Advance Methodologies in Linear and Nonlinear Quantitative Structure-Activity Relationships (QSARs): from Drug Design to In Silico Toxicology Applications

Direttore della Scuola: Prof. Maurizio Casarin
Coordinatore d'Indirizzo: Prof.ssa Adriana Chilin
Supervisore: Prof. Stefano Moro

Dottoranda: Lisa Michielan

31 Gennaio 2010
Abstract

Novel computational strategies are continuously being demanded by the pharmaceutical industry to assist, improve, and speed up the drug discovery process. In this scenario, chemoinformatics provide reliable mathematical tools to derive quantitative structure-activity relationships (QSARs), able to describe the correlation between molecular descriptors and various experimental profiles of the compounds. In the last years, nonlinear machine learning approaches have demonstrated a noteworthy predictive capability in several QSAR applications, confirming their superiority over the traditional linear methodologies. Particularly, the feasibility of the classification approach has been highlighted in solving complex tasks. Moreover, the introduction of the autocorrelation concept in chemistry allows the structural comparison of the molecules by using a vectorial fixed-length representation to serve as effective molecular descriptor.

In the present thesis, we have deeply investigated the wide applicability and the potentialities of nonlinear QSAR strategies, especially in combination with autocorrelation molecular electrostatic potential descriptors projected on the molecular surface. Our intent is arranged in six different case studies that focus on crucial problems in pharmacodynamics, pharmacokinetics, and toxicity fields.

The first case study considers the estimation of a physicochemical property, the aqueous solvation free energy, that strictly relates to the pharmacokinetic profile and toxicity of chemicals.

Our discussion on pharmacodynamics deals with the prediction of potency and selectivity of human adenosine receptor antagonists (hAR). The adenosine receptor family belongs to GPCR (G protein-coupled receptors) family A, including four different subtypes, referred to as A1, A2A, A2B, and A3, which are widely distributed in the tissues. They differentiate for both pharmacological profile and effector coupling. Intensive explorative synthesis and pharmacological evaluation are aimed at discovering potent and selective ligands for each adenosine receptor subtype.

In the present thesis, we have considered several pyrazolo-triazolo-pyrimidine and xanthine derivatives, studied as promising adenosine receptor antagonists. Then, a second case study focuses on the comparison and the parallel applicability of linear and nonlinear models to predict the binding affinity of human adenosine receptor A2A antagonists and find a consensus in the prediction results. The following studies evaluate the prediction of both selectivity and binding affinity to A2A and A3 subtypes by combining classification and regression strategies, to finally investigate
the full adenosine receptor potency spectrum and human adenosine receptor subtypes selectivity profile by applying a multilabel classification approach.

In the field of pharmacokinetics, and more specifically in metabolism prediction, the use of multi- and single-label classification strategies is involved to analyze the isoform specificity of cytochrome P450 substrates. The results lead to the identification of the appropriate methodology to interpret the real metabolism information, characterized by xenobiotics potentially transformed by multiple cytochrome P450 isoforms.

As final case study, we present a computational toxicology investigation. The recent regulatory initiatives due to REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) require the ecotoxicological and risk assessment of chemicals for safety. Most of the current evaluation protocols are based on costly animal experiments. So, chemoinformatic tools are heartily recommended to facilitate the toxicity characterization of chemical substances. We describe a novel integrated strategy to predict the acute aquatic toxicity through the combination of both toxicokinetic and toxicodynamic behaviors of chemicals, by using a machine learning classification method. The goal is to assign chemicals to different levels of acute aquatic toxicity, providing an appropriate answer to the new regulatory requirements. As preliminary validation of our approach, two toxicokinetic and toxicodynamic models have been applied in series to inspect both aquatic toxicity hazard and mode of action of a set of chemical substances with unknown or uncertain toxicodynamic information, assessing the potential ecological risk and the toxic mechanism.
Riassunto

Nuove strategie computazionali vengono continuamente richieste dall’industria farmaceutica per assistere, migliorare e velocizzare il processo di scoperta dei farmaci. In questo scenario la chemoinformatica fornisce affidabili strumenti matematici per ottenere relazioni quantitative struttura-attività (QSAR), in grado di descrivere la correlazione tra descrittori molecolari e vari profili sperimentali dei composti. Negli ultimi anni approcci non lineari di machine learning hanno dimostrato una notevole capacità predittiva in diverse applicazioni QSAR, confermando la loro superiorità sulle tradizionali metodologie lineari. È stata evidenziata particolarmente la praticabilità dell’approccio di classificazione nel risolvere compiti complessi.

