
327	Eptifibatide	0	-2.0000	0	S1SCCC(=O)NC(CCCCNC(N)=N)C(=O)NCC(=O)NC(CC(O)=O)C(=O)NC(Cc2c3c([nH]c2)cccc3)C(=O)N2C(CCC2)C(=O)NC(C1)C(=O)N
328	Ergoloid mesylate	25	-0.4771	0	O1C(NC(=O)C2CC3C(N(C2)C)Cc2c4c3cccc4[nH]c2)(C(C)C)C(=O)N2C(C(CC)(C)C)C(=O)N3C(CCC3)C12O
329	Ergonovine	50	0.0000	1	OCC(NC(=O)C1C=C2C(N(C1)C)Cc1c3c2cccc3[nH]c1)C
330	Ergotamine	2	-1.6902	0	O1C(NC(=O)C2C=C3C(N(C2)C)Cc2c4c3cccc4[nH]c2)(C)C(=O)N2C(Cc3ccccc3)C(=O)N3C(CCC3)C12O
331	Erlotinib	59	0.1581	1	O(CCOC)c1cc2c(ncnc2Nc2cc(ccc2)C#C)cc1OCCOC
332	Erythromycin	35	-0.2688	0	O1C(CC)C(O)(C)C(O)C(C)C(=O)C(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C
333	Esomeprazole	53	0.0522	1	S(=O)(Cc1ncc(C)c(OC)c1C)c1[nH]c2cc(OC)ccc2n1
334	Estazolam	93	1.1234	1	Clc1cc2c(-n3c(nnc3)CN=C2c2ccccc2)cc1
335	Estradiol	5	-1.2788	0	OC1CCC2C3C(CCC12C)c1c(cc(O)cc1)CC3
336	Estradiol_17-valerate	3	-1.5097	0	O(C(=O)CCCC)C1CCC2C3C(CCC12C)c1c(cc(O)cc1)CC3
337	Estramustine phosphate	43.7	-0.1100	0	ClCCN(CCCl)C(Oc1cc2CCC3C4CCC(OP(O)(O)=O)C4(CCC3c2cc1)C)=O
338	Ethacrynic acid	99	1.9956	1	Clc1c(Cl)c(OCC(O)=O)ccc1C(=O)C(CC)=C
339	Ethambutol	77	0.5248	1	OCC(NCCNC(CC)CO)CC
340	Ethanol	80	0.602	1	OCC

			1		
341	Ethinylestradiol	40	-0.1761	0	OC1(CCC2C3C(CCC12C)c1c(cc(O)cc1)CC3)C#C
342	Ethionamide	80	0.6021	1	S=C(N)c1cc(ncc1)CC
343	Ethopropazine	5	-1.2788	0	S1c2c(N(c3c1cccc3)CC(N(CC)CC)C)cccc2
344	Ethosuximide	93	1.1234	1	O=C1NC(=O)CC1(CC)C
345	Ethoxzolamide	65	0.2688	1	s1c2cc(OCC)ccc2nc1S(=O)(=O)N
346	Etidocaine	0	-2.0000	0	O=C(Nc1c(cccc1C)C)C(N(CCC)CC)CC
347	Etidronic Acid	5	-1.2788	0	P(O)(O)(=O)C(P(O)(O)=O)(O)C
348	Etilefrine	55	0.0872	1	Oc1cc(ccc1)C(O)CNCC
349	Etodolac	90	0.9542	1	O1CCc2c([nH]c3c2cccc3CC)C1(CC(O)=O)CC
350	Etofylline	80	0.6021	1	O=C1N(C)C(=O)N(c2ncn(c12)CCO)C
351	Etoposide	52	0.0348	1	O1C2C(OC(OC2)C)C(O)C(O)C1OC1C2C(C(c3c1cc1OCOc1c3)c1cc(OC)c(O)c(OC)c1)C(OC2)=O
352	Etoricoxib	99	1.9956	1	Clc1cc(c(nc1)-c1ccc(nc1)C)-c1ccc(S(=O)(=O)C)cc1
353	Etretinate	40	-0.1761	0	O(C)c1cc(C)c(C=CC(=CC=CC(=CC(OCC)=O)C)C)c(C)c1C
354	Exemestane	42	-0.1402	0	O=C1CCC2C3C(CCC12C)C1(C(=CC(=O)C=C1)C(C3)=C)C
355	Ezetimibe	50	0.0000	1	Fc1ccc(cc1)C(O)CCC1C(N(C1=O)c1ccc(F)cc1)c1ccc(O)cc1
356	Famciclovir	77	0.5248	1	O(C(=O)C)CC(CCn1c2nc(ncc2nc1)N)COC(=O)C
357	Famotidine	42	-0.1402	0	s1cc(nc1N=C(N)N)CSCCC(=NS(=O)(=O)N)N
358	Felbamate	90	0.9542	1	O(CC(COC(=O)N)c1ccccc1)C(=O)N
359	Felodipine	15	-0.7533	0	Clc1c(cccc1Cl)C1C(C(OCC)=O)=C(NC(C)=C1C(OC)=O)C

360	Femoxetine	8	-1.0607	0	O(CC1CN(CCC1c1ccccc1)C)c1ccc(OC)cc1
361	Fenbufen	78	0.5497	1	OC(=O)CCC(=O)c1ccc(cc1)-c1ccccc1
362	Fenflumizole	50	0.0000	1	Fc1cc(F)ccc1-c1[nH]c(c(n1)-c1ccc(OC)cc1)-c1ccc(OC)cc1
363	Fenoldopam	6	-1.1950	0	Clc1c2c(cc(O)c1O)C(CNCC2)c1ccc(O)cc1
364	Fenoterol	2	-1.6902	0	Oc1cc(cc(O)c1)CC(O)NC(Cc1ccc(O)cc1)C
365	Fenspiride	90	0.9542	1	O1C2(CCN(CC2)CCc2ccccc2)CNC1=O
366	Fentanyl	1	-1.9956	0	O=C(N(C1CCN(CC1)CCc1ccccc1)c1ccccc1)CC
367	Fexofenadine	70	0.3680	1	OC(C1CCN(CC1)CCCC(O)c1ccc(cc1)C(C(O)=O)(C)C)(c1ccccc1)c1ccccc1
368	Finasteride	63	0.2311	1	O=C1NC2CCC3C4CCC(C(=O)NC(C)(C)C)C4(CCC3C2(C=C1)C)C
369	Flecainide	95	1.2788	1	FC(F)(F)COc1ccc(OCC(F)(F)F)cc1C(=O)NCC1NCCCC1
370	Fleroxacin	99	1.9956	1	Fc1c2N(C=C(C(O)=O)C(=O)c2cc(F)c1N1CCN(CC1)C)CCF
371	Floxuridine	90	0.9542	1	FC1=CN(C2OC(CO)C(O)C2)C(=O)NC1=O
372	FLUCLOXACILLIN	60	0.1761	1	Clc1cccc(F)c1-c1noc(C)c1C(=O)NC1C2SC(C)(C)C(N2C1=O)C(O)=O
373	Fluconazole	90	0.9542	1	Fc1cc(F)ccc1C(O)(Cn1ncnc1)Cn1ncnc1
374	Flucytosine	84	0.7202	1	FC1=CNC(=O)N=C1N
375	Fludarabine	57.5	0.1313	1	P(OCC1OC(n2c3nc(F)nc(N)c3nc2)C(O)C1O)(O)(O)=O
376	Flumazenil	16	-0.7202	0	Fc1cc2c(-n3c(CN(C)C2=O)c(nc3)C(OCC)=O)cc1
377	Flunisolide	6.7	-1.1438	0	FC1C2=CC(=O)C=CC2(C2C(C3CC4OC(OC4(C(=O)CO)C3(CC2O)C)(C)C)C1)C
378	Flunitrazepam	80	0.6021	1	Fc1ccccc1C1=NCC(=O)N(c2c1cc([N+](=O)[O-])cc2)C

379	Fluocortolone	90	0.9542	1	FC1C2=CC(=O)C=CC2(C2C(C3CC(C)C(C(=O)CO)C3(CC2O)C)C1)C
380	Fluorescein	99	1.9956	1	O1C2(c3c(cccc3)C1=O)c1c(Oc3c2ccc(O)c3)cc(O)cc1
381	FLUOROURACIL	28	-0.4102	0	FC1=CNC(=O)NC1=O
382	Fluoxetine	60	0.1761	1	FC(F)(F)c1ccc(OC(CCNC)c2ccccc2)cc1
383	Fluoxymesterone	99	1.9956	1	FC12C(C3CCC(O)(C)C3(CC1O)C)CCC1=CC(=O)CCC12C
384	Flupenthixol	50	0.0000	1	S1c2c(cc(cc2)C(F)(F)F)C(c2c1cccc2)=CCCN1CCN(CC1)CCO
385	Fluphenazine	2.7	-1.5567	0	S1c2c(N(c3c1cccc3)CCCN1CCN(CC1)CCO)cc(cc2)C(F)(F)F
386	Flupirtine	90	0.9542	1	Fc1ccc(cc1)CNc1nc(N)c(NC(OCC)=O)cc1
387	Flurazepam	83	0.6886	1	Clc1cc2c(N(CCN(CC)CC)C(=O)CN=C2c2ccccc2F)cc1
388	Flurbiprofen	92	1.0607	1	Fc1cc(ccc1-c1ccccc1)C(C(O)=O)C
389	Flutamide	90	0.9542	1	FC(F)(F)c1cc(NC(=O)C(C)C)ccc1[N+](=O)[O-]
390	Fluticasone_propionate	1	-1.9956	0	S(C(=O)C1(OC(=O)CC)C2(CC(O)C3(F)C(C2CC1C)CC(F)C1=CC(=O)C=CC13C)C)CF
391	Fluvastatin	24	-0.5006	0	Fc1ccc(cc1)-c1c2c(n(C(C)C)c1C=CC(O)CC(O)CC(O)=O)cccc2
392	Fluvoxamine	53	0.0522	1	FC(F)(F)c1ccc(cc1)C(=NOCCN)CCCCOC
393	Folic Acid	75	0.4771	1	O=C1NC(=Nc2ncc(nc12)CNc1ccc(cc1)C(=O)NC(CCC(O)=O)C(O)=O)N
394	Folinic acid	4	-1.3802	0	O=C1N=C(NC=2NCC(N(C1=2)C=O)CNc1ccc(cc1)C(=O)NC(CCC(O)=O)C(O)=O)N
395	Foscarnet	9	-1.0048	0	P(O)(O)(=O)C(O)=O
396	Fosfomycin	31	-0.3475	0	P(O)(O)(=O)C1OC1C
397	Fosinopril	34	-0.2881	0	P(OC(OC(=O)CC)C(C)C)(=O)(CCCCc1ccccc1)CC(=O)N1CC(CC1C(O)=O)C1CCCCC1

398	Frovatriptan	27	-0.4320	0	O=C(N)c1cc2c3CC(NC)CCc3[nH]c2cc1
399	Fructose	50	0.0000	1	O1CC(O)C(O)C(O)C1(O)CO
400	Furosemide	71	0.3889	1	Clc1cc(NCc2occc2)c(C(O)=O)c(S(=O)(=O)N)c1
401	Fusidic acid	91	1.0048	1	O(C(=O)C)C1CC2(C(CC(O)C3C2(CCC2C(C)C(O)CCC23C)C)C1=C(CCC=C(C)C)C(O)=O)C
402	Gabapentin	60	0.1761	1	OC(=O)CC1(CCCCC1)CN
403	Galantamine	99	1.9956	1	O1c2c3C4(C1CC(O)C=C4)CCN(Cc3ccc2OC)C
404	Ganciclovir	4	-1.3802	0	O=C1NC(=Nc2n(cnc12)COC(CO)CO)N
405	Gatifloxacin	96	1.3802	1	Fc1cc2c(N(C=C(C(O)=O)C2=O)C2CC2)c(OC)c1N1CC(NCC1)C
406	Gefitinib	60	0.1761	1	Clc1cc(Nc2ncnc3c2cc(OCCCN2CCOCC2)c(OC)c3)ccc1F
407	Gemfibrozil	98	1.6902	1	O(CCCC(C(O)=O)(C)C)c1cc(ccc1C)C
408	Genaconazole	99	1.9956	1	S(=O)(=O)(C(C(O)(Cn1ncnc1)c1ccc(F)cc1F)C)C
409	Gentamicin_C1a	0	-2.0000	0	O1C(OC2C(O)C(OC3OCC(O)(C)C(NC)C3O)C(N)CC2N)C(N)CCC1CC
410	Gestodene	93	1.1234	1	OC1(C=CC2C3C(C4C(=CC(=O)CC4)CC3)CCC12CC)C#C
411	Ginkgolide A	80	0.6021	1	O1C2CC34C56C(OC(=O)C5O)OC3(C2(O)C(C)C1=O)C(OC4CC6C(C)(C)C)=O
412	Ginkgolide B	88	0.8653	1	O1C2C(O)(C34OC5OC(=O)C(O)C56C3(C(OC4=O)CC6C(C)(C)C)C2O)C(C)C1=O
413	Gliclazide	97	1.5097	1	S(=O)(=O)(NC(=O)NN1CC2C(CCC2)C1)c1ccc(cc1)C
414	Glimepiride	99	1.9956	1	S(=O)(=O)(NC(=O)NC1CCC(CC1)C)c1ccc(cc1)CCNC(=O)N1CC(C)=C(CC)C1=O
415	Glipizide	95	1.2788	1	S(=O)(=O)(NC(=O)NC1CCCCC1)c1ccc(cc1)CCNC(=O)c1ncc(nc1)C
416	Glucosamine	26	-0.4543	0	O1C(CO)C(O)C(O)C(N)C1O

417	Glyburide (Glibenclamide)	82	0.6585	1	COC1=C(C=C(C=C1)Cl)C(=O)NCCC2=CC=C(C=C2)S(=O)(=O)NC(=O)NC3CCCCC3
418	Glycopyrrolate	17	-0.6886	0	O(C(=O)C(O)(C1CCCC1)c1ccccc1)C1CC[N+](C1)(C)C
419	Granisetron	60	0.1761	1	O=C(NC1CC2N(C(C1)CCC2)C)c1nn(c2c1cccc2)C
420	Grepafloxacin	70	0.3680	1	Fc1c(c2c(N(C=C(C(O)=O)C2=O)C2CC2)cc1N1CC(NCC1)C)C
421	Griseofulvin	47.5	-0.0435	0	Clc1c2OC3(C(CC(=O)C=C3OC)C)C(=O)c2c(OC)cc1OC
422	Guanabenz	75	0.4771	1	Clc1cccc(Cl)c1C=NN=C(N)N
423	Guanadrel	85	0.7533	1	O1C(COC12CCCCC2)CN=C(N)N
424	Guanethidine	20	-0.6021	0	N(CCN1CCCCCCC1)C(N)=N
425	Guanfacine	81	0.6297	1	Clc1cccc(Cl)c1CC(=O)N=C(N)N
426	Halofantrine	6.3	-1.1724	0	Clc1cc(Cl)cc2c1cc(c1c2cc(cc1)C(F)(F)F)C(O)CCN(CCCC)CCCC
427	Haloperidol	60	0.1761	1	Clc1cc(ccc1)C1(O)CCN(CC1)CCCC(=O)c1ccc(F)cc1
428	Hesperetin	20	-0.6021	0	O1c2c(C(=O)CC1c1cc(O)c(OC)cc1)c(O)cc(O)c2
429	Hexobarbital	95	1.2788	1	CN1C(=O)NC(=O)C(C)(C2=CCCCC2)C1=O
430	Homocysteine	53	0.0522	1	SCCC(N)C(O)=O
431	Hydralazine	23	-0.5248	0	n1ncc2c(cccc2)c1NN
432	Hydrochlorothiazide	71	0.3889	1	Clc1cc2NCNS(=O)(=O)c2cc1S(=O)(=O)N
433	Hydrocodone	80	0.6021	1	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC)cc3)CCC2=O

434	Hydroflumethiazide	50	0.0000	1	S(=O)(=O)(N)c1cc2S(=O)(=O)NCNc2cc1C(F)(F)F
435	Hydromorphone	42	-0.1402	0	O1C2C34C(C(N(CC3)C)Cc3c4c1c(O)cc3)CCC2=O
436	Hydroxychloroquine	74	0.4543	1	Clc1cc2nccc(NC(CCCN(CCO)CC)C)c2cc1
437	Hydroxyurea	79	0.5754	1	O=C(NO)N
438	Hydroxyzine	80	0.6021	1	Clc1ccc(cc1)C(N1CCN(CC1)CCOCCO)c1ccccc1
439	Hypericin	20	-0.6021	0	Oc1c2c3c(c4c5c(C(=O)c6c7c(c8c(c(C2=O)c(O)cc8O)c3c57)c(O)cc6O)c(O)cc4C)c(c1)C
440	Ibandronate	1	-1.9956	0	P(O)(O)(=O)C(P(O)(O)=O)(O)CCN(CCCCC)C
441	Ibuprofen	85	0.7533	1	OC(=O)C(C)c1ccc(cc1)CC(C)C
442	Ibutilide	8	-1.0607	0	S(=O)(=O)(Nc1ccc(cc1)C(O)CCCN(CCCCCCC)CC)C
443	Idarubicin	28	-0.4102	0	O1C(C)C(O)C(N)CC1OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(cccc3)C1=O)c2O)C(=O)C
444	Idazoxan	34	-0.2881	0	O1c2c(OCC1C=1NCCN=1)cccc2
445	Ifosfamide	92	1.0607	1	ClCCN1P(OCCC1)(=O)NCCCl
446	Iloprost	16	-0.7202	0	OC1CC2C(CC(C2)=CCCCC(O)=O)C1C=CC(O)C(CC#CC)C
447	Imatinib	98	1.6902	1	O=C(Nc1cc(Nc2nc(ccn2)-c2cccnc2)c(cc1)C)c1ccc(cc1)CN1CCN(CC1)C
448	Imidapril	20	-0.6021	0	O=C1N(C(=O)C(NC(CCc2ccccc2)C(OCC)=O)C)C(CN1C)C(O)=O
449	Imipenem	1	-1.9956	0	S(CCN=CN)C=1CC2N(C(=O)C2C(O)C)C=1C(O)=O
450	Imipramine	42	-0.1402	0	N(CCCN1c2c(CCc3c1cccc3)cccc2)(C)C
451	Incadronate	1	-1.995	0	P(O)(O)(=O)C(P(O)(O)=O)NC1CCCCCC1

			6		
452	Indapamide	90	0.9542	1	Clc1cc(S(=O)(=O)N)c(cc1)C(=O)NC1c2c(NC1C)cccc2
453	Indinavir	65	0.2688	1	OC1Cc2c(cccc2)C1NC(=O)C(Cc1ccccc1)CC(O)CN1CCN(CC1C(=O)NC(C)(C)C)Cc1cccnc1
454	Indobufen	85	0.7533	1	O=C1N(Cc2c1cccc2)c1ccc(cc1)C(CC)C(O)=O
455	Indomethacin	99	1.9956	1	Clc1ccc(cc1)C(=O)n1c2c(cc(OC)cc2)c(CC(O)=O)c1C
456	Indoramin	18	-0.6585	0	O=C(NC1CCN(CC1)CCc1c2c([nH]c1)cccc2)c1ccccc1
457	Ipratropium	3	-1.5097	0	O(C(=O)C(CO)c1ccccc1)C1CC2[N+](C(C1)CC2)(C(C)C)C
458	Irbesartan	70	0.3680	1	O=C1N(Cc2ccc(cc2)-c2ccccc2-c2nn[nH]n2)C(=NC12CCCC2)CCCC
459	Irinotecan	8	-1.0607	0	O1CC2=C(C=C3N(Cc4c3nc3c(cc(OC(=O)N5CCC(N6CCCCC6)CC5)cc3)c4CC)C2=O)C(O)(CC)C1=O
460	Isoniazid	80	0.6021	1	O=C(NN)c1ccncc1
461	Isoproterenol	0	-2.0000	0	Oc1cc(ccc1O)C(O)CNC(C)C
462	Isosorbide 2-Mononitrate	93	1.1234	1	O1C2C(OCC2O)C(O[N+](=O)[O-])C1
463	Isosorbide Dinitrate	22	-0.5497	0	C1C(C2C(O1)C(CO2)O[N+](=O)[O-])O[N+](=O)[O-]
464	Isotretinoin	40	-0.1761	0	OC(=O)C=C(C=CC=C(C=CC=1C(CCCC=1C)(C)C)C)C
465	Isoxicam	97	1.5097	1	S1(=O)(=O)N(C)C(=C(O)Nc2noc(c2)C)C(=O)c2c1cccc2
466	Isradipine	19.5	-0.6158	0	o1nc2c(n1)cccc2C1C(C(OC(C)C)=O)=C(NC(C)=C1C(OC)=O)C
467	Itraconazole	55	0.0872	1	Clc1cc(Cl)ccc1C1(OC(CO1)COc1ccc(N2CCN(CC2)c2ccc(N3C=NN(C(CC)C)C3=O)cc2)cc1)Cn1ncnc1
468	Kanamycin	1	-1.9956	0	O1C(CN)C(O)C(O)C(O)C1OC1C(O)C(OC2OC(CO)C(O)C(N)C2O)C(N)CC1N

469	KETAMINE	20	-0.6021	0	CNC1(CCCCC1=O)C2=CC=CC=C2Cl
470	Ketanserin	50	0.0000	1	Fc1ccc(cc1)C(=O)C1CCN(CC1)CCN1C(=O)c2c(NC1=O)cccc2
471	Ketobemidone	34	-0.2881	0	Oc1cc(ccc1)C1(CCN(CC1)C)C(=O)CC
472	Ketoconazole	75	0.4771	1	CC(=O)N1CCN(CC1)c2ccc(OCC3COC(Cn4ccnc4)(O3)c5ccc(Cl)cc5Cl)cc2
473	Ketoprofen	90	0.9542	1	OC(=O)C(C)c1cc(ccc1)C(=O)c1ccccc1
474	Ketorolac	99	1.9956	1	OC(=O)C1CCn2c1ccc2C(=O)c1ccccc1
475	Ketotifen	60	0.1761	1	s1c2c(cc1)C(c1c(CC2=O)cccc1)=C1CCN(CC1)C
476	KNI-272, Kynostatin-272	40	-0.1761	0	S1CC(N(C1)C(=O)C(O)C(NC(=O)C(NC(=O)COc1nccc2c1cccc2)CSC)Cc1ccccc1)C(=O)NC(C)(C)C
477	L-5-hydroxytryptophan	70	0.3680	1	Oc1cc2c([nH]cc2CC(N)C(O)=O)cc1
478	Labetalol	25	-0.4771	0	Oc1ccc(cc1C(=O)N)C(O)CNC(CCc1ccccc1)C
479	Lacidipine	18	-0.6585	0	O(C(=O)C=1C(C(C(OCC)=O)=C(NC=1C)C)c1ccccc1C=CC(OC(C)(C)C)=O)CC
480	Lactulose	3	-1.5097	0	O1C(O)(CO)C(O)C(OC2OC(CO)C(O)C(O)C2O)C1CO
481	Lamivudine	86	0.7884	1	S1CC(OC1CO)N1C=CC(=NC1=O)N
482	Lamotrigine	97.6	1.6092	1	Clc1c(cccc1Cl)-c1nnc(nc1N)N
483	Lansoprazole	81	0.6297	1	S(=O)(Cc1ncc(C)c(OCC(F)(F)F)c1)c1[nH]c2c(n1)cccc2
484	Lapatinib	0	-2.0000	0	Clc1cc(Nc2ncnc3c2cc(cc3)-c2oc(cc2)CNCCS(=O)(=O)C)ccc1OCc1cc(F)ccc1
485	L-Arginine	68	0.3274	1	OC(=O)C(N)CCCN=C(N)N
486	Leflunomide	80	0.6021	1	FC(F)(F)c1ccc(NC(=O)c2cnoc2C)cc1

487	Lercanidipine	10	-0.9542	0	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)C(CN(CCC(c1ccccc1)c1ccccc1)C)(C)C
488	Letrozole	99	1.9956	1	n1cn(nc1)C(c1ccc(cc1)C#N)c1ccc(cc1)C#N
489	Levamisole	65	0.2688	1	S1CCN2CC(N=C12)c1ccccc1
490	Levetiracetam	99	1.9956	1	O=C1N(CCC1)C(CC)C(=O)N
491	Levobunolol	75	0.4771	1	O(CC(O)CNC(C)(C)C)c1c2c(ccc1)C(=O)CCC2
492	Levocabastine	99	1.9956	1	Fc1ccc(cc1)C1(CCC(N2CC(C)C(CC2)(C(O)=O)c2ccccc2)CC1)C#N
493	Levocarnitine	15	-0.7533	0	C[N+](C)(C)CC(CC(=O)[O-])O
494	Levodopa	41	-0.1581	0	Oc1cc(ccc1O)CC(N)C(O)=O
495	Levofloxacin	97.5	1.5911	1	Fc1cc2c3N(C=C(C(O)=O)C2=O)C(COc3c1N1CCN(CC1)C)C
496	Levomepromazine	50	0.0000	1	S1c2c(N(c3c1cccc3)CC(CN(C)C)C)cc(OC)cc2
497	Levoprotiline	40	-0.1761	0	OC(CC12CCC(c3c1cccc3)c1c2cccc1)CNC
498	Levosimendan	85	0.7533	1	O=C1NN=C(C(C1)C)c1ccc(NN=C(C#N)C#N)cc1
499	Lidocaine	35	-0.2688	0	O=C(Nc1c(cccc1C)C)CN(CC)CC
500	lincomycin	25	-0.4771	0	S(C)C1OC(C(NC(=O)C2N(CC(C2)CCC)C)C(O)C)C(O)C(O)C1O
501	Linezolid	99	1.9956	1	Fc1cc(N2CC(OC2=O)CNC(=O)C)ccc1N1CCOCC1
502	Lipoic Acid	29	-0.3889	0	S1SCCC1CCCCC(O)=O
503	Lisinopril	25	-0.4771	0	OC(=O)C1N(CCC1)C(=O)C(NC(CCc1ccccc1)C(O)=O)CCCCN
504	Lisuride	14	-0.7884	0	O=C(NC1C=C2C(N(C1)C)Cc1c3c2cccc3[nH]c1)N(CC)CC

505	Lofexidine	90	0.9542	1	Clc1cccc(Cl)c1OC(C)C=1NCCN=1
506	Lomefloxacin	97	1.5097	1	Fc1c2N(C=C(C(O)=O)C(=O)c2cc(F)c1N1CC(NCC1)C)CC
507	Lomustine	99	1.9956	1	ClCCN(N=O)C(=O)NC1CCCCC1
508	Loperamide	40	-0.1761	0	Clc1ccc(cc1)C1(O)CCN(CC1)CCC(C(=O)N(C)C)(c1ccccc1)c1ccccc1
509	Loracarbef	94	1.1950	1	ClC=1CCC2N(C(=O)C2NC(=O)C(N)c2ccccc2)C=1C(O)=O
510	Loratadine	40	-0.1761	0	Clc1cc2c(cc1)C(c1ncccc1CC2)=C1CCN(CC1)C(OCC)=O
511	Lorazepam	93	1.1234	1	Clc1ccccc1C1=NC(O)C(=O)Nc2c1cc(Cl)cc2
512	Lormetazepam	75	0.4771	1	Clc1ccccc1C1=NC(O)C(=O)N(c2c1cc(Cl)cc2)C
513	Losartan	35	-0.2688	0	Clc1nc(n(Cc2ccc(cc2)-c2ccccc2-c2[nH]nnn2)c1CO)CCCC
514	Loteprednol Etabonate	0	-2.0000	0	ClCOC(=O)C1(OC(OCC)=O)CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C
515	Lovastatin	5	-1.2788	0	O1C(CC(O)CC1=O)CCC1C2C(=CC(CC2OC(=O)C(CC)C)C)C=CC1C
516	Loxapine	30	-0.3680	0	Clc1cc2c(Oc3c(N=C2N2CCN(CC2)C)cccc3)cc1
517	Lubiprostone	0	-2.0000	0	FC(F)(CCCC)C1(OC2C(CC1)C(CCCCCCC(O)=O)C(=O)C2)O
518	Lumiracoxib	74	0.4543	1	Clc1cccc(F)c1Nc1ccc(cc1CC(O)=O)C
519	lurtotecan	12	-0.8653	0	O1CC2=C(C=C3N(Cc4c3nc3c(cc5OCCOc5c3)c4CN3CCN(CC3)C)C2=O)C(CC)C1=O
520	Lymecycline	99	1.9956	1	CC1(C2CC3C(C(=O)C(=C(C3(C(=O)C2=C(C4=C1C=CC=C4O)O)O)O)C(=O)NCNCCCCC(C(=O)O)N)N(C)C)O
521	Lynestrenol	30	-0.3680	0	OC1(CCC2C3C(C4C(CC3)=CCCC4)CCC12C)C#C
522	Mannitol	7	-1.1234	0	OC(C(O)C(O)CO)C(O)CO

523	Maprotiline	68	0.3274	1	N(CCCC12CCC(c3c1cccc3)c1c2cccc1)C
524	Maraviroc	23	-0.5248	0	FC1(F)CCC(CC1)C(=O)NC(CCN1C2CC(n3c(nnc3C)C(C)C)CC1CC2)c1ccccc1
525	Mazindol	93	1.1234	1	Clc1ccc(cc1)C1(O)N2C(=NCC2)c2c1cccc2
526	Mebendazole	22	-0.5497	0	O(C(=O)Nc1[nH]c2cc(ccc2n1)C(=O)c1ccccc1)C
527	Mecamylamine	99	1.9956	1	N(C)C1(C2CC(CC2)C1(C)C)C
528	Meclofenamic acid	65.1	0.2708	1	Clc1c(Nc2ccccc2C(O)=O)c(Cl)ccc1C
529	Medifoxamine	21	-0.5754	0	O(C(Oc1ccccc1)CN(C)C)c1ccccc1
530	Medroxalol	38	-0.2126	0	O1c2cc(ccc2OC1)CCC(NCC(O)c1cc(C(=O)N)c(O)cc1)C
531	Medroxyprogesterone	15	-0.7533	0	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)C(C3)C)C)C(=O)C
532	Medroxyprogesterone_acetate	10	-0.9542	0	O(C(=O)C)C1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)C(C3)C)C)C(=O)C
533	Mefenamic Acid	90	0.9542	1	OC(=O)c1ccccc1Nc1cccc(C)c1C
534	Mefloquine	80	0.6021	1	FC(F)(F)c1c2nc(cc(c2ccc1)C(O)C1NCCCC1)C(F)(F)F
535	Megestrol Acetate	90	0.9542	1	O(C(=O)C)C1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)C(=C3)C)C)C(=O)C
536	Melagatran	5	-1.2788	0	OC(=O)CNC(C(=O)N1CCC1C(=O)NCc1ccc(cc1)C(N)=N)C1CCCCC1
537	Melatonin	12	-0.8653	0	O(C)c1cc2c(N=CC2CCNC(=O)C)cc1
538	Meloxicam	97	1.5097	1	s1c(cnc1NC(=O)C=1N(S(=O)(=O)c2c(cccc2)C=1O)C)C
539	Melphalan	71	0.3889	1	ClCCN(CCCl)c1ccc(cc1)CC(N)C(O)=O

540	Memantine	99	1.9956	1	NC12CC3(CC(C1)(CC(C3)C2)C)C
541	Meperidine	52	0.0348	1	O(C(=O)C1(CCN(CC1)C)c1ccccc1)CC
542	Mepindolol	82	0.6585	1	O(CC(O)CNC(C)C)c1c2cc([nH]c2ccc1)C
543	Mepivacaine	55	0.0872	1	O=C(Nc1c(cccc1C)C)C1N(CCCC1)C
544	Meptazinol	9	-1.0048	0	Oc1cc(ccc1)C1(CCCCN(C1)C)CC
545	Mercaptopurine	12	-0.8653	0	S=C1N=CNc2nc[nH]c12
546	Meropenem	0	-2.0000	0	CC(O)C1C2C(C)C(SC3CNC(C3)C(=O)N(C)C)=C(N2C1=O)C(O)=O
547	Mesna	62	0.2126	1	S(O)(=O)(=O)CCS
548	Mesterolone	3	-1.5097	0	OC1CCC2C3C(CCC12C)C1(C(CC(=O)CC1C)CC3)C
549	Metergoline	23	-0.5248	0	O(Cc1ccccc1)C(=O)NCC1CC2C(N(C1)C)Cc1c3c2cccc3n(c1)C
550	Metformin	52	0.0348	1	N(C(NC(N)=N)=N)(C)C
551	Methacycline	58	0.1402	1	OC12C(C(N(C)C)C(O)=C(C(=O)N)C1=O)C(O)C1C(=C2O)C(=O)c2c(cccc2O)C1=C
552	Methadone	92	1.0607	1	O=C(C(CC(N(C)C)C)(c1ccccc1)c1ccccc1)CC
553	Methadyl Acetate	47	-0.0522	0	O(C(C(CC(N(C)C)C)(c1ccccc1)c1ccccc1)CC)C(=O)C
554	Methamphetamine	67	0.3076	1	N(C(Cc1ccccc1)C)C
555	Methapyrilene	20	-0.6021	0	s1cccc1CN(CCN(C)C)c1ncccc1
556	Methimazole	93	1.1234	1	S=C1NC=CN1C
557	Methotrexate	33	-0.3076	0	OC(=O)C(NC(=O)c1ccc(N(Cc2nc3c(nc(nc3N)N)nc2)C)cc1)CCC(O)=O
558	Methyldopa	42	-0.140	0	Oc1cc(ccc1O)CC(N)(C(O)=O)C

			2		
559	Methylergonovine	60	0.1761	1	OCC(NC(=O)C1C=C2C(N(C1)C)Cc1c3c2cccc3[nH]c1)CC
560	Methylphenidate	22	-0.5497	0	O(C(=O)C(C1NCCCC1)c1ccccc1)C
561	Methylphenobarbital	73	0.4320	1	O=C1N(C)C(=O)NC(=O)C1(CC)c1ccccc1
562	Methylprednisolone	82	0.6585	1	OC1(CCC2C3C(C4(C(=CC(=O)C=C4)C(C3)C)C)C(O)CC12C)C(=O)CO
563	Methyltestosterone	50	0.0000	1	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)CC3)C)C
564	Methysergide	13	-0.8256	0	OCC(NC(=O)C1C=C2C(N(C1)C)Cc1c3c2cccc3n(c1)C)CC
565	Meticillin	0	-2.0000	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)c1c(OC)cccc1OC
566	Metoclopramide	76	0.5006	1	O=C(NCCN(CC)CC)c1cc(Cl)c(N)cc1OC
567	Metolazone	65	0.2688	1	Clc1cc2NC(N(c3ccccc3C)C(=O)c2cc1S(=O)(=O)N)C
568	Metopimazine	20	-0.6021	0	S1c2c(N(c3c1cccc3)CCCN1CCC(CC1)C(=O)N)cc(S(=O)(=O)C)cc2
569	Metoprolol	38	-0.2126	0	O(CC(O)CNC(C)C)c1ccc(cc1)CCOC
570	Metronidazole	99	1.9956	1	OCCn1c(ncc1[N+](=O)[O-])C
571	Mexiletine	87	0.8256	1	O(CC(N)C)c1c(cccc1C)C
572	Mianserin	22	-0.5497	0	N12C(c3c(Cc4c1cccc4)cccc3)CN(CC2)C
573	Mibefradil	75	0.4771	1	Fc1cc2c(cc1)C(C(C)C)C(OC(=O)COC)(CC2)CCN(CCCc1[nH]c2c(n1)cccc2)C
574	Miconazole	25	-0.4771	0	Clc1cc(Cl)ccc1C(OCc1ccc(Cl)cc1Cl)Cn1ccnc1

575	Midazolam	44	-0.1047	0	Clc1cc2c(-n3c(CN=C2c2ccccc2F)cnc3C)cc1
576	Midodrine	93	1.1234	1	O(C)c1ccc(OC)cc1C(O)CNC(=O)CN
577	Mifepristone	69	0.3475	1	OC1(CCC2C3C(=C4C(=CC(=O)CC4)CC3)C(CC12C)c1ccc(N(C)C)cc1)C#CC
578	Miglitol	60	0.1761	1	OC1C(O)C(O)CN(CCO)C1CO
579	Miglustat	97	1.5097	1	OC1C(O)C(O)CN(CCCC)C1CO
580	Milnacipran	84	0.7202	1	O=C(N(CC)CC)C1(CC1CN)c1ccccc1
581	Milrinone	86	0.7884	1	O=C1NC(C)=C(C=C1C#N)c1ccncc1
582	Minocycline	97	1.5097	1	CN(C1C2CC3CC4=C(N(C)C)C=CC(O)=C4C(O)=C3C(C2(O)C(O)=C(C(N)=O)C1=O)=O)C
583	Minoxidil	95	1.2788	1	C1CCN(CC1)C2=NC(=N)N(C(=C2)N)O
584	Mirtazapine	50	0.0000	1	n1c2N3C(c4c(Cc2ccc1)cccc4)CN(CC3)C
585	Misoprostol	80	0.6021	1	OC1CC(=O)C(CCCCCCC(OC)=O)C1C=CCC(O)(CCCC)C
586	Mitotane	40	-0.1761	0	Clc1ccccc1C(C(Cl)Cl)c1ccc(Cl)cc1
587	Mitoxantrone	0	-2.0000	0	Oc1c2c(C(=O)c3c(C2=O)c(NCCNCCO)ccc3NCCNCCO)c(O)cc1
588	Mivacurium	0	-2.0000	0	O(C)c1c(OC)cc(cc1OC)CC1[N+](CCc2c1cc(OC)c(OC)c2)(CCCOC(=O)CCC=CCCC(OCCC[N+]1(CCc2c(cc(OC)c(OC)c2)C1Cc1cc(OC)c(OC)c(OC)c1)C)=O)C
589	Mizolastine	65	0.2688	1	Fc1ccc(cc1)Cn1c2c(nc1N1CCC(N(C)C=3NC(=O)C=CN=3)CC1)cccc2
590	Moclobemide	75	0.4771	1	Clc1ccc(cc1)C(=O)NCCN1CCOCC1
591	Modafinil	80	0.6021	1	S(=O)(C(c1ccccc1)c1ccccc1)CC(=O)N
592	Moexipril	13	-0.8256	0	O(C)c1cc2CC(N(Cc2cc1OC)C(=O)C(NC(CCc1ccccc1)C(OCC)=O)C)C(O)=O
593	Molsidomine	48	-0.0348	0	o1n[n+](N2CCOCC2)cc1NC(OCC)=O

594	Montelukast	62	0.2126	1	Clc1cc2nc(ccc2cc1)C=Cc1cc(ccc1)C(SCC1(CC1)CC(O)=O)CCc1ccccc1C(O)(C)C
595	Moricizine	38	-0.2126	0	S1c2c(N(c3c1cccc3)C(=O)CCN1CCOCC1)cc(NC(OCC)=O)cc2
596	Morphine	24	-0.5006	0	O1C2C34C(C(N(CC3)C)Cc3c4c1c(O)cc3)C=CC2O
597	Moxalactam	85	0.7533	1	S(CC=1COC2N(C(=O)C2(OC)NC(=O)C(C(O)=O)c2ccc(O)cc2)C=1C(O)=O)c1[n+](nn[nH]1)C
598	Moxestrol	33	-0.3076	0	O(C)C1CC2(C(C3C1c1c(cc(O)cc1)CC3)CCC2(O)C#C)C
599	Moxifloxacin	86	0.7884	1	Fc1cc2c(N(C=C(C(O)=O)C2=O)C2CC2)c(OC)c1N1CC2C(NCCC2)C1
600	Moxonidine	88	0.8653	1	Clc1nc(nc(OC)c1NC=1NCCN=1)C
601	Nabilone	20	-0.6021	0	O1c2c(C3C(CCC(=O)C3)C1(C)C)c(O)cc(c2)C(CCCCCC)(C)C
602	Nabumetone	35	-0.2688	0	O(C)c1cc2c(cc(cc2)CCC(=O)C)cc1
603	N-ACETYLPROCAINAMIDE	83	0.6886	1	O=C(NCCN(CC)CC)c1ccc(NC(=O)C)cc1
604	Nadolol	30	-0.3680	0	O(CC(O)CNC(C)(C)C)c1c2CC(O)C(O)Cc2ccc1
605	NAFCILLIN	30	-0.3680	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)c1c2c(ccc1OCC)cccc2
606	Nalbuphine	16	-0.7202	0	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2O)CC1CCC1
607	Nalidixic acid	96	1.3802	1	O=C1c2ccc(nc2N(C=C1C(O)=O)CC)C
608	Nalmefene	40	-0.1761	0	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=C)CC1CC1
609	Naloxone	2	-1.6902	0	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=O)CC=C
610	Naltrexone	22	-0.5497	0	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=O)CC1CC1