Inoltre, l’introduzione del concetto di autocorrelazione in chimica permette il confronto strutturale delle molecole attraverso l’uso di una rappresentazione vettoriale di lunghezza fissa che serve da efficace descrittore molecolare.

Nella presente tesi abbiamo studiato approfonditamente l’ampia applicabilità e le potenzialità delle strategie QSAR non lineari, soprattutto in combinazione con i descrittori autocorrelati potenziale elettrostatico molecolare proiettato sulla superficie molecolare. Il nostro intento si articola in sei differenti casi studio, che si concentrano su problemi cruciali nei campi della farmacodinamica, farmacocinetica e tossicologia.

Il primo caso studio considera la valutazione di una proprietà fisico- chimica, l’energia libera di solvazione acquosa, che è strettamente connessa con il profilo farmacocinetico e la tossicità dei composti chimici.

La nostra discussione in farmacodinamica riguarda la predizione di potenza e selettività di antagonisti del recettore adenosinico umano (hAR). La famiglia del recettore adenosinico appartiene alla famiglia A di GPCR (recettori accoppiati a proteine G), che include quattro diversi sottotipi, cui si riferisce come A$_1$, A$_2$A, A$_2$B e A$_3$, ampiamente distribuiti nei tessuti. Si differenziano sia per profilo farmacologico che per effettore cui sono accoppiati. Le intense sintesi esplorative e valutazione farmacologica hanno lo scopo di scoprire ligandi potenti e selettivi per ogni sottotipo del recettore adenosinico. Nella presente tesi abbiamo considerato diversi derivati pirazolo-triazolo-pirimidinici e xanthenici, studiati come promettenti antagonisti del recettore adenosinico. Quindi, un secondo caso studio si focalizza sul confronto e l’applicabilità in parallelo di modelli lineari e non lineari per predire l’affinità di legame di antagonisti del recettore adenosinico A$_2$A umano e trovare un consenso nei risultati di predizione. Gli studi successivi valutano la predizione
sia della selettività che dell'affinità di legame ai sottotipi $A_2A$ e $A_3R$ combinando strategie di classificazione e regressione, per studiare infine il completo spettro di potenza del recettore adenosinico e il profilo di selettività per i sottotipi hAR mediante l'applicazione di un approccio di classificazione multilabel.

Nel campo della farmacocinetica, e più specificamente nella predizione del metabolismo, è coinvolto l'uso di strategie di classificazione multi- e single-label per analizzare la specificità di isoforma di substrati del citocromo P450. I risultati conducono all'identificazione della metodologia appropriata per interpretare la reale informazione sul metabolismo, caratterizzata da xenobiotici potenzialmente trasformati da multiple isoforme del citocromo P450.

Come caso studio finale, presentiamo un'indagine in tossicologia computazionale. Le recenti iniziative regolatorie dovute al REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) richiedono l'accertamento ecotossicologico e del rischio dei composti chimici per la sicurezza. La maggior parte dei correnti protocolli di valutazione è basata su costosi esperimenti animali. Così, gli strumenti chimoinformatici sono caldamente raccomandati per facilitare la caratterizzazione della tossicità di sostanze chimiche. Noi descriviamo una nuova strategia integrata per predire la tossicità acquatica acuta attraverso la combinazione di entrambi i comportamenti tossicocinetico e tossicodinamico dei composti chimici, utilizzando un metodo di classificazione machine learning. L'obiettivo è assegnare i composti chimici a diversi livelli di tossicità acquatica acuta, fornendo un'appropriata risposta alle nuove esigenze regolatorie. Come validazione preliminare del nostro approccio, due modelli tossicocinetico e tossicodinamico sono stati applicati in serie per esaminare sia il rischio di tossicità acquatica che il modo d'azione di un set di sostanze chimiche con informazione tossicodinamica sconosciuta o incerta, valutandone il potenziale rischio ecologico ed il meccanismo tossico.
List of Papers


4 Parallel application of linear and nonlinear QSAR methodologies 55
4.1 Introduction ......................................................... 56
4.2 Human A_{2}A receptors dataset ................................. 56
4.3 Results and discussion .............................................. 59
  4.3.1 PLS and RSA models ........................................ 59
  4.3.2 External test set prediction .................................. 63
4.4 Final remarks ......................................................... 65

5 Isoform specificity of cytochrome P450 substrates 67
5.1 Cytochrome P450 in drug metabolism ......................... 68
5.2 Computational approaches to CYP450 ......................... 69
5.3 Dataset ................................................................. 71
  5.3.1 Data set 1 ...................................................... 72
  5.3.2 Data set 2 ...................................................... 72
5.4 Results ................................................................. 73
  5.4.1 Multilabel classification with Data set 1 ................. 73
  5.4.2 Single-label classification with Data set 2 ............... 77
  5.4.3 Validation of the models with an external test set ...... 79
5.5 Discussion ............................................................... 81
  5.5.1 Aspects related to the data set ............................ 81
  5.5.2 Considerations on the selected descriptors .............. 81
  5.5.3 ct-SVM and CPG NN models ................................. 82
  5.5.4 ct-SVM and SVM models ..................................... 82
  5.5.5 Analysis of some classified compounds ................... 83
5.6 Final remarks .......................................................... 85

6 Classification and regression to investigate selectivity and
  binding affinity ......................................................... 87
6.1 Introduction ............................................................ 88
6.2 Dataset ................................................................. 89
6.3 Results and discussion .............................................. 91
  6.3.1 SVM classification model .................................. 93
  6.3.2 SVM regression models ...................................... 94
  6.3.3 Validation of in series SVMclass and SVR models .... 96
6.4 Final remarks .......................................................... 100

7 Exploring potency and selectivity of hAR antagonists 101
7.1 Introduction ............................................................ 102
7.2 Adenosine receptor antagonists ................................ 104
7.3 Dataset ................................................................. 105
7.4 Results and discussion .............................................. 108
  7.4.1 autoMEP/ct-SVM classification models ................. 109
8 Prediction of toxicodynamic and toxicokinetic properties 117
  8.1 Introduction ............................................. 118
  8.2 Dataset .................................................. 121
  8.3 Results and discussion ................................. 123
    8.3.1 TOX* class model .................................. 124
    8.3.2 MOA* class model ................................. 126
    8.3.3 Applicability of TOX* class and MOA* class models .. 128
  8.4 Final remarks ........................................... 131

9 General conclusions ......................................... 133

Bibliography .................................................. 135

Supporting Information ....................................... 154
List of Abbreviations

ADMET  Absorption, Distribution, Metabolism, Elimination, Toxicity
ANN    Artificial Neural Networks
AR     Adenosine Receptor
autoMEP Autocorrelation Molecular Electrostatic Potential
CoMFA  Comparative Molecular Field Analysis
CPG NN Counter-Propagation Neural Network
cv-SVM Cross-training with Support Vector Machine
CYP450 Cytochrome P450
DOE    Design of Experiment
EPAFHM Environmental Protection Agency Fathead Minnow Acute Toxicity
GPCR   G Protein-Coupled Receptor
HOMO   Highest Occupied Molecular Orbital
LUMO   Lowest Unoccupied Molecular Orbital
MCC    Matthews Correlation Coefficient
MEP    Molecular Electrostatic Potential
MLR    Multiple Linear Regression
MOA    Mode of Action
OECD   Organization of Economic Cooperation and Development
PEOE   Partial Equalization of Orbital Electronegativity
PC     Principal Component
PCA    Principal Component Analysis
PCR    Principal Component Regression
PLS    Projection to Latent Structures by means of Partial Least Square
QSAR   Quantitative Structure-Activity Relationship
QSPR   Quantitative Structure-Property Relationship
RBF    Radial Basis Function
REACH Registration, Evaluation, Authorization and Restriction of Chemicals
RSA    Response Surface Analysis
SOM    Self Organizing Map
SVM    Support Vector Machine
SVR    Support Vector Regression
Preface

Chemoinformatics can provide both description and understanding of various pharmacodynamic, pharmacokinetic and toxicity properties of the compounds by quantitative structure-activity relationships (QSARs). In such applications machine learning methods represent valuable mathematical tools able to solve complex tasks, overcoming the potentialities offered by the classical linear QSAR strategies. The six case studies presented in this thesis constitute interesting examples of the use of both regression and classification models in combination with particular 3D autocorrelated molecular descriptors, to predict different experimental properties.

The first chapter comprises an introductory overview on the modern critical aspects of pharmaceutical research. In this context the recent driving forces due to the international regulatory initiatives are also introduced.

The second chapter focuses on the basic concepts of QSAR analysis and includes the techniques applied in the six case studies, that are described in the following chapters. The details of the calculations are also reported.

The third chapter summarizes the results of the application of a nonlinear QSAR strategy, by using Response Surface Analysis, to predict the aqueous solvation free energy of organic compounds (Paper I).

The fourth chapter represents an evaluation of the use of linear and nonlinear approaches in parallel, to obtain a consensus in the prediction of the binding affinity of human A$_{2A}$ adenosine receptor antagonists (Paper II).

The fifth chapter analyzes the isoform specificity of cytochrome P450 substrates by comparing the multi- with the single-label classification methods, to find the best model able to interpret an important phase of the metabolism (Paper III).