611	Nandrolone	2.24	-1.6399	0	OC1CCC2C3C(C4C(=CC(=O)CC4)CC3)CCC12C
612	Naproxen	99	1.9956	1	O(C)c1cc2c(cc(cc2)C(C(O)=O)C)cc1
613	Naratriptan	69	0.3475	1	S(=O)(=O)(NC)CCc1cc2cc([nH]c2cc1)C1CCN(CC1)C
614	Naringin	20	-0.6021	0	O1C(CO)C(O)C(O)C(OC2OC(C)C(O)C(O)C2O)C1Oc1cc(O)c2c(OC(CC2=O)c2ccc(O)cc2)c1
615	Nateglinide	73	0.4320	1	OC(=O)C(NC(=O)C1CCC(CC1)C(C)C)Cc1ccccc1
616	Nebivolol	50	0.0000	1	Fc1cc2CCC(Oc2cc1)C(O)CNCC(O)C1Oc2c(cc(F)cc2)CC1
617	Nedocromil	3	-1.5097	0	O1c2c(cc3c(N(CC)C(=CC3=O)C(O)=O)c2CCC)C(=O)C=C1C(O)=O
618	Nefazodone	20	-0.6021	0	Clc1cc(N2CCN(CC2)CCCN2N=C(N(CCOc3ccccc3)C2=O)CC)ccc1
619	Nefopam	36	-0.2499	0	O1CCN(Cc2c(cccc2)C1c1ccccc1)C
620	Nelarabine	11	-0.9080	0	O1C(CO)C(O)C(O)C1n1c2nc(nc(OC)c2nc1)N
621	Nelfinavir	50	0.0000	1	S(CC(NC(=O)c1cccc(O)c1C)C(O)CN1CC2C(CC1C(=O)NC(C)(C)C)CCCC2)c1ccccc1
622	Neomycin	3	-1.5097	0	C1C(C(C(C(C1N)OC2C(C(C(C(O2)CN)O)O)N)OC3C(C(C(O3)CO)OC4C(C(C(C(O4)CN)O)O)N)O)O)N
623	Neostigmine	2	-1.6902	0	O(C(=O)N(C)C)c1cc([N+](C)(C)C)ccc1
624	Netilmicin	0	-2.0000	0	O1C(OC2C(O)C(OC3OCC(O)(C)C(NC)C3O)C(NCC)CC2N)C(N)CC=C1CN
625	Nevirapine	93	1.1234	1	O=C1Nc2c(nccc2C)N(c2ncccc12)C1CC1
626	Nicainoprol	70	0.3680	1	O(CC(O)CNC(C)C)c1c2N(CCCc2ccc1)C(=O)c1cccnc1
627	Nicardipine	23	-0.5248	0	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)CCN(Cc1ccccc1)C
628	Nicergoline	5	-1.2788	0	Brc1cc(cnc1)C(OCC1CC2(OC)C(N(C1)C)Cc1c3c2cccc3n(c1)C)=O

629	Nicorandil	75	0.4771	1	O([N+](=O)[O-])CCNC(=O)c1cccnc1
630	Nicotine	30	-0.3680	0	n1cc(ccc1)C1N(CCC1)C
631	Nifedipine	50	0.0000	1	O(C(=O)C)C=1C(C(C(OC)=O)=C(NC=1C)C)c1ccccc1[N+](=O)[O-]
632	Nilotinib	30	-0.3680	0	FC(F)(F)c1cc(NC(=O)c2cc(Nc3nc(ccn3)-c3cccnc3)c(cc2)C)cc(-n2cc(nc2)C)c1
633	Nilutamide	99	1.9956	1	FC(F)(F)c1cc(N2C(=O)C(NC2=O)(C)C)ccc1[N+](=O)[O-]
634	Nilvadipine	14	-0.7884	0	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C#N)c1cc([N+](=O)[O-])ccc1)C(C)C
635	Nimodipine	11	-0.9080	0	O(C(=O)C=1C(C(C(OCCOC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)C(C)C
636	Nisoldipine	5	-1.2788	0	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1ccccc1[N+](=O)[O-])CC(C)C
637	Nitazoxanide	70	0.3680	1	s1c([N+](=O)[O-])cnc1NC(=O)c1ccccc1OC(=O)C
638	Nitisinone	90	0.9542	1	FC(F)(F)c1cc([N+](=O)[O-])c(cc1)C(=O)C1C(=O)CCCC1=O
639	Nitrazepam	78	0.5497	1	O=C1Nc2c(cc([N+](=O)[O-])cc2)C(=NC1)c1ccccc1
640	Nitrendipine	16.5	-0.7042	0	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)CC
641	Nitrofurantoin	87	0.8256	1	o1c(ccc1[N+](=O)[O-])C=NN1CC(=O)NC1=O
642	Nitroglycerin	1	-1.9956	0	O=N(=O)OCC(CON(=O)=O)ON(=O)=O
643	Nizatidine	70	0.3680	1	s1cc(nc1CN(C)C)CSCCNC(NC)=C[N+](=O)[O-]
644	Nomifensine	27	-0.4320	0	Nc1c2c(ccc1)C(CN(C2)C)c1ccccc1
645	Norepinephrine	3	-1.5097	0	Oc1cc(ccc1O)C(O)CN
646	Norethindrone	64	0.2499	1	OC1(CCC2C3C(C4C(=CC(=O)CC4)CC3)CCC12C)C#C
647	Norfenefrine	5	-1.2788	0	Oc1cc(ccc1)C(O)CN

648	Norfenfluramine	85	0.7533	1	FC(F)(F)c1cc(ccc1)CC(N)C
649	Norfloxacin	35	-0.2688	0	Fc1cc2c(N(C=C(C(O)=O)C2=O)CC)cc1N1CCNCC1
650	Norgestrel	94	1.1950	1	OC1(CCC2C3C(C4C(=CC(=O)CC4)CC3)CCC12CC)C#C
651	Nortilidine	99	1.9956	1	O(C(=O)C1(CCC=CC1NC)c1ccccc1)CC
652	NORTRIPTYLINE	51	0.0174	1	N(CCC=C1c2c(CCc3c1cccc3)cccc2)C
653	Norzimelidine	66	0.2881	1	Brc1ccc(cc1)C(=CCNC)c1cccnc1
654	Noscapine	30	-0.3680	0	O1C(c2c(c(OC)c(OC)cc2)C1=O)C1N(CCc2c1c(OC)c1OCOc1c2)C
655	Novobiocin	0	-2.0000	0	O1C(C)(C)C(OC)C(OC(=O)N)C(O)C1Oc1ccc2c(OC(O)=C(NC(=O)c3cc(CC=C(C)C)c(O)cc3)C2=O)c1C
656	Nystatin	1	-1.9956	0	O1C(C)C(O)C(N)C(O)C1OC1C=CC=CC=CC=CCCC=CC=CC(C)C(O)C(C)C(OC(=O)CC(O)CC(O)CC(O)CCC(O)C(O)CC2(OC(C1)C(C(O)=O)C(O)C2)O)C
657	octreotide	2	-1.6902	0	S1SCC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(N)Cc2ccccc2)C1)Cc1ccccc1)Cc1c2c([nH]c1)cccc2)CCCCN)C(O)C)C(=O)NC(C(O)C)CO
658	Olanzapine	60	0.1761	1	s1c2Nc3c(N=C(N4CCN(CC4)C)c2cc1C)cccc3
659	Olmesartan	26	-0.4543	0	O1C(COC(=O)c2n(Cc3ccc(cc3)-c3ccccc3-c3nn[nH]n3)c(nc2C(O)(C)C)CCC)=C(OC1=O)C
660	Olsalazine	2	-1.6902	0	Oc1ccc(N=Nc2cc(C(O)=O)c(O)cc2)cc1C(O)=O
661	Ondansetron	62	0.2126	1	O=C1c2c(n(c3c2cccc3)C)CCC1Cn1ccnc1C
662	Orlistat	1	-1.9956	0	O1C(CC(OC(=O)C(NC=O)CC(C)C)CCCCCCCCCCC)C(CCCCCC)C1=O
663	Orphenadrine	90	0.9542	1	O(C(c1ccccc1C)c1ccccc1)CCN(C)C
664	Oseltamivir	75	0.4771	1	O(C(CC)CC)C1C=C(CC(N)C1NC(=O)C)C(OCC)=O
665	Oseltamivir acid	79	0.5754	1	O(C(CC)CC)C1C=C(CC(N)C1NC(=O)C)C(O)=O

666	Ouabain	1	-1.9956	0	O1C(C)C(O)C(O)C(O)C1OC1CC2(O)CCC3C(C2(CO)C(O)C1)C(O)CC1(C)C(CCC13O)C1=CC(OC1)=O
667	OXACILLIN	33	-0.3076	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)c1c(noc1C)-c1ccccc1
668	Oxalic_acid	7	-1.1234	0	OC(=O)C(O)=O
669	Oxandrolone	97	1.5097	1	O1CC2(C(CC1=O)CCC1C3CCC(O)(C)C3(CCC12)C)C
670	Oxaprozin	98	1.6902	1	o1c(c(nc1CCC(O)=O)-c1ccccc1)-c1ccccc1
671	Oxazepam	97	1.5097	1	Clc1cc2c(NC(=O)C(O)N=C2c2ccccc2)cc1
672	Oxcarbazepine	95	1.2788	1	O=C1Cc2c(N(c3c1cccc3)C(=O)N)cccc2
673	Oxiracetam	75	0.4771	1	OC1CC(=O)N(C1)CC(=O)N
674	Oxitropium	1	-1.9956	0	O1C2C3[N+](C(CC(OC(=O)C(CO)c4ccccc4)C3)C12)(CC)C
675	Oxprenolol	44	-0.1047	0	O(CC(O)CNC(C)C)c1ccccc1OCC=C
676	Oxybutynin	6	-1.1950	0	O(C(=O)C(O)(C1CCCCC1)c1ccccc1)CC#CCN(CC)CC
677	Oxycodone	42	-0.1402	0	O1C2C34CCN(C(Cc5c3c1c(OC)cc5)C4(O)CCC2=O)C
678	Oxyfedrine	85	0.7533	1	O(C)c1cc(ccc1)C(=O)CCNC(C(O)c1ccccc1)C
679	Oxymetazoline	11	-0.9080	0	Oc1c(C)c(CC=2NCCN=2)c(cc1C(C)(C)C)C
680	Oxymorphone	10	-0.9542	0	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=O)C
681	Oxytetracycline	58	0.1402	1	CC1(C2C(C3C(C(=O)C(=C(C3(C(=O)C2=C(C4=C1C=CC=C4O)O)O)O)C(=O)N)N(C)C)O)O
682	Paclitaxel	6.5	-1.1579	0	O1C2CC(O)C3(C(C(OC(=O)c4ccccc4)C4(O)CC(OC(=O)C(O)C(NC(=O)c5ccccc5)c5ccccc5)C(=C(C4(C)C)C(OC(=O)C)C3=O)C)C2(OC(=O)C)C1)C
683	Paliperidone	28	-0.4102	0	Fc1cc2onc(c2cc1)C1CCN(CC1)CCC=1C(=O)N2C(=NC=1C)C(O)CCC2

684	Palonosetron	97	1.5097	1	O=C1N(CC2CCCc3c2c1ccc3)C1C2CCN(C1)CC2
685	Pamidronate	1	-1.9956	0	P(O)(O)(=O)C(P(O)(O)=O)(O)CCN
686	Pancuronium	0	-2.0000	0	O(C(=O)C)C1C2(C(CC1[N+]1(CCCCC1)C)C1C(CC2)C2(CC([N+]3(CCCCC3)C)C(OC(=O)C)CC2CC1)C)C
687	Papaverine	28	-0.4102	0	O(C)c1cc(ccc1OC)Cc1nccc2c1cc(OC)c(OC)c2
688	Paricalcitol	72	0.4102	1	OC1CC(O)CC(C1)=CC=C1C2CCC(C(C=CC(C(O)(C)C)C)C)C2(CCC1)C
689	Paromomycin	2	-1.6902	0	O1C(CN)C(O)C(O)C(N)C1OC1C(O)C(OC1CO)OC1C(OC2OC(CO)C(O)C(O)C2N)C(N)CC(N)C1O
690	Paroxetine	50	0.0000	1	Fc1ccc(cc1)C1CCNCC1COc1cc2OCOc2cc1
691	Pefloxacin	90	0.9542	1	Fc1cc2c(N(C=C(C(O)=O)C2=O)CC)cc1N1CCN(CC1)C
692	Penbutolol	93	1.1234	1	O(CC(O)CNC(C)(C)C)c1ccccc1C1CCCC1
693	Penciclovir	77	0.5248	1	O=C1N=C(Nc2n(cnc12)CCC(CO)CO)N
694	Penclomedine	49	-0.0174	0	Clc1c(OC)c(Cl)c(OC)nc1C(Cl)(Cl)Cl
695	Penicillamine	50	0.0000	1	SC(C(N)C(O)=O)(C)C
696	Penicillin_G	30	-0.3680	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)Cc1ccccc1
697	Pentaerythritol Tetranitrate	60	0.1761	1	C(C(CO[N+](=O)[O-])(CO[N+](=O)[O-])CO[N+](=O)[O-])O[N+](=O)[O-]
698	pentamidine	0	-2.0000	0	O(CCCCCOc1ccc(cc1)C(N)=N)c1ccc(cc1)C(N)=N
699	Pentazocine	18	-0.6585	0	Oc1cc2c(CC3N(CCC2(C)C3C)CC=C(C)C)cc1
700	Pentobarbital	80	0.6021	1	O=C1NC(=O)NC(=O)C1(C(CCC)C)CC
701	Pentopril	58	0.1402	1	O(C(=O)C(CC(C(=O)N1c2c(CC1C(O)=O)cccc2)C)C)CC

702	Pentosan	3	-1.5097	0	C1C(C(C(C(O1)OC2COC(C(C2O)O)OC3COC(C(C3O)O)O)O)O)O
703	Pentosan Polysulfate	3	-1.5097	0	S(OC1C(OS(O)(=O)=O)C(OCC1OC1OCC(O)C(OS(O)(=O)=O)C1OS(O)(=O)=O)O)(O)(=O)=O
704	Pentoxifylline	25	-0.4771	0	O=C1N(CCCCC(=O)C)C(=O)N(c2ncn(c12)C)C
705	Pergolide	38	-0.2126	0	S(CC1CC2C(N(C1)CCC)Cc1c3c2cccc3[nH]c1)C
706	Perhexiline	73.5	0.4430	1	N1CCCCC1CC(C1CCCCC1)C1CCCCC1
707	Perindopril	25	-0.4771	0	O(C(=O)C(NC(C(=O)N1C2C(CC1C(O)=O)CCCC2)C)CCC)CC
708	Perindoprilat	25	-0.4771	0	OC(=O)C1N(C2C(C1)CCCC2)C(=O)C(NC(CCC)C(O)=O)C
709	Permethrin	60	0.1761	1	ClC(Cl)=CC1C(C)(C)C1C(OCc1cc(Oc2ccccc2)ccc1)=O
710	Perphenazine	40	-0.1761	0	Clc1cc2N(c3c(Sc2cc1)cccc3)CCCN1CCN(CC1)CCO
711	Phenacetin	25	-0.4771	0	O(CC)c1ccc(NC(=O)C)cc1
712	Phencyclidine	72	0.4102	1	N1(CCCCC1)C1(CCCCC1)c1ccccc1
713	Phenformin	55	0.0872	1	N(=C(N=C(N)N)N)CCc1ccccc1
714	Phenindione	97	1.5097	1	O=C1c2c(cccc2)C(=O)C1c1ccccc1
715	Pheniramine	99	1.9956	1	CN(C)CCC(C1=CC=CC=C1)C2=CC=CC=N2
716	Phenobarbital	99	1.9956	1	O=C1NC(=O)NC(=O)C1(CC)c1ccccc1
717	Phenoperidine	20	-0.6021	0	O(C(=O)C1(CCN(CC1)CCC(O)c1ccccc1)c1ccccc1)CC
718	Phenoxybenzamine	25	-0.4771	0	ClCCN(Cc1ccccc1)C(COc1ccccc1)C
719	Phenylbutazone	90	0.9542	1	O=C1C(N(N(C1=O)c1ccccc1)c1ccccc1)CCCC

720	Phenylbutyrate	78	0.5497	1	OC(=O)CCCc1ccccc1
721	Phenylephrine	35	-0.2688	0	Oc1cc(ccc1)C(O)CNC
722	phenylpropanolamine	95	1.2788	1	OC(C(N)C)c1ccccc1
723	Phenytoin	90	0.9542	1	O=C1NNC(=O)C1(c1ccccc1)c1ccccc1
724	Physostigmine	10	-0.9542	0	O(C(=O)NC)c1cc2c(N(C3N(CCC23C)C)C)cc1
725	Pimecrolimus	0	-2.0000	0	ClC1CCC(CC1OC)C=C(C)C1OC(=O)C2N(CCCC2)C(=O)C(=O)C2(OC(C(OC)CC2C)C(OC)CC(CC(=CC(CC)C(=O)CC(O)C1C)C)C)O
726	Pimexone	65	0.2688	1	O1c2c(ccc(OC)c2CN2CCCCC2)C(=O)c2c1cccc2
727	Pimozide	40	-0.1761	0	Fc1ccc(cc1)C(CCCN1CCC(N2c3c(NC2=O)cccc3)CC1)c1ccc(F)cc1
728	Pinacidil	57	0.1224	1	n1ccc(NC(=NC#N)NC(C(C)(C)C)C)cc1
729	Pindolol	88	0.8653	1	O(CC(O)CNC(C)C)c1c2c([nH]cc2)ccc1
730	Pioglitazone	83	0.6886	1	S1C(Cc2ccc(OCCc3ncc(cc3)CC)cc2)C(=O)NC1=O
731	Pipemidic_acid	93	1.1234	1	O=C1c2c(nc(nc2)N2CCNCC2)N(C=C1C(O)=O)CC
732	Piperacillin	0	-2.0000	0	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(NC(=O)N1CCN(CC)C(=O)C1=O)c1ccccc1
733	Pipotiazine	26	-0.4543	0	S1c2c(N(c3c1cccc3)CCCN1CCC(CC1)CCO)cc(S(=O)(=O)N(C)C)cc2
734	Pirazolac	93.5	1.1579	1	Clc1ccc(cc1)-c1cn(nc1CC(O)=O)-c1ccc(F)cc1
735	Pirenzepine	14	-0.7884	0	O=C1Nc2cccnc2N(c2c1cccc2)C(=O)CN1CCN(CC1)C
736	Piretanide	92	1.0607	1	S(=O)(=O)(N)c1cc(cc(N2CCCC2)c1Oc1ccccc1)C(O)=O
737	Piritrexim	68	0.3274	1	O(C)c1ccc(OC)cc1Cc1nc2c(nc1)c(C)c(nc2N)N
738	Pirmenol	83	0.6886	1	OC(CCCN1C(CCCC1C)C)(c1ccccc1)c1ncccc1

739	Piroxicam	99	1.9956	1	S1(=O)(=O)N(C)C(C(=O)Nc2ncccc2)=C(O)c2c1cccc2
740	Piroximone	81	0.6297	1	O=C1NC(C(=O)c2ccncc2)=C(N1)CC
741	Pivampicillin	30.9	-0.3495	0	S1C2N(C(C(OCOC(=O)C(C)(C)C)=O)C1(C)C)C(=O)C2NC(=O)C(N)c1ccccc1
742	Pizotyline	78	0.5497	1	s1c2c(cc1)C(c1c(CC2)cccc1)=C1CCN(CC1)C
743	Porfimer	0	-2.0000	0	O(C(C)c1c2[nH]c(C=c3[nH]c(=CC4=NC(=CC5=NC(=C2)C(C)=C5CCC(O)=O)C(CCC(O)=O)=C4C)c(C)c3C(O)C)c1C)C(C)c1c=2[nH]c(=CC3=NC(=CC4=NC(=Cc5[nH]c(C=2)c(C)c5C(O)C)C(C)=C4CCC(O)=O)C(CCC(O)=O)=C3C)c1C
744	Posaconazole	96	1.3802	1	Fc1cc(F)ccc1C1(OCC(C1)COc1ccc(N2CCN(CC2)c2ccc(N3C=NN(C(C(O)C)CC)C3=O)cc2)cc1)Cn1ncnc1
745	Practolol	99	1.9956	1	O(CC(O)CNC(C)C)c1ccc(NC(=O)C)cc1
746	Pramipexole	93	1.1234	1	s1c2CC(NCCC)CCc2nc1N
747	Pravastatin	18	-0.6585	0	O(C(=O)C(CC)C)C1C2C(=CC(O)C1)C=CC(C)C2CCC(O)CC(O)CC(O)=O
748	Praziquantel	7	-1.1234	0	O=C1N2C(c3c(CC2)cccc3)CN(C1)C(=O)C1CCCCC1
749	Prazosin	68	0.3274	1	o1cccc1C(=O)N1CCN(CC1)c1nc(N)c2cc(OC)c(OC)cc2n1
750	Prednisolone	82	0.6585	1	OC1(CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)C(=O)CO
751	Prednisolone phosphate	82	0.6585	1	P(OCC(=O)C1(O)CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)(O)(O)=O
752	Prednisone	70	0.3680	1	OC1(CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(=O)CC12C)C(=O)CO
753	Pregabalin	90	0.9542	1	OC(=O)CC(CC(C)C)CN
754	Prenalterol	33	-0.3076	0	O(CC(O)CNC(C)C)c1ccc(O)cc1
755	Primaquine	96	1.3802	1	O(C)c1cc(NC(CCCN)C)c2ncccc2c1
756	Primidone	92	1.0607	1	O=C1NCNC(=O)C1(CC)c1ccccc1

757	Probenecid	99	1.9956	1	S(=O)(=O)(N(CCC)CCC)c1ccc(cc1)C(O)=O
758	Probucol	5	-1.2788	0	S(C(Sc1cc(C(C)(C)C)c(O)c(c1)C(C)(C)C)(C)C)c1cc(C(C)(C)C)c(O)c(c1)C(C)(C)C
759	Procainamide	83	0.6886	1	O=C(NCCN(CC)CC)c1ccc(N)cc1
760	Procaine	0	-2.0000	0	O(C(=O)c1ccc(N)cc1)CCN(CC)CC
761	Procaterol	40	-0.1761	0	Oc1c2NC(=O)C=Cc2c(cc1)C(O)C(NC(C)C)CC
762	Prochlorperazine	15	-0.7533	0	Clc1cc2N(c3c(Sc2cc1)cccc3)CCCN1CCN(CC1)C
763	Progabide	60	0.1761	1	Clc1ccc(cc1)C(NCCCC(=O)N)=C1C=C(F)C=CC1=O
764	Progesterone	25	-0.4771	0	O=C1CCC2(C3C(C4CCC(C(=O)C)C4(CC3)C)CCC2=C1)C
765	Proguanil	60	0.1761	1	Clc1ccc(NC(=NC(=NC(C)C)N)N)cc1
766	Promazine	16.25	-0.7121	0	S1c2c(N(c3c1cccc3)CCCN(C)C)cccc2
767	Promethazine	25	-0.4771	0	S1c2c(N(c3c1cccc3)CC(N(C)C)C)cccc2
768	Propafenone	12	-0.8653	0	O(CC(O)CNCCC)c1ccccc1C(=O)CCc1ccccc1
769	Propiomazine	33	-0.3076	0	S1c2c(N(c3c1cccc3)CC(N(C)C)C)cc(cc2)C(=O)CC
770	Propranolol	26	-0.4543	0	O(CC(O)CNC(C)C)c1c2c(ccc1)cccc2
771	Propylthiouracil	85	0.7533	1	S=C1NC(=CC(=O)N1)CCC
772	Proquazone	7	-1.1234	0	O=C1N=C(c2c(N1C(C)C)cc(cc2)C)c1ccccc1
773	Proscillaridin	7	-1.1234	0	O1C(C)C(O)C(O)C(O)C1OC1CCC2(C3C(CCC2=C1)C1(O)CCC(C1(CC3)C)C=1C=CC(OC=1)=O)C
774	Prothionamide	99	1.9956	1	S=C(N)c1cc(ncc1)CCC

775	Protriptyline	85	0.7533	1	N(CCCC1c2c(C=Cc3c1cccc3)cccc2)C
776	Proxyphylline	99	1.9956	1	O=C1N(C)C(=O)N(c2ncn(c12)CC(O)C)C
777	Pyrazinamide	90	0.9542	1	O=C(N)c1nccnc1
778	Pyridostigmine	7.6	-1.0849	0	O(C(=O)N(C)C)c1ccc[n+](c1)C
779	Pyrimethamine	90	0.9542	1	Clc1ccc(cc1)-c1c(nc(nc1N)N)CC
780	Pyrvinium	5	-1.2788	0	[n+]1(c2c(cc(N(C)C)cc2)ccc1C=Cc1cc(n(c1C)-c1ccccc1)C)C
781	Quazepam	32	-0.3274	0	Clc1cc2c(N(CC(F)(F)F)C(=S)CN=C2c2ccccc2F)cc1
782	Quetiapine	9	-1.0048	0	S1c2c(cccc2)C(=Nc2c1cccc2)N1CCN(CC1)CCOCCO
783	Quinagolide	99	1.9956	1	S(=O)(=O)(NC1CC2C(N(C1)CCC)Cc1c(C2)c(O)ccc1)N(CC)CC
784	Quinapril	60	0.1761	1	O(C(=O)C(NC(C(=O)N1CCc2c(cccc2)C1C(O)=O)C)CCc1ccccc1)CC
785	Quinaprilat	52	0.0348	1	OC(=O)C1N(c2c(CC1)cccc2)C(=O)C(NC(CCc1ccccc1)C(O)=O)C
786	Quinidine	75	0.4771	1	O(C)c1cc2c(nccc2C(O)C2N3CC(C(C2)CC3)C=C)cc1
787	Rabeprazole	52	0.0348	1	S(=O)(Cc1nccc(OCCCOC)c1C)c1[nH]c2c(n1)cccc2
788	Raffinose	0	-2.0000	0	O1C(OC2OC(COC3OC(CO)C(O)C(O)C3O)C(O)C(O)C2O)(CO)C(O)C(O)C1CO
789	Raloxifene	2	-1.6902	0	s1c2c(ccc(O)c2)c(C(=O)c2ccc(OCCN3CCCCC3)cc2)c1-c1ccc(O)cc1
790	Raltegravir	31.8	-0.3314	0	Fc1ccc(cc1)CNC(O)=C1N=C(N(C)C(=O)C1=O)C(NC(=O)c1oc(nn1)C)(C)C
791	Raltitrexed	15	-0.7533	0	s1c(ccc1N(Cc1cc2c(NC(=NC2=O)C)cc1)C)C(=O)NC(CCC(O)=O)C(O)=O
792	Ramelteon	1.8	-1.7368	0	O1CCc2c3c(CCC3CCNC(=O)CC)ccc12
793	Ramipril	28	-0.4102	0	O(C(=O)C(NC(C(=O)N1C2C(CC1C(O)=O)CCC2)C)CCc1ccccc1)CC

794	Ramiprilat	44	-0.1047	0	OC(=O)C1N(C2C(C1)CCC2)C(=O)C(NC(CCc1ccccc1)C(O)=O)C
795	Ramixotidine	60	0.1761	1	O=n1ccccc1C(=O)NCCSCc2ccc(CN(C)C)o2
796	Ranitidine	52	0.0348	1	S(Cc1oc(cc1)CN(C)C)CCNC(NC)=C[N+](=O)[O-]
797	Ranolazine	42.5	-0.1313	0	O(CC(O)CN1CCN(CC1)CC(=O)Nc1c(cccc1C)C)c1ccccc1OC
798	Rasagiline	36	-0.2499	0	N(CC#C)C1CCc2c1cccc2
799	Reboxetine	94	1.1950	1	O1CCNCC1C(Oc1ccccc1OCC)c1ccccc1
800	Recainam	73	0.4320	1	O=C(Nc1c(cccc1C)C)NCCCNC(C)C
801	Remikiren	1	-1.9956	0	S(=O)(=O)(C(C)(C)C)CC(Cc1ccccc1)C(=O)NC(Cc1[nH]cnc1)C(=O)NC(CC1CCCCC1)C(O)C(O)C1CC1
802	Remoxipride	90	0.9542	1	Brc1ccc(OC)c(C(=O)NCC2N(CCC2)CC)c1OC
803	Repaglinide	56	0.1047	1	O(CC)c1cc(ccc1C(O)=O)CC(=O)NC(CC(C)C)c1ccccc1N1CCCCC1
804	Reproterol	50	0.0000	1	Oc1cc(cc(O)c1)C(O)CNCCCn1c2c(nc1)N(C)C(=O)N(C)C2=O
805	Reserpine	50	0.0000	1	O(C)C1C(C2C(CC1OC(=O)c1cc(OC)c(OC)c(OC)c1)CN1C(C3=Nc4c(C3CC1)cccc4OC)C2)C(OC)=O
806	Ribavirin	60	0.1761	1	O1C(CO)C(O)C(O)C1n1nc(nc1)C(=O)N
807	Rifabutin	20	-0.6021	0	O1c2c3c4c(c(O)c2C)C(=O)C(NC(=O)C(=CC=CC(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)C=COC1(C)C3=O)C)=C1NC2(N=C14)CCN(CC2)CC(C)C
808	Rifampin	70	0.3680	1	O1c2c3c4c(c(O)c2C)c(O)c(NC(=O)C(=CC=CC(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)C=COC1(C)C3=O)C)c(C=NN1CCN(CC1)C)c4O
809	Rifapentine	70	0.3680	1	O1c2c3c4c(c(O)c2C)c(O)c(NC(=O)C(=CC=CC(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)C=COC1(C)C3=O)C)c(C=NN1CCN(CC1)C1CCCC1)c4O
810	Rifaximin	1	-1.9956	0	O1c2c3c4c(c(O)c2C)c(O)c(NC(=O)C(=CC=CC(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)C=COC1(C)C3=O)C)c1n2c(nc14)C=C(C=C2)C
811	Riluzole	64	0.2499	1	s1c2cc(OC(F)(F)F)ccc2nc1N

812	Risedronate	1	-1.9956	0	P(O)(O)(=O)C(P(O)(O)=O)(O)Cc1cccnc1
813	Risperidone	66	0.2881	1	Fc1cc2onc(c2cc1)C1CCN(CC1)CCC=1C(=O)N2C(=NC=1C)CCCC2
814	Ritodrine	30	-0.3680	0	Oc1ccc(cc1)C(O)C(NCCc1ccc(O)cc1)C
815	Ritonavir	70	0.3680	1	s1cncc1COC(=O)NC(Cc1ccccc1)C(O)CC(NC(=O)C(NC(=O)N(Cc1nc(sc1)C(C)C)C)C(C)C)Cc1ccccc1
816	Rivastigmine	72	0.4102	1	O(C(=O)N(CC)C)c1cc(ccc1)C(N(C)C)C
817	Rizatriptan	45	-0.0872	0	[nH]1cc(c2cc(ccc12)Cn1ncnc1)CCN(C)C
818	Rofecoxib	93	1.1234	1	S(=O)(=O)(C)c1ccc(cc1)C=1COC(=O)C=1c1ccccc1
819	Rolipram	75	0.4771	1	O(c1cc(ccc1OC)C1CC(=O)NC1)C1CCCC1
820	Ropinirole	55	0.0872	1	O=C1Nc2c(C1)c(ccc2)CCN(CCC)CCC
821	Roquinimex	99	1.9956	1	OC=1N(c2c(cccc2)C(=O)C=1C(=O)N(C)c1ccccc1)C
822	Rosaramicin	32	-0.3274	0	O1C(OC2C(C)C(O)CC(OC(CC)C(C3OC3(C=CC(=O)C(CC2CC=O)C)C)C)=O)C(O)C(N(C)C)CC1C
823	Rosiglitazone	99	1.9956	1	S1C(Cc2ccc(OCCN(C)c3ncccc3)cc2)C(=O)NC1=O
824	Rosuvastatin	20	-0.6021	0	S(=O)(=O)(N(C)c1nc(-c2ccc(F)cc2)c(C=CC(O)CC(O)CC(O)=O)c(n1)C(C)C)C
825	Roxithromycin	50	0.0000	1	O1C(CC)C(O)(C)C(O)C(C)C(=NOCOCCOC)C(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(O)C)(C2)C)C(C)C1=O)C
826	Rufinamide	85	0.7533	1	Fc1cccc(F)c1Cn1nnc(c1)C(=O)N
827	Saccharin	84	0.7202	1	S1(=O)(=O)NC(=O)c2c1cccc2
828	S-Adenosylmethionine	5	-1.2788	0	S(=[CH]C1OC(n2c3ncnc(N)c3nc2)C(O)C1O)(CCC(N)C(O)=O)C
829	Salicylic Acid	99	1.9956	1	Oc1ccccc1C(O)=O

830	Scopolamine	27	-0.4320	0	O1C2C3N(C(CC(OC(=O)C(CO)c4ccccc4)C3)C12)C
831	Selegiline; (-)-Deprenil	4.4	-1.3370	0	N(C(Cc1ccccc1)C)(CC#C)C
832	Sematilide	61.3	0.1997	1	S(=O)(=O)(Nc1ccc(cc1)C(=O)NCCN(CC)CC)C
833	Sertaconazole	0	-2.0000	0	Clc1cc(Cl)ccc1C(OCc1c2c(sc1)c(Cl)ccc2)Cn1ccnc1
834	Sertindole	74	0.4543	1	Clc1cc2c(n(cc2C2CCN(CC2)CCN2CCNC2=O)-c2ccc(F)cc2)cc1
835	Sertraline	44	-0.1047	0	Clc1cc(ccc1Cl)C1CCC(NC)c2c1cccc2
836	Sildenafil	38	-0.2126	0	S(=O)(=O)(N1CCN(CC1)C)c1cc(C=2NC(=O)c3n(nc(c3N=2)CCC)C)c(OCC)cc1
837	Silodosin	32	-0.3274	0	FC(F)(F)COc1ccccc1OCCNC(Cc1cc(c2N(CCc2c1)CCCO)C(=O)N)C
838	Simvastatin	3	-1.5097	0	O1C(CC(O)CC1=O)CCC1C2C(=CC(CC2OC(=O)C(CC)(C)C)C)C=CC1C
839	Sirolimus	15	-0.7533	0	O1C(CC(=O)C(C=C(C)C(O)C(OC)C(=O)C(CC(C=CC=CC=C(C)C(OC)CC2OC(O)(C(=O)C(=O)N3C(CCCC3)C1=O)C(CC2)C)C)C)C)C(CC1CC(OC)C(O)CC1)C
840	Sitafloxacin	89	0.9080	1	Clc1c2N(C=C(C(O)=O)C(=O)c2cc(F)c1N1CC2(CC2)C(N)C1)C1CC1F
841	Sitagliptin	87	0.8256	1	Fc1cc(F)c(F)cc1CC(N)CC(=O)N1CCn2c(nnc2C(F)(F)F)C1
842	Solifenacin	90	0.9542	1	O(C(=O)N1CCc2c(cccc2)C1c1ccccc1)C1C2CCN(C1)CC2
843	Sorafenib	39	-0.1943	0	Clc1ccc(NC(=O)Nc2ccc(Oc3cc(ncc3)C(=O)NC)cc2)cc1C(F)(F)F
844	Sotalol	95	1.2788	1	CC(C)NCC(C1=CC=C(C=C1)NS(=O)(=O)C)O
845	Sparfloxacin	92	1.0607	1	Fc1c(N2CC(NC(C2)C)C)c(F)c2N(C=C(C(O)=O)C(=O)c2c1N)C1CC1
846	Spiramycin	35	-0.2688	0	O1C(C)C(OC2OC(C)C(O)C(O)(C2)C)C(N(C)C)C(O)C1OC1C(OC)C(O)CC(OC(CC=CC=CC(OC2OC(C)C(N(C)C)CC2)C(CC1CC=O)C)C)=O
847	Spironolactone	25	-0.4771	0	S(C(=O)C)C1C2C3CCC4(OC(=O)CC4)C3(CCC2C2(C(C1)=CC(=O)CC2)C)C

848	Stavudine	82	0.6585	1	O1C(C=CC1N1C=C(C)C(=O)NC1=O)CO
849	Streptomycin	1	-1.9956	0	O1C(CO)C(O)C(O)C(NC)C1OC1C(O)(C=O)C(OC1OC1C(NC(N)=N)C(O)C(NC(N)=N)C(O)C1O)C
850	Streptozocin	21	-0.5754	0	O1C(CO)C(O)C(O)C(NC(=O)N(N=O)C)C1O
851	Succimer	20	-0.6021	0	SC(C(S)C(O)=O)C(O)=O
852	Succinylcholine	0	-2.0000	0	O(C(=O)CCC(OCC[N+](C)(C)C)=O)CC[N+](C)(C)C
853	Sufentanil	0	-2.0000	0	s1cccc1CCN1CCC(N(C(=O)CC)c2ccccc2)(CC1)COC
854	Sulfadiazine	90	0.9542	1	S(=O)(=O)(Nc1ncccn1)c1ccc(N)cc1
855	Sulfadimethoxine	59.1	0.1599	1	S(=O)(=O)(Nc1nc(OC)nc(OC)c1)c1ccc(N)cc1
856	Sulfamerazine	81	0.6297	1	S(=O)(=O)(Nc1nc(ccn1)C)c1ccc(N)cc1
857	Sulfamethazine	86.5	0.8067	1	S(=O)(=O)(Nc1nc(cc(n1)C)C)c1ccc(N)cc1
858	Sulfamethoxazole	99	1.9956	1	S(=O)(=O)(Nc1noc(c1)C)c1ccc(N)cc1
859	Sulfasalazine	15	-0.7533	0	S(=O)(=O)(Nc1ncccc1)c1ccc(N=Nc2cc(C(O)=O)c(O)cc2)cc1
860	Sulfinpyrazone	99	1.9956	1	S(=O)(CCC1C(=O)N(N(C1=O)c1ccccc1)c1ccccc1)c1ccccc1
861	Sulfisoxazole	96	1.3802	1	CC1=C(ON=C1C)NS(=O)(=O)C2=CC=C(C=C2)N
862	Sulindac	88	0.8653	1	S(=O)(C)c1ccc(cc1)C=C1c2c(cc(F)cc2)C(CC(O)=O)=C1C
863	Sulpiride	27	-0.4320	0	S(=O)(=O)(N)c1cc(C(=O)NCC2CCN(C2)CC)c(OC)cc1
864	Sumatriptan	14	-0.7884	0	S(=O)(=O)(NC)Cc1cc2c([nH]cc2CCN(C)C)cc1
865	Sunitinib	54.5	0.0784	1	Fc1cc2c(NC(=O)C2=Cc2[nH]c(C)c(C(=O)NCCN(CC)CC)c2C)cc1
866	Suprofen	92	1.0607	1	s1cccc1C(=O)c1ccc(cc1)C(C(O)=O)C