The sixth chapter presents an introduction to the prediction of the receptor subtype selectivity task. In more detail, we consider A$_{2A}$ and A$_{3}$
adenosine receptor antagonists to derive an integrated strategy based on a first classifier and a second regression model by applying Support Vector Machine. The goal is to simultaneously discriminate $A_{2A}R$ versus $A_{3}R$ antagonists and to predict the binding affinity to the corresponding receptor subtype for unknown compounds used as test set (Paper IV).

The seventh chapter deeply discusses the selectivity of human adenosine receptor antagonists by extending the case study reported in the sixth chapter to the whole adenosine receptor family ($A_{1}$, $A_{2A}$, $A_{2B}$ and $A_{3}$ subtypes). More specifically, we present a novel application of the multilabel classification approach. After introducing three classifiers, based on decreasing thresholds of potency, both potency profile and selectivity are predicted by applying the classification models as in series quantitative sieves (Paper V).

The eight chapter focuses on the estimation of ecotoxicological endpoints and investigates the classification approach as alternative tool to predict toxicokinetic and toxicodynamic properties of chemicals. In particular, a first model is derived to assign chemicals to different levels of acute aquatic toxicity; a second classifier provides the prediction of the mode of action (MOA) of toxic compounds (Paper VI).

In light of these investigations, we have draft the final conclusions, that emphasize the appreciable performances of nonlinear QSAR techniques to predict several pharmacodynamic, pharmacokinetic and toxicity profiles.
Chapter 1

Introduction

1.1 Challenges in pharmaceutical research

Drug discovery process is aimed at bringing to market new therapeutic agents with desirable pharmacodynamic profile and favourable ADMET (Absorption, Distribution, Metabolism, Elimination and Toxicity) properties. The target selectivity is a further crucial requirement for drugs to avoid efficacy problems and limiting side-effects, incurring when the compounds do not differentiate between different receptors. The goal is to design drugs without, or with minimum, side-effects while retaining the desired function.

Nowadays, the pharmaceutical research has to face many obstacles with the result of a very low success rate, regardless the extremely growing employed resources. [1] According to recent Tufts Center for the Study of Drug Development data, drug development, starting from the clinical trials to the final approval, is about 8.5 years long with a cost exceeding $40 billion, and only 21.5% of clinical success rate. In fact, the pharmaceutical research is actually involved in the study of more complex diseases, while the increasing size and costs of the clinical trials, the candidate high attrition rates and the late occurrence of failures in the clinical studies are emphasized as the main negative contributions to the economic profile of the research in the pharmaceutical companies. Various aspects have been identified and reported as the causes of the high level of attrition undergone by the compounds during the developmental stages. [2] The reasons for attrition have changed over time and in 2000 some problems of efficacy, safety or toxicological effects were recognized as highly responsible for the failures, covering more than 50% of the
1.1 Challenges in pharmaceutical research

causes for abandoning, as shown in Figure 1.1.1. Clearly, most of the efforts are directed to unproductive clinical trials, since most drug candidates are eliminated late in the clinical development without recovering the starting investment. [2]

![Figure 1.1.1: Reasons for attrition (2000). [2]](image)

In the last decades the potentialities of the combinatorial chemistry have provided new large databases with unknown compounds. Therefore, at the early stage of drug discovery suitable computational approaches are needed to shorten the time and increase the success rate by deriving *in silico* models for the prediction of some corresponding desirable properties. Then, several computational tools have been developed to eliminate lead compounds with undesirable profiles, before they enter the costly late phases of drug development, and to let compounds to proceed in the optimization step. The result would be the reduction of the attrition rate in drug discovery. [3, 4]

However, in the initial stage of drug development the optimization of the properties related to absorption, distribution, metabolism and elimination is expected as well as the study of the pharmacodynamic profile of novel chemical entities. [5, 6] Recent *in silico* methods have focused on the metabolic endpoints and the prediction of drug metabolism directly from structure represents an advanced approach integrated into expert systems. [7, 8] Moreover, the computational tools are suggested to successfully assist the *in vitro* methods for studying the human drug metabolism in order to compensate the limitation of the use of each of these approaches alone. [9]

Quantitative Structure-Activity Relationships (QSARs) or Quantitative Structure-Property Relationships (QSPRs) approaches represent probably the most robust well-known tools to mathematically analyze the correlation between the molecular properties and an experimental endpoint.
Among various algorithms available, novel nonlinear machine learning methods have been applied for the prediction of pharmacodynamic and AD-MET properties. [10-12] In more detail, many QSARs have been attempted to correlate molecular descriptors with druglikeness, activity, selectivity, toxicity and several pharmacokinetic properties, such as aqueous solubility, blood-brain barrier and human intestinal absorption, plasma protein binding, oral bioavailability or steady-state volume of distribution. [12-23]