867	Suramin	0	-2.0000	0	S(O)(=O)(=O)c1c2c(cc(S(O)(=O)=O)c1)c(S(O)(=O)=O)ccc2NC(=O)c1cc(NC(=O)c2cc(NC(=O)Nc3cc(ccc3)C(=O)Nc3cc(ccc3C)C(=O)Nc3c4c(cc(S(O)(=O)=O)cc4S(O)(=O)=O)c(S(O)(=O)=O)cc3)ccc2)c(cc1)C
868	Tacrine	17	-0.6886	0	n1c2c(CCCC2)c(N)c2c1cccc2
869	Tacrolimus	25	-0.4771	0	O1C(C(=CC2CC(OC)C(O)CC2)C)C(C)C(O)CC(=O)C(C=C(CC(CC(OC)C2OC(O)(C(=O)C(=O)N3C(CCCC3)C1=O)C(CC2OC)C)C)C)CC=C
870	Tadalafil	16	-0.7202	0	O1c2cc(ccc2OC1)C1N2C(Cc3c1[nH]c1c3cccc1)C(=O)N(CC2=O)C
871	Talinolol	55	0.0872	1	O(CC(O)CNC(C)(C)C)c1ccc(NC(=O)NC2CCCCC2)cc1
872	Tamsulosin	99	1.9956	1	S(=O)(=O)(N)c1cc(ccc1OC)CC(NCCOc1ccccc1OCC)C
873	Tasosartan	50	0.0000	1	O=C1N(c2nc(nc(c2CC1)C)C)Cc1ccc(cc1)-c1ccccc1-c1nn[nH]n1
874	Tegaserod	11	-0.9080	0	O(C)c1cc2c([nH]cc2C=NNC(NCCCCC)=N)cc1
875	Telbivudine	68	0.3274	1	O1C(CO)C(O)CC1N1C=C(C)C(=O)NC1=O
876	Telenzepine	54	0.0696	1	s1cc2c(N(c3c(NC2=O)cccc3)C(=O)CN2CCN(CC2)C)c1C
877	Telithromycin	57	0.1224	1	O1C(CC)C2(OC(=O)N(C2C(C)C(=O)C(CC(OC)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(=O)C(C)C1=O)C)CCCCn1cc(nc1)-c1cccnc1)C
878	Telmisartan	50	0.0000	1	CCCc1nc2c(C)cc(cc2n1Cc3ccc(cc3)c4ccccc4C(=O)O)c5nc6ccccc6n5C
879	Temafloxacin	95	1.2788	1	Fc1cc(F)ccc1N1C=C(C(O)=O)C(=O)c2cc(F)c(N3CC(NCC3)C)cc12
880	Temazepam	91	1.0048	1	Clc1cc2c(N(C)C(=O)C(O)N=C2c2ccccc2)cc1
881	Temozolomide	96	1.3802	1	O=C1N(N=Nc2c1[nH]nc2C(=O)N)C
882	Teniposide	41	-0.1581	0	s1cccc1C1OC2C(OC(OC3C4C(C(c5c3cc3OCOc3c5)c3cc(OC)c(O)c(OC)c3)C(OC4)=O)C(O)C2O)CO1
883	Tenofovir	25	-0.4771	0	P(O)(O)(=O)COC(Cn1c2ncnc(N)c2nc1)C
884	Tenoxicam	99	1.9956	1	s1c2c(S(=O)(=O)N(C)C(C(=O)Nc3ncccc3)=C2O)cc1

885	Terazosin	90	0.9542	1	O1CCCC1C(=O)N1CCN(CC1)c1nc(N)c2cc(OC)c(OC)cc2n1
886	Terbinafine	40	-0.1761	0	N(Cc1c2c(ccc1)cccc2)(CC=CC#CC(C)(C)C)C
887	Terbutaline	14	-0.7884	0	Oc1cc(cc(O)c1)C(O)CNC(C)(C)C
888	Terfenadine	1	-1.9956	0	OC(C1CCN(CC1)CCCC(O)c1ccc(cc1)C(C)(C)C)(c1ccccc1)c1ccccc1
889	Terguride	30	-0.3680	0	O=C(NC1CC2C(N(C1)C)Cc1c3c2cccc3[nH]c1)N(CC)CC
890	Terodiline	82	0.6585	1	N(C(C)(C)C)C(CC(c1ccccc1)c1ccccc1)C
891	Tertatolol	60	0.1761	1	S1c2c(CCC1)cccc2OCC(O)CNC(C)(C)C
892	Tesaglitazar	99	1.9956	1	S(Oc1ccc(cc1)CCOc1ccc(cc1)CC(OCC)C(O)=O)(=O)(=O)C
893	Testosterone	4	-1.3802	0	OC1CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)CC3)C
894	Tetrabenazine	5	-1.2788	0	O(C)c1cc2C3N(CC(CC(C)C)C(=O)C3)CCc2cc1OC
895	Tetracycline	77	0.5248	1	CC1(O)C2CC3C(N(C)C)C(C(C(N)=O)=C(O)C3(O)C(C2=C(O)C4=C1C=CC=C4O)=O)=O
896	Thiamine	4.5	-1.3268	0	s1c[n+](Cc2cnc(nc2N)C)c(C)c1CCO
897	Thiocyanate	99	1.9956	1	S=C=Nc1c2c(ccc1)cccc2
898	Thioguanine	30	-0.3680	0	S=C1N=C(Nc2nc[nH]c12)N
899	Thioridazine	40	-0.1761	0	S1c2c(N(c3c1cccc3)CCC1N(CCCC1)C)cc(SC)cc2
900	Tiagabine	90	0.9542	1	s1ccc(C)c1C(=CCCN1CC(CCC1)C(O)=O)c1sccc1C
901	Tianeptine	99	1.9956	1	Clc1cc2S(=O)(=O)N(c3c(cccc3)C(NCCCCCCC(O)=O)c2cc1)C
902	Tiapamil	42.5	-0.1313	0	S1(=O)(=O)CCCS(=O)(=O)C1(CCCN(CCc1cc(OC)c(OC)cc1)C)c1cc(OC)c(OC)cc1
903	Tiaprofenic Acid	90	0.9542	1	s1c(ccc1C(C(O)=O)C)C(=O)c1ccccc1

904	Ticarcillin	0	-2.0000	0	CC1(C)SC2C(NC(=O)C(C(O)=O)c3cc[s]c3)C(=O)N2C1C(O)=O
905	Ticlopidine	80	0.6021	1	Clc1ccccc1CN1CCc2sccc2C1
906	Tigecycline	0	-2.0000	0	CC(C)(C)NCC(=O)NC1=C(C2=C(CC3CC4C(C(=O)C(=C(C4(C(=O)C3=C2O)O)O)C(=O)N)N(C)C)C(=C1)N(C)C)O
907	Tilidine	90	0.9542	1	O(C(=O)C1(CCC=CC1N(C)C)c1ccccc1)CC
908	Tiludronic Acid	6	-1.1950	0	Clc1ccc(SC(P(O)(O)=O)P(O)(O)=O)cc1
909	Timolol	61	0.1943	1	s1nc(N2CCOCC2)c(OCC(O)CNC(C)(C)C)n1
910	Tinidazole	98	1.6902	1	S(=O)(=O)(CCn1c(ncc1[N+](=O)[O-])C)CC
911	Tiopronin	63	0.2311	1	SC(C(=O)NCC(O)=O)C
912	Tiotropium	2.5	-1.5911	0	s1cccc1C(O)(C(OC1CC2[N+](C(C1)C1OC12)(C)C)=O)c1sccc1
913	Tirofiban	0	-2.0000	0	S(=O)(=O)(NC(Cc1ccc(OCCCCC2CCNCC2)cc1)C(O)=O)CCCC
914	Tizanidine	21	-0.5754	0	Clc1ccc2nsnc2c1NC=1NCCN=1
915	Tobramycin	1	-1.9956	0	O1C(CO)C(O)C(N)C(O)C1OC1C(O)C(OC2OC(CN)C(O)CC2N)C(N)CC1N
916	Tocainide	89	0.9080	1	O=C(Nc1c(cccc1C)C)C(N)C
917	Tolbutamide	85	0.7533	1	S(=O)(=O)(NC(=O)NCCCC)c1ccc(cc1)C
918	Tolcapone	62	0.2126	1	Oc1c([N+](=O)[O-])cc(cc1O)C(=O)c1ccc(cc1)C
919	Tolfenamic Acid	60	0.1761	1	Clc1cccc(Nc2ccccc2C(O)=O)c1C
920	Toliprolol	90	0.9542	1	O(CC(O)CNC(C)C)c1cc(ccc1)C
921	Tolmesoxide	85	0.7533	1	S(=O)(C)c1cc(OC)c(OC)cc1C
922	Tolmetin	95	1.2788	1	OC(=O)Cc1n(C)c(cc1)C(=O)c1ccc(cc1)C

923	Toloxatone	56	0.1047	1	O1C(CN(c2cc(ccc2)C)C1=O)CO
924	Topiramate	70	0.3680	1	S(OCC12OC(OC1C1OC(OC1CO2)(C)C)(C)C)(=O)(=O)N
925	Topotecan	32	-0.3274	0	O1CC2=C(C=C3N(Cc4c3nc3c(c4)c(CN(C)C)c(O)cc3)C2=O)C(O)(CC)C1=O
926	Toremifene	99	1.9956	1	ClCCC(=C(c1ccc(OCCN(C)C)cc1)c1ccccc1)c1ccccc1
927	Torsemide	91	1.0048	1	S(=O)(=O)(NC(=O)NC(C)C)c1cnccc1Nc1cc(ccc1)C
928	Tramadol	72.5	0.4210	1	O(C)c1cc(ccc1)C1(O)CCCCC1CN(C)C
929	Trandolapril	10	-0.9542	0	O(C(=O)C(NC(C(=O)N1C2C(CC1C(O)=O)CCCC2)C)CCc1ccccc1)CC
930	Tranexamic_acid	34	-0.2881	0	OC(=O)C1CCC(CC1)CN
931	Tranylcypromine	50	0.0000	1	NC1CC1c1ccccc1
932	Trapidil	96	1.3802	1	n12ncnc1N=C(C=C2N(CC)CC)C
933	Trazodone	81	0.6297	1	Clc1cc(N2CCN(CC2)CCCN2N=C3N(C=CC=C3)C2=O)ccc1
934	Treosulfan	97	1.5097	1	S(OCC(O)C(O)COS(=O)(=O)C)(=O)(=O)C
935	TRH	2	-1.6902	0	O=C1NC(CC1)C(=O)NC(Cc1nc[nH]c1)C(=O)N1CCCC1C(=O)N
936	Triamcinolone	23	-0.5248	0	FC12C(C3CC(O)C(O)(C(=O)CO)C3(CC1O)C)CCC1=CC(=O)C=CC12C
937	Triamcinolone Acetonide	23	-0.5248	0	FC12C(C3CC4OC(OC4(C(=O)CO)C3(CC1O)C)(C)C)CCC1=CC(=O)C=CC12C
938	Triamterene	51	0.0174	1	n1c(N)c2nc(-c3ccccc3)c(nc2nc1N)N
939	Triazolam	44	-0.1047	0	Clc1ccccc1C1=NC=C2N(c3c1cc(Cl)cc3)C(=NN2)C
940	Trichlormethiazide	60	0.1761	1	Clc1cc2NC(NS(=O)(=O)c2cc1S(=O)(=O)N)C(Cl)Cl
941	Trientine	92.5	1.091	1	N(CCNCCN)CCN

			1		
942	Trihexyphenidyl	99	1.9956	1	OC(CCN1CCCCC1)(C1CCCCC1)c1ccccc1
943	Trimazosin	61	0.1943	1	O(C)c1c(OC)c2nc(nc(N)c2cc1OC)N1CCN(CC1)C(OCC(O)(C)C)=O
944	Trimethobenzamide	80	0.6021	1	O(C)c1c(OC)cc(cc1OC)C(=O)NCc1ccc(OCCN(C)C)cc1
945	Trimethoprim	98	1.6902	1	O(C)c1c(OC)cc(cc1OC)Cc1cnc(nc1N)N
946	Trimetrexate	45	-0.0872	0	O(C)c1c(OC)cc(NCc2cnc3nc(nc(N)c3c2C)N)cc1OC
947	Trimipramine	40	-0.1761	0	N(CC(CN1c2c(CCc3c1cccc3)cccc2)C)(C)C
948	Triprolidine	4	-1.3802	0	n1ccccc1C(=CCN1CCCC1)c1ccc(cc1)C
949	Trofosfamide	99	1.9956	1	ClCCN1P(OCCC1)(=O)N(CCCl)CCCl
950	Troglitazone	45	-0.0872	0	S1C(Cc2ccc(OCC3(Cc4c(OC3)c(C)c(C)c(O)c4C)C)cc2)C(=O)NC1=O
951	Tropisetron	66	0.2881	1	O(C(=O)c1c2c([nH]c1)cccc2)C1CC2N(C(C1)CC2)C
952	Trospium Chloride	10	-0.9542	0	C1CC[N+]2(C1)C3CCC2CC(C3)OC(=O)C(C4=CC=CC=C4)(C5=CC=CC=C5)O
953	Trovafloxacin	91	1.0048	1	Fc1cc(F)ccc1N1C=C(C(O)=O)C(=O)c2cc(F)c(nc12)N1CC2C(C1)C2N
954	Tulobuterol	48	-0.0348	0	Clc1ccccc1C(O)CNC(C)(C)C
955	Urapidil	78	0.5497	1	O(C)c1ccccc1N1CCN(CC1)CCCNC=1N(C)C(=O)N(C)C(=O)C=1
956	Uridine	8	-1.0607	0	OCC1OC(C(O)C1O)N2C=CC(=O)NC2=O
957	Valdecoxib	83	0.6886	1	S(=O)(=O)(N)c1ccc(cc1)-c1c(noc1C)-c1ccccc1
958	Valganciclovir	61	0.1943	1	O=C1N=C(Nc2n(cnc12)COC(COC(=O)C(N)C(C)C)CO)N
959	Valproic Acid	99	1.9956	1	OC(=O)C(CCC)CCC

960	Valrubicin	0	-2.0000	0	FC(F)(F)C(=O)NC1CC(OC(C)C1O)OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(C1=O)c(OC)ccc3)c2O)C(=O)COC(=O)CCCC
961	Valsartan	23	-0.5248	0	OC(=O)C(N(Cc1ccc(cc1)-c1ccccc1-c1[nH]nnn1)C(=O)CCCC)C(C)C
962	Vancomycin	5	-1.2788	0	Clc1c2Oc3cc4C(NC(=O)C(NC(=O)C(NC(=O)C(NC)CC(C)C)C(O)c(c1)cc2)CC(=O)N)C(=O)NC1c2cc(-c5c(cc(O)cc5O)C(NC(=O)C(NC1=O)C(O)c1cc(Cl)c(Oc(c4)c3OC3OC(CO)C(O)C(O)C3OC3OC(C)C(O)C(N)(C3)C)cc1)C(O)=O)c(O)cc2
963	Vardenafil	15	-0.7533	0	S(=O)(=O)(N1CCN(CC1)CC)c1cc(C2=NC(=O)c3n(N2)c(nc3C)CCC)c(OCC)cc1
964	Varenicline	87	0.8256	1	n1c2c(ncc1)cc1C3CC(c1c2)CNC3
965	Vecuronium	0	-2.0000	0	O(C(=O)C)C1C2(C(CC1[N+]1(CCCCC1)C)C1C(CC2)C2(CC(N3CCCCC3)C(OC(=O)C)CC2CC1)C)C
966	Venlafaxine	45	-0.0872	0	O(C)c1ccc(cc1)C(CN(C)C)C1(O)CCCCC1
967	Verapamil	22	-0.5497	0	O(C)c1cc(ccc1OC)C(C(C)C)(CCCN(CCc1cc(OC)c(OC)cc1)C)C#N
968	Vidarabine	0	-2.0000	0	O1C(CO)C(O)C(O)C1n1c2ncnc(N)c2nc1
969	Vigabatrin	99	1.9956	1	OC(=O)CCC(N)C=C
970	Vildagliptin	85	0.7533	1	OC12CC3(NCC(=O)N4CCCC4C#N)CC(C1)CC(C3)C2
971	Viloxazine (R)	85	0.7533	1	O1CCNCC1COc1ccccc1OCC
972	Vinblastine	0	-2.0000	0	O(C(=O)C)C1C2(C3N(CCC34C(N(c3cc(OC)c(cc34)C3(CC4CC(O)(CN(C4)CCc4c3[nH]c3c4cccc3)CC)C(OC)=O)C)C1(O)C(OC)=O)CC=C2)CC
973	Vincamine	20	-0.6021	0	OC1(n2c3C4N(CCCC4(C1)CC)CCc3c1c2cccc1)C(OC)=O
974	Vinorelbine	27	-0.4320	0	O(C(=O)C)C1C2(C3N(CCC34C(N(c3cc(OC)c(cc34)C3(CC4C=C(CN(C4)Cc4c3[nH]c3c4cccc3)CC)C(OC)=O)C)C1(O)C(OC)=O)CC=C2)CC
975	Vinpocetine	6.7	-1.1438	0	O(C(=O)C=1n2c3C4N(CCCC4(C=1)CC)CCc3c1c2cccc1)CC

976	Vitamin C	99	1.9956	1	O1C(C(O)CO)C(=O)C(O)=C1O
977	Voglibose	6	-1.1950	0	OC1C(O)C(O)(CC(NC(CO)CO)C1O)CO
978	Voriconazole	96	1.3802	1	Fc1cc(F)ccc1C(O)(C(C)c1ncncc1F)Cn1ncnc1
979	Vorinostat	43	-0.1224	0	O=C(Nc1ccccc1)CCCCCCC(=O)NO
980	Warfarin	93	1.1234	1	O1c2c(cccc2)C(O)=C(C(CC(=O)C)c2ccccc2)C1=O
981	Xamoterol	7	-1.1234	0	O1CCN(CC1)C(=O)NCCNCC(O)COc1ccc(O)cc1
982	Ximelagatran	21	-0.5754	0	O(C(=O)CNC(C(=O)N1CCC1C(=O)NCc1ccc(cc1)C(NO)=N)C1CCCCC1)CC
983	Xipamide	70	0.3680	1	Clc1cc(O)c(cc1S(=O)(=O)N)C(=O)Nc1c(cccc1C)C
984	Yohimbine	33	-0.3076	0	OC1CCC2C(CC3N(C2)CCc2c3[nH]c3c2cccc3)C1C(OC)=O
985	Zalcitabine	88	0.8653	1	O1C(CCC1N1C=CC(=NC1=O)N)CO
986	Zaleplon	31	-0.3475	0	O=C(N(CC)c1cc(ccc1)C=1n2ncc(c2N=CC=1)C#N)C
987	Zanamivir	2	-1.6902	0	O1C(C(O)C(O)CO)C(NC(=O)C)C(NC(N)=N)C=C1C(O)=O
988	Zidovudine	63	0.2311	1	CC1=CN(C(=O)NC1=O)C2CC(C(O2)CO)N=[N+]=[N-]
989	Zimelidine	29	-0.3889	0	Brc1ccc(cc1)C(=CCN(C)C)c1cccnc1
990	Ziprasidone	60	0.1761	1	Clc1cc2NC(=O)Cc2cc1CCN1CCN(CC1)c1nsc2c1cccc2
991	Zolmitriptan	45	-0.0872	0	O1CC(NC1=O)Cc1cc2c([nH]cc2CCN(C)C)cc1
992	Zolpidem	72	0.4102	1	O=C(N(C)C)Cc1n2C=C(C=Cc2nc1-c1ccc(cc1)C)C
993	Zonisamide	99	1.9956	1	S(=O)(=O)(N)Cc1noc2c1cccc2
994	Zotepine	10	-0.9542	0	Clc1cc2c(Sc3c(C=C2OCCN(C)C)cccc3)cc1

995	Zuclopenthixol	49	-0.0174	0	Clc1cc2c(Sc3c(cccc3)C2=CCCN2CCN(CC2)CCO)cc1

Table A2. qHTS ARE-*bla* combinatorial QSAR consensus model predictions and experimental qHTS ARE-*bla* testing results for FDA liver damage compounds, which are aggregated activity values derived from the concentration-response curves and CurveP values

No.	CID	Liver Damage Activity	Within AD Model (1=yes; 0=no)	Consensus Prediction	Phase I and Phase II Library	SMILES
1	4761	0	0	0.69	known Marginal	n1ccccc1C(CCN(C)C)c1ccccc1
2	1065	1	0	0.83	known Marginal	O(C)c1cc2c(nccc2C(O)C2N3CC(C(C2)CC3)C=C)cc1
3	3121	1	0	0.31	known Marginal	OC(=O)C(CCC)CCC
4	4781	1	0	0.86	known Marginal	OC=1N(N(C(=O)C=1CCCC)c1ccccc1)c1ccccc1
5	4139	0	0	1.03	known Marginal	S1C2=C\C(=[N+](/C)\C)\C=CC2=Nc2c1cc(N(C)C)cc2
6	3962	1	0	0.69	known Marginal	O1C(CC(O)CC1=O)CCC1C2C(=CC(CC2OC(=O)C(CC)C)C)C=CC1C
7	4112	1	0	1.05	known Marginal	OC(=O)C(NC(=O)c1ccc(N(Cc2nc3c(nc(nc3N)N)nc2)C)cc1)CCC(O)=O
8	864	0	1	0.63	known Marginal	S1SCCC1CCCCC(O)=O
9	887	0	0	0.11	known Marginal	OC
10	2337	0	0	0.43	known Marginal	O(C(=O)c1ccc(N)cc1)CC
11	2793	0	0	1.11	known Marginal	Clc1cc2N(c3c(CCc2cc1)cccc3)CCCN1CCC(N2CCCCC2)(CC1)C(=O)N
12	3042	0	0	0.67	known Marginal	O(C(=O)C1(CCCCC1)C1CCCCC1)CCN(CC)CC
13	3156	0	0	0.84	known Marginal	O1CCN(CC1)CCC1CN(CC)C(=O)C1(c1ccccc1)c1ccccc1
14	3157	0	0	1.05	known Marginal	O1c2c(OCC1C(=O)N1CCN(CC1)c1nc(N)c3cc(OC)c(OC)cc3n1)cccc2
15	3195	0	0	0.74	known Marginal	O=C1n2c3C4N(CCCC4(C1)CC)CCc3c1c2cccc1
16	3290	0	0	0.89	known Marginal	S1c2c(N(c3c1cccc3)CC(N(CC)CC)C)cccc2
17	3485	0	0	0.35	known Marginal	O=CCCCC=O
18	3743	0	0	0.99	known Marginal	Ic1c(C(=O)NCC(O)CO)c(I)c(N(C(=O)C)CC(O)CO)c(I)c1C(=O)NCCO
19	3981	0	0	0.90	known Marginal	O=C(N(CC)CC)C1C=C2C(N(C1)C)Cc1c3c2cccc3[nH]c1
20	4103	0	0	0.94	known Marginal	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)c1c(OC)cccc1OC
21	4870	0	0	0.95	known Marginal	Clc1cc2NC(N(S(=O)(=O)c2cc1S(=O)(=O)N)C)CSCC(F)(F)F
22	4926	0	0	0.83	known Marginal	S1c2c(N(c3c1cccc3)CCCN(C)C)cccc2

23	4935	0	0	0.74	known Marginal	O(CCC)c1ccc(cc1N)C(OCCN(CC)CC)=O
24	4989	0	0	0.68	known Marginal	s1cccc1\C=C\C1=NCCCN1C
25	5381	0	0	0.89	known Marginal	S1CCC(c2cc(ccc12)C#Cc1ncc(cc1)C(OCC)=O)(C)C
26	5586	0	0	0.97	known Marginal	Clc1ccc(cc1S(=O)(=O)N)C(=O)NN1CC2C3CC(C2C1)CC3
27	5733	0	0	0.83	known Marginal	Clc1ccc(cc1)C(=O)c1n(C)c(cc1C)CC(O)=O
28	7618	0	0	0.57	known Marginal	OCCN(CCO)CCO
29	72106	0	0	0.89	known Marginal	FC(F)(F)c1cc(Nc2ncccc2C(OCCN2CCOCC2)=O)ccc1
30	121829	0	0	1.05	known Marginal	S(=O)(=O)(N)c1cc(ccc1OC)CC(NCCOc1ccccc1OCC)C
31	253083	0	0	0.73	known Marginal	FC(F)(F)C1=CN(C2OC(CO)C(O)C2)C(=O)NC1=O
32	6321265	0	0	1.00	known Marginal	S(O)(=O)(=O)c1cc2c(N\C(=C\3/Nc4c(cc(S(O)(=O)=O)cc4)C/3=O)\C2=O)cc1
33	137	1	0	0.50	known Marginal	OC(=O)CCC(=O)CN
34	2250	1	0	0.99	known Marginal	Fc1ccc(cc1)-c1n(CCC(O)CC(O)CC(O)=O)c(C(C)C)c(C(=O)Nc2ccccc2)c1-c1ccccc1
35	2282	1	0	0.99	known Marginal	S1C2N(C(C(OC(OC(OCC)=O)C)=O)C1(C)C)C(=O)C2NC(=O)C(N)c1ccccc1
36	2476	1	0	0.85	known Marginal	O1C2C34C5(CC(C(O)(C(C)(C)C)C)C2(OC)CC5)C(N(CC3)CC2CC2)Cc2c4c1c(O)cc2
37	2626	1	0	1.04	known Marginal	S1C2N(C(=O)C2(OC)NC(=O)CSCC#N)C(C(O)=O)=C(C1)CSc1nnnn1C
38	2678	1	0	0.91	known Marginal	Clc1ccc(cc1)C(N1CCN(CC1)CCOCC(O)=O)c1ccccc1
39	3191	1	0	0.87	known Marginal	O(C(c1ccccc1)c1ccccc1)C1CCN(CC1)CCCC(=O)c1ccc(cc1)C(C)(C)C
40	3333	1	0	0.93	known Marginal	Clc1c(cccc1Cl)C1C(C(OCC)=O)=C(NC(C)=C1C(OC)=O)C
41	3385	1	0	0.65	known Marginal	FC1=CNC(=O)NC1=O
42	3478	1	0	1.09	known Marginal	S(=O)(=O)(NC(=O)NC1CCCCC1)c1ccc(cc1)CCNC(=O)c1ncc(nc1)C
43	3488	1	0	1.16	known Marginal	Clc1cc(C(=O)NCCc2ccc(S(=O)(=O)NC(=O)NC3CCCCC3)cc2)c(OC)cc1
44	3637	1	0	0.70	known Marginal	n1ncc2c(cccc2)c1NN
45	3715	1	0	0.91	known Marginal	Clc1ccc(cc1)C(=O)n1c2c(cc(OC)cc2)c(CC(O)=O)c1C
46	4428	1	0	0.85	known Marginal	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=O)CC1CC1
47	4451	1	0	0.93	known Marginal	S(CC(NC(=O)c1cccc(O)c1C)C(O)CN1CC2C(CC1C(=O)NC(C)(C)C)CCCC2)c1ccccc1
48	4488	1	0	0.74	known Marginal	FC(F)(F)c1cc(Nc2ncccc2C(O)=O)ccc1
49	4497	1	0	0.94	known Marginal	O(C(=O)C=1C(C(C(OCCOC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)C(C)C

50	4614	1	0	0.82	known Marginal	o1c(c(nc1CCC(O)=O)-c1ccccc1)-c1ccccc1
51	4641	1	0	0.92	known Marginal	OC=1N(N(C(=O)C=1CCCC)c1ccccc1)c1ccc(O)cc1
52	4731	1	0	0.94	known Marginal	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)COc1ccccc1
53	5002	1	0	1.03	known Marginal	S1c2c(cccc2)C(=Nc2c1cccc2)N1CCN(CC1)CCOCCO
54	5076	1	0	0.96	known Marginal	s1cncc1COC(=O)NC(Cc1ccccc1)C(O)CC(NC(=O)C(NC(=O)N(Cc1nc(sc1)C(C)C)C)C)C)Cc1ccccc1
55	5161	1	1	0.65	known Marginal	O(C(=O)c1ccccc1O)c1ccccc1C(O)=O
56	5399	1	0	0.80	known Marginal	O=C(CC\C=C(\CC\C=C(\CC\C=C(\CC\C=C(\C)/C)/C)/C)/C)/C)C
57	5468	1	0	0.75	known Marginal	s1c(ccc1C(C(O)=O)C)C(=O)c1ccccc1
58	5677	1	0	0.69	known Marginal	OC1C(O)C(O)(CC(NC(CO)CO)C1O)CO
59	5717	1	0	1.01	known Marginal	S(=O)(=O)(NC(=O)c1cc(OC)c(cc1)Cc1c2cc(NC(OC3CCCC3)=O)ccc2n(c1)C)c1ccccc1C
60	37393	1	0	1.03	known Marginal	Clc1cc(Cl)cc2c1cc(c1c2cc(cc1)C(F)(F)F)C(O)CCN(CCCC)CCCC
61	123631	1	0	1.00	known Marginal	Clc1cc(Nc2ncnc3c2cc(OCCCN2CCOCC2)c(OC)c3)ccc1F
62	151166	1	0	0.90	known Marginal	Clc1cccc(F)c1Nc1ccc(cc1CC(O)=O)C
63	216235	1	0	1.07	known Marginal	Clc1c(noc1NS(=O)(=O)c1ccsc1C(=O)Cc1cc2OCOc2cc1C)C
64	657298	1	0	0.59	known Marginal	S=C1NC(=CC(=O)N1)CCC
65	3086686	1	0	0.98	known Marginal	Fc1cc\2c(NC(=O)/C/2=C/c2[nH]c(C)c(C(=O)NCCN(CC)CC)c2C)cc1
66	4475886	1	0	0.83	known Marginal	O(C(=O)C)C1C2(C(CC1[N+]1(CCCCC1)C)C1C(CC2)C2(CC(N3CCCCC3)C(OC(=O)C)CC2CC1)C)C
67	54676860	1	1	0.47	known Marginal	O1C(C(O)CO)C(O)C(=O)C1=O
68	10783	0	0	0.52	known Marginal	ClCCc1scnc1C
69	41684	0	1	0.94	known Marginal	s1c([N+](=O)[O-])cnc1NC(=O)c1ccccc1OC(=O)C
70	60149	0	0	1.04	known Marginal	Clc1cc2c(n(cc2C2CCN(CC2)CCN2CCNC2=O)-c2ccc(F)cc2)cc1
71	110634	0	0	0.92	known Marginal	S(=O)(=O)(N1CCN(CC1)CC)c1cc(C=2NC(=O)c3n(N=2)c(nc3C)CCC)c(OCC)cc1
72	730059958	0	0	0.99	known Marginal	O1CC1(C(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)CN1CCOCC1)CCc1ccccc1)CC(C)C)Cc1ccccc1)CC(C)C)C
73	2375	1	0	0.93	known Marginal	S(=O)(=O)(CC(O)(C(=O)Nc1cc(C(F)(F)F)c(cc1)C#N)C)c1ccc(F)cc1
74	3308	1	0	0.81	known Marginal	O1CCc2c([nH]c3c2cccc3CC)C1(CC(O)=O)CC

75	5453	1	0	0.74	known Marginal	S=P(N1CC1)(N1CC1)N1CC1
76	5625	1	0	1.08	known Marginal	S(=O)(=O)(Nc1cc2cc([nH]c2cc1)C(=O)N1CCN(CC1)c1ncccc1NC(C)C)C
77	158	0	0	0.72	known Marginal	OC1CC(=O)C(C\C=C\CCCC(O)=O)C1\C=C\C(O)CCCCC
78	836	0	0	0.62	known Marginal	Oc1cc(ccc1O)CC(N)C(O)=O
79	853	0	1	1.05	known Marginal	Ic1cc(cc(I)c1Oc1cc(I)c(O)c(I)c1)CC(N)C(O)=O
80	861	0	0	1.04	known Marginal	Ic1cc(cc(I)c1Oc1cc(I)c(O)cc1)CC(N)C(O)=O
81	1935	0	1	0.79	known Marginal	n1c2c(CCCC2)c(N)c2c1cccc2
82	2391	0	0	0.83	known Marginal	O(C(=O)C)c1ccc(cc1)C(c1ccc(OC(=O)C)cc1)c1ncccc1
83	2616	0	0	0.93	known Marginal	S1C2N(C(=O)C2NC(=O)CSc2ccncc2)C(C(O)=O)=C(C1)COC(=O)C
84	2638	0	0	1.02	known Marginal	s1cccc1CC(=O)NC1(OC)C2SCC(COC(=O)N)=C(N2C1=O)C(O)=O
85	3000	0	0	0.75	known Marginal	FC12C(C3CC(C)C(C(=O)CO)C3(CC1O)C)CCC1=CC(=O)C=CC12C
86	3125	0	0	0.55	known Marginal	Oc1ccc(cc1)CC(N)(C(O)=O)C
87	3149	0	0	0.82	known Marginal	O(CC[N+](CCCCCCCCCCCC)(C)C)c1ccccc1
88	3281	0	0	0.47	known Marginal	Cl\C=C\C(O)(CC)C#C
89	3332	0	1	0.60	known Marginal	OC(=O)Cc1ccc(cc1)-c1ccccc1
90	3350	0	0	0.80	known Marginal	O=C1NC2CCC3C4CCC(C(=O)NC(C)(C)C)C4(CCC3C2(C=C1)C)C
91	3360	0	0	0.94	known Marginal	FC(F)(F)c1c2nccc(Nc3ccccc3C(OCC(O)CO)=O)c2ccc1
92	3553	0	0	0.97	known Marginal	ClCC(=O)C12OC(OC1CC1C3CCC4=CC(=O)CCC4(C)C3(F)C(O)CC12C)(C)C
93	3889	0	0	1.01	known Marginal	S(CC=1COC2N(C(=O)C2(OC)NC(=O)C(C(O)=O)c2ccc(O)cc2)C=1C(O)=O)c1nnnn1C
94	4037	0	0	0.85	known Marginal	Clc1c(Nc2ccccc2C(O)=O)c(Cl)ccc1C
95	4046	0	0	0.96	known Marginal	FC(F)(F)c1c2nc(cc(c2ccc1)C(O)C1NCCCC1)C(F)(F)F
96	4177	0	0	1.00	known Marginal	Clc1ccccc1C12OCC(N1CC(=O)Nc1c2cc(Cl)cc1)C
97	4768	0	0	0.88	known Marginal	ClCCN(Cc1ccccc1)C(COc1ccccc1)C
98	4865	0	0	0.82	known Marginal	O1CC2C(C(c3c(cc4OCOc4c3)C2O)c2cc(OC)c(OC)c(OC)c2)C1=O
99	5035	0	0	0.96	known Marginal	s1c2c(ccc(O)c2)c(C(=O)c2ccc(OCCN3CCCCC3)cc2)c1-c1ccc(O)cc1
100	5318	0	0	1.01	known Marginal	Clc1cc(Cl)ccc1C(SCc1ccc(Cl)cc1)Cn1ccnc1
101	5470	0	1	0.77	known Marginal	OC1(CCC2C3C(C4=C(CC3C)CC(=O)CC4)CCC12C)C#C
102	5510	0	0	0.94	known Marginal	S=C(Oc1cc2c(cc1)cccc2)N(C)c1cc(ccc1)C

103	5538	0	1	0.83	known Marginal	OC(=O)\C=C(\C=C\C=C(\C=C\C=1C(CCCC=1C)(C)C)/C)/C
104	5566	0	0	0.98	known Marginal	S1c2c(N(c3c1cccc3)CCCN1CCN(CC1)C)cc(cc2)C(F)(F)F
105	7108	0	0	0.83	known Marginal	S1c2c(Nc3c1cccc3)cccc2
106	16363	0	0	0.89	known Marginal	Fc1ccc(cc1)C(=O)CCCN1CCC(N2c3c(NC2=O)cccc3)CC1
107	26709	0	0	0.90	known Marginal	O1C2CC3C4C(C5(C(=CC(=O)C=C5)CC4)C)C(O)CC3(C)C2(N=C1C)C(=O)COC(=O)C
108	27991	0	0	1.18	known Marginal	S1SCCC(=O)NC(Cc2ccc(O)cc2)C(=O)NC(Cc2ccccc2)C(=O)NC(CCC(=O)N)C(=O)NC(CC(=O)N)C(=O)NC(C1)C(=O)N1CCCC1C(=O)NC(CCCNC(N)=N)C(=O)NCC(=O)N
109	33630	0	0	0.98	known Marginal	Clc1ccc(cc1C(F)(F)F)C1(O)CCN(CC1)CCCC(c1ccc(F)cc1)c1ccc(F)cc1
110	47472	0	0	0.99	known Marginal	Clc1cccc(Cl)c1SC(CCc1ccc(Cl)cc1)Cn1ccnc1
111	60907	0	0	1.01	known Marginal	O1c2c3c4c(c(O)c2C)c(O)c(NC(=O)/C(=C\C=C/C(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)\C=C/OC1(C)C3=O)/C)cc4O
112	72474	0	0	0.90	known Marginal	Fc1c(c2c(N(C=C(C(O)=O)C2=O)C2CC2)cc1N1CC(NCC1)C)C
113	72938	0	0	1.02	known Marginal	Fc1cc2c(nc1N1CC(CN)/C(=N/OC)/C1)N(C=C(C(O)=O)C2=O)C1CC1
114	114866	0	0	0.92	known Marginal	S(=O)(=O)(C)c1ccc(cc1)\C=C/1\c2c(cc(F)cc2)C(CC(O)=O)=C\1C
115	521928	0	1	0.75	known Marginal	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)CC3)C)C(=O)C
116	541103	0	1	0.78	known Marginal	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)C(C3)C)C)C(=O)C
117	3787925	0	0	0.71	known Marginal	OC1(CCC2C3C(C4C(CC3)=CCCC4)CCC12C)C#C
118	76	1	1	0.58	known Marginal	OC1CC2=CCC3C4CCC(=O)C4(CCC3C2(CC1)C)C
119	1972	1	0	1.02	known Marginal	O1C(C)C(O)C(N)C(O)C1OC1\C=C/C=C/C=C/C=C/C=C\C=C\C=C\C(C)C(O)C(C)C(OC(=O)CC(O)CC(O)CCC(O)C(O)CC(O)CC2(OC(C1)C(C(O)=O)C(O)C2)O)C
120	2094	1	0	0.55	known Marginal	O=C1NC=Nc2[nH]ncc12
121	2240	1	0	0.63	known Marginal	O1C2OC3(OOC24C(CCC(C4CC3)C)C(C)C1=O)C
122	2333	1	0	0.96	known Marginal	Brc1cc(cc(Br)c1O)C(=O)c1c2c(oc1CC)ccccc2
123	2381	1	0	0.69	known Marginal	OC(CCN1CCCCC1)(C1C2CC(C1)C=C2)c1ccccc1
124	2635	1	0	1.03	known Marginal	s1cc(nc1N)CC(=O)NC1C2SCC(CSc3nnnn3CCN(C)C)=C(N2C1=O)C(O)=O
125	2657	1	0	1.06	known Marginal	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SC

						C(CSC3=NC(=O)C(=O)NN3C)=C(N2C1=O)C(O)=O
126	2673	1	0	0.93	known Marginal	S1C2N(C(=O)C2NC(=O)C(N)C=2CC=CCC=2)C(C(O)=O)=C(C1)C
127	2708	1	0	0.88	known Marginal	ClCCN(CCCl)c1ccc(cc1)CCCC(O)=O
128	2732	1	0	0.97	known Marginal	Clc1ccc(cc1S(=O)(=O)N)C1(O)NC(=O)c2c1cccc2
129	2733	1	0	0.62	known Marginal	Clc1cc2NC(Oc2cc1)=O
130	2752	1	0	0.97	known Marginal	O(C(=O)C=1C(C(C(OC\C=C\c2ccccc2)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)CCOC
131	3033	1	0	0.90	known Marginal	Clc1cccc(Cl)c1Nc1ccccc1CC(O)=O
132	3117	1	0	0.88	known Marginal	S(SC(=S)N(CC)CC)C(=S)N(CC)CC
133	3143	1	0	0.94	known Marginal	O1C2CC(O)C3(C(C(OC(=O)c4ccccc4)C4(O)CC(OC(=O)C(O)C(NC(OC(C)(C)C)=O)c5ccccc5)C(=C(C4(C)C)C(O)C3=O)C)C2(OC(=O)C)C1)C
134	3159	1	0	0.69	known Marginal	FC1=CN(C2OC(C)C(O)C2O)C(=O)NC1=O
135	3198	1	0	1.03	known Marginal	Clc1cc(Cl)ccc1C(OCc1ccc(Cl)cc1)Cn1ccnc1
136	3203	1	0	0.96	known Marginal	Clc1cc2c(NC(OC2(C#CC2CC2)C(F)(F)F)=O)cc1
137	3310	1	0	0.87	known Marginal	O1C2C(OC(OC2)C)C(O)C(O)C1OC1C2C(C(c3c1cc1OCOc1c3)c1cc(OC)c(O)c(OC)c1)C(OC2)=O
138	3339	1	0	0.79	known Marginal	Clc1ccc(cc1)C(=O)c1ccc(OC(C(OC(C)C)=O)(C)C)cc1
139	3348	1	0	0.90	known Marginal	OC(C1CCN(CC1)CCCC(O)c1ccc(cc1)C(C(O)=O)(C)C)(c1ccccc1)c1ccccc1
140	3387	1	0	0.70	known Marginal	FC12C(C3CCC(O)(C)C3(CC1O)C)CCC1=CC(=O)CCC12C
141	3403	1	0	0.93	known Marginal	Fc1ccc(cc1)-c1c2c(n(C(C)C)c1\C=C\C(O)CC(O)CC(O)=O)cccc2
142	3476	1	0	1.10	known Marginal	S(=O)(=O)(NC(=O)NC1CCC(CC1)C)c1ccc(cc1)CCNC(=O)N1CC(C)=C(CC)C1=O
143	3955	1	0	0.99	known Marginal	Clc1ccc(cc1)C1(O)CCN(CC1)CCC(C(=O)N(C)C)(c1ccccc1)c1ccccc1
144	4021	1	0	0.55	known Marginal	O=C1N(NC(=C1)C)c1ccccc1
145	4053	1	0	0.94	known Marginal	ClCCN(CCCl)c1ccc(cc1)CC(N)C(O)=O
146	4138	1	0	0.63	known Marginal	Oc1cc(ccc1O)CC(N)(C(O)=O)C
147	4160	1	1	0.65	known Marginal	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)CC3)C)C
148	4211	1	0	0.96	known Marginal	Clc1ccccc1C(C(Cl)Cl)c1ccc(Cl)cc1
149	4252	1	0	0.99	known Marginal	S(CC=1COC2N(C(=O)C2(OC)NC(=O)CSC(F)F)C=1C(O)=O)c1nnnn1CCO
150	4272	1	1	0.74	known Marginal	O1Cc2c(c(O)c(C\C=C(\CCC(O)=O)/C)c(OC)c2C)C1=O