1.2 Predictive toxicology

The regulatory frame is considered an additional obstacle in the drug discovery process, since a very accurate risk evaluation is required to assess the safety of the drug once on the marketplace. [1] Recently, the poorly efficient risk assessment process and the uncomplete information on hazard properties of chemicals has driven the need for new regulatory dispositions, that have been introduced in European Community on June 1 2007 with the chemical management system REA CH (Registration, Evaluation, Authorization and Restriction of Chemicals). [24, 25] The immediate objective of REA CH, in a relatively short time period (11 years), is to characterize the toxicological properties of a large group of substances, manufactured or imported in quantities in excess of 1 ton per year. The attempt of this regulation is the increase in the production of useful data for the decisions involving the improvement of the protection of human health and environment, through a better identification and understanding of the chemical properties hazardous to safety. Diverse expensive animal testing experiments are usually expected for in vivo toxicological data requirements, as shown in Figure 1.2.1.

![Figure 1.2.1: Classical toxicological testing procedures in the drug discovery process.](image-url)
The experimental toxicity assessment is relevant for human health. Unfortunately, the huge resource demand deals with the large amount of chemicals needed in the experiments and the cost of animals. Thus, very recently, a paradigm shift has been suggested in toxicology with a specific reference to the computational methods as reliable support in the toxicity assessment. In particular, the predictive toxicology represents an attractive tool to investigate the effects on human health and the potential ecotoxicological risk of chemical substances in the drug discovery process as well as in the environmental hazard assessment. In this context, pharmaceuticals, personal health care products, nutritional ingredients and products of the chemical industries are all potentially dangerous and need to be assessed. Then, the aim of the computational toxicology is to assist their evaluation through *in silico* models, by assigning a priority for the traditional toxicological tests and providing information about the consequences to their exposition, as graphically represented in Figure 1.2.2.1 [27]

![Figure 1.2.2: Roles of computational toxicology, that yields data predictive of results from animal toxicity studies. This discipline will allow prioritization of chemicals for further testing and can assist in prediction of risks to humans.](image)

Chapter 1

1.3 Motivation

The same introduction of REACH should speed up the risk assessment process by prioritizing compounds for traditional toxicity testing and providing information on the Exposure Scenarios (ESs) concerning the chemical safety profile. [27] In fact, REACH promotes alternative tools to collect extensive information on hazards of chemicals in order to reduce animal use in toxicology. As a consequence, several Intelligent or Integrated Testing Strategies (ITS) have been proposed as rapid, efficient approaches to obtain exposure and effects data and identify different modes of toxic action. [28, 29] Moreover, in vitro or computational methods, optimized in vivo studies, chemical categories, read-across analysis and thresholds of toxicological concern (TTCs) are admitted non-testing strategies to replace missing data or endpoints, and profitably reduce costly animal experiments. [27]

So far, powerful computational toxicology prediction systems have been developed for the exposure and hazard assessment to satisfy the new regulatory requests. [30] In drug discovery the in silico approaches, and especially machine learning methodologies, for the toxicity prediction of safety-relevant endpoints are precious contributions to early discovery of adverse drug reactions. [31, 32] A brief overview of both tools and models in computational toxicology have been considered. [33] Furthermore, a recent review about the toxicity databases available, in silico toxicology tools together with their advantages and limitations has been published. [34]

In toxicology QSARs are widely used approaches to infer the toxicological properties of compounds from their molecular structure. [35] Several studies have focused on the prediction of the environmental toxicity properties of drugs. [36] Aquatic toxicity of chemical substances is lately investigated as basic information in the hazard and environmental risk assessment. [37-41]

1.3 Motivation

Nonlinear strategies offer a useful tool by deriving quantitative structure-activity relationships for the investigation of new molecular structures with the goal to facilitate their evaluation at the early stage of drug discovery process. The present thesis aims to demonstrate the satisfactory predictive capability and the enormous potentialities of several nonlinear QSAR approaches for the prediction of properties ranging from the pharmacodynamics to the ADMET profiles. The chapters 3-8 separately explore six case stud-
ies to evaluate or compare the performances of linear and nonlinear QSAR strategies. We discuss various models by combining different descriptor sets with several algorithms and we have validated our results by introducing new compounds as test set. In more detail, we have predicted the aqueous solvation free energy as physicochemical property (chapter 3). In three case studies the pharmacodynamic property investigated is the binding affinity to the different human adenosine receptor subtypes, by focusing on potency to $A_2A$ (chapter 4) or $A_{2A}/A_3$ subtypes (chapter 6) and on selectivity (chapters 6 and 7) predictions. The chapter 8 relates to the recent debates on the limited toxicological information and propose a novel strategy to predict toxicokinetic and toxicodynamic behaviors of chemicals.