151	4413	1	0	1.03	known Marginal	O(C(=O)c1ccc(NC(N)=N)cc1)c1cc2c(cc(cc2)C(N)=N)cc1
152	4495	1	0	0.97	known Marginal	S(=O)(=O)(Nc1ccc([N+](=O)[O-])cc1Oc1ccccc1)C
153	4615	1	0	0.89	known Marginal	O=C1Nc2c(N1CCCN1CCN(CC1)C(c1ccccc1)c1ccccc1)cccc2
154	4680	1	0	0.93	known Marginal	O(C)c1cc(ccc1OC)Cc1nccc2c1cc(OC)c(OC)c2
155	4900	1	1	0.66	known Marginal	OC1(CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(=O)CC12C)C(=O)CO
156	4993	1	0	0.89	known Marginal	Clc1ccc(cc1)-c1c(nc(nc1N)N)CC
157	5042	1	0	0.95	known Marginal	Clc1ccc(cc1)C(=O)NC(CC1=CC(=O)Nc2c1cccc2)C(O)=O
158	5155	1	0	0.56	known Marginal	O1C(C=CC1N1C=C(C)C(=O)NC1=O)CO
159	5160	1	0	0.91	known Marginal	O(CC(OC(=O)CCC(O)=O)CN(C)C)c1ccccc1CCc1cc(OC)ccc1
160	5249	1	0	0.83	known Marginal	O(CC(O)=O)c1cc(OC\C=C(\C)/C)ccc1C(=O)\C=C\c1ccc(OC\C=C(\C)/C)cc1
161	5352	1	0	0.92	known Marginal	S(=O)(C)c1ccc(cc1)\C=C/1\c2c(cc(F)cc2)C(CC(O)=O)=C\1C
162	5376	1	0	0.94	known Marginal	O(CCN(C)C)c1ccc(cc1)\C(=C(/CC)\c1ccccc1)\c1ccccc1
163	5378	1	0	0.83	known Marginal	O=C1N(CCCCN2CCN(CC2)c2ncccn2)C(=O)C2C1C1CC2CC1
164	5391	1	0	0.95	known Marginal	Clc1cc2c(N(C)C(=O)C(O)N=C2c2ccccc2)cc1
165	5402	1	0	0.94	known Marginal	N(Cc1c2c(ccc1)cccc2)(C\C=C\C#CC(C)(C)C)C
166	5516	1	0	1.02	known Marginal	ClCC\C(=C(/c1ccc(OCCN(C)C)cc1)\c1ccccc1)\c1ccccc1
167	5560	1	0	1.05	known Marginal	Clc1cc2NC(NS(=O)(=O)c2cc1S(=O)(=O)N)C(Cl)Cl
168	5639	1	0	0.97	known Marginal	O(C)c1ccccc1N1CCN(CC1)CCCNC=1N(C)C(=O)N(C)C(=O)C=1
169	16231	1	0	0.91	known Marginal	Clc1nc(C(=O)NC(N)=N)c(nc1N)N
170	53276	1	0	1.00	known Marginal	Fc1ccc(cc1)CNc1nc(N)c(NC(OCC)=O)cc1
171	71774	1	0	0.85	known Marginal	[S+](CCC(=O)Nc1ccc(OCC(O)COCC)cc1)(C)C
172	77999	1	1	0.96	known Marginal	S1C(Cc2ccc(OCCN(C)c3ncccc3)cc2)=C(O)NC1=O
173	134018	1	0	0.86	known Marginal	s1c(nc(C)c1C(O)=O)-c1cc(C#N)c(OCC(C)C)cc1
174	208898	1	0	0.95	known Marginal	S(=O)(=O)(Nc1cc2c(oc(CCCC)c2C(=O)c2ccc(OCCCN(CCCC)CCCC)cc2)cc1)C
175	220503	1	1	0.56	known Marginal	OC1CCC2C3C(C4C(=CC(=O)CC4)CC3)CCC12C
176	1349907	1	0	0.54	known Marginal	S=C1NC=CN1C
177	3478	1	0	0.86	known Marginal	S(=O)(CCCC(F)(F)C(F)(F)F)CCCCCCCCCC

	439					1C2C3CCC(O)C3(CCC2c2c(C1)cc(O)cc2)C
178	3717450	1	0	0.97	known Marginal	O(C(=O)C)C1C2(C3N(CCC34C(N(c3cc(OC)c(cc34)C3(CC4CC(O)(CN(C4)CCc4c3[nH]c3c4cccc3)CC)C(OC)=O)C=O)C1(O)C(OC)=O)CC=C2)CC
179	4979942	1	0	0.96	known Marginal	Fc1ccc(cc1)C(O)CCC1C(N(C1=O)c1ccc(F)cc1)c1ccc(O)cc1
180	6828193	1	0	0.94	known Marginal	O1c2c3c4c(c(O)c2C)C(O)=C(NC(=O)/C(=C\C=C/C(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)\C=C/OC1(C)C3=O)/C)C1=NC2(N=C14)CCN(CC2)CC(C)C
181	6842115	1	0	1.08	known Marginal	O1c2c3c4c(c(O)c2C)c(O)c(NC(=O)/C(=C\C=C/C(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)\C=C/OC1(C)C3=O)/C)c(\C=N\N1CCN(CC1)C)c4O
182	10113978	1	0	1.14	known Marginal	S(=O)(=O)(N)c1cc(Nc2nc(N(C)c3cc4nn(C)c(c4cc3)C)ccn2)ccc1C
183	232	0	0	0.67	known Negative	OC(=O)C(N)CCCNC(N)=N
184	237	0	0	1.06	known Negative	Clc1cc2nc3c(cc(OC)cc3)c(NC(CCCN(CC)CC)C)c2cc1
185	604	0	0	0.53	known Negative	OC(C(O)C(O)CO)C(O)C(O)=O
186	1072	0	0	0.98	known Negative	O=C1NC(=O)N=C2N(c3cc(C)c(cc3N=C12)C)CC(O)C(O)C(O)CO
187	1153	0	0	0.52	known Negative	Oc1ccc(cc1)CC(N)C(O)=O
188	1993	0	0	0.52	known Negative	O(C(C[N+](C)(C)C)C)C(=O)C
189	2122	0	0	0.94	known Negative	Clc1cc2NC(NS(=O)(=O)c2cc1S(=O)(=O)N)CSCC=C
190	2345	0	1	0.44	known Negative	O(C(=O)c1ccccc1)Cc1ccccc1
191	2349	0	0	0.95	known Negative	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)Cc1ccccc1
192	2435	0	0	0.92	known Negative	Brc1c2nccnc2ccc1NC=1NCCN=1
193	2512	0	0	0.99	known Negative	O=C(N(CCCN(C)C)C(=O)NCC)C1CC2C(N(C1)CC=C)Cc1c3c2cccc3[nH]c1
194	2680	0	0	0.77	known Negative	O(C(=O)C1CCC(CC1)CN)c1ccc(cc1)CCC(O)=O
195	3052	0	0	0.63	known Negative	O=C(N(CC)CC)N1CCN(CC1)C
196	3344	0	0	0.79	known Negative	O1C2(CCN(CC2)CCc2ccccc2)CNC1=O
197	3378	0	0	0.97	known Negative	Fc1ccc(cc1)C(N1CCN(CC1)C\C=C\c1ccccc1)c1ccc(F)cc1
198	3741	0	0	0.99	known Negative	Ic1c(C(=O)NCC(O)CO)c(I)c(N(C(=O)CO)CCO)c(I)c1C(=O)NCC(O)CO
199	3914	0	0	0.70	known Negative	O(CC(O)CNC(C)(C)C)c1c2c(ccc1)C(=O)CCC2
200	4031	0	0	0.80	known Negative	O(C)c1cc(ccc1OC)C(OCCCCN(C(Cc1ccc(OC)cc1)C)CC)=O
201	4057	0	0	0.82	known Negative	O(C(=O)C(O)(c1ccccc1)c1ccccc1)C1CCC[N+](C1)(C)C
202	4078	0	0	0.94	known Negative	S1c2c(N(c3c1cccc3)CCC1N(CCCC1)C)cc(S(=O)C)cc2

203	4107	0	0	0.66	known Negative	O(CC(O)COC(=O)N)c1ccccc1OC
204	4170	0	0	0.99	known Negative	Clc1cc2NC(N(c3ccccc3C)C(=O)c2cc1S(=O)(=O)N)C
205	4195	0	0	0.77	known Negative	O(C)c1ccc(OC)cc1C(O)CNC(=O)CN
206	4418	0	0	0.94	known Negative	O(C)c1ccccc1N1CCN(CC1)CC(O)COc1c2c(ccc1)cccc2
207	4422	0	0	0.81	known Negative	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=C)CC1CC1
208	4678	0	0	0.56	known Negative	OC(C(CO)(C)C)C(=O)NCCCO
209	4914	0	0	0.66	known Negative	O(C(=O)c1ccc(N)cc1)CCN(CC)CC
210	4923	0	0	0.84	known Negative	Clc1ccc(NC(NC(NC(C)C)=N)=N)cc1
211	4934	0	0	0.90	known Negative	O1c2c(cccc2)C(c2c1cccc2)C(OCC[N+](C(C)C)(C(C)C)C)=O
212	4991	0	0	0.63	known Negative	O(C(=O)N(C)C)c1ccc[n+](c1)C
213	5095	0	0	0.73	known Negative	O=C1Nc2c(C1)c(ccc2)CCN(CCC)CCC
214	5344	0	0	0.91	known Negative	S(=O)(=O)(Nc1onc(C)c1C)c1ccc(N)cc1
215	5401	0	0	0.99	known Negative	O1CCCC1C(=O)N1CCN(CC1)c1nc(N)c2cc(OC)c(OC)cc2n1
216	5476	0	0	0.81	known Negative	s1cccc1\C(=C\1/CC(OC)C[N+](C/1)(C)C)\c1sccc1
217	5500	0	0	0.74	known Negative	O1c2c(CCC1(CCCC(CCCC(CCCC(C)C)C)C)C)c(C)c(OC(=O)c1cccnc1)c(C)c2C
218	5647	0	0	0.85	known Negative	O=C1NC(=Nc2n(cnc12)COCCOC(=O)C(N)C(C)C)N
219	5736	0	0	1.02	known Negative	Clc1cc2c(Sc3c(C=C2OCCN(C)C)cccc3)cc1
220	6058	0	0	0.42	known Negative	SCCN
221	6726	0	0	0.73	known Negative	N1(CCN(CC1)C)C(c1ccccc1)c1ccccc1
222	8400	0	1	0.50	known Negative	OC(C(=O)c1ccccc1)c1ccccc1
223	39524	0	0	0.63	known Negative	O1c2cc(ccc2OC1)\C=C\C(O)C(C)(C)C
224	510400	0	0	0.73	known Negative	O(C)c1ccc(cc1)C(=O)CC(=O)c1ccc(cc1)C(C)(C)C
225	54766	0	0	0.93	known Negative	O(C(=O)N(C)C)c1cc(cc(OC(=O)N(C)C)c1)C(O)CNC(C)(C)C
226	60879	0	0	0.94	known Negative	s1cccc1C\C(=C\c1n(Cc2ccc(cc2)C(O)=O)c(nc1)CCCC)\C(O)=O
227	123606	0	0	0.87	known Negative	S(=O)(=O)(N1CCCC1)Cc1cc2c([nH]cc2CCN(C)C)cc1
228	578798	0	1	0.76	known Negative	O1C2(C(OC1(C)C)CC1C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)C(=O)CO
229	3033538	0	0	0.88	known Negative	n12CCCCc1nnc2CCN1CCN(CC1)c1ccccc1C
230	4474936	0	0	0.97	known Negative	O(C)c1cc2CC(N(Cc2cc1OC)C(=O)C(NC(CCc1ccccc1)C(OCC)=O)C)C(O)=O
231	4687521	0	0	0.93	known Negative	O1c2cc(ccc2OC1)C1N2C(Cc3c1[nH]c1c3cccc1)C(=O)N(CC2=O)C
232	6101856	0	0	0.95	known Negative	FC(F)(F)c1cc(ccc1)CCCNC(C)c1c2c(ccc1)cccc2

233	191	1	0	0.76	known Negative	O1C(CO)C(O)C(O)C1n1c2ncnc(N)c2nc1
234	753	1	1	0.34	known Negative	OC(CO)CO
235	1054	1	1	0.45	known Negative	Oc1c(CO)c(cnc1C)CO
236	1546	1	0	0.89	known Negative	Clc1nc(N)c2ncn(c2n1)C1OC(CO)C(O)C1
237	1615	1	0	0.48	known Negative	O1c2cc(ccc2OC1)CC(NC)C
238	2022	1	1	0.62	known Negative	O=C1NC(=Nc2n(cnc12)COCCO)N
239	2170	1	0	0.98	known Negative	Clc1cc2c(Oc3c(N=C2N2CCNCC2)cccc3)cc1
240	2442	1	0	0.93	known Negative	Brc1cc(Br)cc(CN(C)C2CCCCC2)c1N
241	2478	1	0	0.57	known Negative	S(OCCCCOS(=O)(=O)C)(=O)(=O)C
242	2676	1	0	0.99	known Negative	Fc1ccc(cc1)-c1c(COC)c(nc(C(C)C)c1\C=C\C(O)CC(O)CC(O)=O)C(C)C
243	2805	1	0	0.88	known Negative	Clc1ccc(cc1)C(OCCN1CCCCC1)c1ccccc1
244	2895	1	0	0.77	known Negative	N(CC\C=C/1\c2c(C=Cc3c\1cccc3)cccc2)(C)C
245	2955	1	0	0.84	known Negative	S(=O)(=O)(c1ccc(N)cc1)c1ccc(N)cc1
246	3062	1	0	0.86	known Negative	O1C(C)C(OC2OC(C)C(O)C(O)C2)C(O)CC1OC1C(OC(OC2CC3CCC4C(CC(O)C5(C)C(CCC45O)C4=CC(OC4)=O)C3(CC2)C)CC1O)C
247	3279	1	0	0.61	known Negative	OCC(NCCNC(CC)CO)CC
248	3307	1	0	1.05	known Negative	Clc1ccccc1C1=NCc2n(-c3sc(cc13)CC)c(nn2)C
249	3312	1	0	0.86	known Negative	O(C)c1cc(C)c(\C=C\C(=C\C=C\C(=C\C(OCC)=O)\C)\C)c(C)c1C
250	3512	1	0	0.94	known Negative	Clc1c2OC3(C(CC(=O)C=C3OC)C)C(=O)c2c(OC)cc1OC
251	3634	1	0	0.93	known Negative	S1c2c(C(=O)c3c1cccc3)c(NCCN(CC)CC)ccc2CO
252	3652	1	0	1.00	known Negative	Clc1cc2nccc(NC(CCCN(CCO)CC)C)c2cc1
253	3672	1	1	0.42	known Negative	OC(=O)C(C)c1ccc(cc1)CC(C)C
254	3690	1	0	0.89	known Negative	ClCCN1P(OCCC1)(=O)NCCCl
255	3696	1	1	0.73	known Negative	N(CCCN1c2c(CCc3c1cccc3)cccc2)(C)C
256	3767	1	0	0.57	known Negative	O=C(NN)c1ccncc1
257	3784	1	0	0.89	known Negative	o1nc2c(n1)cccc2C1C(C(OC(C)C)=O)=C(NC(C)=C1C(OC)=O)C
258	3869	1	0	0.79	known Negative	Oc1ccc(cc1C(=O)N)C(O)CNC(CCc1ccccc1)C
259	3873	1	0	0.90	known Negative	S(=O)(Cc1occc1)CC(=O)NC\C=C\COc1nccc(c1)CN1CCCCC1
260	3878	1	0	0.94	known Negative	Clc1c(cccc1Cl)-c1nnc(nc1N)N
261	4163	1	0	0.92	known Negative	OCC(NC(=O)C1C=C2C(N(C1)C)Cc1c3c2cccc3n(c1)C)CC

262	4260	1	0	0.71	known Negative	O(C(=O)C)c1cc(C(C)C)c(OCCN(C)C)cc1C
263	4417	1	0	0.87	known Negative	O1CCCC1CC(Cc1c2c(ccc1)cccc2)C(OCCN(CC)CC)=O
264	4450	1	0	0.75	known Negative	O1CCN(Cc2c(cccc2)C1c1ccccc1)C
265	4601	1	0	0.73	known Negative	O(C(c1ccccc1C)c1ccccc1)CCN(C)C
266	4646	1	0	0.68	known Negative	OC(=O)\C=C\c1ccc(cc1)Cn1ccnc1
267	4723	1	0	0.61	known Negative	o1c(-c2ccccc2)c(O)nc1N
268	4746	1	0	0.74	known Negative	N1CCCCC1CC(C1CCCCC1)C1CCCCC1
269	4763	1	1	0.67	known Negative	O=C1NC(=O)NC(=O)C1(CC)c1ccccc1
270	4782	1	1	0.45	known Negative	Oc1cc(ccc1)C(O)CNC
271	4942	1	0	0.79	known Negative	O(C(=O)C(OCCC)(c1ccccc1)c1ccccc1)C1CCN(CC1)C
272	5005	1	0	0.96	known Negative	O(C(=O)C(NC(C(=O)N1Cc2c(CC1C(O)=O)cccc2)C)CCc1ccccc1)CC
273	5039	1	0	0.88	known Negative	S(Cc1oc(cc1)CN(C)C)CCN\C(=N\C)\C[N+](=O)[O-]
274	5071	1	0	0.45	known Negative	NC(C)C12CC3CC(C1)CC(C2)C3
275	5257	1	0	0.97	known Negative	Fc1c(N2CC(NC(C2)C)C)c(F)c2N(C=C(C(O)=O)C(=O)c2c1N)C1CC1
276	5267	1	0	0.80	known Negative	S(C(=O)C)C1C2C3CCC4(OC(=O)CC4)C3(CCC2C2(C(C1)=CC(=O)CC2)C)C
277	5355	1	0	0.87	known Negative	S(=O)(=O)(N)c1cc(C(=O)NCC2N(CCC2)CC)c(OC)cc1
278	5467	1	0	0.76	known Negative	S(=O)(=O)(C)c1cc(C(=O)NCCN(CC)CC)c(OC)cc1
279	5472	1	0	0.83	known Negative	Clc1ccccc1CN1CCc2sccc2C1
280	5503	1	0	0.98	known Negative	S(=O)(=O)(NC(=O)NN1CCCCCC1)c1ccc(cc1)C
281	5656	1	0	0.69	known Negative	O(C)c1ccc(cc1)C(CN(C)C)C1(O)CCCCC1
282	5718	1	1	0.51	known Negative	O1C(CCC1N1C=CC(=NC1=O)N)CO
283	19675	1	0	0.94	known Negative	S1c2c(N(c3c1cccc3)CCCN1CCC(CC1)CCO)cc(cc2)C(=O)C
284	36339	1	0	0.69	known Negative	O(C(=O)c1n(cnc1)C(C)c1ccccc1)CC
285	65866	1	0	0.89	known Negative	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)C(CN(CCC(c1ccccc1)c1ccccc1)C)(C)C
286	99937	1	0	0.46	known Negative	O1C2C(OCC2O)C(O)C1
287	362789	1	0	0.56	known Negative	S(CC(N)C(O)=O)CCCO
288	5231054	1	0	0.95	known Negative	Fc1cc(F)ccc1C(O)(C(C)c1ncncc1F)Cn1ncnc1
289	6850813	1	0	0.70	known Negative	O(C(CCNC)c1ccccc1)c1ccccc1C
290	366877 67	1	0	0.94	known Negative	Oc1ccccc1-c1nc(nn1-c1ccc(cc1)C(O)=O)-c1ccccc1O

291	4537 1287	1	0	1.01	known Negative	S(C=1C(C2N(C(=O)C2C(O)C)C=1C(O)=O)C)C1CC(NC1)CNS(=O)(=O)N
292	85	0	0	0.56	known Negative	OC(CC(O)=O)C[N+](C)(C)C
293	180	0	0	0.26	known Negative	O=C(C)C
294	244	0	1	0.24	known Negative	OCc1ccccc1
295	564	0	0	0.40	known Negative	OC(=O)CCCCCN
296	594	0	0	0.65	known Negative	SCC(N)C(O)=O
297	611	0	1	0.52	known Negative	OC(=O)C(N)CCC(O)=O
298	700	0	0	0.35	known Negative	OCCN
299	815	0	0	0.93	known Negative	O1C(CN)C(O)C(O)C(O)C1OC1C(O)C(OC2OC(CO)C(O)C(N)C2O)C(N)CC1N
300	896	0	0	0.77	known Negative	O(C)c1cc2c([nH]cc2CCNC(=O)C)cc1
301	936	0	1	0.37	known Negative	O=C(N)c1cccnc1
302	1030	0	1	0.34	known Negative	OC(CO)C
303	1130	0	0	0.82	known Negative	s1c[n+](Cc2cnc(nc2N)C)c(C)c1CCO
304	1981	0	0	0.96	known Negative	Clc1ccc(cc1)C(=O)n1c2c(cc(OC)cc2)c(CC(OCC(O)=O)=O)c1C
305	2140	0	0	0.91	known Negative	Ic1c(C(O)=O)c(I)c(NC(=O)C)c(I)c1NC(=O)C
306	2266	0	1	0.38	known Negative	OC(=O)CCCCCCCC(O)=O
307	2318	0	0	0.86	known Negative	FC(F)(F)c1cc(ccc1)CC(NCCOC(=O)c1ccccc1)C
308	2342	0	0	0.85	known Negative	O(C(=O)C)C1CC2N(CC1C(=O)N(CC)CC)CCc1cc(OC)c(OC)cc12
309	2369	0	0	0.81	known Negative	O(CC(O)CNC(C)C)c1ccc(cc1)CCOCC1CC1
310	2536	0	0	0.99	known Negative	O(C(=O)c1ccc(NC(N)=N)cc1)c1ccc(cc1)CC(OCC(=O)N(C)C)=O
311	2537	0	0	0.34	known Negative	O=C1CC2CCC1(C)C2(C)C
312	2560	0	0	0.93	known Negative	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(C(O)=O)c1ccccc1
313	2564	0	0	0.85	known Negative	Clc1ccc(cc1)C(OCCN(C)C)c1ncccc1
314	2583	0	0	0.86	known Negative	O(CC(O)CNC(C)(C)C)c1c2CCC(=O)Nc2ccc1
315	2720	0	0	0.92	known Negative	Clc1cc2N=CNS(=O)(=O)c2cc1S(=O)(=O)N
316	2729	0	0	0.97	known Negative	Clc1cc\2c(Sc3c(cccc3)/C/2=C\CCN(C)C)cc1
317	2781	0	0	0.89	known Negative	Clc1ccc(cc1)C(OCCC1N(CCC1)C)(C)c1ccccc1
318	2905	0	0	0.70	known Negative	OC1(CCCC1)C(C(OCCN(C)C)=O)c1ccccc1
319	2995	0	1	0.77	known Negative	N(CCCN1c2c(CCc3c1cccc3)cccc2)C
320	3025	0	0	0.89	known Negative	O(CCCC)c1nc2c(cccc2)c(c1)C(=O)NCCN(CC)CC
321	3041	0	0	1.06	known Negative	Clc1cccc(Cl)c1-c1noc(C)c1C(=O)NC1C2SC(C)(C)C(N2C1=O)C(O)=O

322	3055	0	0	0.79	known Negative	OC(CCCN1CCCCC1)(c1ccccc1)c1ccccc1
323	3100	0	0	0.71	known Negative	O(C(c1ccccc1)c1ccccc1)CCN(C)C
324	3162	0	0	0.72	known Negative	O(C(C)(c1ccccc1)c1ncccc1)CCN(C)C
325	3182	0	0	0.65	known Negative	O=C1N(C)C(=O)N(c2ncn(c12)CC(O)CO)C
326	3337	0	0	0.70	known Negative	FC(F)(F)c1cc(ccc1)CC(NCC)C
327	3342	0	1	0.67	known Negative	O(c1cc(ccc1)C(C(O)=O)C)c1ccccc1
328	3382	0	0	0.89	known Negative	FC12C(C3CC4OC(OC4(C(=O)COC(=O)C)C3(CC1O)C)(C)C)CC(F)C1=CC(=O)C=CC12C
329	3446	0	0	0.45	known Negative	OC(=O)CC1(CCCCC1)CN
330	3494	0	0	0.74	known Negative	O(C(=O)C(O)(C1CCCC1)c1ccccc1)C1CC[N+](C1)(C)C
331	3516	0	0	0.47	known Negative	O(CC(O)CO)c1ccccc1OC
332	3517	0	0	0.84	known Negative	Clc1cccc(Cl)c1\C=N\NC(N)=N
333	3518	0	0	0.65	known Negative	N(CCN1CCCCCCC1)C(N)=N
334	3519	0	0	0.81	known Negative	Clc1cccc(Cl)c1CC(=O)NC(N)=N
335	3623	0	0	0.69	known Negative	O(C(=O)C(O)c1ccccc1)C1CC2N(C(C1)CC2)C
336	3661	0	0	0.73	known Negative	O(C(=O)C(CO)c1ccccc1)C1CC2N(C(C1)CC2)C
337	3687	0	0	0.77	known Negative	IC1=CN(C2OC(CO)C(O)C2)C(=O)NC1=O
338	3734	0	0	0.99	known Negative	Ic1c(C(=O)NC(CO)CO)c(I)c(NC(=O)C(O)C)c(I)c1C(=O)NC(CO)CO
339	3736	0	0	1.01	known Negative	Ic1c(C(=O)N(CC(O)CO)C)c(I)c(NC(=O)COC)c(I)c1C(=O)NCC(O)CO
340	3776	0	0	0.20	known Negative	OC(C)C
341	3778	0	0	0.72	known Negative	O=C1N(N(C)C(C)=C1C(C)C)c1ccccc1
342	3871	0	0	0.69	known Negative	O1C(CO)C(O)C(O)C(O)C1OC(C(O)C(O)CO)C(O)CO
343	3998	0	0	0.77	known Negative	S(=O)(=O)(N)c1ccc(cc1)CN
344	4011	0	0	0.79	known Negative	N(CCCC12CCC(c3c1cccc3)c1c2cccc1)C
345	4049	0	0	0.59	known Negative	OC(C(O)C(O)CO)C(O)CNC
346	4054	0	0	0.44	known Negative	NC12CC3(CC(C1)(CC(C3)C2)C)C
347	4086	0	0	0.58	known Negative	Oc1cc(cc(O)c1)C(O)CNC(C)C
348	4087	0	0	0.51	known Negative	Oc1cc(ccc1)C(O)C(N)C
349	4133	0	1	0.35	known Negative	Oc1ccccc1C(OC)=O
350	4167	0	0	0.84	known Negative	S1c2c(cccc2)C(c2c1cccc2)CC1CCCN(C1)C
351	4257	0	0	1.09	known Negative	Clc1cc2N(c3c(CCc2cc1)cccc3)CCCN1CCC2(N3C(NC2=O)CCCC3)CC1
352	4442	0	0	0.87	known Negative	O=C(C1CCc2nc[nH]c2C1)c1c2c(n(c1)C)cccc2
353	4567	0	0	0.79	known Negative	Oc1ccc(cc1)C(O)C(NC(CCc1ccccc1)C)C
354	4636	0	0	0.79	known Negative	Oc1c(C)c(CC=2NCCN=2)c(cc1C(C)(C)C)C

355	4688	0	1	0.39	known Negative	N(Cc1ccccc1)(CC#C)C
356	4745	0	0	0.83	known Negative	S(CC1CC2C(N(C1)CCC)Cc1c3c2cccc3[nH]c1)C
357	4753	0	0	0.54	known Negative	O=C(NC(=O)N)Cc1ccccc1
358	4762	0	0	0.51	known Negative	O1CCNC(C)C1c1ccccc1
359	4775	0	1	0.33	known Negative	OC(=O)CCCc1ccccc1
360	4828	0	0	0.79	known Negative	O(CC(O)CNC(C)C)c1c2c([nH]cc2)ccc1
361	4830	0	0	0.90	known Negative	Fc1ccc(cc1)C(=O)CCCN1CCC(N2CCCCC2)(CC1)C(=O)N
362	4836	0	0	0.81	known Negative	S1(=O)(=O)C2N(C(C(O)=O)C1(Cn1nncc1)C)C(=O)C2
363	4849	0	0	1.04	known Negative	S(=O)(=O)(N)c1cc(cc(N2CCCC2)c1Oc1ccccc1)C(O)=O
364	4850	0	0	0.84	known Negative	O1c2cc(ccc2OC1)CN1CCN(CC1)c1ncccn1
365	4853	0	0	0.81	known Negative	OC(CCCN1C(CCCC1C)C)(c1ccccc1)c1nccc1
366	4885	0	0	0.81	known Negative	s1c2CC(NCCC)CCc2nc1N
367	4916	0	0	0.86	known Negative	Oc1c2NC(=O)C=Cc2c(cc1)C(O)C(NC(C)C)CC
368	4919	0	0	0.68	known Negative	OC(CCN1CCCC1)(C1CCCCC1)c1ccccc1
369	5077	0	0	0.66	known Negative	O(C(=O)N(CC)C)c1cc(ccc1)C(N(C)C)C
370	5193	0	0	0.67	known Negative	O=C1NC(=O)NC(=O)C1(C(CCC)C)CC=C
371	5300	0	0	0.76	known Negative	O1C(CO)C(O)C(O)C(NC(=O)N(N=O)C)C1O
372	5320	0	0	0.79	known Negative	S(=O)(=O)(NC(=O)C)c1ccc(N)cc1
373	5466	0	0	0.87	known Negative	s1ccc(C)c1\C(=C\CCN1CC(CCC1)C(O)=O)\c1sccc1C
374	5479	0	0	0.74	known Negative	S(=O)(=O)(CCn1c(ncc1[N+](=O)[O-])C)CC
375	5504	0	1	0.55	known Negative	N1CCN=C1Cc1ccccc1
376	5509	0	0	0.74	known Negative	OC(=O)Cc1n(C)c(cc1)C(=O)c1ccc(cc1)C
377	5565	0	0	0.58	known Negative	N(CCNCCN)CCN
378	5572	0	0	0.69	known Negative	OC(CCN1CCCCC1)(C1CCCCC1)c1ccccc1
379	5587	0	0	0.75	known Negative	n1ccccc1N(Cc1ccccc1)CCN(C)C
380	5593	0	0	0.76	known Negative	OCC(C(=O)N(Cc1ccncc1)CC)c1ccccc1
381	5595	0	0	0.81	known Negative	O(C(=O)c1c2c([nH]c1)cccc2)C1CC2N(C(C1)CC2)C
382	5606	0	0	0.71	known Negative	Clc1ccccc1C(O)CNC(C)(C)C
383	5634	0	1	0.36	known Negative	OC(=O)CCCCCCCCC=C
384	5666	0	0	0.68	known Negative	O1CCNCC1COc1ccccc1OCC
385	5673	0	0	0.87	known Negative	O(C(=O)C=1n2c3C4N(CCCC4(C=1)CC)CCc3c1c2cccc1)CC
386	5775	0	0	0.86	known Negative	Oc1cc(N(CC=2NCCN=2)c2ccc(cc2)C)ccc1
387	6049	0	0	0.67	known Negative	OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CC(O)=O

388	6476	0	0	0.55	known Negative	O=C1N(C)C(=O)CC1(C)c1ccccc1
389	6503	0	0	0.60	known Negative	OCC(N)(CO)CO
390	6806	0	0	0.83	known Negative	Ic1c(C(O)=O)c(I)cc(I)c1NC(=O)C
391	7029	0	0	0.50	known Negative	O=C(C(N(CC)CC)C)c1ccccc1
392	7077	0	0	0.73	known Negative	O(CCN(C)C)c1ccccc1Cc1ccccc1
393	7302	0	0	0.31	known Negative	O1CCCC1=O
394	8285	0	0	0.59	known Negative	OCC(CO)(CO)CO
395	16229	0	0	0.91	known Negative	O=C1N(C)C(=O)N(c2nc(n(c12)CCN(CCO)CC)Cc1ccccc1)C
396	16351	0	0	1.07	known Negative	Clc1cc2c(Sc3c(N=C2N2CCN(CC2)C)cccc3)cc1
397	21855	0	0	0.75	known Negative	n1ccccc1C(C)C=1c2c(CC=1CCN(C)C)cccc2
398	23897	0	0	0.69	known Negative	O1CCN(CC1)CC1CCc2[nH]c(C)c(c2C1=O)CC
399	25229	0	0	1.13	known Negative	Ic1c(C(=O)NC)c(I)c(NC(=O)CCCCC(=O)Nc2c(I)c(C(=O)NC)c(I)c(C(O)=O)c2I)c(I)c1C(O)=O
400	32170	0	0	0.82	known Negative	s1cc(C)c(NC(=O)C(NCCC)C)c1C(OC)=O
401	39214	0	0	0.94	known Negative	ClCCN(N=O)C(=O)NCc1cnc(nc1N)C
402	44564	0	0	0.98	known Negative	Clc1c(NC(=O)C(O)=O)cc(cc1NC(=O)C(O)=O)C#N
403	59768	0	0	0.80	known Negative	O(CC(O)CNC(C)C)c1ccc(cc1)CCC(OC)=O
404	71273	0	0	0.75	known Negative	O=C(Nc1c(cccc1C)C)C1N(CCCC1)CCC
405	72054	0	0	0.93	known Negative	O1CCc2cc(ccc12)CCN1CC(CC1)C(C(=O)N)(c1ccccc1)c1ccccc1
406	151075	0	0	0.86	known Negative	O=C(c1cccc(CC(=O)N)c1N)c1ccccc1
407	610682	0	0	0.78	known Negative	N1CCN=C1Cc1cc2c(cc1)cccc2
408	3022645	0	0	0.81	known Negative	O1CCNCC1C(Oc1ccccc1OCC)c1ccccc1
409	10127622	0	0	1.00	known Negative	Brc1cc(Cl)c(Nc2c(cc3n(cnc3c2F)C)C(=O)NOCCO)cc1
410	22184409	0	0	0.83	known Negative	O(C(=O)CC)C1C2(C(CC1[N+]1(CCCCC1)CC=C)C1C(CC2)C2(CC(N3CCCCC3)C(OC(=O)C)CC2CC1)C)C
411	54680103	0	0	1.06	known Negative	Clc1c2c(C(=O)C3C(CC4C(O)(C(=O)C(C(=O)N)C(=O)C4N(C)C)C3=O)C2O)c(O)cc1
412	174	1	1	0.24	known Negative	OCCO
413	338	1	1	0.32	known Negative	Oc1ccccc1C(O)=O
414	581	1	0	0.66	known Negative	SCC(NC(=O)C)C(O)=O
415	598	1	0	0.58	known Negative	S(O)(=O)(=O)CCS
416	745	1	0	0.93	known Negative	SCC(NC(=O)CCC(N)C(O)=O)C(=O)NCC(

						O)=O
417	938	1	1	0.30	known Negative	OC(=O)c1cccnc1
418	978	1	1	0.35	known Negative	OC(=O)c1ccc(N)cc1
419	1046	1	1	0.46	known Negative	O=C(N)c1nccnc1
420	1775	1	0	0.79	known Negative	O=C1NC(=O)NC1(c1ccccc1)c1ccccc1
421	1978	1	0	0.82	known Negative	O(CC(O)CNC(C)C)c1ccc(NC(=O)CCC)cc1C(=O)C
422	2019	1	0	1.22	known Negative	O1C(C)C(NC(=O)C=2C3=Nc4c(OC3=C(C)C(=O)C=2N)c(ccc4C(=O)NC2C(OC(=O)C(N(C)C(=O)CN(C)C(=O)C3N(CCC3)C(=O)C(NC2=O)C(C)C)C(C)C)C)C(=O)NC(C(C)C)C(=O)N2C(CCC
423	2130	1	0	0.39	known Negative	NC12CC3CC(C1)CC(C2)C3
424	2142	1	0	0.97	known Negative	O1C(CN)C(O)C(O)C(O)C1OC1C(O)C(OC2OC(CO)C(O)C(N)C2O)C(NC(=O)C(O)CCN)CC1N
425	2157	1	0	0.99	known Negative	Ic1cc(cc(I)c1OCCN(CC)CC)C(=O)c1c2c(oc1CCCC)cccc2
426	2159	1	0	0.84	known Negative	S(=O)(=O)(CC)c1cc(C(=O)NCC2N(CCC2)CC)c(OC)cc1N
427	2160	1	0	0.76	known Negative	N(CC\C=C/1\c2c(CCc3c\1cccc3)cccc2)(C)C
428	2176	1	0	1.07	known Negative	S1(=O)(=O)N(C)C(C(=O)Nc2ncccc2)=C(OC(OC(OCC)=O)C)c2c1cccc2
429	2232	1	0	1.14	known Negative	S(=O)(=O)(NC(CCCNC(N)=N)C(=O)N1CCC(CC1C(O)=O)C)c1c2NCC(Cc2ccc1)C
430	2244	1	1	0.38	known Negative	O(C(=O)C)c1ccccc1C(O)=O
431	2249	1	0	0.78	known Negative	O(CC(O)CNC(C)C)c1ccc(cc1)CC(=O)N
432	2324	1	0	0.93	known Negative	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)C1CCCN(C1)Cc1ccccc1
433	2351	1	0	0.82	known Negative	O(CC(C)C)CC(N1CCCC1)CN(Cc1ccccc1)c1ccccc1
434	2405	1	0	0.80	known Negative	O(CC(O)CNC(C)C)c1ccc(cc1)COCCOC(C)C
435	2471	1	0	1.06	known Negative	S(=O)(=O)(N)c1cc(cc(NCCCC)c1Oc1ccccc1)C(O)=O
436	2520	1	0	0.91	known Negative	O(C)c1cc(ccc1OC)C(C(C)C)(CCCN(CCc1cc(OC)c(OC)cc1)C)C#N
437	2541	1	0	0.96	known Negative	O(CC)c1nc2c(n1Cc1ccc(cc1)-c1ccccc1-c1nn[nH]n1)c(ccc2)C(O)=O
438	2663	1	0	0.87	known Negative	O(CC(O)CNC(C)(C)C)c1ccc(NC(=O)N(CC)CC)cc1C(=O)C
439	2719	1	0	0.93	known Negative	Clc1cc2nccc(NC(CCCN(CC)CC)C)c2cc1
440	2726	1	0	0.98	known Negative	Clc1cc2N(c3c(Sc2cc1)cccc3)CCCN(C)C
441	2727	1	0	0.92	known Negative	Clc1ccc(S(=O)(=O)NC(=O)NCCC)cc1
442	2756	1	0	0.85	known Negative	S(Cc1[nH]cnc1C)CCN\C(=N/C)\NC#N
443	2771	1	0	0.91	known Negative	Fc1ccc(cc1)C1(OCc2c1ccc(c2)C#N)CCCN(C)C