Moreover, we would like to evaluate the efficiency of the molecular descriptors selected in rationalizing the chemical structures to derive robust regression or classification models for the experimental properties in analysis. The autocorrelation seems to represent an efficient strategy to develop QSAR models for structurally different compounds. More specifically, we consider the autocorrelated descriptors encoding for molecular electrostatic potential computed on the molecular surface. Finally, while investigating some of our case studies, we present the results of the further introduction of descriptors (global, topological, quantum chemical, etc.) to interpret cytochrome P450 isoform specificity and toxicity mechanisms.
In the field of chemoinformatics, quantitative structure-activity relationships (QSAR) have demonstrated to be powerful tools in the prediction of simple chemical-physical properties as well as complex pharmacodynamic, pharmacokinetic and toxicological profiles. Several key steps are involved in any QSAR approach: data collection for the molecular structure building and the calculation of suitable molecular descriptors, data pretreatment, model generation and optimization, and finally, the statistical validation and evaluation of the model. Machine learning represents a well-known family of algorithms based on a solid statistical theory able to handle complex problems, and especially developed as modeling methods. Lately, among non-linear strategies, the machine learning methodologies have been applied as robust alternatives to the traditional linear QSAR techniques, such as the Partial Least Squares (PLS) analysis. To date, they have shown promising potentialities in many scientific studies. Very recently, the current mathematical techniques applied in QSAR approaches have been reviewed. [42]

Artificial neural networks or support vector machines have gained interesting progresses in both classification and quantitative prediction of different endpoints. Based also on the increasing availability of experimental data, these techniques have been further applied in the assessment of several pharmacokinetic properties. [43] Some valid machine learning applications in the prediction of the cytochrome P450 isoform specificity, interactions and inhibition were summarized. [44, 45] In the present thesis, the classification approach is demonstrated to be able to investigate the selectivity problem. However, the model validity has to deal with the recent regulatory context created by REACH law and some reference principles need to be satisfied.
2.1 QSAR

Most drugs attain the therapeutic activity through a specific target recognition process. In the optimization step of drug candidates, whether the information on the target is not available, the ligand-based drug design approach might be applied for the evaluation of new compounds.

Chemoinformatics, combining knowledge from different fields, is characterized by the interest on the chemical structures to extract information on the corresponding activity or properties. Among the chemoinformatic methods, the quantitative structure-activity relationships (QSAR) or quantitative structure-activity relationships (QSPR) relate molecular descriptors to the quantitative measure of a property. [46] QSAR are based on the general principle that the chemical structures can be mathematically codified as distributions of molecular properties, or molecular descriptors. Then, an appropriate statistical modeling method is used to achieve the correlation between the molecular descriptors (X variables) and the defined property (Y variable), such as biological activity, volume of distribution, toxicity, to predict the corresponding property of unknown compounds. [46-49] In QSAR analysis, a training set and a test set are selected from a starting collection of data (Figure 2.1.1).

Figure 2.1.1: Training and test set selection procedure. The X matrix contains molecular descriptors, the property data are included in the Y matrix.

The training set is used to generate the model, that is then statistically evaluated on its ability to predict the property values of a test set. Finally, the QSAR model can be applied to the prediction of the property of new chemical structures. The main processes involved in a QSAR analysis are represented in Figure 2.1.2.

Moreover, based on the nature of the relationship between the molecular descriptors and the property in analysis, linear and nonlinear strategies can be distinguished. In particular, multiple regression, principal component
analysis, projection to latent structures by means of partial least square will be discussed as linear strategies, while we will describe response surface analysis, support vector machine and artificial neural networks as nonlinear techniques.

Considering the input data, the regression and classification approaches can be defined. If the considered property is represented by continuous data, the QSAR model is referred to as regression regardless whether the relationship is linear or not; if binary data are introduced as qualitative measure of the property for the model generation, discrete classification models, or SARs, are derived with the aim to separate the compounds into different classes (Figure 2.1.3). [50]

The classification allows to assign a sample to one class (single-label) or to more classes (multilabel) to derive a qualitative prediction. Moreover, in the traditional single-label classification the classes are considered mutually exclusive; when the samples belong to multiple classes, the multilabel classification analysis seems to be more appropriate.
2.2 Molecular structure building

The database generation step is required to order the selected molecules and visualize their three-dimensional structures. In the present thesis, 3D models of all molecules in the training sets, validation sets, internal and external test sets were obtained using the 3D structure generator Corina, setting parameters to standard values. Corina is an integral part of ADRIANA Code suite. [51]

Conformer selection is a crucial step in the approaches considering 3D molecular descriptors. If the information about the possible binding mode of the compounds to the corresponding target is limited, we have decided to select the energetically most stable conformers produced by the software conformational analysis. We verified that the conformations derived by Corina are reasonably similar to the poses obtained by docking experiments. Protonation states are selected in agreement with the corresponding pKₐ at the physiological pH value (7.4 unit).