444	2801	1	0	0.85	known Negative	Clc1cc2N(c3c(CCc2cc1)cccc3)CCCN(C)C
445	2806	1	0	0.89	known Negative	Clc1ccccc1C(N1CCc2sccc2C1)C(OC)=O
446	2814	1	0	1.04	known Negative	Clc1ccccc1-c1noc(C)c1C(=O)NC1C2SC(C)(C)C(N2C1=O)C(O)=O
447	2973	1	0	0.96	known Negative	O=C(N(O)CCCCCNC(=O)CCC(=O)N(O)CCCCCN)CCC(=O)NCCCCCN(O)C(=O)C
448	3016	1	0	0.94	known Negative	Clc1cc2c(N(C)C(=O)CN=C2c2ccccc2)cc1
449	3111	1	0	0.81	known Negative	S(O)(=O)(=O)CN(C)C=1C(=O)N(N(C)C=1C)c1ccccc1
450	3114	1	0	0.85	known Negative	O=C(N)C(CCN(C(C)C)C(C)C)(c1ccccc1)c1ncccc1
451	3222	1	0	0.87	known Negative	O(C(=O)C(NC(C(=O)N1CCCC1C(O)=O)C)CCc1ccccc1)CC
452	3241	1	0	0.84	known Negative	N12C(c3c(Cc4c1cccc4)cccc3)CN=C2N
453	3291	1	1	0.37	known Negative	O=C1NC(=O)CC1(CC)C
454	3331	1	0	0.79	known Negative	O(CC(COC(=O)N)c1ccccc1)C(=O)N
455	3366	1	0	0.67	known Negative	FC1=CNC(=O)N=C1N
456	3404	1	0	0.92	known Negative	FC(F)(F)c1ccc(cc1)/C(=N\OCCN)/CCCCOC
457	3440	1	0	0.84	known Negative	Clc1cc(NCc2occc2)c(cc1S(=O)(=O)N)C(O)=O
458	3447	1	0	0.81	known Negative	O(C(=O)CCCCCNC(N)=N)c1ccc(cc1)C(OCC)=O
459	3454	1	1	0.67	known Negative	O=C1NC(=Nc2n(cnc12)COC(CO)CO)N
460	3475	1	0	0.96	known Negative	S(=O)(=O)(NC(=O)NN1CC2C(CCC2)C1)c1ccc(cc1)C
461	3590	1	0	0.46	known Negative	OC(CCCC(N)C)(C)C
462	3657	1	0	0.56	known Negative	O=C(NO)N
463	3658	1	0	0.90	known Negative	Clc1ccc(cc1)C(N1CCN(CC1)CCOCCO)c1ccccc1
464	3671	1	0	0.66	known Negative	O=C(C(C)C)c1c2n(nc1C(C)C)C=CC=C2
465	3706	1	0	1.00	known Negative	OC1Cc2c(cccc2)C1NC(=O)C(Cc1ccccc1)CC(O)CN1CCN(CC1C(=O)NC(C)(C)C)Cc1cccnc1
466	3749	1	0	0.98	known Negative	O=C1N(Cc2ccc(cc2)-c2ccccc2-c2nn[nH]n2)C(=NC12CCCC2)CCCC
467	3752	1	0	0.89	known Negative	Clc1ccc(Cl)cc1-c1nc(nc(n1)N)N
468	3825	1	1	0.57	known Negative	OC(=O)C(C)c1cc(ccc1)C(=O)c1ccccc1
469	3872	1	0	0.69	known Negative	O1C(O)(CO)C(O)C(OC2OC(CO)C(O)C(O)C2O)C1CO
470	3902	1	0	0.99	known Negative	n1cn(nc1)C(c1ccc(cc1)C#N)c1ccc(cc1)C#N
471	3913	1	0	0.75	known Negative	S1CCN2CC(N=C12)c1ccccc1
472	3937	1	0	0.85	known Negative	OC(=O)C1N(CCC1)C(=O)C(NC(CCc1ccccc1)C(O)=O)CCCCN

473	3961	1	0	1.07	known Negative	Clc1nc(n(Cc2ccc(cc2)-c2ccccc2-c2nn[nH]n2)c1CO)CCCC
474	3964	1	0	0.98	known Negative	Clc1cc2c(Oc3c(N=C2N2CCN(CC2)C)cccc3)cc1
475	3965	1	1	0.61	known Negative	O=C1CCCC1Cc1ccc(cc1)C(C(O)=O)C
476	4064	1	0	0.71	known Negative	O(CC(CCC)(COC(=O)N)C)C(=O)N
477	4066	1	0	0.89	known Negative	S1c2c(N(c3c1cccc3)CC1C3CCN(C1)CC3)cccc2
478	4158	1	0	0.62	known Negative	O(C(=O)C(C1NCCCC1)c1ccccc1)C
479	4171	1	0	0.75	known Negative	O(CC(O)CNC(C)C)c1ccc(cc1)CCOC
480	4173	1	0	0.64	known Negative	OCCn1c(ncc1[N+](=O)[O-])C
481	4178	1	0	0.51	known Negative	O(CC(N)C)c1c(cccc1C)C
482	4184	1	0	0.76	known Negative	N12C(c3c(Cc4c1cccc4)cccc3)CN(CC2)C
483	4192	1	0	1.05	known Negative	Clc1cc2c(-n3c(CN=C2c2ccccc2F)cnc3C)cc1
484	4213	1	0	0.55	known Negative	O1C(CO)C(O)C(O)C1n1cnc(C(=O)N)c1O
485	4235	1	0	0.79	known Negative	Clc1ccc(cc1)C(=O)NCCN1CCOCC1
486	4236	1	0	0.80	known Negative	S(=O)(C(c1ccccc1)c1ccccc1)CC(=O)N
487	4443	1	0	0.77	known Negative	OC(=O)C(NC(=O)C1CCC(CC1)C(C)C)Cc1ccccc1
488	4474	1	0	0.91	known Negative	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)CCN(Cc1ccccc1)C
489	4475	1	0	1.04	known Negative	Brc1cc(cnc1)C(OCC1CC2(OC)C(N(C1)C)Cc1c3c2cccc3n(c1)C)=O
490	4507	1	0	0.92	known Negative	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)CC
491	4528	1	0	0.81	known Negative	Nc1c2c(ccc1)C(CN(C2)C)c1ccccc1
492	4585	1	0	1.01	known Negative	s1c2Nc3c(N=C(N4CCN(CC4)C)c2cc1C)cccc3
493	4595	1	0	0.87	known Negative	O=C1c2c(n(c3c2cccc3)C)CCC1Cn1ccnc1C
494	4810	1	0	0.92	known Negative	Clc1nc(nc(OC)c1NC=1NCCN=1)C
495	4820	1	0	0.74	known Negative	O=C(Nc1c(cccc1C)C)CC12N(CCC1)CCC2
496	4835	1	0	1.07	known Negative	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(NC(=O)N1CCN(CC)C(=O)C1=O)c1ccccc1
497	4843	1	0	0.50	known Negative	O=C1N(CCC1)CC(=O)N
498	4888	1	0	0.77	known Negative	O1c2c(cc(cc2)C(C(O)=O)C)Cc2cccnc12
499	4915	1	0	0.67	known Negative	O=C(NC(C)C)c1ccc(cc1)CNNC
500	4932	1	0	0.75	known Negative	O(CC(O)CNCCC)c1ccccc1C(=O)CCc1ccccc1
501	5032	1	0	0.43	known Negative	OC(C(NC)C)c1ccccc1
502	5038	1	0	0.90	known Negative	O(C(=O)C(NC(C(=O)N1C2C(CC1C(O)=O)CCC2)C)CCc1ccccc1)CC
503	5064	1	1	0.52	known Negative	O1C(CO)C(O)C(O)C1n1nc(nc1)C(=O)N

504	5393	1	0	0.96	known Negative	S1CC(NC(CCc2ccccc2)C(OCC)=O)C(=O)N(CC1c1sccc1)CC(O)=O
505	5394	1	0	0.74	known Negative	O=C1n2c(N=NN1C)c(nc2)C(=O)N
506	5530	1	0	0.39	known Negative	NC1CC1c1ccccc1
507	5531	1	0	0.66	known Negative	n12ncnc1N=C(C=C2N(CC)CC)C
508	5574	1	0	0.84	known Negative	S1c2c(N(c3c1cccc3)CC(CN(C)C)C)cccc2
509	5584	1	1	0.75	known Negative	N(CC(CN1c2c(CCc3c1cccc3)cccc2)C)(C)C
510	5596	1	0	0.83	known Negative	O(C(=O)C(O)(c1ccccc1)c1ccccc1)C1CC2[N+]3(C(C1)CC2)CCCC3
511	5709	1	0	0.75	known Negative	N1CCN=C1Cc1c(cc(cc1C)C(C)(C)C)C
512	15387	1	0	0.69	known Negative	Fc1ccc(cc1)C(=O)CCCN1CCC(CC1)C
513	21109	1	0	0.75	known Negative	O(C)c1c(OC)c(OC)ccc1CN1CCNCC1
514	31018	1	1	0.47	known Negative	n1cc(ccc1)CN
515	34870	1	0	0.84	known Negative	OC(=O)CCCCCCNC1c2c(CCc3c1cccc3)ccccc2
516	38945	1	0	0.60	known Negative	O=C(Nc1c(cccc1C)C)C(N)C
517	41781	1	0	0.99	known Negative	S(=O)(=O)(NC(=O)NC(C)C)c1cnccc1Nc1cc(ccc1)C
518	51577	1	0	0.56	known Negative	OC1C(O)C(O)CN(CCO)C1CO
519	56339	1	0	0.91	known Negative	P(OCN1C(=O)C(NC1=O)(c1ccccc1)c1ccccc1)(O)(O)=O
520	56959	1	0	0.91	known Negative	O(CC(O)CN1CCN(CC1)CC(=O)Nc1c(cccc1C)C)c1ccccc1OC
521	59708	1	0	0.51	known Negative	O=C1N(CCC1)C(CC)C(=O)N
522	60490	1	0	0.85	known Negative	s1c2c(cc1C(N(O)C(=O)N)C)cccc2
523	60865	1	0	0.82	known Negative	O1Cc2c(cccc2)/C(/c2cc(ccc12)CC(O)=O)=C\CCN(C)C
524	71301	1	0	1.00	known Negative	Fc1cc2CCC(Oc2cc1)C(O)CNCC(O)C1Oc2c(cc(F)cc2)CC1
525	71771	1	0	0.99	known Negative	Clc1cccc(Cl)c1Nc1ccccc1CC(OCC(O)=O)=O
526	104865	1	0	1.06	known Negative	S(=O)(=O)(Nc1nc(nc(OCCO)c1Oc1ccccc1OC)-c1ncccn1)c1ccc(cc1)C(C)(C)C
527	107751	1	0	0.91	known Negative	S(C(=O)C)CC(Cc1ccccc1)C(=O)NCC(OCc1ccccc1)=O
528	123619	1	0	1.03	known Negative	Clc1cc(c(nc1)-c1ccc(nc1)C)-c1ccc(S(=O)(=O)C)cc1
529	216237	1	0	1.03	known Negative	Clc1cc2c(N(CCCC2O)C(=O)c2ccc(NC(=O)c3ccccc3C)cc2C)cc1
530	3000715	1	0	0.74	known Negative	S=C1NC(=O)C(C(CCC)C)(CC)C(=O)N1
531	3062	1	0	1.13	known Negative	Clc1cccc(C)c1NC(=O)c1sc(nc1)Nc1nc(n

	316					c(N2CCN(CC2)CCO)c1)C
532	3874387	1	0	0.91	known Negative	S(C=1C(C2N(C(=O)C2C(O)C)C=1C(O)=O)C)C1CC(NC1)C(=O)N(C)C
533	4169159	1	0	0.84	known Negative	O(C(=O)C(NC(C(=O)N1C2C(CC1C(O)=O)CCCC2)C)CCC)CC
534	5281007	1	0	0.72	known Negative	O=C(N)c1[nH]cnc1N=NN(C)C
535	54671008	1	0	1.04	known Negative	Fc1ccc(cc1)CNC(=O)C=1N=C(N(C)C(=O)C=1O)C(NC(=O)c1oc(nn1)C)(C)C
536	54681041	1	0	1.10	known Negative	OC12C(CC3C(C1=O)C(=O)c1c(C3)c(N(C)C)cc(NC(=O)CNC(C)(C)C)c1O)C(N(C)C)C(=O)C(C(=O)N)C2=O
537	54681536	1	0	1.00	known Negative	OC12C(C(N(C)C)C(=O)C(C(=O)N)C1=O)C(O)C1C(C2=O)C(=O)c2c(cccc2O)C1C
538	54685734	1	0	0.99	known Negative	OC12C(CC3C(C1=O)C(=O)c1c(cccc1O)C3(O)C)C(N(C)C)C(=O)C(C(=O)N)C2=O
539	54687237	1	0	1.05	known Negative	OC12C(CC3C(C1=O)C(=O)c1c(C3)c(N(C)C)ccc1O)C(N(C)C)C(=O)C(C(=O)N)C2=O
540	401	0	0	0.54	known Negative	O1NC(=O)C(N)C1
541	2141	0	0	0.76	known Negative	S(P(O)(O)=O)CCNCCCN
542	3154	0	0	0.93	known Negative	s1c2S(=O)(=O)C(CC(NCC)c2cc1S(=O)(=O)N)C
543	3202	0	0	0.48	known Negative	Oc1cc([N+](CC)(C)C)ccc1
544	3415	0	0	0.64	known Negative	P(O)(O)(=O)C(O)=O
545	3928	0	0	0.82	known Negative	S(C)C1OC(C(NC(=O)C2N(CC(C2)CCC)C)C(O)C)C(O)C(O)C1O
546	5184	0	0	0.83	known Negative	O1C2C3N(C(CC(OC(=O)C(CO)c4ccccc4)C3)C12)C
547	5358	0	0	0.86	known Negative	S(=O)(=O)(NC)Cc1cc2c([nH]cc2CCN(C)C)cc1
548	156850	0	0	0.99	known Negative	S(=O)(=O)(CCc1cc2c([nH]cc2CC2N(CCC2)C)cc1)c1ccccc1
549	3504707	0	0	0.97	known Negative	s1c2S(=O)(=O)N(CC(NCC)c2cc1S(=O)(=O)N)CCCOC
550	128495515	0	0	0.92	known Negative	S(C1CC(N(C1)C(=O)C(CSC(=O)c1ccccc1)C)C(O)=O)c1ccccc1
551	444	1	0	0.73	known Negative	Clc1cc(ccc1)C(=O)C(NC(C)(C)C)C
552	702	1	0	0.10	known Negative	OCC
553	2088	1	0	0.88	known Negative	P(O)(O)(=O)C(P(O)(O)=O)(O)CCCN
554	2717	1	0	0.81	known Negative	Clc1ccc(cc1)C1S(=O)(=O)CCC(=O)N1C
555	3305	1	0	0.84	known Negative	P(O)(O)(=O)C(P(O)(O)=O)(O)C
556	3325	1	0	0.91	known Negative	s1cc(nc1NC(N)=N)CSCC/C(=N/S(=O)(=O)N)/N
557	3698	1	0	0.64	known Negative	O=C1NC=C(C=C1N)c1ccncc1
558	3702	1	0	0.97	known Negative	Clc1ccc(cc1S(=O)(=O)N)C(=O)NN1c2c(CC1C)cccc2
559	3748	1	0	0.64	known Negative	O=C(NNC(C)C)c1ccncc1

560	4091	1	0	0.64	known Negative	N(C(NC(N)=N)=N)(C)C
561	4440	1	0	0.99	known Negative	S(=O)(=O)(NC)CCc1cc2c([nH]cc2C2CCN(CC2)C)cc1
562	4599	1	0	0.83	known Negative	O1C(CC(OC(=O)C(NC=O)CC(C)C)CCCCCCCCCCC)C(CCCCCC)C1=O
563	4674	1	0	0.87	known Negative	P(O)(O)(=O)C(P(O)(O)=O)(O)CCN
564	4909	1	1	0.57	known Negative	O=C1NCNC(=O)C1(CC)c1ccccc1
565	5245	1	0	0.87	known Negative	P(O)(O)(=O)C(P(O)(O)=O)(O)Cc1cccnc1
566	5487	1	0	1.00	known Negative	Clc1ccc2nsnc2c1NC=1NCCN=1
567	5726	1	0	0.86	known Negative	O1C(CO)C([N-][N+]#N)CC1N1C=C(C)C(=O)NC1=O
568	5734	1	0	0.82	known Negative	S(=O)(=O)(N)Cc1noc2c1cccc2
569	47528	1	0	0.60	known Negative	O([N+](=O)[O-])CCNC(=O)c1cccnc1
570	351727	1	0	1.00	known Negative	O(C)C12N(C3=C(C1COC(=O)N)C(=O)C(=N)C(C)C3=O)CC1NC12
571	311	0	0	0.53	known Negative	OC(CC(O)=O)(CC(O)=O)C(O)=O
572	727	0	0	0.83	known Negative	ClC1C(Cl)C(Cl)C(Cl)C(Cl)C1Cl
573	2018	0	0	0.45	known Negative	OC(=O)Cc1ccc(NC(=O)C)cc1
574	2021	0	0	0.83	known Negative	O1C2C(OC3OC(CC(=O)C13O)C)C(O)C(NC)C(O)C2NC
575	2083	0	0	0.64	known Negative	Oc1ccc(cc1CO)C(O)CNC(C)(C)C
576	2123	0	0	0.66	known Negative	n1c(nc(nc1N(C)C)N(C)C)N(C)C
577	2131	0	0	1.01	known Negative	Clc1ccccc1C[N+](CCNC(=O)C(=O)NCC[N+](Cc1ccccc1Cl)(CC)CC)(CC)CC
578	2193	0	1	0.59	known Negative	O=C1CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)CC3)C
579	2196	0	0	0.59	known Negative	O(C)c1ccc(cc1)C(=O)N1CCCC1=O
580	2206	0	0	0.59	known Negative	O=C1N(N(C)C(=C1)C)c1ccccc1
581	2218	0	0	0.79	known Negative	N(CCCN(CC)CC)(C1Cc2c(C1)cccc2)c1ccccc1
582	2242	0	0	0.81	known Negative	O(C(=O)C(NC(=O)C(N)CC(O)=O)Cc1ccccc1)C
583	2247	0	0	1.00	known Negative	Fc1ccc(cc1)Cn1c2c(nc1NC1CCN(CC1)CCc1ccc(OC)cc1)cccc2
584	2267	0	0	0.95	known Negative	Clc1ccc(cc1)CC1=NN(C2CCCN(CC2)C)C(=O)c2c1cccc2
585	2308	0	0	0.88	known Negative	ClC12C(C3CC(C)C(O)(C(=O)CO)C3(CC1O)C)CCC1=CC(=O)C=CC12C
586	2327	0	0	0.78	known Negative	Oc1c(O)c(O)ccc1CNNC(=O)C(N)CO
587	2448	0	0	0.91	known Negative	Brc1ccc(cc1)C1(O)CCN(CC1)CCCC(=O)c1ccc(F)cc1
588	2466	0	0	0.62	known Negative	O(CCCC)c1ccc(cc1)CC(=O)NO
589	2749	0	0	0.61	known Negative	O=C1N(O)C(=CC(=C1)C)C1CCCCC1
590	2783	0	0	0.88	known Negative	Clc1cc(cc(Cl)c1N)C(O)CNC(C)(C)C
591	2816	0	0	1.00	known Negative	Clc1ccccc1C12OCCN1CC(=O)Nc1c2cc(Cl)cc1

592	2882	0	0	0.85	known Negative	O1c2c(C(=O)C=C1C(O)=O)c(OCC(O)CO c1c3c(OC(=CC3=O)C(O)=O)ccc1)ccc2
593	2883	0	0	0.52	known Negative	O=C(N(CC)c1ccccc1C)\C=C\C
594	3152	0	0	0.81	known Negative	O(C)c1cc2c(CC(CC3CCN(CC3)Cc3ccccc 3)C2=O)cc1OC
595	3285	0	1	0.75	known Negative	OC1(CCC2C3C(CCC12C)c1c(cc(O)cc1)C C3)C#C
596	3341	0	0	0.96	known Negative	Clc1c2c(cc(O)c1O)C(CNCC2)c1ccc(O)cc 1
597	3343	0	0	0.83	known Negative	Oc1cc(cc(O)c1)C(O)CNC(Cc1ccc(O)cc1) C
598	3372	0	0	1.01	known Negative	S1c2c(N(c3c1cccc3)CCCN1CCN(CC1)C CO)cc(cc2)C(F)(F)F
599	3379	0	0	0.82	known Negative	FC1C2=CC(=O)C=CC2(C2C(C3CC4OC(O C4(C(=O)CO)C3(CC2O)C)(C)C)C1)C
600	3384	0	0	0.79	known Negative	FC12C(C3CCC(O)(C(=O)C)C3(CC1O)C)C C(C1=CC(=O)C=CC12C)C
601	3393	0	0	1.04	known Negative	Clc1cc2c(N(CCN(CC)CC)C(=O)CN=C2c2 ccccc2F)cc1
602	3775	0	0	0.91	known Negative	O=C(N)C(CC[N+](C(C)C)(C(C)C)C)(c1ccc cc1)c1ccccc1
603	3938	0	0	0.92	known Negative	O=C(NC1C=C2C(N(C1)C)Cc1c3c2cccc3[nH]c1)N(CC)CC
604	3948	0	0	0.87	known Negative	Fc1c2N(C=C(C(O)=O)C(=O)c2cc(F)c1N 1CC(NCC1)C)CC
605	4043	0	1	0.75	known Negative	OC1C2C(C3CCC(C(=O)C)C3(C1)C)CC(C 1=CC(=O)CCC12C)C
606	4063	0	1	0.70	known Negative	OC1(C(=O)CO)C2(CC(=O)C3C(C2CC1C) CCC1=CC(=O)C=CC13C)C
607	4101	0	0	0.57	known Negative	N12CN3CN(C1)CN(C2)C3
608	4140	0	0	0.92	known Negative	OCC(NC(=O)C1C=C2C(N(C1)C)Cc1c3c2 cccc3[nH]c1)CC
609	4197	0	0	0.75	known Negative	O=C1NC(C)=C(C=C1C#N)c1ccncc1
610	4534	0	0	0.79	known Negative	Oc1cc(ccc1O)CC(C(Cc1cc(O)c(O)cc1)C) C
611	4623	0	0	1.10	known Negative	Clc1cc(Cl)ccc1CO\N=C(/Cn1ccnc1)\c1c cc(Cl)cc1Cl
612	4812	0	0	0.72	known Negative	O=C1c2c(cccc2)C(=O)C(C)=C1C\C=C(\ CCCC(CCCC(CCCC(C)C)C)C)/C
613	4848	0	0	0.95	known Negative	O=C1Nc2cccnc2N(c2c1cccc2)C(=O)CN 1CCN(CC1)C
614	4906	0	0	0.69	known Negative	O=C(Nc1ccccc1C)C(NCCC)C
615	5152	0	0	0.82	known Negative	Oc1ccc(cc1CO)C(O)CNCCCCCCOCCCCc 1ccccc1
616	5314	0	0	0.67	known Negative	O(C(=O)CCC(OCC[N+](C)(C)C)=O)CC[N +](C)(C)C
617	5356	0	0	0.86	known Negative	S(=O)(=O)(N)c1ccc(N2S(=O)(=O)CCCC2)cc1
618	5430	0	0	0.82	known Negative	s1cc(nc1)-c1[nH]c2c(n1)cccc2
619	5486	0	0	0.94	known Negative	S(=O)(=O)(NC(Cc1ccc(OCCCCC2CCNCC

						2)cc1)C(O)=O)CCCC
620	5515	0	0	0.98	known Negative	O1CC2=C(C=C3N(Cc4c3nc3c(c4)c(CN(C)C)c(O)cc3)C2=O)C(O)(CC)C1=O
621	5719	0	0	0.95	known Negative	O=C(N(CC)c1cc(ccc1)C=1n2ncc(c2N=CC=1)C#N)C
622	5721	0	0	0.80	known Negative	O1C(C(O)C(O)CO)C(NC(=O)C)C(NC(N)=N)C=C1C(O)=O
623	5731	0	0	0.83	known Negative	O1CC(NC1=O)Cc1cc2c([nH]cc2CCN(C)C)cc1
624	9417	0	0	0.88	known Negative	OCCN1CCN(CC1)CCCN1c2c(C=Cc3c1cccc3)cccc2
625	9419	0	0	0.81	known Negative	O=C1N(c2c(N(c3c1cccc3)C)cccc2)CCN(C)C
626	11289	0	0	0.96	known Negative	Cl\C(=C(/c1ccc(OC)cc1)\c1ccc(OC)cc1)\c1ccc(OC)cc1
627	14955	0	0	1.02	known Negative	O1c2cc3C([N+](CCc3cc2OC)(C)C)Cc2ccc(Oc3c4C([N+](CCc4cc(OC)c3OC)(C)C)Cc3cc1c(OC)cc3)cc2
628	15459	0	0	0.60	known Negative	O1C(CNC1=O)COc1cc(cc(c1)C)C
629	17012	0	0	1.03	known Negative	S1c2c(cc(cc2)C(F)(F)F)\C(\c2c1cccc2)=C\CCN1CCN(CC1)CCO
630	19861	0	0	0.75	known Negative	n1c/2c(CCc3c(cccc3)\C\2=C\2/CCN(CC/2)C)ccc1
631	27400	0	0	0.84	known Negative	s1c2c(cc1)/C(/c1c(CC2)cccc1)=C\1/CCN(CC/1)C
632	31264	0	0	0.39	known Negative	O1C(OC(OC1C)C)C
633	33625	0	0	0.83	known Negative	O=C(NC1CCN(CC1)CCc1c2c([nH]c1)cccc2)c1ccccc1
634	50294	0	0	0.85	known Negative	O1c2c(cc3c(N(CC)C(=CC3=O)C(O)=O)c2CCC)C(=O)C=C1C(O)=O
635	51081	0	0	0.83	known Negative	Fc1cc2c(N(C=C(C(O)=O)C2=O)CC)cc1N1CCN(CC1)C
636	51263	0	0	1.06	known Negative	O=C1N(N=NN1CC)CCN1CCC(N(C(=O)CC)c2ccccc2)(CC1)COC
637	60795	0	0	1.07	known Negative	Clc1c(N2CCN(CC2)CCCCOc2cc3NC(=O)CCc3cc2)cccc1Cl
638	60854	0	0	1.14	known Negative	Clc1cc2NC(=O)Cc2cc1CCN1CCN(CC1)c1nsc2c1cccc2
639	71329	0	0	1.00	known Negative	S(=O)(=O)(Nc1ccc(cc1)CCN(CCOc1ccc(NS(=O)(=O)C)cc1)C)C
640	73342	0	0	0.85	known Negative	N(Cc1c2c(ccc1)cccc2)(C\C=C\c1ccccc1)C
641	107706	0	0	0.99	known Negative	S(=O)(=O)(Nc1ccccc1C(=O)NCC(O)=O)c1ccc(OC(=O)C(C)(C)C)cc1
642	125564	0	0	0.99	known Negative	Fc1ccc(cc1)-c1cc(nc2c1CCCCCC2)N1CCN(CC1)CC
643	176167	0	0	0.99	known Negative	O=C1NC(=O)C(=C1c1c2c(n(c1)C1CCN(CC1)Cc1ncccc1)cccc2)c1c2c(n(c1)C)cccc2

644	249187	0	0	0.85	known Negative	O1C(COC(=O)C)C(OC(=O)C)C(OC(=O)C)C1N1N=CC(=O)NC1=O
645	495574	0	0	0.82	known Negative	S(O)(=O)(=O)c1cc2c(cc1C(C)C)CCC1C(CCCC12C)(C(O)=O)C
646	521440	0	1	0.61	known Negative	OC1(CCC2C3C(C4(C(=CC(=O)CC4)CC3)C)C(=O)CC12C)C(=O)CO
647	5230270	0	0	0.96	known Negative	Fc1cc2c(cc1)C(C(C)C)C(OC(=O)COC)(C2)CCN(CCCc1[nH]c2c(n1)cccc2)C
648	5466794	0	0	1.00	known Negative	Clc1ccccc1C=1Oc2c(C(=O)C=1)c(O)cc(O)c2C1CCN(CC1O)C
649	6844184	0	0	0.94	known Negative	O(C)c1cc2c([nH]cc2\C=N\NC(NCCCCC)=N)cc1
650	9882672	0	0	0.72	known Negative	O=C1N(CC2CCCc3c2c1ccc3)C1C2CCN(C1)CC2
651	40469209	0	0	0.93	known Negative	Oc1ccc(N=Nc2ccc(cc2)C(=O)NCCC(O)=O)cc1C(O)=O
652	54486399	0	0	1.12	known Negative	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC(Cn3nc(nn3)C)=C(N2C1=O)C(O)=O
653	596	1	1	0.63	known Negative	O1C(CO)C(O)C(O)C1N1C=CC(=NC1=O)N
654	1302	1	0	0.70	known Negative	O(C)c1cc2c(cc(cc2)C(C(O)=O)C)cc1
655	1983	1	1	0.39	known Negative	Oc1ccc(NC(=O)C)cc1
656	2187	1	0	1.00	known Negative	n1cn(nc1)Cc1cc(cc(c1)C(C#N)(C)C)C(C#N)(C)C
657	2269	1	0	0.86	known Negative	O1C(CC)C(O)(C)C(O)C(N(CC(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C)C)C
658	2274	1	0	0.99	known Negative	s1cc(nc1N)/C(=N\OC(C(O)=O)(C)C)/C(=O)NC1C(N(S(O)(=O)=O)C1=O)C
659	2315	1	0	0.92	known Negative	S(=O)(=O)(N)c1cc2S(=O)(=O)NC(Nc2cc1C(F)(F)F)Cc1ccccc1
660	2467	1	0	0.73	known Negative	O(C)c1cc(OC)cc(OC)c1C(=O)CCCN1CCCC1
661	2519	1	0	0.61	known Negative	O=C1N(C)C(=O)N(c2ncn(c12)C)C
662	2554	1	0	0.84	known Negative	O=C(N)N1c2c(C=Cc3c1cccc3)cccc2
663	2609	1	0	1.03	known Negative	ClC=1CSC2N(C(=O)C2NC(=O)C(N)c2cccc2)C=1C(O)=O
664	2630	1	0	1.12	known Negative	S1C2N(C(=O)C2NC(=O)C(NC(=O)N2CCN(CC)C(=O)C2=O)c2ccc(O)cc2)C(C(O)=O)=C(C1)CSc1nnnn1C
665	2640	1	0	1.12	known Negative	s1nc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC(C[n+]3c4n(N=CC=C4)cc3)=C(N2C1=O)C(O)=O
666	2725	1	0	0.78	known Negative	Clc1ccc(cc1)C(CCN(C)C)c1ncccc1
667	2763	1	0	0.85	known Negative	ClC1(Cl)CC1c1ccc(OC(C(O)=O)(C)C)cc1
668	2907	1	0	0.88	known Negative	ClCCN(P1(OCCCN1)=O)CCCl
669	2910	1	0	0.95	known Negative	Clc1cc2NC(NS(=O)(=O)c2cc1S(=O)(=O)N)C1C2CC(C1)C=C2
670	2913	1	0	0.78	known Negative	N1(CCC(CC1)=C1c2c(C=Cc3c1cccc3)cccc2)C

671	3043	1	1	0.50	known Negative	O1C(CCC1n1c2N=CNC(=O)c2nc1)CO
672	3063	1	0	0.78	known Negative	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC)cc3)CCC2O
673	3076	1	0	1.01	known Negative	S1c2c(N(CCN(C)C)C(=O)C(OC(=O)C)C1c1ccc(OC)cc1)cccc2
674	3151	1	0	1.06	known Negative	Clc1cc2NC(=O)N(c2cc1)C1CCN(CC1)CCCN1c2c(NC1=O)cccc2
675	3226	1	0	0.73	known Negative	ClC(F)C(F)(F)OC(F)F
676	3255	1	0	0.93	known Negative	O1C(CC)C(O)(C)C(O)C(C)C(=O)C(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C
677	3324	1	0	0.88	known Negative	O(C(=O)C)CC(CCn1c2nc(ncc2nc1)N)COC(=O)C
678	3354	1	0	0.78	known Negative	O1c2c(cccc2C(OCCN2CCCCC2)=O)C(=O)C(C)=C1c1ccccc1
679	3365	1	0	0.93	known Negative	Fc1cc(F)ccc1C(O)(Cn1ncnc1)Cn1ncnc1
680	3386	1	0	0.86	known Negative	FC(F)(F)c1ccc(OC(CCNC)c2ccccc2)cc1
681	3394	1	0	0.75	known Negative	Fc1cc(ccc1-c1ccccc1)C(C(O)=O)C
682	3461	1	0	0.70	known Negative	FC1(F)C(O)C(OC1N1C=CC(=NC1=O)N)CO
683	3463	1	0	0.53	known Negative	O(CCCC(C(O)=O)(C)C)c1cc(ccc1C)C
684	3495	1	0	0.88	known Negative	O1C(C(O)=O)C(O)C(O)C(OC2OC(C(O)=O)C(O)C(O)C2O)C1OC1CCC2(C3C(CCC2C1(C)C)(C)C1(C(C2CC(CCC2(CC1)C)(C(O)=O)C)=CC3=O)C)C
685	3562	1	1	0.76	known Negative	BrC(Cl)C(F)(F)F
686	3639	1	0	0.91	known Negative	Clc1cc2NCNS(=O)(=O)c2cc1S(=O)(=O)N
687	3640	1	1	0.62	known Negative	OC1(CCC2C3C(C4(C(=CC(=O)CC4)CC3)C)C(O)CC12C)C(=O)CO
688	3689	1	0	0.77	known Negative	Oc1ccc(cc1)C(O)C(N1CCC(CC1)Cc1ccccc1)C
689	3695	1	0	0.81	known Negative	S(CCNC=N)C=1CC2N(C(=O)C2C(O)C)C=1C(O)=O
690	3763	1	0	0.67	known Negative	ClC(OC(F)F)C(F)(F)F
691	3792	1	0	0.86	known Negative	O(C)c1cc(ccc1OC)C(=O)NCc1ccc(OCCN(C)C)cc1
692	3827	1	0	0.86	known Negative	s1c2c(cc1)\C(\c1c(CC2=O)cccc1)=C\1/CCN(CC/1)C
693	3877	1	0	0.70	known Negative	S1CC(OC1CO)N1C=CC(=NC1=O)N
694	3950	1	0	0.79	known Negative	ClCCN(N=O)C(=O)NC1CCCCC1
695	4075	1	1	0.39	known Negative	Oc1ccc(N)cc1C(O)=O
696	4159	1	1	0.70	known Negative	OC1(CCC2C3C(C4(C(=CC(=O)C=C4)C(C3)C)C)C(O)CC12C)C(=O)CO
697	4259	1	0	0.93	known Negative	Fc1cc2c(N(C=C(C(O)=O)C2=O)C2CC2)c(OC)c1N1CC2C(NCCC2)C1
698	4460	1	0	0.95	known Negative	O1C(OC2C(O)C(OC3OCC(O)(C)C(NC)C3O)C(NCC)CC2N)C(N)CC=C1CN
699	4463	1	0	0.87	known Negative	O=C1Nc2c(nccc2C)N(c2ncccc12)C1CC

						1
700	4506	1	0	0.98	known Negative	O=C1Nc2c(cc([N+](=O)[O-])cc2)C(=NC1)c1ccccc1
701	4539	1	0	0.83	known Negative	Fc1cc2c(N(C=C(C(O)=O)C2=O)CC)cc1N1CCNCC1
702	4583	1	0	0.86	known Negative	Fc1cc2c3N(C=C(C(O)=O)C2=O)C(COc3c1N1CCN(CC1)C)C
703	4616	1	0	0.95	known Negative	Clc1cc2c(NC(=O)C(O)N=C2c2ccccc2)cc1
704	4691	1	0	0.88	known Negative	Fc1ccc(cc1)C1CCNCC1COc1cc2OCOc2cc1
705	4889	1	0	0.70	known Negative	O(C(=O)C(CC)C)C1C2C(=CC(O)C1)C=CC(C)C2CCC(O)CC(O)CC(O)=O
706	4894	1	1	0.65	known Negative	OC1(CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)C(=O)CO
707	4943	1	0	0.39	known Negative	Oc1c(cccc1C(C)C)C(C)C
708	5070	1	0	0.85	known Negative	s1c2cc(OC(F)(F)F)ccc2nc1N
709	5106	1	0	0.87	known Negative	O1C(CC)C(O)(C)C(O)C(C)\C(=N/OCOCCOC)\C(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C
710	5203	1	0	0.94	known Negative	Clc1cc(ccc1Cl)C1CCC(NC)c2c1cccc2
711	5329	1	0	0.90	known Negative	S(=O)(=O)(Nc1noc(c1)C)c1ccc(N)cc1
712	5379	1	0	0.91	known Negative	Fc1cc2c(N(C=C(C(O)=O)C2=O)C2CC2)c(OC)c1N1CC(NCC1)C
713	5429	1	0	0.58	known Negative	O=C1NC(=O)N(c2ncn(c12)C)C
714	5514	1	0	0.86	known Negative	S(OCC12OC(OC1C1OC(OC1CO2)(C)C)(C)C)(=O)(=O)N
715	5523	1	0	0.69	known Negative	O(C)c1cc(ccc1)C1(O)CCCCC1CN(C)C
716	5577	1	0	0.89	known Negative	O(C)c1c(OC)cc(cc1OC)C(=O)NCc1ccc(OCCN(C)C)cc1
717	5578	1	0	0.89	known Negative	O(C)c1c(OC)cc(cc1OC)Cc1cnc(nc1N)N
718	5597	1	0	0.79	known Negative	O(C)c1c(OC)cc(cc1OC)C(=O)NC1CCCNC1
719	5720	1	0	0.81	known Negative	S1c2c(cccc2)C(=O)Cc2cc(ccc12)C(C(O)=O)C
720	5732	1	0	0.92	known Negative	O=C(N(C)C)Cc1n2C=C(C=Cc2nc1-c1ccc(cc1)C)C
721	5735	1	0	1.06	known Negative	Clc1ccc(nc1)N1C(=O)C2=NC=CNC2=C1OC(=O)N1CCN(CC1)C
722	7703	1	1	0.31	known Negative	O(C)c1ccc(cc1)\C=C\C
723	31072	1	0	0.64	known Negative	S=C1N(C=CN1C)C(OCC)=O
724	39042	1	0	0.88	known Negative	Clc1ccc(cc1)C(=O)NCCc1ccc(OC(C(O)=O)(C)C)cc1
725	60021	1	0	0.95	known Negative	Fc1cc(F)ccc1N1C=C(C(O)=O)C(=O)c2cc(F)c(N3CC(NCC3)C)cc12
726	68865	1	0	0.78	known Negative	O(C(=O)CN(CC)CC)c1ccc(NC(=O)C)cc1

727	119607	1	0	1.08	known Negative	S(=O)(=O)(N)c1ccc(cc1)-c1c(noc1C)-c1ccccc1
728	124087	1	0	0.85	known Negative	Clc1cc2c(cc1)\C(\c1ncccc1CC2)=C\1/CCNCC/1
729	197712	1	0	0.89	known Negative	O(C(C(OC)(c1ccccc1)c1ccccc1)C(O)=O)c1nc(cc(n1)C)C
730	216239	1	0	1.00	known Negative	Clc1ccc(NC(=O)Nc2ccc(Oc3cc(ncc3)C(=O)NC)cc2)cc1C(F)(F)F
731	368341	1	0	1.02	known Negative	s1c(ccc1N(Cc1cc2c(N=C(NC2=O)C)cc1)C)C(=O)NC(CCC(O)=O)C(O)=O
732	400633	1	0	0.90	known Negative	FC1=CN(C2OC(C)C(O)C2O)C(=O)N=C1NC(OCCCCC)=O
733	465466	1	0	0.85	known Negative	Fc1cc2c3N(C=C(C(O)=O)C2=O)C(COc3c1C1(N)CC1)C
734	479930	1	0	0.99	known Negative	FC(F)Oc1c2N(C=C(C(O)=O)C(=O)c2ccc1-c1cc2c(cc1)C(NC2)C)C1CC1
735	2761171	1	0	0.62	known Negative	S=C(N)c1cc(ncc1)CC
736	4659569	1	0	0.89	known Negative	Oc1c([N+](=O)[O-])cc(cc1O)C(=O)c1ccc(cc1)C
737	4663848	1	0	0.93	known Negative	O1C(CC)C(O)(C)C(O)C(C)C(=O)C(CC(OC)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C
738	5384001	1	0	1.04	known Negative	S(=O)(=O)(Nc1ncccc1)c1ccc(N=Nc2cc(C(O)=O)c(O)cc2)cc1
739	6433119	1	0	1.07	known Negative	Clc1sc(cc1)C(=O)NCC1OC(=O)N(C1)c1ccc(N2CCOCC2=O)cc1
740	216963 52	1	0	1.10	known Negative	O(C(=O)NC(C(C)(C)C)C(=O)NC(Cc1ccccc1)C(O)CN(NC(=O)C(NC(OC)=O)C(C)(C)C)Cc1ccc(cc1)-c1ncccc1)C
741	218655 26	1	0	0.76	known Negative	O1C2(CCC1=O)CCC1C3C4(OC4CC12C)C1(C(CC3C(OC)=O)=CC(=O)CC1)C
742	546765 37	1	0	0.97	known Negative	O1c2c(cccc2)C(O)=C(C(CC(=O)C)c2ccc([N+](=O)[O-])cc2)C1=O
743	546774 70	1	0	0.90	known Negative	s1c(cnc1NC(=O)C1N(S(=O)(=O)c2c(ccc2)C1=O)C)C
744	665831 67	1	0	1.04	known Negative	Oc1c(cccc1N=NC=1C(=O)N(NC=1C)c1cc(C)c(cc1)C)-c1cc(ccc1)C(O)=O
745	2522	0	0	0.88	known Positive	OC1CC(O)C\C(=C/C=C\2/C3CCC(C(\C=C\C(O)C4CC4)C)C3(CCC/2)C)\C1=C
746	2612	0	0	0.95	known Positive	s1cccc1CC(=O)NC1C2SCC(COC(=O)C)=C(N2C1=O)C(O)=O
747	2683	0	0	0.80	known Positive	[n+]1(ccccc1)CCCCCCCCCCCCCCCC
748	2713	0	0	1.10	known Positive	Clc1ccc(NC(NC(NCCCCCCNC(NC(Nc2ccc(Cl)cc2)=N)=N)=N)=N)cc1
749	2722	0	0	0.83	known Positive	Clc1cc(Cl)c2c(nccc2)c1O
750	2761	0	0	0.89	known Positive	N1(CCN(CC1)C\C=C\c1ccccc1)C(c1ccccc1)c1ccccc1
751	3049	0	1	0.87	known Positive	Oc1ccc(cc1)/C(/C(=C/C)/c1ccc(O)cc1)=C/C
752	3180	0	0	0.64	known Positive	O(CCCC)c1ccc(cc1)C(=O)CCN1CCCCC1

753	3245	0	0	0.77	known Positive	O(C(CN1CCN(CC1)CC(C(=O)c1ccccc1)C)c1ccccc1)CC
754	3249	0	1	0.90	known Positive	OC1C\C(=C/C=C\2/C3CCC(C(\C=C\C(C(C)C)C)C)C3(CCC/2)C)\C(CC1)=C
755	3278	0	1	0.86	known Positive	Clc1c(Cl)c(OCC(O)=O)ccc1C(=O)C(CC)=C
756	3735	0	0	0.98	known Positive	Ic1c(CC(CC)C(O)=O)c(I)cc(I)c1N
757	4033	0	1	0.61	known Positive	ClCCN(CCCl)C
758	4080	0	0	0.78	known Positive	O(C)c1cc2CCC3C4CCC(O)(C#C)C4(CCC3c2cc1)C
759	4121	0	0	1.00	known Positive	Clc1cc2NC(N(S(=O)(=O)c2cc1S(=O)(=O)N)C)CCl
760	4756	0	0	0.98	known Positive	n1c(N)c(N=Nc2ccccc2)ccc1N
761	5517	0	0	0.95	known Positive	Fc1cc(F)ccc1N1C=C(C(O)=O)C(=O)c2cc(F)c(nc12)N1CC(N)CC1
762	5546	0	1	1.02	known Positive	n1c(N)c2nc(-c3ccccc3)c(nc2nc1N)N
763	8808	0	0	0.94	known Positive	S(O)(=O)(=O)N=Nc1ccc(N(C)C)cc1
764	13505	0	0	1.00	known Positive	O(C(=O)C1(CCN(CC1)CCC(C#N)(c1ccccc1)c1ccccc1)c1ccccc1)CC
765	13764	0	0	0.88	known Positive	o1c(ccc1[N+](=O)[O-])\C=N\NC(=O)c1ccc(O)cc1
766	23568	0	0	1.00	known Positive	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)c1c2c(ccc1OCC)cccc2
767	35330	0	0	0.83	known Positive	O(C(=O)c1ccc(cc1)C)c1cc(ccc1OC(=O)c1ccc(cc1)C)C(O)CNC(C)(C)C
768	57469	0	0	0.90	known Positive	n1c2c(c3n(cnc3c1N)CC(C)C)cccc2
769	60164	0	0	0.88	known Positive	O(C)c1ccc(cc1C12CC3CC(C1)CC(C2)C3)-c1cc2c(cc(cc2)C(O)=O)cc1
770	68733	0	0	0.88	known Positive	O(C(=O)C=1C(C(C(OCC)=O)=C(NC=1C)C)c1ccccc1\C=C\C(OC(C)(C)C)=O)CC
771	82146	0	1	0.87	known Positive	OC(=O)c1ccc(cc1)C(=C)c1cc2c(cc1C)C(CCC2(C)C)(C)C
772	206044	0	0	0.92	known Positive	o1c(nc(CCOc2ccc(cc2)CN(CC(O)=O)C(Oc2ccc(OC)cc2)=O)c1C)-c1ccccc1
773	522463	0	1	0.67	known Positive	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)C=C1)CC3)C)C
774	4460696	0	0	0.79	known Positive	OC1C2C(C3CC(C)C(C(=O)CC)(C)C3(C1)C)CCC1=CC(=O)C=CC12C
775	5280443	0	1	0.89	known Positive	O1c2c(C(=O)C=C1c1ccc(O)cc1)c(O)cc(O)c2
776	140966625	0	0	1.05	known Positive	s1nc(N2CCN(CC2)CCCCN2C(=O)C3C(CCCC3)C2=O)c2c1cccc2
777	546760 38	0	1	0.87	known Positive	O1c2c(cccc2)C(O)=C(CC=2C(Oc3c(cccc3)C=2O)=O)C1=O
778	1071	1	1	0.83	known Positive	OC\C=C(\C=C\C=C(\C=C\C=1C(CCCC=1C)(C)C)/C)/C
779	1691	1	1	0.98	known Positive	O1C(C)C(O)C(N)CC1OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(C1=O)c(OC)ccc3)c2O)C(=O)CO

780	1805	1	1	0.72	known Positive	O1C(CO)C(O)C(O)C1N1C=NC(=NC1=O)N
781	2092	1	0	0.99	known Positive	O1CCCC1C(=O)NCCCN(C)c1nc(N)c2cc(OC)c(OC)cc2n1
782	2447	1	0	0.77	known Positive	BrC(C(C)C)C(=O)NC(=O)N
783	2578	1	0	0.81	known Positive	ClCCN(N=O)C(=O)NCCCl
784	2610	1	0	0.95	known Positive	S1C2N(C(=O)C2NC(=O)C(N)c2ccc(O)cc2)C(C(O)=O)=C(C1)C
785	2666	1	0	0.96	known Positive	S1C2N(C(=O)C2NC(=O)C(N)c2ccccc2)C(C(O)=O)=C(C1)C
786	2754	1	0	0.90	known Positive	O(CCCCc1nnnn1C1CCCCC1)c1cc2CCC(=O)Nc2cc1
787	2764	1	0	0.87	known Positive	Fc1cc2c(N(C=C(C(O)=O)C2=O)C2CC2)cc1N1CCNCC1
788	2828	1	0	0.79	known Positive	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC)cc3)C=CC2O
789	2833	1	0	0.98	known Positive	O(C)C1=CC=C2C(=CC1=O)C(NC(=O)C)CCc1c2c(OC)c(OC)c(OC)c1
790	2889	1	1	0.94	known Positive	O(C)c1cc(ccc1O)\C=C\C(=O)CC(=O)\C=C\c1cc(OC)c(O)cc1
791	2949	1	0	0.87	known Positive	o1ncc2CC3(C4C(C5CCC(O)(C#C)C5(CC4)C)CCC3=Cc12)C
792	3181	1	1	0.75	known Positive	O=C1CCC2(C3C(C4CCC(C(=O)C)C4(CC3)C)C=CC2=C1)C
793	3236	1	0	0.60	known Positive	O=C(C(CN1CCCCC1)C)c1ccc(cc1)CC
794	3363	1	0	0.65	known Positive	FC1=CN(C2OC(CO)C(O)C2)C(=O)NC1=O
795	3397	1	0	0.83	known Positive	FC(F)(F)c1cc(NC(=O)C(C)C)ccc1[N+](=O)[O-]
796	3685	1	0	1.00	known Positive	O1C(C)C(O)C(N)CC1OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(cccc3)C1=O)c2O)C(=O)C
797	3793	1	0	1.05	known Positive	Clc1cc(Cl)ccc1C1(OC(CO1)COc1ccc(N2CCN(CC2)c2ccc(N3C=NN(C(CC)C)C3=O)cc2)cc1)Cn1ncnc1
798	3817	1	0	1.05	known Positive	O(CC(=O)NC(Cc1ccccc1)C(O)CC(NC(=O)C(N1CCCNC1=O)C(C)C)Cc1ccccc1)c1c(cccc1C)C
799	3823	1	0	1.07	known Positive	Clc1cc(Cl)ccc1C1(OC(CO1)COc1ccc(N2CCN(CC2)C(=O)C)cc1)Cn1ccnc1
800	3883	1	0	1.02	known Positive	S(=O)(Cc1nccc(OCC(F)(F)F)c1C)c1[nH]c2c(n1)cccc2
801	3899	1	0	0.80	known Positive	FC(F)(F)c1ccc(NC(=O)c2cnoc2C)cc1
802	3957	1	0	0.96	known Positive	Clc1cc2c(cc1)\C(\c1ncccc1CC2)=C\1/CCN(CC/1)C(OCC)=O
803	4044	1	0	0.70	known Positive	OC(=O)c1ccccc1Nc1cccc(C)c1C
804	4212	1	0	1.04	known Positive	Oc1c2c(C(=O)c3c(C2=O)c(NCCNCCO)ccc3NCCNCCO)c(O)cc1
805	4449	1	0	1.06	known Positive	Clc1cc(N2CCN(CC2)CCCN2N=C(N(CCOc3ccccc3)C2=O)CC)ccc1

806	4493	1	0	0.86	known Positive	FC(F)(F)c1cc(N2C(=O)C(NC2=O)(C)C)ccc1[N+](=O)[O-]
807	4536	1	1	0.74	known Positive	OC1(CCC2C3C(C4C(=CC(=O)CC4)CC3)CCC12C)C#C
808	4638	1	1	0.76	known Positive	OC1(CCC2C3C(CCC12C)C1(CC(C=O)C(=O)CC1CC3)C)C
809	4829	1	1	0.95	known Positive	S1C(Cc2ccc(OCCc3ncc(cc3)CC)cc2)=C(O)NC1=O
810	4842	1	0	0.85	known Positive	BrCCC(=O)N1CCN(CC1)C(=O)CCBr
811	4844	1	1	0.96	known Positive	O1C(C)C(OC2OCCCC2)C(N)CC1OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(C1=O)c(OC)ccc3)c2O)C(=O)CO
812	4887	1	0	0.97	known Positive	O1c2c(cccc2NC(=O)c2ccc(OCCCCc3ccccc3)cc2)C(=O)C=C1c1nn[nH]n1
813	4890	1	0	0.97	known Positive	Clc1cc2c(N(CC3CC3)C(=O)CN=C2c2ccccc2)cc1
814	5029	1	0	1.01	known Positive	S(=O)(Cc1nccc(OCCCOC)c1C)c1[nH]c2c(n1)cccc2
815	5291	1	0	0.93	known Positive	O=C(Nc1cc(Nc2nc(ccn2)-c2cccnc2)c(cc1)C)c1ccc(cc1)CN1CCN(CC1)C
816	5408	1	1	0.59	known Positive	OC1CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)CC3)C
817	5573	1	0	0.83	known Positive	O(C)c1c(OC)cc(cc1OC)C(OCC(N(C)C)(CC)c1ccccc1)=O
818	5591	1	1	1.02	known Positive	S1C(Cc2ccc(OCC3(Oc4c(CC3)c(C)c(O)c(C)c4C)C)cc2)=C(O)NC1=O
819	5645	1	0	0.74	known Positive	OC1C2C3CCC(C(CCC(O)=O)C)C3(CCC2C2(C(C1)CC(O)CC2)C)C
820	8249	1	0	0.68	known Positive	N(CCc1ccccc1)C(NC(N)=N)=N
821	9354	1	0	0.73	known Positive	SC(C(S)C(O)=O)C(O)=O
822	26248	1	1	0.85	known Positive	O(C(=O)C)c1c2c(cc(c1)C(O)=O)C(=O)c1c(C2=O)c(OC(=O)C)ccc1
823	41693	1	0	0.91	known Positive	s1cccc1CCN1CCC(N(C(=O)CC)c2ccccc2)(CC1)COC
824	65948	1	0	0.94	known Positive	O(C(=O)C=1C(C(C(OC(C)C)=O)=C(NC=1N)C)c1cc([N+](=O)[O-])ccc1)C1CN(C1)C(c1ccccc1)c1ccccc1
825	71158	1	0	0.55	known Positive	S(O)(=O)(=O)CCCNC(=O)C
826	74989	1	0	0.88	known Positive	Clc1ccc(cc1)C1CCC(CC1)C1C(=O)C(=O)c2c(cccc2)C1=O
827	119584	1	0	1.02	known Positive	Clc1cc(C(=O)NCC2OCCN(C2)Cc2ccc(F)cc2)c(OCC)cc1N
828	354624	1	0	0.89	known Positive	Clc1nc(N)c2ncn(c2n1)C1OC(CO)C(O)C1F
829	481696	1	0	1.03	known Positive	Fc1cc(F)ccc1N1C=C(C(O)=O)C(=O)c2cc(F)c(nc12)N1CC2C(C1)C2N
830	2723601	1	1	0.83	known Positive	S=C1NC(=Nc2nc[nH]c12)N
831	3086	1	0	1.02	known Positive	S(c1ccccc1C(=O)NC)c1cc2[nH]nc(c2cc

	685					1)\C=C\c1ncccc1
832	4659568	1	0	0.95	known Positive	Oc1c([N+](=O)[O-])cc(cc1O)\C=C(\C(=O)N(CC)CC)/C#N
833	4677798	1	0	0.74	known Positive	O1C(CC(O)CC1=O)CCC1C2C(=CC(CC2OC(=O)C(CC)(C)C)C)C=CC1C
834	21910730	1	0	0.97	known Positive	Fc1cc(F)ccc1C1(OCC(C1)COc1ccc(N2CCN(CC2)c2ccc(N3C=NN(C(C(O)C)CC)C3=O)cc2)cc1)Cn1ncnc1
835	7699	0	0	0.82	known Positive	O(C(=O)c1ccc(NCCCC)cc1)CCOCCOCCOCCOCCOCCOCCOCCOCCOC
836	4008	1	0	0.90	known Positive	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)CCN1CCN(CC1)C(c1ccccc1)c1ccccc1
837	2273	0	0	1.02	known Positive	Clc1cc(NCc2sccc2)c(cc1S(=O)(=O)N)-c1nn[nH]n1
838	2735	0	1	0.88	known Positive	OC1C\C(=C/C=C\2/C3CCC(C(CCCC(C)C)C)C3(CCC/2)C)\C(CC1)=C
839	2800	0	0	0.95	known Positive	Cl\C(=C(/c1ccc(OCCN(CC)CC)cc1)\c1ccccc1)\c1ccccc1
840	4735	0	1	0.98	known Positive	O(CCCCCOc1ccc(cc1)C(N)=N)c1ccc(cc1)C(N)=N
841	5482	0	0	1.06	known Positive	Clc1cc(Cl)ccc1C(OCc1ccsc1Cl)Cn1ccnc1
842	34633	0	0	1.03	known Positive	S1c2c(N(c3c1cccc3)C(=O)CCN1CCOCC1)cc(NC(OCC)=O)cc2
843	2618	1	0	1.08	known Positive	s1c(nnc1SCC=1CSC2N(C(=O)C2NC(=O)Cn2nnnc2)C=1C(O)=O)C
844	2659	1	0	1.02	known Positive	S1C2N(C(=O)C2NC(=O)\C(=N\OC)\c2occc2)C(C(O)=O)=C(C1)COC(=O)N
845	2662	1	0	0.98	known Positive	S(=O)(=O)(N)c1ccc(-n2nc(cc2-c2ccc(cc2)C)C(F)(F)F)cc1
846	2958	1	1	1.00	known Positive	O1C(C)C(O)C(N)CC1OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(C1=O)c(OC)ccc3)c2O)C(=O)C
847	3059	1	0	0.77	known Positive	Fc1cc(F)ccc1-c1cc(C(O)=O)c(O)cc1
848	4679	1	0	0.98	known Positive	S(=O)(Cc1nccc(OC)c1OC)c1[nH]c2cc(OC(F)F)ccc2n1
849	16362	1	1	1.01	known Positive	Fc1ccc(cc1)C(CCCN1CCC(N2c3c(NC2=O)cccc3)CC1)c1ccc(F)cc1
850	65999	1	0	0.97	known Positive	OC(=O)c1ccccc1-c1ccc(cc1)Cn1c2c(nc1CCC)c(cc(c2)-c1nc2c(n1C)cccc2)C
851	208908	1	0	1.02	known Positive	Clc1cc(Nc2ncnc3c2cc(cc3)-c2oc(cc2)CNCCS(=O)(=O)C)ccc1OCc1cc(F)ccc1
852	3579812	1	0	1.01	known Positive	OC(C(NC(=O)C(NC(=O)c1nc2c(cc1)cccc2)CC(=O)N)Cc1ccccc1)CN1CC2C(CC1C(=O)NC(C)(C)C)CCCC2
853	4460995	1	1	0.69	known Positive	O=C1CCC2C3C(CCC12C)C1(C(=CC(=O)C=C1)C(C3)=C)C

854	1870 7085	1	0	1.04	known Positive	Fc1cc(F)ccc1N1C=C(C(O)=O)C(=O)c2cc(F)c(nc12)N1CC2C(C1)C2NC(=O)C(NC(=O)C(N)C)C
855	774	1	0	0.54	known Negative	[nH]1cncc1CCN
856	2265	1	0	0.93	known Positive	S(c1n(cnc1[N+](=O)[O-])C)c1ncnc2[nH]cnc12
857	4030	1	0	0.88	known Positive	O(C(=O)Nc1[nH]c2cc(ccc2n1)C(=O)c1ccccc1)C
858	6674 90	1	1	0.77	known Positive	S=C1NC=Nc2[nH]cnc12
859	4509	1	0	0.88	known Positive	o1c(ccc1[N+](=O)[O-])\C=N\N1CC(=O)NC1=O
860	4594	1	0	0.98	known Positive	S(=O)(Cc1ncc(C)c(OC)c1C)c1[nH]c2cc(OC)ccc2n1
861	2952	1	0	0.97	Untested	o1c(ccc1\C=N\N1CC(=O)NC1=O)-c1ccc([N+](=O)[O-])cc1
862	784	0	0	0.28	Untested	OO
863	866	0	0	0.53	Untested	OC(=O)C(N)CCCCN
864	1125	0	0	0.78	Untested	O=C1NC(=NC=2NCC(NC1=2)C(O)C(O)C)N
865	1614	0	0	0.46	Untested	O1c2cc(ccc2OC1)CC(N)C
866	1783	0	0	1.13	Untested	Clc1cc(Sc2n(Cc3ccncc3)c(nc2C(C)C)COC(=O)N)cc(Cl)c1
867	2099	0	0	0.89	Untested	O=C1N(CCc2n(c3c(c12)cccc3)C)Cc1[nH]cnc1C
868	2133	0	0	0.85	Untested	FC12C(C3CC4OC5(OC4(C(=O)COC(=O)C)C3(CC1O)C)CCCC5)CCC1=CC(=O)C=CC12C
869	2175	0	0	1.02	Untested	S1C2N(C(C(OCOC(=O)C3N4C(S(=O)(=O)C3(C)C)CC4=O)=O)C1(C)C)C(=O)C2NC(=O)C(N)c1ccccc1
870	2239	0	0	1.03	Untested	s1c(ccc1C(=O)N)-c1nc(sc1)SCC(O)CNC(C)(C)C
871	2264	0	0	0.99	Untested	Clc1cc(c2OCC(=O)N(c2c1)C)C(=O)NC1C2CCN(C1)CC2
872	2296	0	0	0.91	Untested	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C)c1cc([N+](=O)[O-])ccc1)C1CCN(C1)Cc1ccccc1
873	2311	0	0	0.95	Untested	O=C1N(c2c(CCC1NC(CCc1ccccc1)C(OC)=O)cccc2)CC(O)=O
874	2340	0	0	0.71	Untested	O(C(=O)c1ccccc1)C1CC2N(C(CC2)C1C(O)=O)C
875	2341	0	0	0.66	Untested	N(Cc1ccccc1)(C(Cc1ccccc1)C)C
876	2344	0	0	0.77	Untested	O(C(c1ccccc1)c1ccccc1)C1CC2N(C(C1)CC2)C
877	2350	0	0	0.91	Untested	Clc1ccc(cc1)C(OC1CCN(CC1)CCCC(O)=O)c1ncccc1
878	2370	0	0	0.65	Untested	O(C(C[N+](C)(C)C)C)C(=O)N
879	2431	0	0	0.66	Untested	Brc1ccccc1C[N+](CC)(C)C
880	2479	0	0	0.57	Untested	O=C1NC(=O)NC(=O)C1(C(CC)C)CC

881	2993	0	0	0.99	Untested	N=C1C=C(N(c2c1cccc2)CCCCCCCCCCN1c2c(cccc2)C(=N)C=C1C)C
882	3060	0	0	0.89	Untested	FC12C(C3CCC(OC(=O)CCC)(C(=O)COC(=O)C)C3(CC1O)C)CC(F)C1=CC(=O)C=CC12C
883	3116	0	0	0.88	Untested	O(C(=O)N(CCCCCCN(C(Oc1ccc[n+](c1)C)=O)C)C)c1ccc[n+](c1)C
884	3148	0	0	0.88	Untested	O(C(=O)c1c2c([nH]c1)cccc2)C1CC2N3CC(=O)C(C2)CC3C1
885	3155	0	0	0.87	Untested	S1Cc2c(cccc2)\C(\c2c1cccc2)=C/CCN(C)C
886	3306	0	0	0.48	Untested	Oc1cc(ccc1)C(O)CNCC
887	3364	0	0	1.03	Untested	Clc1cccc(F)c1-c1noc(C)c1C(=O)NC1C2SC(C)(C)C(N2C1=O)C(O)=O
888	3648	0	0	0.80	Untested	O1C2C34C(C(N(CC3)C)Cc3c4c1c(O)cc3)CCC2=O
889	3723	0	0	0.96	Untested	Ic1c(CNC(=O)C)c(I)c(NC(=O)C)c(I)c1C(O)=O
890	3724	0	0	1.15	Untested	Ic1c(C(=O)NCC(O)CO)c(I)c(N(C(=O)C)CC(O)CN(C(=O)C)c2c(I)c(C(=O)NCC(O)CO)c(I)c(C(=O)NCC(O)CO)c2I)c(I)c1C(=O)NCC(O)CO
891	3731	0	0	1.00	Untested	Ic1c(C(=O)NCC(O)CO)c(I)c(N(C(=O)CO)C)c(I)c1C(=O)NCC(O)CO
892	3762	0	1	0.68	Untested	Oc1cc(ccc1O)C(O)C(NC(C)C)CC
893	3826	0	0	0.74	Untested	OC(=O)C1CCn2c1ccc2C(=O)c1ccccc1
894	3918	0	0	0.66	Untested	Oc1cc2C34C(C(N(CC3)C)Cc2cc1)CCCC4
895	4041	0	0	0.93	Untested	Clc1cc2c(N(CCN=C2c2ccccc2)C)cc1
896	4494	0	0	1.02	Untested	O(C(=O)C=1C(C(C(OC)=O)=C(NC=1C)C#N)c1cc([N+](=O)[O-])ccc1)C(C)C
897	4635	0	0	0.84	Untested	O1C2C34CCN(C(Cc5c3c1c(OC)cc5)C4(O)CCC2=O)C
898	4677	0	0	0.95	Untested	S(SCCNC(=O)CCNC(=O)C(O)C(CO)(C)C)CCNC(=O)CCNC(=O)C(O)C(CO)(C)C
899	4689	0	0	0.98	Untested	O1C(CN)C(O)C(O)C(N)C1OC1C(O)C(OC1CO)OC1C(OC2OC(CO)C(O)C(O)C2N)C(N)CC(N)C1O
900	4724	0	0	0.73	Untested	O(CC(O)CNC(C)(C)C)c1ccccc1C1CCCC1
901	4741	0	0	1.36	Untested	s1cc(nc1-c1nc(sc1)CCNC(=O)C(NC(=O)C(C(O)C(NC(=O)C(NC(=O)c1nc(nc(N)c1C)C(NCC(N)C(=O)N)CC(=O)N)C(OC1OC(CO)C(O)C(O)C1OC1OC(CO)C(O)C(OC(=O)N)C1O)c1[nH]cnc
902	4747	0	0	1.08	Untested	S1c2c(N(c3c1cccc3)CCCN1CCC(O)CC1)cc(cc2)C#N
903	4771	0	0	0.39	Untested	NC(Cc1ccccc1)(C)C
904	4786	0	0	0.42	Untested	OC(C(N)C)c1ccccc1

905	4845	0	0	0.66	Untested	Oc1ccc(nc1CO)C(O)CNC(C)(C)C
906	4940	0	0	0.89	Untested	S1c2c(N(c3c1cccc3)CC(N(C)C)C)cc(cc2)C(=O)CC
907	5241	0	0	0.96	Untested	Ic1c(CCC(O)=O)c(I)cc(I)c1\N=C\N(C)C
908	5440	0	0	0.97	Untested	S1c2c(N(c3c1cccc3)CCCN1CCN(CC1)C)cc(SCC)cc2
909	5454	0	0	0.98	Untested	S1c2c(cc(S(=O)(=O)N(C)C)cc2)\C(\c2c1cccc2)=C\CCN1CCN(CC1)C
910	5471	0	0	0.87	Untested	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(C(O)=O)c1ccsc1
911	5534	0	0	0.65	Untested	N1(CCCC1)CC#CCN1CCCC1
912	6471	0	0	0.66	Untested	O(C(=O)CC)C1(CCN(CC1C)C)c1ccccc1
913	7558	0	0	0.37	Untested	N(C(CC1CCCCC1)C)C
914	8612	0	0	0.82	Untested	Clc1cc(N)ccc1C(OCCN(CC)CC)=O
915	9296	0	0	0.64	Untested	S(OCC(O)C(O)COS(=O)(=O)C)(=O)(=O)C
916	9331	0	0	1.02	Untested	O=C(N)C1(N2CCCCC2)CCN(CC1)CCC(C#N)(c1ccccc1)c1ccccc1
917	9991	0	0	0.81	Untested	N(/NC(N)=N)=C(\C=N\NC(N)=N)/C
918	10413	0	1	0.36	Untested	OC(=O)CCCO
919	10517	0	0	0.80	Untested	O(C(C(CC(N(C)C)C)(c1ccccc1)c1ccccc1)CC)C(=O)C
920	12454	0	0	1.04	Untested	Clc1cc\2c(Sc3c(cccc3)/C/2=C\CCN2CCN(CC2)CCO)cc1
921	12460	0	0	0.55	Untested	O1CCN(C)C(C)C1c1ccccc1
922	12620	0	0	0.68	Untested	OCCCCCCCCCCCCCCCCCCCCCC
923	14670	0	0	0.87	Untested	S1c2c(N(c3ncccc13)CCCN(C)C)cccc2
924	18804	0	0	0.76	Untested	O(C(Cc1ccccc1)(C(CNC)C)c1ccccc1)C(=O)CC
925	22297	0	0	0.37	Untested	N(C(CC\C=C(\C)/C)C)C
926	26394	0	0	0.89	Untested	O(CC)c1cc(ccc1OCC)CC1=NCCc2c1cc(OCC)c(OCC)c2
927	26937	0	0	0.76	Untested	Clc1ccc(cc1)CC(O)(C(CN(C)C)C)C
928	28061	0	0	0.80	Untested	ClCC(O)Cn1c(ncc1[N+](=O)[O-])C
929	31477	0	0	0.77	Untested	O(C(=O)C)c1c(C)c(C)c(OCC(O)CNC(C)C)cc1C
930	32281	0	0	0.94	Untested	O=C1NC(CC1)C(=O)NC(Cc1[nH]cnc1)C(=O)N1CCCC1C(=O)N
931	36523	0	0	1.14	Untested	Oc1ccc(cc1)CC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc1[nH]cnc1)Cc1c2c([nH]c1)cccc2)CO)C(=O)NCC(=O)NC(CC(C)C)C(=O)NC(CCCNC(N)=N)C(=O)N1CCCC1C(=O)N

932	37497	0	0	0.75	Untested	O=C(Nc1c(cccc1C)C)C(N(CCC)CC)CC
933	43008	0	0	0.97	Untested	O(C)c1cc2nn[nH]c2cc1C(=O)NCC1N(CCC1)CC=C
934	47979	0	0	0.93	Untested	S(=O)(=O)(N)c1cc(C(=O)NCC2N(CCC2)CC=C)c(OC)c(OC)c1
935	54910	0	0	0.89	Untested	Clc1ccc(cc1)C1OCc2c1cnc(C)c2O
936	55645	0	0	0.93	Untested	S(=O)(=O)(NC1CC2C(N(C1)CCC)Cc1c(C2)c(O)ccc1)N(CC)CC
937	65015	0	0	0.89	Untested	N1CCCNCCNCCCN(CC1)Cc1ccc(cc1)CN1CCCNCCNCCCNCC1
938	65985	0	0	0.99	Untested	Ic1c(C(=O)N(CC(O)CO)C)c(I)c(NC(=O)C(CO)CO)c(I)c1C(=O)N(CC(O)CO)C
939	68602	0	0	0.67	Untested	[nH]1cncc1C(C)c1cccc(C)c1C
940	104867	0	0	0.95	Untested	S(CC(N)C(O)=O)CCCC\C=C(\NC(=O)C1CC1(C)C)/C(O)=O
941	157922	0	0	0.50	Untested	O(C(=O)CCC(=O)CN)C
942	216210	0	0	1.05	Untested	OC(=O)CCN(C(=O)c1cc2nc(n(c2cc1)C)CNc1ccc(cc1)C(N)=N)c1ncccc1
943	342467	0	0	1.03	Untested	Ic1c(C(=O)NC2C(O)C(O)C(OC2O)CO)c(I)c(NC(=O)C)c(I)c1N(C(=O)C)C
944	362949	0	0	0.63	Untested	O1C(CO)C(O)C(O)C1n1nc(nc1)C(N)=N
945	3032285	0	0	0.76	Untested	S=C1NC(=O)C(C(CCC)C)(CC=C)C(=O)N1
946	3033226	0	0	0.83	Untested	S(=O)(C(c1ccccc1)c1ccccc1)CC(=O)NO
947	4460698	0	0	0.86	Untested	O1CCN(CC1)C1CC2(C(CC1O)CCC1C3CC([N+]4(CCCC4)CC=C)C(OC(=O)C)C3(CCC12)C)C
948	4630917	0	0	0.72	Untested	Oc1cc2c(CC3CCCCCC2(C)C3N)cc1
949	5251896	0	0	0.81	Untested	OC12CC3(NCC(=O)N4CCCC4C#N)CC(C1)CC(C3)C2
950	5284561	0	0	0.88	Untested	O=C(N)c1cc2c3CC(NC)CCc3[nH]c2cc1
951	5310993	0	0	0.48	Untested	OC(=O)c1ncc([n+]([O-])c1)C
952	9818231	0	0	0.92	Untested	O=C(N1CC(N(C)c2ncnc3[nH]ccc23)C(CC1)C)CC#N
953	9881626	0	0	0.69	Untested	O1CCc2c3c(CCC3CCNC(=O)CC)ccc12
954	9885319	0	0	0.86	Untested	O(C(=O)N1CCc2c(cccc2)C1c1ccccc1)C1C2CCN(C1)CC2
955	112462 84	0	0	0.73	Untested	O(CC(O)=O)c1c2CC3C(Cc2ccc1)C(CCC(O)CCCCC)C(O)C3
956	147244 82	0	0	1.28	Untested	s1cccc1CC(NC(=O)CNC(=O)C1N(CC(O)C1)C(=O)C1N(CCC1)C(=O)C(NC(=O)C(N)CCCNC(N)=N)CCCNC(N)=N)C(=O)NC

						(C(=O)N1Cc2c(CC1C(=O)N1C3C(CC1C(=O)NC(CCCNC(N)=N)C(O)=O
957	24776445	0	0	1.03	Untested	Clc1ccc(NC(=O)c2ccc(S(=O)(=O)C)cc2Cl)cc1-c1ncccc1
958	44208978	0	0	1.58	Untested	Clc1ccc(cc1)CC(NC(=O)C(NC(=O)C)Cc1cc2c(cc1)cccc2)C(=O)NC(Cc1cccnc1)C(=O)NC(C(=O)NC(Cc1ccc(NC(=O)C2NC(=O)NC(=O)C2)cc1)C(=O)NC(Cc1ccc(NC(=O)N)cc1)C(=O)N
959	54724371	0	0	1.03	Untested	OC12C(CC3C(C1=O)C(=O)c1c(cccc1O)C3(O)C)C(N(C)C)C(=O)C(C(=O)NCNCCCCC(N)C(O)=O)C2=O
960	74387779	0	0	0.94	Untested	Clc1cc\2c(Sc3c(cccc3)/C/2=C\CCC2CCC(CC2)CCO)cc1
961	389	1	0	0.59	Untested	OC(=O)C(N)CCCN
962	1134	1	0	0.59	Untested	O1C(CO)C(O)CC1N1C=C(C)C(=O)NC1=O
963	1206	1	0	0.37	Untested	N(C(Cc1ccccc1)C)C
964	2132	1	0	0.93	Untested	Brc1cc(Br)cc(CNC2CCC(O)CC2)c1N
965	2162	1	0	0.96	Untested	Clc1ccccc1C1C(C(OCC)=O)=C(NC(C)=C1C(OC)=O)COCCN
966	2226	1	0	0.99	Untested	O1C(CO)C(O)C(N)C(O)C1OC1C(O)C(OC2OC(CCC2N)CN)C(N)CC1NC(=O)C(O)CCN
967	2451	1	0	1.09	Untested	Brc1sc2-n3c(nnc3C)CN=C(c2c1)c1ccccc1Cl
968	2481	1	0	0.61	Untested	O=C1NC(=O)NC(=O)C1(CC(C)C)CC=C
969	2786	1	0	0.95	Untested	ClC(C(NC(=O)C1N(CC(C1)CCC)C)C1OC(SC)C(O)C(O)C1O)C
970	2789	1	0	0.94	Untested	Clc1cc2N(C(=O)CC(=O)N(c2cc1)C)c1ccccc1
971	2811	1	0	1.00	Untested	Clc1ccccc1C1=NCC(=O)N(c2sc(cc12)CC)C
972	2826	1	0	0.70	Untested	O(C(=O)c1ccccc1)C1CC2N(C(CC2)C1C(OC)=O)C
973	3007	1	0	0.36	Untested	NC(Cc1ccccc1)C
974	3380	1	0	1.02	Untested	Fc1ccccc1C1=NCC(=O)N(c2c1cc([N+](=O)[O-])cc2)C
975	3911	1	0	1.15	Untested	Oc1ccc(cc1)CC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc1nc[nH]c1)Cc1c2c([nH]c1)cccc2)CO)C(=O)NC(CC(C)C)C(=O)NC(CC(C)C)C(=O)NC(CCCNC(N)=N)C(=O)N1CCC
976	3916	1	0	0.94	Untested	S1c2c(N(c3c1cccc3)CC(CN(C)C)C)cc(OC)cc2
977	4020	1	0	0.88	Untested	Clc1ccc(cc1)C1(O)N2C(=NCC2)c2c1cccc2
978	4205	1	0	0.81	Untested	n1c2N3C(c4c(Cc2ccc1)cccc4)CN(CC3)C
979	4547	1	0	0.95	Untested	O(CC)c1cc(ccc1C(O)=O)CC(=O)NC(CC(C)C)c1ccccc1N1CCCCC1

980	4603	1	0	0.82	Untested	O(C(CC)CC)C1C=C(CC(N)C1NC(=O)C)C(OCC)=O
981	4639	1	0	0.83	Untested	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2=O)C
982	4675	1	0	0.86	Untested	O(C(=O)C)C1C2(C(CC1[N+]1(CCCCC1)C)C1C(CC2)C2(CC([N+]3(CCCCC3)C)C(OC(=O)C)CC2CC1)C)C
983	5215	1	0	0.92	Untested	S(=O)(=O)(Nc1ncccn1)c1ccc(N)cc1
984	5372	1	0	0.96	Untested	O1C(/C(=C/C2CC(OC)C(O)CC2)/C)C(C)C(O)CC(=O)C(\C=C(\CC(CC(OC)C2OC(O)(C(=O)C(=O)N3C(CCCC3)C1=O)C(CC2OC)C)C)/C)CC=C
985	5526	1	0	0.41	Untested	OC(=O)C1CCC(CC1)CN
986	5556	1	0	1.05	Untested	Clc1ccccc1C1=NCc2n(-c3c1cc(Cl)cc3)c(nn2)C
987	8271	1	0	0.70	Untested	O=C1N(C)C(=O)NC(=O)C1(CC)c1ccccc1
988	15330	1	0	0.77	Untested	O(C(Cc1ccccc1)(C(CN(C)C)C)c1ccccc1)C(=O)CC
989	25215	1	0	0.91	Untested	Clc1cc2c(N(C)C(=O)CN=C2C=2CCCCC=2)cc1
990	26388	1	0	1.03	Untested	S1c2c(N(c3c1cccc3)CCCN1CCC(CC1)C(=O)N)cc(S(=O)(=O)C)cc2
991	40704	1	0	0.93	Untested	Brc1cc(OC)c(OC)cc1C[N+]1(CCOCC1)CCOCCC1C2CC(CC1)C2(C)C
992	41317	1	0	0.87	Untested	O(C)c1cc(C)c(\C=C\C(=C\C=C\C(=C\C(O)=O)\C)\C)c(C)c1C
993	47725	1	0	1.18	Untested	Oc1ccc(cc1)CC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc1nc[nH]c1)Cc1c2c([nH]c1)cccc2)CO)C(=O)NC(C(=O)NC(CC(C)C)C(=O)NC(CCCNC(N)=N)C(=O)N1CCCC1C(=O)
994	60726	1	0	0.89	Untested	Brc1ccc(cc1)C(=O)c1cccc(CC(O)=O)c1N
995	62865	1	0	1.01	Untested	S1c2c(N(c3c1cccc3)CC(CN(C)C)C)cc(cc2)C#N
996	68712	1	0	0.54	Untested	O1CCN=C1NC(C1CC1)C1CC1
997	68870	1	0	1.08	Untested	Clc1cc2S(=O)(=O)N(c3c(cccc3)C(NCCCCCCC(O)=O)c2cc1)C
998	71113	1	0	0.38	Untested	O=C(N)C(CCC)CCC
999	91276	1	0	0.79	Untested	O(CCCNC(=O)CO)c1cc(ccc1)CN1CCCCC1
1000	110633	1	0	0.84	Untested	O=C1NC(=Nc2n(cnc12)COC(COC(=O)C(N)C(C)C)CO)N
1001	115355	1	0	0.79	Untested	FC(F)(F)c1cc([N+](=O)[O-])c(cc1)C(=O)C1C(=O)CCCC1=O
1002	122252	1	0	0.88	Untested	s1cccc1C(Oc1c2c(ccc1)cccc2)CCNC
1003	383414	1	0	1.12	Untested	S1SCC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(N)Cc2ccccc2)