The final molecules are globally neutral, so, they can be used for the calculation of the molecular descriptors.
2.3 Molecular descriptors

In a QSAR study we need for each molecule some numerical properties, through a mathematical description of their structure. [46, 49] The molecular descriptors are numerical representations of physicochemical or topological properties. As anticipated, they can be used as independent variables in QSARs and the descriptors selection should be accurate according to the type of experimental data for achieving satisfactory modeling results.

The descriptors derive from experimental measurements, theoretical calculations or mathematical operations, and they may refer to the whole molecule or one molecular fragment. Moreover, they can be represented as scalars or vectors and they are defined according to the number of dimensions they require for the computation. Consequently, their complexity is related to the dimensionality of the molecular representation (Figure 2.3.1).

Figure 2.3.1: Dimensional levels of structural information.

Simple 1D descriptors need knowledge of only the code of a molecule and consider the presence of a particular element. More complex global molecular properties or functional-group counts require the connection table to be computed (2D descriptors). The 3D descriptors reflecting molecular shape or the distribution of a property on the molecular surface need the previous computation of the three-dimensional molecular structure.

We have selected different molecular descriptors for our analysis, as described in the following paragraphs.

1Adapted from Gasteiger, J.; Engel, T. Chemoinformatics, Wiley-VHC, 2003.
2.3 Molecular descriptors

2.3.1 Molecular Electrostatic Potential (MEP)

Autocorrelation molecular electrostatic potential (MEP) vectors have been introduced by Gasteiger and collaborators as molecular descriptors computed on the molecular surface (Figure 2.3.2a). [52] In our models MEPs derive from a classical point charge model: the electrostatic potential for each molecule is obtained by moving a unit positive point charge across the molecular surface, and it is calculated at various points \( j \) on this surface by applying the following equation:

\[
V_i = \frac{1}{4\pi\varepsilon_0} \sum_{j}^{\text{atoms}} \frac{q_i}{r_{ji}}
\]

(2.3.1)

where \( q_i \) represents the partial charge of each atom \( i \) and \( r_{ji} \) is the distance between points \( j \) and atom \( i \). Starting from the 3D model of a molecule and its partial atomic charges, the electrostatic potential is calculated for points on the molecular surface.

\( V_i = \frac{1}{4\pi\varepsilon_0} \sum_{j}^{\text{atoms}} \frac{q_i}{r_{ji}} \)

Figure 2.3.2: a) Representation of the molecular electrostatic potential; b) references for the surface calculation.

Partial atomic charges were calculated by the PEOE (Partial Equalization of Orbital Electronegativity) method and its extension to conjugated systems implemented in ADRIANA. Code. [51, 53, 54] As reference for the surface calculation, we have considered Connolly’s solvent accessible surface, obtained by moving a probe sphere on the van der Waals surface, as shown in Figure 2.3.2b\(^2\). Connolly’s solvent accessible surface with a solvent radius of 2.0 Å and van der Waals radius reduction factor\(^3\) of 1.00 have been used


\(^3\)Factor of reduction, which the van der Waals radius is multiplied for.
to project the corresponding MEP. Once the autocorrelation function has been applied, the autocorrelation vector is derived. [51, 53, 54]

### 2.3.2 Autocorrelation Vectors

The autocorrelation function transforms the constitution of a molecule into a fixed length representation. In fact, as originally computed, the MEP properties depend on the spatial orientation of the molecule and a previous alignment is needed to compare different molecular structures. The mathematical notion of autocorrelation is schematically reported in Figure 2.3.3.

![Figure 2.3.3: a) Effect of the translation of f(x) on x axis; b) geometrical meaning of the autocorrelated function F(t).](image)

Given the function $f(x)$ measuring a property in AB domain, if the inner variable $t$ is introduced, the new autocorrelated function $F(t)$ is defined as:

$$F(t) = \int_{AB} f(x)f(x+t)dx$$  \hspace{1cm} (2.3.2)

where $F(t)$ is an intrinsic descriptor, not relying on the external reference systems and on the translation of the function $f(x)$ on the $x$ axis. Conversely, the translation changes $f(x)$ values for any $x \in AB$ domain.