						C1)Cc1ccccc1)Cc1c2c([nH]c1)cccc2)CCCCN)C(O)C)C(=O)NC(C(O)C)CO
1004	4715169	1	0	0.47	Untested	OC(=O)CC(CC(C)C)CN
1005	5281051	1	0	1.04	Untested	Oc1c2c3c(c4c5c(C(=O)c6c7c(c8c(c(C2=O)c(O)cc8O)c3c57)c(O)cc6O)c(O)cc4C)c(c1)C
1006	6331630	1	0	0.91	Untested	O1C2C(CC(=O)CC3OC4C5OC6(OC7C(OC4CC3)C5OC7C6)CCC3OC(CCC4OC(C2)C(=C)C(C4)C)C(C3)=C)C(OC)C1CC(O)CN
1007	9837769	1	0	0.79	Untested	S1CC(OC1CO)N1C=C(F)C(=NC1=O)N
1008	9891967	1	0	0.98	Untested	FC(F)(F)COc1ccccc1OCCNC(Cc1cc(c2N(CCc2c1)CCCO)C(=O)N)C
1009	9912811	1	0	0.95	Untested	S(C=1C(C2N(C(=O)C2C(O)C)C=1C(O)=O)C)C1CC(NC1)C(=O)Nc1cc(ccc1)C(O)=O
1010	21923359	1	0	1.01	Untested	Fc1ccc(cc1)C1N(CCOC1OC(C)c1cc(cc(c1)C(F)(F)F)C(F)(F)F)CC=1NC(=O)NN=1
1011	22570509	1	0	0.95	Untested	FC1(F)CCC(CC1)C(=O)NC(CCN1C2CC(n3c(nnc3C)C(C)C)CC1CC2)c1ccccc1
1012	25077179	1	0	1.17	Untested	Oc1ccc(cc1)CC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc1[nH]cnc1)Cc1c2c([nH]c1)cccc2)CO)C(=O)NC(Cc1c2c([nH]c1)cccc2)C(=O)NC(CC(C)C)C(=O)NC(CCCNC(N)
1013	54680692	1	0	0.76	Untested	O1c2c(cccc2)C(O)=C(C(CC)c2ccccc2)C1=O
1014	119171	0	0	0.95	Untested	P1(OCC(CO1)(C)C)(=O)c1c(-c2cc([N+](=O)[O-])ccc2)c(C(OCCN(Cc2ccccc2)c2ccccc2)=O)c(nc1C)C
1015	2747	1	0	0.79	Untested	N1CCN=C1C1CC1(c1ccccc1)c1ccccc1
1016	9034	1	0	0.81	Untested	O=C1N(C)C(=O)NC(=O)C1(C(C#CCC)C)CC=C
1017	43085	1	0	1.03	Untested	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(NC(=O)N1CCN(S(=O)(=O)C)C1=O)c1ccccc1
1018	2118	0	0	1.02	Untested	Clc1cc2c(-n3c(nnc3C)CN=C2c2ccccc2)cc1
1019	2216	0	0	0.91	Untested	Clc1cc(N)cc(Cl)c1NC=1NCCN=1
1020	2461	0	0	0.72	Untested	OC=1N(C2CCCCC2)C(=O)NC(=O)C=1CCCC
1021	2472	0	0	0.97	Untested	O(C)c1cc2c(nc(nc2N)N2CCCN(CC2)C(=O)CCC)cc1OC
1022	3742	0	0	1.18	Untested	Ic1c(C(=O)NCC(=O)Nc2c(I)c(C(=O)NCCO)c(I)c(C(O)=O)c2I)c(I)c(N(C(=O)C)C)c(I)c1C(=O)NC

1023	4034	0	0	0.94	Untested	Clc1ccc(cc1)C(N1CCN(CC1)Cc1cc(ccc1)C)c1ccccc1
1024	4166	0	0	0.88	Untested	O1C(C)C(OC2OC(C)C(OC)C(O)C2)C(O)CC1OC1C(OC(OC2CC3CCC4C(CC(O)C5(C)C(CCC45O)C4=CC(OC4)=O)C3(CC2)C)CC1O)C
1025	4419	0	0	0.81	Untested	O1C2C34CCN(C(Cc5c3c1c(O)cc5)C4(O)CCC2O)CC1CCC1
1026	4634	0	0	0.84	Untested	O(C(=O)C(O)(C1CCCCC1)c1ccccc1)CC#CCN(CC)CC
1027	5069	0	0	1.05	Untested	Clc1ccccc1C(=O)c1cc(Cl)ccc1-n1nc(nc1CNC(=O)CN)C(=O)N(C)C
1028	30768	0	0	0.99	Untested	Clc1cc2c(N=C(OC2(C)c2ccccc2)NCC)cc1
1029	33746	0	0	1.00	Untested	Clc1cc2c(N(C)C(=O)CN3C2(OC(=CC3=O)C)c2ccccc2)cc1
1030	38521	0	0	0.69	Untested	O1C(COC12CCCCC2)CNC(N)=N
1031	6426711	0	0	0.92	Untested	O(C)c1cc2c(CC2CN(CCCN2CCc3cc(OC)c(OC)cc3CC2=O)C)cc1OC
1032	143	1	0	1.03	Untested	O=C1NC(=NC=2NCC(N(C1=2)C=O)CNc1ccc(cc1)C(=O)NC(CCC(O)=O)C(O)=O)N
1033	2487	1	0	0.76	Untested	OC12C3(CCCC1)CCN(C2Cc1c3cc(O)cc1)CC1CCC1
1034	3443	1	0	0.82	Untested	O(C(=O)C)C/1CC2(C(CC(O)C3C2(CCC2C(C)C(O)CCC23C)C)\C\1=C(\CC\C=C(\C)/C)/C(O)=O)C
1035	3510	1	0	0.88	Untested	O=C(NC1CC2N(C(C1)CCC2)C)c1nn(c2c1cccc2)C
1036	5650	1	0	0.95	Untested	OC(=O)C(N(Cc1ccc(cc1)-c1ccccc1-c1nn[nH]n1)C(=O)CCCC)C(C)C
1037	158781	1	0	0.96	Untested	OC(C)(C)c1nc(n(Cc2ccc(cc2)-c2ccccc2-c2nn[nH]n2)c1C(O)=O)CCC
1038	193962	1	0	1.19	Untested	Brc1c(nc(nc1Oc1c(cc(cc1C)C#N)C)Nc1ccc(cc1)C#N)N
1039	3465281	1	0	0.84	Untested	OC1C(C(NC(=O)C)C(CC)CC)C(NC(N)=N)CC1C(O)=O
1040	11622909	1	0	0.77	Untested	O=C(N(CC)CC)C1(CC1CN)c1ccccc1
1041	21889836	1	0	0.93	Untested	O=C(NC(C(=O)C(=O)NC1CC1)CCC)C1N(CC2C1CCC2)C(=O)C(NC(=O)C(NC(=O)c1nccnc1)C1CCCCC1)C(C)(C)C
1042	235943 67	1	0	1.04	Untested	O=C(NC(CC1CCC1)C(=O)C(=O)N)C1N(CC2C1C2(C)C)C(=O)C(NC(=O)NC(C)(C)C)C(C)(C)C
1043	734589 70	1	0	1.03	Untested	s1cc(nc1C)\C=C(/C)\C1NC(=O)CC(O)C(C)(C)C(=O)C(C)C(O)C(CCCC2(OC2C1)C)C
1044	3746	0	0	0.78	Untested	O(C(=O)C(CO)c1ccccc1)C1CC2[N+](C(C1)CC2)(C(C)C)C

1045	5357	0	0	0.78	Untested	S(=O)(=O)(CC)c1cc(C(=O)NCC2N(CCC2)CC)c(OC)cc1
1046	60831	0	0	0.95	Untested	s1cccc1C(O)(C(OC1CC2[N+](C(C1)C1OC12)(C)C)=O)c1sccc1
1047	60852	0	0	0.87	Untested	P(O)(O)(=O)C(P(O)(O)=O)(O)CCN(CCCCC)C
1048	3052774	0	0	0.86	Untested	s1c(C(O)=O)c(CC(O)=O)c(C#N)c1N(CC(O)=O)CC(O)=O
1049	9890723	0	0	1.00	Untested	S1CCSC12CC(N(C2)C(=O)C(NC(CCc1ccccc1)C(OCC)=O)C)C(O)=O
1050	11451527	0	0	1.03	Untested	Fc1c2cc([nH]c2ccc1OC1=NC=Nn2c1c(C)c(OCC(O)C)c2)C
1051	17854871	0	0	1.07	Untested	S(O)(=O)(=O)NC1C(OS(O)(=O)=O)C(OC2OC(C(O)=O)C(OC3OC(COS(O)(=O)=O)C(O)C(O)C3NS(O)(=O)=O)C(O)C2O)C(OC1OC1C(O)C(OS(O)(=O)=O)C(OC1C(O)=O)OC1C(O)C(NS(O)(=O
1052	21865524	0	0	1.00	Untested	S1C2N(C(=O)C(NC(=O)C(S)Cc3ccccc3)CC1)C(CCC2)C(O)=O
1053	2684	1	0	0.66	Untested	S1CC2(OC1C)C1CCN(C2)CC1
1054	3821	1	0	0.77	Untested	Clc1ccccc1C1(NC)CCCCC1=O
1055	4510	1	0	0.69	Untested	O([N+](=O)[O-])C(CO[N+](=O)[O-])CO[N+](=O)[O-]
1056	4727	1	0	0.62	Untested	SC(C(N)C(O)=O)(C)C
1057	4739	1	0	0.73	Untested	O1C(CO)C(O)CC1n1c2NC=NCC(O)c2nc1
1058	4857	1	0	0.89	Untested	S1C2N(C(C(OCOC(=O)C(C)(C)C)=O)C1(C)C)C(=O)C2\N=C\N1CCCCCC1
1059	42113	1	0	0.57	Untested	FC(OC(F)F)C(F)(F)F
1060	54684452	1	0	1.05	Untested	S(=O)(=O)(Nc1cc(ccc1)C(CC)C1C(=O)CC(OC1=O)(CCc1ccccc1)CCC)c1ncc(cc1)C(F)(F)F
1061	3699	0	0	0.88	Untested	P(O)(O)(=O)\C(\P(O)(O)=O)=N\C1CCCCCC1
1062	4100	0	0	0.91	Untested	S1\C(=N\C(=O)C)\N(N=C1S(=O)(=O)N)C
1063	235776	0	0	0.90	Untested	[S+]12C(C3N(Cc4ccccc4)C(=O)N(C3C1)Cc1ccccc1)CCC2
1064	609377	0	0	0.96	Untested	Clc1ccc(SC(P(O)(O)=O)P(O)(O)=O)cc1
1065	2459	1	0	0.74	Untested	SC(C(=O)NC(C(O)=O)CS)(C)C
1066	2751	1	0	0.92	Untested	O=C1N2N(CCCC1NC(CCc1ccccc1)C(OC)=O)CCCC2C(O)=O
1067	3417	1	0	0.66	Untested	P(O)(O)(=O)C1OC1C
1068	5206	1	0	0.63	Untested	FC(F)(F)C(OCF)C(F)(F)F

1069	25419	1	0	1.03	Untested	ClC(Cl)(P(O)(O)=O)P(O)(O)=O
1070	176870	0	0	0.96	Untested	O(CCOC)c1cc2c(ncnc2Nc2cc(ccc2)C#C)cc1OCCOC
1071	9933475	0	0	0.97	Untested	Fc1c2cc([nH]c2ccc1Oc1ncnc2c1cc(OC)c(OCCCN1CCCC1)c2)C
1072	17754772	0	0	0.98	Untested	[nH]1c2ncnc(c2cc1)-c1cn(nc1)C(CC#N)C1CCCC1
1073	160	0	0	0.70	Untested	OC1CC(O)C(\C=C\C(O)CCCCC)C1C\C=C\CCCC(O)=O
1074	214	0	0	0.71	Untested	OC1CC(=O)C(CCCCCCC(O)=O)C1\C=C\C(O)CCCCC
1075	291	0	0	0.97	Untested	P(OCC1OC(N2C=CC(=NC2=O)N)C(O)C1O)(OP(OCC[N+](C)(C)C)(O)=O)(O)=O
1076	424	0	0	0.56	Untested	OC(=O)C(N)CC(O)=O
1077	1181	0	0	0.96	Untested	P(OCC1OC(N2C=CC(=O)NC2=O)C(O)C1O)(OP(OP(O)(O)=O)(O)=O)(O)=O
1078	2182	0	0	0.90	Untested	Clc1c2CN3CC(=O)NC3=Nc2ccc1Cl
1079	2245	0	0	1.09	Untested	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(NC(=O)C(N)CC(=O)NC)c1ccc(O)cc1
1080	2551	0	0	0.63	Untested	O(CC[N+](C)(C)C)C(=O)N
1081	2614	0	0	1.00	Untested	S1C2N(C(=O)C2NC(=O)C(O)c2ccccc2)C(C(O)=O)=C(C1)CSc1nnnn1C
1082	2654	0	0	0.96	Untested	s1cc(nc1N)/C(=C\CC(O)=O)/C(=O)NC1C2SCC=C(N2C1=O)C(O)=O
1083	2802	0	0	1.05	Untested	Clc1ccccc1C1=NCC(=O)Nc2c1cc([N+](=O)[O-])cc2
1084	2972	0	0	0.43	Untested	OC=1C(=O)C=CN(C)C=1C
1085	2978	0	0	0.79	Untested	O1c2c(C3C(CCC(=C3)C)C1(C)C)c(O)cc(c2)CCCCC
1086	3038	0	0	0.97	Untested	Clc1c(S(=O)(=O)N)cc(S(=O)(=O)N)cc1Cl
1087	3079	0	0	0.67	Untested	N1(C2C3C(CCCC3)(CC1)c1cc(ccc1C2)C)C
1088	3105	0	0	0.78	Untested	O(C(=O)C(C)(C)C)c1cc(ccc1OC(=O)C(C)(C)C)C(O)CNC
1089	3228	0	0	0.89	Untested	O1C(CO)C(O)C(O)C1N1C=CC(=NC1=O)NC(=O)CCCCCCCCCCCCCCCCCCCCC
1090	3235	0	0	0.94	Untested	S1/C(=C\C(=C\c2ccccc2)\C)/C(=O)N(CC(O)=O)C1=S
1091	3292	0	0	0.60	Untested	OC=1N(CC)C(=O)NC=1c1ccccc1
1092	3449	0	0	0.79	Untested	O1c2c3C4(C1CC(O)C=C4)CCN(Cc3ccc2OC)C
1093	3592	0	0	0.81	Untested	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OC(=O)C)cc3)C=CC2OC(=O)C

1094	3759	0	0	0.70	Untested	o1nc(cc1C)C(=O)NNCc1ccccc1
1095	3890	0	0	0.76	Untested	OC1CC(O)C(C\C=C\CCCC(OC(C)C)=O)C1CCC(O)CCc1ccccc1
1096	3956	0	0	1.00	Untested	ClC=1CCC2N(C(=O)C2NC(=O)C(N)c2ccccc2)C=1C(O)=O
1097	3958	0	0	0.97	Untested	Clc1ccccc1C1=NC(O)C(=O)Nc2c1cc(Cl)cc2
1098	4058	0	0	0.65	Untested	O(C(=O)C1(CCN(CC1)C)c1ccccc1)CC
1099	4264	0	0	0.81	Untested	O1C(CC2COC(C\C(=C\C(OCCCCCCCCC(O)=O)=O)\C)C(O)C2O)C1C(C(O)C)C
1100	4578	0	0	0.89	Untested	O(Cc1ccccc1)c1nc(nc2[nH]cnc12)N
1101	4676	0	0	0.84	Untested	S(C=1CC2N(C(=O)C2C(O)C)C=1C(O)=O)C1CCN(C1)C(=N)C
1102	4823	0	0	1.02	Untested	O(C)c1ccc(cc1)-c1[nH]c2cc(ccc2n1)C1=NNC(=O)CC1C
1103	4948	0	0	1.03	Untested	S1C2N(C(=O)C2NC(=O)C(N)c2ccc(O)cc2)C(C(O)=O)=C(C1)CSc1[nH]nnc1
1104	4976	0	0	0.75	Untested	N(CCCC1c2c(C=Cc3c1cccc3)cccc2)C
1105	5078	0	0	0.84	Untested	[nH]1cc(c2cc(ccc12)Cn1ncnc1)CCN(C)C
1106	5205	0	0	0.76	Untested	N1(CC2=C(CC1)c1c(Cc3c2cccc3)cccc1)C
1107	5407	0	0	0.67	Untested	O1C2(C(C3C(CC2)C2(C(=CC(=O)C=C2)CC3)C)CCC1=O)C
1108	5484	0	0	0.74	Untested	s1cccc1\C(=C\1/CCCN(C/1)C)\c1sccc1
1109	5583	0	0	1.07	Untested	O(C)c1c(OC)cc(NCc2ccc3nc(nc(N)c3c2C)N)cc1OC
1110	5636	0	0	0.72	Untested	OC1CC(O)C(C\C=C\CCCC(O)=O)C1CCC(=O)CCCCCCC
1111	5771	0	0	1.15	Untested	S1SCC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(N)C1)Cc1ccc(O)cc1)C(CC)C)CCC(=O)N)CC(=O)N)C(=O)N1CCCC1C(=O)NC(CC(C)C)C(=O)NCC(=O)N
1112	5772	0	0	1.17	Untested	S1SCC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(N)C1)Cc1ccc(O)cc1)Cc1ccccc1)CCC(=O)N)CC(=O)N)C(=O)N1CCCC1C(=O)NC(CCCCN)C(=O)NCC(=O)N
1113	6468	0	0	0.61	Untested	N1(CCCCC1)C1(CCCCC1)c1ccccc1
1114	7458	0	0	0.81	Untested	Ic1ccc(cc1)C(CCCCCCCCC(OCC)=O)C
1115	7697	0	0	0.54	Untested	Clc1ccc(OCC(O)CO)cc1
1116	10101	0	0	0.72	Untested	Oc1cc(ccc1)C1(CCN(CC1)C)C(=O)CC
1117	10548	0	0	0.78	Untested	S(P(OCC)(OCC)=O)CC[N+](C)(C)C

1118	16850	0	0	0.87	Untested	O1C2(c3c(cccc3)C1=O)c1c(Oc3c2ccc(O)c3)cc(O)cc1
1119	17925	0	0	0.97	Untested	Clc1ccccc1C1=NCC(=O)Nc2c1cc(Cl)cc2
1120	27648	0	0	1.01	Untested	Ic1c(N(C(=O)C)CC(C(O)=O)C)c(I)cc(I)c1N
1121	31640	0	0	1.00	Untested	Clc1cc2c(N(CC(F)(F)F)C(=O)CN=C2c2ccccc2)cc1
1122	32329	0	0	0.68	Untested	O(C(=O)C1(CCC=CC1N(C)C)c1ccccc1)CC
1123	32800	0	0	1.05	Untested	S(CCC(NC(=O)C(NC(=O)CNC(=O)C(NC(=O)C(NC(=O)C(N)CC(O)=O)Cc1ccc(OS(O)(=O)=O)cc1)CCSC)Cc1c2c([nH]c1)cccc2)C(=O)NC(CC(O)=O)C(=O)NC(Cc1ccccc1)C(=O)N)C
1124	33649	0	0	0.95	Untested	S(=O)(=O)(NC(=O)NC1C2CCC(C)(C1O)C2(C)C)c1ccc(cc1)C
1125	39860	0	0	0.86	Untested	O1c2c(C3C(CCC(=O)C3)C1(C)C)c(O)cc(c2)C(CCCCCC)(C)C
1126	47319	0	0	1.03	Untested	O(C)c1cc(ccc1OC)CC1[N+](CCc2c1cc(OC)c(OC)c2)(CCC(OCCCCCOC(=O)CC[N+]1(CCc2c(cc(OC)c(OC)c2)C1Cc1cc(OC)c(OC)cc1)C)=O)C
1127	52421	0	1	0.81	Untested	O(C(OCC)=O)C1(CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C)C(=O)COC(=O)CC
1128	55756	0	0	0.86	Untested	OC(=O)\C=C\c1nc(ccc1)/C(=C\CN1CCCC1)/c1ccc(cc1)C
1129	60169	0	0	1.04	Untested	O(C)c1c(OC)cc(cc1OC)CC1[N+](CCc2c1c(OC)c(OC)c(OC)c2)(CCCOC(=O)CCC(OCCC[N+]1(CCc2c(c(OC)c(OC)c(OC)c2)C1Cc1cc(OC)c(OC)c(OC)c1)C)=O)C
1130	60753	0	0	0.90	Untested	S(=O)(=O)(Nc1ccc(cc1)C(O)CCCN(CCCCCCC)CC)C
1131	62179	0	0	1.12	Untested	Clc1ccccc1C1=NCC=2N(c3c1cc([N+](=O)[O-])cc3)C(=O)/C(/N=2)=C\N1CCN(CC1)C
1132	68778	0	0	0.76	Untested	N1(CCC(CC1)(c1ccccc1)c1ccccc1)C(C)(C)C
1133	104803	0	0	1.07	Untested	O(C)c1c(OC)cc(cc1OC)CC1[N+](CCc2c1cc(OC)c(OC)c2)(CCCOC(=O)CC\C=C\CCC(OCCC[N+]1(CCc2c(cc(OC)c(OC)c2)C1Cc1cc(OC)c(OC)c(OC)c1)C)=O)C
1134	119828	0	0	1.11	Untested	S(=O)(=O)(NC(=O)CC)c1ccc(cc1)-c1c(noc1C)-c1ccccc1
1135	121892	0	0	1.01	Untested	Fc1ccc(cc1)CNc1cc(N)c(NC(OCC)=O)cc1
1136	123610	0	0	1.07	Untested	S1SCCC(=O)NC(CCCCNC(N)=N)C(=O)NCC(=O)NC(CC(O)=O)C(=O)NC(Cc2c3c([nH]c2)cccc3)C(=O)N2C(CCC2)C(=O)NC(C1)C(=O)N

1137	123908	0	0	0.58	Untested	OC(C(O)C(O)=O)(CC(O)=O)C(O)=O
1138	208951	0	0	0.97	Untested	O1c2c(NC1=O)cccc2N1CCN(CC1)Cc1cc(ccc1)-c1ccccc1
1139	321132	0	0	0.95	Untested	ClCCN(N=O)C(=O)NCC1OC(OC)C(O)C(O)C1O
1140	394493	0	0	0.93	Untested	O=C1NC(=Nc2[nH]cc(c12)CCc1ccc(cc1)C(=O)NC(CCC(O)=O)C(O)=O)N
1141	416706	0	0	0.95	Untested	OC1(C(=O)COC(=O)C)C2(CC(O)C3C(C2CC1C)C=C(C1=Cc2n(ncc2CC13C)-c1ccccc1)C)C
1142	521308	0	0	0.80	Untested	O1C2C34C(C(N(CC3)C)Cc3c4c1c(OCCN1CCOCC1)cc3)C=CC2O
1143	541123	0	0	0.76	Untested	FC12C(C3CCC(O)(C(=O)CO)C3(CC1O)C)CCC1=CC(=O)CCC12C
1144	3033151	0	0	0.94	Untested	S=C1Nc2c(N1C1CCN(CC1)CCCC(=O)c1ccc(F)cc1)cccc2
1145	3794430	0	0	0.94	Untested	FC(F)(F)C(=O)NC1CC(OC(C)C1O)OC1CC(O)(Cc2c1c(O)c1c(C(=O)c3c(C1=O)c(OC)ccc3)c2O)C(=O)COC(=O)CCCC
1146	4488115	0	0	0.87	Untested	S(C=1C(C2N(C(=O)C2C(O)C)C=1C(O)=O)C)C1Cn2[n+](C1)cnc2
1147	4630253	0	0	0.85	Untested	ClC1C2C(C3(C(C1)=CC(=O)C=C3)C)C(O)CC1(C2CC(C)C1(O)C(=O)CO)C
1148	4630261	0	0	0.98	Untested	S1C2N(C(=O)C2NC(=O)C(O)c2ccccc2)C(C(O)=O)=C(C1)CSc1nnnn1CS(O)(=O)=O
1149	4659387	0	0	0.87	Untested	S(C(=O)C1(O)C2(CC(O)C3(F)C(C2CC1C)CC(F)C1=CC(=O)C=CC13C)C)CF
1150	4659388	0	0	0.95	Untested	ClC12C(C3CC(C)C(O)(C(=O)CCl)C3(CC1O)C)CCC1=CC(=O)C=CC12C
1151	5229711	0	0	0.72	Untested	OC1CC(O)C(\C=C\C(O)(CCCCC)C)C1C\C=C\CCCC(O)=O
1152	5229907	0	0	0.83	Untested	FC12C(C3CC(O)C(O)(C(=O)CO)C3(CC1O)C)CC(F)C1=CC(=O)C=CC12C
1153	5284548	0	0	0.81	Untested	S(O)(=O)(=O)C(CC(OCC(CCCCC)CC)=O)C(OCC(CCCC)CC)=O
1154	5311167	0	0	0.97	Untested	ClCC(=O)C1(O)C2(CC(O)C3(F)C(C2CC1C)CC(F)C1=CC(=O)C=CC13C)C
1155	5360237	0	0	0.86	Untested	O1C(COC2OC(C)C(O)C(O)C2O)C(O)C(O)C(O)C1OC1=C(OC=2C(=C1O)C(=O)C=C(OCCO)C=2)c1cc(OCCO)c(OCCO)cc1
1156	5478929	0	0	0.94	Untested	Oc1cc2N(CC(O)c2cc1N=NC(=O)N)C
1157	6785321	0	0	0.85	Untested	Oc1c2c(ccc1N=NC(=O)N)cccc2
1158	9865442	0	0	0.84	Untested	ClCOC(=O)C1(O)CCC2C3C(C4(C(=CC(=O)C=C4)CC3)C)C(O)CC12C
1159	10182969	0	0	1.07	Untested	O(C)c1ccc(-n2nc(c3c2C(=O)N(CC3)c2ccc(N3CCCCC3=O)cc2)C(=O)N)cc1

1160	11163584	0	0	0.90	Untested	O=C1N(C)C(=O)N(Cc2ccccc2C#N)C(N2CC(N)CCC2)=C1
1161	11667893	0	0	0.95	Untested	O=C(Nc1cc2NCC(c2cc1)(C)C)c1cccnc1NCc1ccncc1
1162	16132921	0	0	1.15	Untested	S(CCC(NC(=O)C(NC(=O)CNC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)CCC(=O)N)CC(O)=O)Cc1ccc(OS(O)(=O)=O)cc1)C(O)C)Cc1c2c([nH]c1)cccc2)C(=O)NC(CC(O)
1163	19371515	0	0	0.77	Untested	O=Cc1n2c(nc1)/C(/c1c(CC2)cccc1)=C\1/CCN(CC/1)C
1164	20980918	0	0	0.99	Untested	S(=O)(=O)(NC(=O)CCC\C=C\CC1C(\C=C\C(O)COc2ccccc2)C(O)CC1=O)C
1165	21014972	0	0	1.32	Untested	Clc1c2Oc3cc4C(NC(=O)C(NC(=O)C(NC(=O)C(NC)CC(C)C)C(O)c(c1)cc2)CC(=O)N)C(=O)NC1c2cc(-c5c(cc(O)c(CNCP(O)(O)=O)c5O)C(NC(=O)C(NC1=O)C(O)c1cc(Cl)c(Oc(c4)c3O
1166	21183636	0	0	1.14	Untested	S1SCCC(=O)NC(Cc2ccc(OCC)cc2)C(=O)NC(C(CC)C)C(=O)NC(C(O)C)C(=O)NC(CC(=O)N)C(=O)NC(C1)C(=O)N1CCCC1C(=O)NC(CCCN)C(=O)NCC(=O)N
1167	21879626	0	0	1.21	Untested	Clc1ccc(cc1)CC(NC(=O)C(NC(=O)C)Cc1cc2c(cc1)cccc2)C(=O)NC(Cc1cccnc1)C(=O)NC(C(=O)N(C(Cc1ccc(O)cc1)C(=O)NC(CC(=O)N)C(=O)NC(CC(C)C)C(=O)NC(CCCCNC(C)C)C(=
1168	21977877	0	0	0.73	Untested	SCC(=O)C1(O)CCC2C3C(C4(C(=CC(=O)CC4)CC3)C)C(O)CC12C
1169	25074996	0	0	1.27	Untested	Clc1ccc(cc1)CC(NC(=O)C(NC(=O)C)Cc1cc2c(cc1)cccc2)C(=O)NC(Cc1cccnc1)C(=O)NC(C(=O)NC(Cc1ccc(O)cc1)C(=O)NC(CCCNC(=O)N)C(=O)NC(CC(C)C)C(=O)NC(CCCNC(N)=N)C
1170	25078009	0	0	1.15	Untested	Oc1ccc(cc1)CC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C1NC(=O)CC1)Cc1nc[nH]c1)Cc1c2c([nH]c1)cccc2)CO)C(=O)NC(Cc1ncn(c1)Cc1ccccc1)C(=O)NC(CC(C)C)C(=O)NC(CCCNC(N)
1171	44208957	0	0	1.30	Untested	Clc1ccc(cc1)CC(NC(=O)C(NC(=O)C)Cc1cc2c(cc1)cccc2)C(=O)NC(Cc1cccnc1)C(=O)NC(C(=O)NC(Cc1ccc(O)cc1)C(=O)NC(CCCCN\C(=N/CC)\NCC)C(=O)NC(CC(C)C)C(=O)NC(CCCC
1172	53393954	0	0	0.85	Untested	OC1CC(O)C(\C=C\C(O)CCc2ccccc2)C1C\C=C\CCCC(=O)NCC
1173	53394208	0	0	0.89	Untested	FC(F)(F)c1cc(OCC(O)\C=C\C2C(C\C=C\CCCC(OC(C)C)=O)C(O)CC2O)ccc1
1174	53394480	0	0	0.83	Untested	FC(F)(\C=C\C1C(C\C=C\CCCC(OC(C)C)=O)C(O)CC1O)COc1ccccc1
1175	57519507	0	0	0.95	Untested	O=C(NO)\C=C\c1ccc(cc1)CNCCc1c2c([nH]c1C)cccc2

1176	74084286	0	0	0.68	Untested	O1CCCC1C=1CC2N(C(=O)C2C(O)C)C=1C(O)=O
1177	359	1	1	0.64	Untested	Oc1cc(O)cc(O)c1
1178	2082	1	0	0.82	Untested	S(CCC)c1cc2[nH]c(nc2cc1)NC(OC)=O
1179	2171	1	0	0.95	Untested	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)C(N)c1ccc(O)cc1
1180	2177	1	0	0.95	Untested	S(=O)(=O)(N(CC(C)C)CC(O)C(NC(OC1CCOC1)=O)Cc1ccccc1)c1ccc(N)cc1
1181	2366	1	1	0.47	Untested	n1ccccc1CCNC
1182	2441	1	0	0.97	Untested	Brc1cc2c(NC(=O)CN=C2c2ncccc2)cc1
1183	2623	1	0	1.03	Untested	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC(C[N+]3(CCCC3)C)=C(N2C1=O)C(O)=O
1184	2632	1	0	1.00	Untested	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC(COC(=O)C)=C(N2C1=O)C(O)=O
1185	2651	1	0	1.05	Untested	s1cc(nc1N)/C(=N\OC(C(O)=O)(C)C)/C(=O)NC1C2SCC(C[n+]3ccccc3)=C(N2C1=O)C(O)=O
1186	2655	1	0	1.00	Untested	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC=C(N2C1=O)C(O)=O
1187	2809	1	0	0.97	Untested	Clc1cc2c(NC(=O)C(N=C2c2ccccc2)C(O)=O)cc1
1188	2997	1	0	0.94	Untested	Clc1cc2c(NC(=O)CN=C2c2ccccc2)cc1
1189	3171	1	0	0.65	Untested	Oc1cc(ccc1O)C(O)C(N)C(O)=O
1190	3199	1	0	0.97	Untested	S1CC2(NCCc3c2cc(OC)c(O)c3)C(OCC2N3C(C4N(C(Cc5c4c(O)c(OC)c(c5)C)C3O)C)C1c1c2c2OCOc2c(C)c1OC(=O)C)=O
1191	3299	1	0	1.02	Untested	Clc1cc2c(NC(=O)C(N=C2c2ccccc2F)C(OCC)=O)cc1
1192	3367	1	0	0.78	Untested	Fc1nc(N)c2ncn(c2n1)C1OC(CO)C(O)C1O
1193	3407	1	0	0.94	Untested	Clc1cccc(NC(=O)c2ccccc2)c1CN(CC(=O)N1CCOCC1)C
1194	3755	1	0	0.94	Untested	O1C(CN)C(O)C(O)C(O)C1OC1C(O)C(OC2OCC(O)(C)C(NC)C2O)C(NC(=O)C(O)CN)CC1N
1195	3804	1	0	0.88	Untested	O1C(C)C(OC2OC(C)C(OC(=O)CC(C)C)C(O)(C2)C)C(N(C)C)C(O)C1OC1C(OC)C(OC(=O)C)CC(OC(C/C=C\C=C\C(O)C(CC1CC=O)C)C)=O
1196	3929	1	0	0.93	Untested	Fc1cc(N2CC(OC2=O)CNC(=O)C)ccc1N1CCOCC1
1197	4116	1	0	0.76	Untested	ClC(Cl)C(F)(F)OC

1198	4608	1	0	1.01	Untested	S1C2N(C(C(O)=O)C1(C)C)C(=O)C2NC(=O)c1c(noc1C)-c1ccccc1
1199	4999	1	0	1.08	Untested	Clc1cc2c(N(CC(F)(F)F)C(=S)CN=C2c2ccccc2F)cc1
1200	5090	1	0	0.82	Untested	S(=O)(=O)(C)c1ccc(cc1)C=1COC(=O)C=1c1ccccc1
1201	5199	1	0	0.82	Untested	O1C(CO)C(O)C(O)C(O)C1Oc1c2c(ccc1)C(c1c(C2=O)c(O)cc(c1)C(O)=O)C1c2c(C(=O)c3c1cccc3OC1OC(CO)C(O)C(O)C1O)c(O)cc(c2)C(O)=O
1202	5485	1	0	0.80	Untested	s1cccc1\C(=C\1/CCC2[N+](C/1)(CCCC2)C)\c1sccc1
1203	13314	1	0	0.97	Untested	Clc1ccccc1C1=NC(O)C(=O)N(c2c1cc(Cl)cc2)C
1204	31729	1	0	0.82	Untested	Oc1ccc(cc1)C(O)C(NCCc1ccc(O)cc1)C
1205	38409	1	0	0.89	Untested	Clc1c(Cl)c(OCC(O)=O)ccc1C(=O)c1sccc1
1206	39941	1	0	0.89	Untested	Clc1ccc(cc1)-c1oc2c(n1)cc(cc2)C(C(O)=O)C
1207	54547	1	0	1.04	Untested	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC(COC)=C(N2C1=O)C(O)=O
1208	60172	1	0	0.90	Untested	P(O)(O)(=O)COCCn1c2ncnc(N)c2nc1
1209	60815	1	0	0.86	Untested	O(C(=O)C1(N(C(=O)CC)c2ccccc2)CCN(CC1)CCC(OC)=O)C
1210	64987	1	0	0.92	Untested	P(O)(O)(=O)COC(Cn1c2ncnc(N)c2nc1)C
1211	68740	1	0	0.87	Untested	P(O)(O)(=O)C(P(O)(O)=O)(O)Cn1ccnc1
1212	68942	1	0	0.82	Untested	Fc1ccc(cc1)C1C(=O)c2c(cccc2)C1=O
1213	71349	1	0	1.14	Untested	S1SCC(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)C(N)Cc2cc3c(cc2)cccc3)C1)Cc1ccc(O)cc1)Cc1c2c([nH]c1)cccc2)CCCCN)C(C)C)C(=O)NC(C(O)C)C(=O)N
1214	105031	1	1	0.71	Untested	O1C2OC3(OOC24C(CCC(C4CC3)C)C(C)C1OC(=O)CCC(O)=O)C
1215	155667	1	0	0.99	Untested	Clc1cc2c(NC(=O)C(N=C2c2ccccc2F)C(O)=O)cc1
1216	198068	1	0	0.91	Untested	S1Cc2c(cccc2)\C(\c2c1cccc2)=C\1/CC2N(C(C/1)CC2)C
1217	230076	1	0	0.94	Untested	O1C(C)C(O)C(O)(CC1OC1CC(OC(C)C1O)OC1CC(OC(C)C1O)OC1C(Cc2c(C1=O)c(O)c1c(cc(OC3OC(C)C(O)C(OC4OC(C)C(O)C(O)C4)C3)c(C)c1O)c2)C(OC)C(=O)C(O)C(O)C)C
1218	356786	1	0	0.76	Untested	FC1CC(OC1n1c2ncnc(N)c2nc1)CO
1219	3025811	1	0	1.01	Untested	Fc1ccc(cc1)-c1c2c(nc(C3CC3)c1\C=C\C(O)CC(O)CC(

						O)=O)cccc2
1220	4487901	1	0	1.06	Untested	s1cc(nc1N)/C(=N\OC)/C(=O)NC1C2SCC(C[n+]3c4CCCc4ccc3)=C(N2C1=O)C(O)=O
1221	4580358	1	0	0.88	Untested	O1C(CC)C(O)(C2OC(NC(C2C)C(CC(O)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(OC2OC(C)C(O)C(OC)(C2)C)C(C)C1=O)C)COCCOC)C
1222	4630419	1	0	0.94	Untested	ClCC(=O)C1(O)C2(CC(O)C3(F)C(C2CC1C)CCC1=CC(=O)C=CC13C)C
1223	6803597	1	0	0.96	Untested	s1cc(nc1N)C(N=O)C(=O)NC1C2SCC(C=C)=C(N2C1=O)C(O)=O
1224	9883933	1	0	0.79	Untested	O(C(OC(=O)NCC1(CCCCC1)CC(O)=O)C)C(=O)C(C)C
1225	11284707	1	0	0.97	Untested	S(=O)(=O)(N(CC(C)C)CC(OP(O)(O)=O)C(NC(OC1CCOC1)=O)Cc1ccccc1)c1ccc(N)cc1
1226	16129688	1	0	1.36	Untested	O1C(C)C(NC(=O)C(NC(=O)C(NC(=O)C(NC(=O)CCCCCCCCC)Cc2c3c([nH]c2)cccc3)CC(=O)N)CC(O)=O)C(=O)NCC(=O)NC(CCCN)C(=O)NC(CC(O)=O)C(=O)NC(C)C(=O)NC(CC(O)=O)C(=O
1227	21348140	1	0	0.95	Untested	O1C2CC(OC)C3(C(C(OC(=O)c4ccccc4)C4(O)CC(OC(=O)C(O)C(NC(OC(C)(C)C)=O)c5ccccc5)C(=C(C4(C)C)C(OC)C3=O)C)C2(OC(=O)C)C1)C
1228	21881641	1	0	0.90	Untested	O1C(CC)C2(OC(=O)N(C2C(C)C(=O)C(CC(OC)(C)C(OC2OC(CC(N(C)C)C2O)C)C(C)C(=O)C(C)C1=O)C)CCCCn1cc(nc1)-c1cccnc1)C
1229	21924983	1	0	0.75	Untested	OC1CC(n2c3N=C(NC(=O)c3nc2)N)C(=C)C1CO
1230	23545473	1	0	0.97	Untested	S(=O)(=O)(N(CC(C)C)CC(O)C(NC(OC1C2CCOC2OC1)=O)Cc1ccccc1)c1ccc(N)cc1
1231	42611257	1	0	1.10	Untested	Clc1ccc(cc1)-c1cc2c([nH]cc2C(=O)c2c(F)c(NS(=O)(=O)CCC)ccc2F)nc1
1232	53394893	1	0	1.06	Untested	ClC1CCC(CC1OC)\C=C(/C)\C1OC(=O)C2N(CCCC2)C(=O)C(=O)C2(OC(C(OC)CC2C)C(OC)CC(C\C(=C\C(CC)C(=O)CC(O)C1C)\C)C)O
1233	1210	0	0	1.03	Untested	O1c2cc3C([NH+](CCc3cc2OC)C)Cc2ccc(Oc3c4C([N+](CCc4cc(OC)c3O)(C)C)Cc3cc1c(O)cc3)cc2
1234	2563	0	0	0.77	Untested	Oc1cc(ccc1O)CC(NN)(C(O)=O)C
1235	104799	0	0	1.00	Untested	ClCCN(N=O)C(=O)NC(P(OCC)(OCC)=O)C
1236	9867822	0	0	0.98	Untested	O1C2n3c4c(c5c(CNC5=O)c5c6c(n(c45)C1(C)C(O)(C2)CO)cccc6)c1c3cccc1

1237	4239	1	0	0.73	Untested	o1n[n+](N2CCOCC2)cc1NC(OCC)=O
1238	214356	1	0	0.96	Untested	P(OC(Cn1ncnc1)(Cn1ncnc1)c1ccc(F)cc1F)(O)(O)=O
1239	1593	0	1	0.87	Untested	OC1C\C(=C/C=C\2/C3CCC(C(CCCC(O)(C)C)C)C3(CCC/2)C)\C(CC1)=C
1240	2008	0	0	0.89	Untested	O1C(C)C(OC2OC(C)C(OC3OC(C)C(=O)CC3)C(O)C2)C(N(C)C)CC1OC1CC(O)(CC)C(c2c1c(O)c1c(c2)C(=O)c2c(C1=O)c(O)ccc2)C(OC)=O
1241	2275	0	0	0.83	Untested	S(O)(=O)(=O)c1c2-c(cc(ccc2C)C(C)C)c(c1)C
1242	2611	0	0	1.02	Untested	s1cccc1CC(=O)NC1C2SCC(C[n+]3ccccc3)=C(N2C1=O)C(O)=O
1243	3183	0	0	0.77	Untested	Br\C=C\C1=CN(C2OC(CO)C(O)C2)C(=O)NC1=O
1244	3250	0	0	0.92	Untested	OCC(NC(=O)C1C=C2C(N(C1)C)Cc1c3c2cccc3[nH]c1)C
1245	3425	0	0	0.99	Untested	S1SCC\C=C\C2OC(=O)C(NC(=O)/C(/NC(=O)C(NC(=O)C(NC(=O)C2)C(C)C)C1)=C/C)C(C)C
1246	3561	0	0	0.93	Untested	IC#CCOc1cc(Cl)c(Cl)cc1Cl
1247	4018	0	0	0.83	Untested	OC1CC(O)C\C(=C/C=C\2/C3CCC(C(OCCC(O)(C)C)C)C3(CCC/2)C)\C1=C
1248	4872	0	1	0.71	Untested	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)C=C3)C)CCC(O)=O
1249	5404	0	0	1.01	Untested	Clc1cc(Cl)ccc1C1(OC(CO1)COc1ccc(N2CCN(CC2)C(C)C)cc1)Cn1ncnc1
1250	5469	0	0	0.90	Untested	Clc1cc2N(CC(=O)N3CCN(CC3)CCO)C(Sc2cc1)=O
1251	6199	0	1	0.60	Untested	O1c2c(C=CC1=O)cc1c(occ1)c2
1252	55362	0	0	0.94	Untested	OC(NC(=O)CCCCCCNC(N)=N)C(=O)NCCCCNCCCN
1253	651300	0	0	0.96	Untested	O(CCCCCCOc1ccc(cc1)C(N)=N)c1ccc(cc1)C(N)=N
1254	71927	0	0	0.75	Untested	O(C)C=1N(N=CC(=N)C=1)c1ccccc1
1255	123964	0	0	0.75	Untested	O1c2c(ccc(C)c2C)C(=O)c2c1c(ccc2)CC(O)=O
1256	433902	0	0	0.90	Untested	O1C(CC2OC(O)(C(OC(=O)\C=C\C=C\CCC)/C(/C2)=C/C(OC)=O)C(\C=C/C2OC(CC3(OC(CC(OC(=O)C)C3(C)C)CC(O)CC1=O)O)C\C(\C2)=C\C(OC)=O)(C)C)C(O)C
1257	579666	0	0	0.65	Untested	OC1CCC2C3C(CCC12C)C1(C(CC(=O)C=C1C)CC3)C
1258	3844471	0	0	0.72	Untested	OC1(C)C2(CC2)C(=C2C(=CC(C)=C2CO)C1=O)C
1259	4474932	0	1	0.76	Untested	OC1(CCC2C3C(CCC12C)C1(C(=CC(=O)CC1)C(=C3)C)C)C(=O)C

1260	44790940	0	0	0.90	Untested	OC1CC(O)C\C(=C/C=C\2/C3CCC(C(\C=C\C(C(C)C)C)C)C3(CCC/2)C)\C1=C
1261	46366000	0	0	0.88	Untested	OC1CC(O)C/C(/C1)=C\C=C/1\C2CCC(C(\C=C\C(C(O)(C)C)C)C)C2(CCC\1)C
1262	52297120	0	0	1.02	Untested	S1C2N(C(=O)C2NC(=O)Cc2ccccc2CN)C(C(O)=O)=C(C1)CSc1nnnn1CC(O)=O
1263	69183140	0	0	0.99	Untested	o1c2c(cc1C(=O)N)cc(N1CCN(CC1)CCCCc1c3cc(ccc3[nH]c1)C#N)cc2
1264	98252850	0	0	0.99	Untested	O1NC(=NC1=O)c1ccccc1-c1ccc(cc1)Cn1c2c(nc1OCC)cccc2C(O)=O
1265	13472099	0	1	0.73	Untested	OC1(CCC2C3C=C(C4=CC(=O)CCC4C3CCC12C)C)C(=O)C
1266	13502539	0	0	0.97	Untested	O1CC(O)C(O)CC1OC1CC(N)(Cc2c1c(O)c1c(C(=O)c3c(cccc3)C1=O)c2O)C(=O)C
1267	13559282	0	0	0.94	Untested	O(C(=O)C)C1(CCC2C3C(=C4C(=CC(=O)CC4)CC3)C(CC12C)c1ccc(N(C)C)cc1)C(=O)C
1268	14883207	0	1	0.79	Untested	OC1(CCC2C3C(C4C(=CC(=O)CC4)CC3)C(CC12CC)=C)C#C
1269	16220172	0	0	0.95	Untested	Oc1cc(NC(=O)c2cnc3c(cccc3)c2O)c(cc1C(C)(C)C)C(C)(C)C
1270	18350668	0	0	1.01	Untested	OC(=O)C(NC(=O)c1ccc(cc1)C(Cc1nc2c(nc(nc2N)N)nc1)CC#C)CCC(O)=O
1271	21874792	0	0	0.93	Untested	FC(F)(F)c1ccc(cc1NC(=O)C1CCC2C3C(CCC12C)C1(C(NC(=O)C=C1)CC3)C)C(F)(F)F
1272	21947593	0	0	1.06	Untested	O=C1N(C)C(C(C)C)C(=O)NC(CCCNC(N)=N)C(=O)NCC(=O)NC(CC(O)=O)C(=O)NC1Cc1ccccc1
1273	23560495	0	0	0.94	Untested	Clc1ccc(cc1Cc1ccc(OCC)cc1)C1OC(CO)C(O)C(O)C1O
1274	53398697	0	0	1.14	Untested	Clc1cc(Nc2c3cc(NC(=O)\C=C\CN(C)C)c(OCC)cc3ncc2C#N)ccc1OCc1ncccc1
1275	159	1	0	0.70	Untested	O\1C2C(C/C/1=C\CCCC(O)=O)C(\C=C\C(O)CCCCC)C(O)C2
1276	1971	1	0	0.91	Untested	OCC1CC(n2c3nc(nc(NC4CC4)c3nc2)N)C=C1
1277	2091	1	0	0.89	Untested	OC1CC(O)C\C(=C/C=C\2/C3CCC(C(CCCC(C)C)C)C3(CCC/2)C)\C1=C
1278	2179	1	0	1.12	Untested	S(=O)(=O)(Nc1cc(OC)c(Nc2c3c(nc4c2cccc4)cccc3)cc1)C
1279	2524	1	0	0.88	Untested	OC1CC(O)C\C(=C/C=C\2/C3CCC(C(CCCC(O)(C)C)C)C3(CCC/2)C)\C1=C
1280	2794	1	0	1.20	Untested	Clc1ccc(N2C3=C\C(=N/C(C)C)\C(Nc4ccc(Cl)cc4)=CC3=Nc3c2cccc3)cc1
1281	3015	1	0	0.88	Untested	ClC=1C2=CC(=O)C3C(C3)C2(C2C(C3CCC(O)(C(=O)C)C3(CC2)C)C=1)C
1282	3261	1	0	1.01	Untested	Clc1cc2c(-n3c(nnc3)CN=C2c2ccccc2)cc1
1283	4056	1	0	0.75	Untested	O=C1c2c(cccc2)C(=O)C(C)=C1C\C=C(\CC\C=C(\CC\C=C(\CC\C=C(\C)/C)/C)/C)/C

						)/C
1284	4568	1	0	1.01	Untested	O1C(C)C(O)C(N)C(O)C1OC1\C=C\C=C\C=C\C=C\CC\C=C\C=C\C(C)C(O)C(C)C(OC(=O)CC(O)CC(O)CC(O)CCC(O)C(O)CC2(OC(C1)C(C(O)=O)C(O)C2)O)C
1285	5040	1	0	1.01	Untested	O1C(CC(=O)C(\C=C(/C)\C(O)C(OC)C(=O)C(CC(\C=C/C=C/C=C(/C)\C(OC)CC2OC(O)(C(=O)C(=O)N3C(CCCC3)C1=O)C(CC2)C)C)C)C)C(CC1CC(OC)C(O)CC1)C
1286	5396	1	0	0.87	Untested	s1cccc1C1OC2C(OC(OC3C4C(C(c5c3cc3OCOc3c5)c3cc(OC)c(O)c(OC)c3)C(OC4)=O)C(O)C2O)CO1
1287	5527	1	0	0.86	Untested	O(C)c1cc(ccc1OC)\C=C\C(=O)Nc1ccccc1C(O)=O
1288	5651	1	0	1.34	Untested	Clc1c2Oc3cc4C(NC(=O)C(NC(=O)C(NC(=O)C(NC)CC(C)C)C(O)c(c1)cc2)CC(=O)N)C(=O)NC1c2cc(-c5c(cc(O)cc5O)C(NC(=O)C(NC1=O)C(O)c1cc(Cl)c(Oc(c4)c3OC3OC(CO)C(O)C(
1289	249098	1	0	0.63	Untested	O1CC2(C(CC1=O)CCC1C3CCC(O)(C)C3(CCC12)C)C
1290	287691	1	0	0.98	Untested	S(C)C1=CC=C2C(=CC1=O)C(NC(=O)C)CCc1c2c(OC)c(OC)c(OC2OC(CO)C(O)C(O)C2O)c1
1291	418931	1	0	0.92	Untested	O1C(C)C(C)C(OC(=O)C)C(C)C(=O)C2(OC2)CC(C)C(OC2OC(CC(N(C)C)C2OC(=O)C)C)C(C)C(OC2OC(C)C(OC(=O)C)C(OC)C2)C(C)C1=O
1292	429017	1	0	1.00	Untested	O(C)c1cc2N(C3C4(C5N(CC4)CC=CC5(CC)C(O)C3(O)C(=O)N)c2cc1C1(CC2CC(O)(CN(C2)CCc2c1[nH]c1c2cccc1)CC)C(OC)=O)C
1293	644241	1	0	1.01	Untested	FC(F)(F)c1cc(NC(=O)c2cc(Nc3nc(ccn3)-c3cccnc3)c(cc2)C)cc(-n2cc(nc2)C)c1
1294	3081921	1	0	1.16	Untested	S(Oc1cc(ccc1O)C(O)C(O)C1NC(=O)C2N(CC(O)C2)C(=O)C(NC(=O)C(NC(=O)c2ccc(cc2)-c2noc(c2)-c2ccc(OCCCCC)cc2)CC(O)C(O)NC(=O)C2N(CC(C)C2O)C(=O)C(NC1=O)C(O)CC(=
1295	4432690	1	0	0.87	Untested	O1C(C)C(NC2C=C(CO)C(O)C(O)C2O)C(O)C(O)C1OC1C(O)C(O)C(OC1CO)OC1C(O)C(O)C(OC1CO)O
1296	4659180	1	0	0.79	Untested	OC1(CCC2C3C(C4C(CC3)=CCCC4)C(CC12CC)=C)C#C
1297	4979943	1	0	1.02	Untested	S(=O)(=O)(N(C)c1nc(-c2ccc(F)cc2)c(\C=C\C(O)CC(O)CC(O)=O)c(n1)C(C)C)C
1298	6816262	1	0	0.89	Untested	Oc1ccc(N=Nc2cc(C(O)=O)c(O)cc2)cc1C(O)=O
1299	11597571	1	0	1.15	Untested	Clc1c(C(Oc2cc(cnc2N)-c2cn(nc2)C2CCNCC2)C)c(Cl)ccc1F

1300	15224271	1	0	1.16	Untested	O(CCCCC)c1ccc(cc1)-c1ccc(cc1)-c1ccc(cc1)C(=O)NC1CC(O)C(O)NC(=O)C2N(CC(C)C2O)C(=O)C(NC(=O)C(NC(=O)C2N(CC(O)C2)C(=O)C(NC1=O)C(O)C)C(O)C(O)c1ccc(O)cc1)C(
1301	21879648	1	0	0.87	Untested	OC1CC2=CCC3C4CC=C(C4(CCC3C2(CC1)C)C)c1cccnc1
1302	23314628	1	0	0.82	Untested	ClC=1C2=CC(=O)CCC2(C2C(C3CCC(O)(C(=O)C)C3(CC2)C)C=1)C
1303	53394043	1	0	0.73	Untested	OC1CC2C(C\C(\C2)=C/CCCC(O)=O)C1\C=C\C(O)C(CC#CC)C
1304	53395233	1	0	1.00	Untested	O1c2c3c4c(c(O)c2C)c(O)c(NC(=O)/C(=C\C=C/C(C)C(O)C(C)C(O)C(C)C(OC(=O)C)C(C)C(OC)\C=C/OC1(C)C3=O)/C)c1n2c(nc14)C=C(C=C2)C
1305	53398658	1	0	1.02	Untested	O1C(CC(=O)C(\C=C(/C)\C(O)C(OC)C(=O)C(CC(\C=C/C=C/C=C(/C)\C(OC)CC2OC(O)(C(=O)C(=O)N3C(CCCC3)C1=O)C(CC2)C)C)C)C(CC1CC(OC)C(OCCO)CC1)C
1306	53420943	1	0	0.98	Untested	s1cc(nc1N)/C(=C\CC)/C(=O)NC1C2SCC(COC(=O)N)=C(N2C1=O)C(O)=O
1307	53432469	1	0	1.03	Untested	s1cnc(C)c1\C=C\C=1CSC2N(C(=O)C2NC(=O)\C(=N\OC)\c2nc(sc2)N)C=1C(O)=O
1308	54690031	1	0	0.98	Untested	Clc1sc2c(S(=O)(=O)N(C)C(C(=O)Nc3nccc3)C2=O)c1

Table A3. Final 20 PubChem assays manually evaluated for consideration into the final biological response profile

	Statistics for relationship to liver damage[a]								Statistics for relationship to ARE-*bla*[a]								
AID	TP[b]	FP	TN	FN	Spec	Sens	CCR[c]	L[c]	TP[b]	FP	TN	FN	Spec	Sens	CCR[c]	L[c]	Title[d]
121	11	2	20	52	0.91	0.17	0.54	1.28	11	2	15	46	0.88	0.19	0.54	1.09	NCI human tumor cell line growth inhibition assay. Data for the K-562 Leukemia cell line
123	11	2	20	52	0.91	0.17	0.54	1.28	11	2	15	46	0.88	0.19	0.54	1.09	NCI human tumor cell line growth inhibition assay. Data for the MOLT-4 Leukemia cell line
248	21	6	54	66	0.90	0.24	0.57	2.07	13	2	33	56	0.94	0.19	0.57	2.20	NCI In Vivo Anticancer Drug Screen. Data for tumor model L1210 Leukemia (intraperitoneal) in B6D2F1 (BDF1) mice
256	22	5	25	34	0.83	0.39	0.61	1.96	12	2	11	32	0.85	0.27	0.56	1.18	NCI In Vivo Anticancer Drug Screen. Data for tumor model L1210 Leukemia (intraperitoneal) in CD2F1 (CDF1) mice
328	23	3	12	14	0.80	0.62	0.71	2.33	15	2	10	19	0.83	0.44	0.64	1.76	NCI In Vivo Anticancer Drug Screen. Data for tumor model P388 Leukemia (intraperitoneal) in B6D2F1 (BDF1) mice
330	27	5	33	45	0.87	0.38	0.62	2.38	15	4	20	43	0.83	0.26	0.55	1.24	NCI In Vivo Anticancer Drug Screen. Data for tumor model P388 Leukemia (intraperitoneal) in CD2F1 (CDF1) mice
589	23	12	70	117	0.85	0.16	0.51	1.04	27	5	35	131	0.88	0.17	0.52	1.14	qHTS Assay for Spectroscopic Profiling in 4-MU Spectral Region
590	22	12	68	112	0.85	0.16	0.51	1.01	27	4	35	124	0.90	0.18	0.54	1.39	qHTS Assay for Spectroscopic Profiling in A350 Spectral Region
1189	28	8	17	31	0.68	0.47	0.58	1.32	48	10	16	44	0.62	0.52	0.57	1.23	DSSTox (CPDBAS) Carcinogenic Potency Database Summary SingleCellCall Results
1199	19	4	12	27	0.75	0.41	0.58	1.32	36	8	10	35	0.56	0.51	0.53	1.01	DSSTox (CPDBAS) Carcinogenic Potency Database Summary Mouse Bioassay Results
1205	21	4	14	26	0.78	0.45	0.61	1.61	39	9	11	34	0.55	0.53	0.54	1.07	DSSTox (CPDBAS) Carcinogenic Potency Database Summary MultiCellCall Results
1996	70	28	3	3	0.10	0.96	0.53	1.03	70	17	5	7	0.23	0.91	0.57	1.11	Aqueous Solubility from MLSMR Stock

																	Solutions
2330	11	6	111	152	0.95	0.07	0.51	1.13	21	5	48	147	0.91	0.13	0.52	1.10	Luminescence Cell-Based Primary HTS to Identify Inhibitors of STK33
686978[e]	74	46	139	200	0.75	0.27	0.51	1.06	99	31	80	218	0.72	0.31	0.52	1.08	qHTS for Inhibitors of human tyrosyl-DNA phosphodiesterase 1 (TDP1): qHTS in cells in absence of CPT
686979	67	31	144	201	0.82	0.25	0.54	1.37	86	29	85	220	0.75	0.28	0.51	1.07	qHTS for Inhibitors of human tyrosyl-DNA phosphodiesterase 1 (TDP1): qHTS in cells in presence of CPT
720532	39	25	159	236	0.86	0.14	0.50	1.00	40	10	102	273	0.91	0.13	0.52	1.30	qHTS for Inhibitors of binding or entry into cells for Marburg Virus
743065	61	44	122	163	0.73	0.27	0.50	1.00	96	27	79	203	0.75	0.32	0.53	1.22	qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway
743067[e]	18	10	117	162	0.92	0.10	0.51	1.15	29	8	76	197	0.90	0.13	0.52	1.20	qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway: Summary[e]
743140[e]	18	3	147	196	0.98	0.08	0.53	3.15	17	3	94	256	0.97	0.06	0.52	1.51	qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway: Summary[e]
743202[e]	20	10	88	142	0.90	0.12	0.51	1.10	26	5	56	174	0.92	0.13	0.52	1.32	qHTS assay for small molecule agonists of the antioxidant response element (ARE) signaling pathway[e]

Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative; Sens, sensitivity; Spec, specificity; CCR, correct classification rate; L, likelihood parameter.
The first three criteria used in the automated selection process were that the assay had to [a]appear in both profile groups, [b]contain > 10 active responses that matched the inputted data, and [c]correlation was better than random (CCR > 0.5 and L ≥ 1). [d]Manual evaluation was required to determine if it was an *in vitro* assay and relevant to liver toxicity or oxidative stress. [e]AIDs 686978, 743067, and 743140, 743202 sufficed the criteria described

References

Adler S, Basketter D, Creton S, Pelkonen O, van Benthem J, Zuang V, et al. 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch. Toxicol. 85:367–485.

Allen TEH, Goodman JM, Gutsell S, Russell PJ. 2014. Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. Chem. Res. Toxicol. 27:2100–2112.

Andrews CW, Bennett L, Yu LX. 2000. Predicting human oral bioavailability of a compound: development of a novel quantitative structure-bioavailability relationship. Pharm. Res. 17: 639–44.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29:730–41.

Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, et al. 2013. The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. Drug Discov. Today 18:716–23.

Baell JB, Holloway GA. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J. Med. Chem. 53:2719–2740.

Ben Hassine S, Arcangioli B. 2009. Tdp1 protects against oxidative DNA damage in non-dividing fission yeast. EMBO J. 28:632–40.

Benet LZ, Broccatelli F, Oprea TI. 2011. BDDCS applied to over 900 drugs. AAPS J. 13:519–47.

Birnbaum LS. 1985. The role of structure in the disposition of halogenated aromatic xenobiotics. Environ. Health Perspect. 61: 11–20.

Borduas N, da Silva G, Murphy JG, Abbatt JPD. 2015. Experimental and Theoretical Understanding of the Gas Phase Oxidation of Atmospheric Amides with OH Radicals: Kinetics, Products, and Mechanisms. J. Phys. Chem. A 119:4298–4308.

Breiman L. 2001. Random forests. Mach. Learn. 45:5–32.

Buxton ILO, Benet LZ. 2011. Chapter 2. Pharmacokinetics: The Dynamics of Drug Absorption, Distribution, Metabolism, and Elimination. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e* (L.L. Brunton, B.A. Chabner, and B.C. Knollmanneds.), The McGraw-Hill Companies, New York, NY.

Chabner BA, Bertino J, Cleary J, Ortiz T, Lane A, Supko JG, et al. 2011. Chapter 61. Cytotoxic Agents. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e* (L.L. Brunton, B.A. Chabner, and B.C. Knollmanneds.), The McGraw-Hill Companies, New York, NY.

Chakravarti SK, Saiakhov RD, Klopman G. 2012. Optimizing predictive performance of CASE Ultra expert system models using the applicability domains of individual toxicity alerts. J. Chem. Inf. Model. 52:2609–18.

Chaudhry Q, Piclin N, Cotterill J, Pintore M, Price NR, Chrétien JR, et al. 2010. Global QSAR models of skin sensitisers for regulatory purposes. Chem. Cent. J. 4 Suppl 1:S5.

Chen M, Borlak J, Tong W. 2014. Predicting idiosyncratic drug-induced liver injury - some recent advances. Expert Rev. Gastroenterol. Hepatol. 8:721–3.

Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. 2014. QSAR modeling: Where have you been? Where are you going to? J. Med. Chem. 57:4977–5010.

Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. Science 319:906–7.

Committee on Toxicity Testing and Assessment of Environmental Agents NRC. 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The National Academies Press, Washington, DC.

Dalgaard P. 2008. *Introductory Statistics with R*. Springer New York.

Daniel WW. 2009. *Biostatistics: A Foundation for Analysis in the Health Sciences*. 9th ed. Wiley, Hoboken, NJ.

Davies DS, Thorgeirsson SS. 1971. Mechanism of hepatic drug oxidation and its relationship to individual differences in rates of oxidation in man. Ann. N. Y. Acad. Sci. 179:411–20.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol. Sci. 95:5–12.

Dubertret L, Alirezai M, Rostain G, Lahfa M, Forsea D, Niculae BD, et al. 2003. The use of lymecycline in the treatment of moderate to severe acne vulgaris: a comparison of the efficacy and safety of two dosing regimens. Eur. J. Dermatol. 13: 44–8.

Eaton DL, Gilbert SG. 2010. Chapter 2. Principles of Toxicology. In *Casarett & Doull's Essentials of Toxicology, 2e* (C.D. Klaassen and J.B. Watkinseds.), The McGraw-Hill Companies, New York, NY.

Ekins S. 2014. Progress in computational toxicology. J. Pharmacol. Toxicol. Methods 69:115–40.

Fielden MR, Brennan R, Gollub J. 2007. A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. Toxicol. Sci. 99:90–100.

Fisher RA. 1935. *The design of experiments*.

Fourches D, Muratov E, Tropsha A. 2010. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J. Chem. Inf. Model. 50:1189–204.

Fujita T, Iwasa J, Hansch C. 1964. A new substituent constant, π, derived from partition coefficients. J. Am. Chem. Soc. 86:5175–5180.

Gaisford W. 1962. Fatality after oxyphenbutazone in Still's disease. Br. Med. J. 2: 1517.

Giacomini KM, Sugiyama Y. 2011. Chapter 5. Membrane Transporters and Drug Response. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e* (L.L. Brunton, B.A. Chabner, and B.C. Knollmanneds.), The McGraw-Hill Companies, New York, NY.

Golbraikh A, Shen M, Xiao Z, Xiao Y-D, Lee K-H, Tropsha A. 2003. Rational selection of training and test sets for the development of validated QSAR models. J. Comput. Aided. Mol. Des. 17:241–53.

Hansch C. 1969. A Quantitative Approach to Biochemical Structure-Activity Relationships. Acc. Chem. Res. 2: 232–&.

Hansch C, Fujita T. 1964. ρ-σ-π Analysis; method for the correlation of biological activity and chemical structure. J. Am. Chem. Soc. 86:1616–1626.

Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M. 1963. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. J. Am. Chem. Soc. 85:2817–2824.

Hartung T. 2009. Toxicology for the twenty-first century. Nature 460:208–12.

Hermann R. 1972. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. J. Phys. Chem. 1465: 2754–2759.

Holland RD, Gehring T, Taylor J, Lake BG, Gooderham NJ, Turesky RJ. 2005. Formation of a mutagenic heterocyclic aromatic amine from creatinine in urine of meat eaters and vegetarians. Chem. Res. Toxicol. 18:579–90.

Hou T, Li Y, Zhang W, Wang J. 2009. Recent developments of in silico predictions of intestinal absorption and oral bioavailability. Comb. Chem. High Throughput Screen. 12:497–506.

Hou T, Wang J, Zhang W, Xu X. 2007. ADME evaluation in drug discovery. 7. prediction of oral absorption by correlation and classification. J. Chem. Inf. Model. 47:208–218.

Huang MJ, Liaw YF. 1995. Clinical associations between thyroid and liver diseases. J. Gastroenterol. Hepatol. 10: 344–50.

Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, et al. 2011. The NCGC Pharmaceutical Collection: A Comprehensive Resource of Clinically Approved Drugs Enabling Repurposing and Chemical Genomics. Sci. Transl. Med. 3:80ps16–80ps16.

Hybertson BM, Gao B, Bose SK, McCord JM. 2011. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. Mol. Aspects Med. 32:234–46.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, et al. 2006. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc. Natl. Acad. Sci. U. S. A. 103:11473–8.

Inoue M, Morikawa M, Tsuboi M, Ito Y, Sugiura M. 1980. Comparative study of human intestinal and hepatic esterases as related to enzymatic properties and hydrolizing activity for ester-type drugs. Jpn. J. Pharmacol. 30: 529–35.

Itoh K, Chiba T, Takahashi S, Ishii T, Igarashi K, Katoh Y, et al. 1997. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem. Biophys. Res. Commun. 236: 313–22.

Izenman AJ. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. 1st ed. Springer Publishing Company, Incorporated.

Jameson JL, Weetman AP. 2012. Chapter 341. Disorders of the Thyroid Gland. In *Harrison's Principles of Internal Medicine, 18e* (D.L. Longo, A.S. Fauci, D.L. Kasper, S.L. Hauser, J.L. Jameson, and J. Loscalzoeds.), The McGraw-Hill Companies, New York, NY.

Judson RS, Martin MT, Egeghy P, Gangwal S, Reif DM, Kothiya P, et al. 2012. Aggregating Data for Computational Toxicology Applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System. Int. J. Mol. Sci. 13:1805–31.

Kensler TW, Wakabayashi N, Biswal S. 2007. Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu. Rev. Pharmacol. Toxicol. 47:89–116.

Kim MT, Huang R, Sedykh A, Wang W, Xia M, Zhu H. 2015. Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using the Antioxidant Response Element Reporter Gene Assay Models and Big Data. Environ. Health Perspect.

Kim MT, Sedykh A, Chakravarti SK, Saiakhov RD, Zhu H. 2014. Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. Pharm. Res. 31:1002–14.

Klopman G. 1984. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. J. Am. Chem. Soc. 106:7315–7321.

Königer C, Wingert I, Marsmann M, Rösler C, Beck J, Nassal M. 2014. Involvement of the host DNA-repair enzyme TDP2 in formation of the covalently closed circular DNA persistence reservoir of hepatitis B viruses. Proc. Natl. Acad. Sci. U. S. A. 111:E4244–53.

Kornbrot D. 2005. Pearson Product Moment Correlation. In *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, Ltd.

Kovatcheva A, Golbraikh A, Oloff S, Feng J, Zheng W, Tropsha A. 2005. QSAR modeling of datasets with enantioselective compounds using chirality sensitive molecular descriptors. SAR QSAR Environ. Res. 16:93–102.

Kovatcheva A, Golbraikh A, Oloff S, Xiao Y-DD, Zheng W, Wolschann P, et al. 2004. Combinatorial QSAR of ambergris fragrance compounds. J. Chem. Inf. Comput. Sci. 44:582–595.

Kruhlak NL, Benz RD, Zhou H, Colatsky TJ. 2012. (Q)SAR modeling and safety assessment in regulatory review. Clin. Pharmacol. Ther. 91:529–34.

Kruhlak NL, Contrera JF, Benz RD, Matthews EJ. 2007. Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. Adv. Drug Deliv. Rev. 59:43–55.

Kwak M-K, Wakabayashi N, Itoh K, Motohashi H, Yamamoto M, Kensler TW. 2003. Modulation of gene expression by cancer chemopreventive dithiolethiones through

the Keap1-Nrf2 pathway. Identification of novel gene clusters for cell survival. J. Biol. Chem. 278:8135–45.

Lafontan M, Girard J. 2008. Impact of visceral adipose tissue on liver metabolism. Part I: heterogeneity of adipose tissue and functional properties of visceral adipose tissue. Diabetes Metab. 34:317–27.

Lipinski C a, Lombardo F, Dominy BW, Feeney PJ. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 46: 3–26.

Low Y, Uehara T, Minowa Y, Yamada H, Ohno Y, Urushidani T, et al. 2011. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. Chem. Res. Toxicol. 24:1251–62.

Ma C-Y, Yang S-Y, Zhang H, Xiang M-L, Huang Q, Wei Y-Q. 2008. Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA-CG-SVM method. J. Pharm. Biomed. Anal. 47:677–682.

Ma Q. 2013. Role of nrf2 in oxidative stress and toxicity. Annu. Rev. Pharmacol. Toxicol. 53:401–26.

MacDonald JS, Robertson RT. 2009. Toxicity testing in the 21st century: a view from the pharmaceutical industry. Toxicol. Sci. 110:40–6.

Malik R, Hodgson H. 2002. The relationship between the thyroid gland and the liver. QJM 95: 559–69.

Mantena SK, King AL, Andringa KK, Eccleston HB, Bailey SM. 2008. Mitochondrial dysfunction and oxidative stress in the pathogenesis of alcohol- and obesity-induced fatty liver diseases. Free Radic. Biol. Med. 44:1259–72.

Martin TM, Harten P, Young DM, Muratov EN, Golbraikh A, Zhu H, et al. 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52:2570–8.

Martin Y. 2005. A bioavailability score. J. Med. Chem. 3164–3170.

Meunier B, de Visser SP, Shaik S. 2004. Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. Chem. Rev. 104:3947–80.

Moda TL, Montanari CA, Andricopulo AD. 2007. Hologram QSAR model for the prediction of human oral bioavailability. Bioorg. Med. Chem. 15:7738–45.

Moeller TA. 2010. From in vitro to in vivo: Preclinical assays to assess drug-induced liver injury demonstrate in vitro-in vivo correlation. Drug Discov. Dev. 13: 10–13.

MultiCASE. MultiCASE Inc. www.multicase.com.

Nielsen AB, Frydenvang K, Liljefors T, Buur A, Larsen C. 2005. Assessment of the combined approach of N-alkylation and salt formation to enhance aqueous solubility of tertiary amines using bupivacaine as a model drug. Eur. J. Pharm. Sci. 24:85–93.

O'Brien PJ, Irwin W, Diaz D, Howard-Cofield E, Krejsa CM, Slaughter MR, et al. 2006. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. Arch. Toxicol. 80:580–604.

Omura K. 2003. Clinical implications of dihydropyrimidine dehydrogenase (DPD) activity in 5-FU-based chemotherapy: Mutations in the DPD gene, and DPD inhibitory fluoropyrimidines. Int. J. Clin. Oncol. 8:132–138.

Paixão P, Gouveia LF, Morais JAG. 2012. Prediction of the human oral bioavailability by using in vitro and in silico drug related parameters in a physiologically based absorption model. Int. J. Pharm. 429:84–98.

Porter NA, Caldwell SE, Mills KA. 1995. Mechanisms of free radical oxidation of unsaturated lipids. Lipids 30:277–290.

Pouliot JJ, Yao KC, Robertson CA, Nash HA. 1999. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. Science 286:552–555.

Saiakhov R, Chakravarti S, Klopman G. 2013. Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs. Mol. Inform. 32:87–97.

Saitoh H, Gerard C, Aungst BJ. 1996. The secretory intestinal transport of some beta-lactam antibiotics and anionic compounds: a mechanism contributing to poor oral absorption. J. Pharmacol. Exp. Ther. 278: 205–11.

Sawa M, Mizuno K, Harada H, Tateishi H, Arai Y, Suzuki S, et al. 2005. Tryptamine-based human beta3-adrenergic receptor agonists. Part 3: improved oral bioavailability via modification of the sulfonamide moiety. Bioorganic Med. Chem. Lett. 15:1061–4.

Schilter B, Benigni R, Boobis A, Chiodini A, Cockburn A, Cronin MTD, et al. 2014. Establishing the level of safety concern for chemicals in food without the need for toxicity testing. Regul. Toxicol. Pharmacol. 68:275–296.

Sedykh A, Fourches D, Duan J, Hucke O, Garneau M, Zhu H, et al. 2013. Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions. Pharm. Res. 30:996–1007.

Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, et al. 2011. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ. Health Perspect. 119:364–70.

Shugarts S, Benet LZ. 2009. The role of transporters in the pharmacokinetics of orally administered drugs. Pharm. Res. 26:2039–54.

Shuhendler AJ, Pu K, Cui L, Uetrecht JP, Rao J. 2014. Real-time imaging of oxidative and nitrosative stress in the liver of live animals for drug-toxicity testing. Nat. Biotechnol. 32:373–80.

Shukla SJ, Huang R, Simmons SO, Tice RR, Witt KL, Vanleer D, et al. 2012. Profiling environmental chemicals for activity in the antioxidant response element signaling pathway using a high throughput screening approach. Environ. Health Perspect. 120:1150–6.

Sies H. 1997. Oxidative stress: oxidants and antioxidants. Exp. Physiol. 82:291–295.

Simmons SO, Fan C-Y, Yeoman K, Wakefield J, Ramabhadran R. 2011. NRF2 Oxidative Stress Induced by Heavy Metals is Cell Type Dependent. Curr. Chem. Genomics 5:1–12.

Spearman C. 1987. The proof and measurement of association between two things. By C. Spearman, 1904. Am. J. Psychol. 100: 441–471.

Stillwell WG, Turesky RJ, Sinha R, Skipper PL, Tannenbaum SR. 1999. Biomonitoring of heterocyclic aromatic amine metabolites in human urine. Cancer Lett. 143: 145–8.

Tai WW. 2012. Chapter 117. Nonsteroidal Anti-Inflammatory Drugs. In *Poisoning & Drug Overdose, 6e.* (K.R. Olsoned.), The McGraw-Hill Companies, New York, NY.

Thomas RS, Black MB, Li L, Healy E, Chu T-M, Bao W, et al. 2012. A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. Toxicol. Sci. 128:398–417.

Thummel KE, Shen DD, Isoherranen N. 2011. Appendix II. Design and Optimization of Dosage Regimens: Pharmacokinetic Data. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e* (L.L. Brunton, B.A. Chabner, and B.C. Knollmanneds.), The McGraw-Hill Companies, New York, NY.

Tian S, Li Y, Wang J, Zhang J, Hou T. 2011. ADME evaluation in drug discovery. 9. Prediction of oral bioavailability in humans based on molecular properties and structural fingerprints. Mol. Pharm. 8:841–51.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: a Tox21 update. Environ. Health Perspect. 121:756–65.

Tontonoz P, Graves RA, Budavari AI, Erdjument-Bromage H, Lui M, Hu E, et al. 1994. Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two nuclear hormone receptors, PPAR7 and RXRa. Nucleic Acids Res. 22:5628–5634.

Tropsha A, Golbraikh A. 2007. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. Curr. Pharm. Des. 13:3494–3504.

Utvik TR, Johnsen S. 1999. Bioavailability of polycyclic aromatic hydrocarbons in the North Sea. Environ. Sci. Technol. 33: 1963–1969.

Vapnik VN. 2000. *The Nature of Statistical Learning Theory*. Springer New York, New York, NY.

Varma DR. 1987. Epidemiological and experimental studies on the effects of methyl isocyanate on the course of pregnancy. Environ. Health Perspect. 72:153–157.

Varma MVS, Obach RS, Rotter C, Miller HR, Chang G, Steyn SJ, et al. 2010. Physicochemical space for optimum oral bioavailability: contribution of human intestinal absorption and first-pass elimination. J. Med. Chem. 53:1098–108.

Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. 2002. Molecular properties that influence the oral bioavailability of drug candidates. J. Med. Chem. 45:2615–2623.

Venugopal R, Jaiswal AK. 1998. Nrf2 and Nrf1 in association with Jun proteins regulate antioxidant response element-mediated expression and coordinated induction of genes encoding detoxifying enzymes. Oncogene 17:3145–56.

Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S, et al. 2006. QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation. J. Med. Chem. 49:7169–81.

Wang Y, Suzek T, Zhang J, Wang J, He S, Cheng T, et al. 2014. PubChem BioAssay: 2014 update. Nucleic Acids Res. 42:D1075–82.

Weininger D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28:31–36.

Willett P, Barnard JJM, Downs GMG. 1998. Chemical Similarity Searching. J. Chem. Inf. Model. 38:983–996.

Worth AP. 2010. 13: The Role of QSAR Methodology in the Regulatory Assessment of Chemicals. In *Media*, Vol. 10 of, pp. 367–382.

Wu C, Decker ER, Blok N, Li J, Bourgoyne a. R, Bui H, et al. 2001. Acyl substitution at the ortho position of anilides enhances oral bioavailability of thiophene sulfonamides: TBC3214, an ETa selective endothelin antagonist1. J. Med. Chem. 44:1211–1216.

Zhang D, Hannink M. 2003. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative. Mol. Cell. Biol. 23:8137–8151.

Zhang J, Hsieh J-H, Zhu H. 2014. Profiling animal toxicants by automatically mining public bioassay data: a big data approach for computational toxicology. PLoS One 9:e99863.

Zheng W, Tropsha A. 2000. Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. J. Chem. Inf. Comput. Sci. 40:185–94.

Zhu H, Kim M, Zhang L, Sedykh A. 2014a. CHAPTER 5. Computers Instead of Cells: Computational Modeling of Chemical Toxicity. In *Reducing, Refining and Replacing the Use of Animals in Toxicity Testing* (D. Allen and M.D. Waterseds.), pp. 163–182, The Royal Society of Chemistry, Cambridge.

Zhu H, Martin TM, Ye L, Sedykh A, Young DM, Tropsha A. 2009. Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. Chem. Res. Toxicol. 22:1913–1921.

Zhu H, Rusyn I, Richard A, Tropsha A. 2008. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. Environ. Health Perspect. 116:506–13.

Zhu H, Zhang J, Kim MT, Boison A, Sedykh A, Moran K. 2014b. Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. Chem. Res. Toxicol. 27:1643–51.

Zhu J, Wang J, Yu H, Li Y, Hou T. 2011. Recent developments of in silico predictions of oral bioavailability. Comb. Chem. High Throughput Screen. 14:362–374.

Zhu X, Kruhlak NL. 2014. Construction and analysis of a human hepatotoxicity database suitable for QSAR modeling using post-market safety data. Toxicology 321:62–72.