Firstly investigated by Moreau and Broto, this concept was introduced in chemistry to analyze the properties of different molecules without molecular superimposition. [55, 56] By formulating a topological definition, they considered that a certain property $p$ of an atom $i$ can be correlated with the corresponding property $p$ of atom $j$ and the products of $p$ values can be summed over all atom pairs having a certain topological distance $d$. 

13
Each component of the autocorrelation vector is consequently calculated as follows:

\[ A(d) = \sum_{i,j}^{N} p_i p_j \delta(d_{ij}, d) \]

where \( A(d) \) is the autocorrelation coefficient referring to atom pairs \( i,j \) at the topological distance \( d \) and \( p_i \) is the atomic property. [55, 56]

The molecular recognition processes and the physicochemical phenomena involve interactions between molecular surfaces and, therefore, representations of molecular surfaces should be appropriate to understand the diversity in the binding affinity and chemical behaviors. We are under the restriction of having to represent molecular surfaces of different size, thus the autocorrelation concept has been extended to 3D structures to achieve this goal. [52, 57, 58] Starting from the topological autocorrelation examples of Moreau and Broto, a set of randomly distributed points on the molecular surface has to be generated first. Then, all distances between the surface points are calculated and sorted into the preset intervals \( d_{lower}-d_{upper} \). Finally, the autocorrelation coefficients are computed:

\[ A(d_{lower}, d_{upper}) = \frac{1}{L} \sum_{j=i}^{N} \sum_{i=1}^{N} p_j p_i \delta(d_{ij}, d_{lower}, d_{upper}) \]  

According to equation 2.3.4, the component of the autocorrelation vector \( A(d_{lower}, d_{upper}) \), referring to the \( i,j \) distance \( d \) in the interval \( d_{lower}-d_{upper} \) is the sum of all products of the properties \( p_i \) and \( p_j \) for atoms \( i \) and \( j \). We consider \( N \) the number of atoms in the molecule and \( L \) a parameter representing the total number of distances in the interval \( d_{lower}-d_{upper} \).

The application of this concept made possible the comparison of different molecular properties, as this 3D descriptor represents a compressed expression of the distribution of the property \( p \) on the molecular surface, as shown in Figure 2.3.4.

For the calculation of the autocorrelation coefficients we have applied the default values for parameter computation, since no improving in statistical model capability was observed by changing them in various way. Default
parameters values are the following: \( d_{\text{lower}} = 1 \, \text{Å} \); \( d_{\text{upper}} = 13 \, \text{Å} \); \( L = 12 \); point density = 10 points/Å\(^2\); vdW radius reduction factor = 1.000. Consequently, we have derived 12 autocorrelation vectors per molecule, computed at the 12 \((L)\) distances in the interval from 1 to 13 Å with a step width of 1 Å. By considering the size of the molecules in our datasets, we decided that the step width of 1 Å was sufficient to describe in an accurate way the distribution of the MEP property on the molecular surface. This transformation produces a molecular descriptor which is a unique fingerprint of each molecule under consideration.

### 2.3.3 Sterimol parameters

3D topological Sterimol descriptors \((B_1, B_2, B_3, B_4\) and \(L\)) have been introduced by Verloop to consider the volumes of the molecular substituents with different geometries (Figure 2.3.5). [59]

*Figure 2.3.5:* Geometrical meaning of Sterimol parameters.

The transposition of this concept to the whole molecule produces global
2.3 Molecular descriptors

Molecular descriptors, that, being intrinsically independent on rotation and translation of the molecule, can be used together with autoMEP vectors. [60] In more detail, L is the length of the molecule and refers to the principal axis, B1 is perpendicular to x axis and it is the smallest distance from L axis to a side of the parallelepiped containing the molecule, B2-B4 have a similar geometrical meaning and they are perpendicular to B1.

2.3.4 Other molecular descriptors

Further descriptors calculated in our studies are listed in Table 2.1. These descriptors are, to a large extent, 2D and 3D molecular descriptors and reflect shape and reactivity properties.

The capability for participating in hydrogen bonding is described directly by the number of hydrogen-bonding acceptors/donors or the hydrogen-bond acceptor/donor potential, or indirectly by the number of basic nitrogen atoms and the number of acidic groups.

The highest hydrogen-bonding acceptor potential is defined as the maximum lone-pair electronegativity on an atom considering all N, O, and F atoms in a compound. The highest hydrogen-bonding donor potential is defined as the most positive charge on the hydrogen atom in the functional groups -OH, -NH, and -SH in a compound.

LogP(o/w), or log of the n-octanol/water partition coefficient, describes the partitioning equilibria. This 2D molecular descriptor was initially used in drug design to quantify lipophilicity. Various methods for the estimation of the experimental logP values have been developed and one of the well-known techniques, published by Nys and Rekker, is based on additive fragment contributions to the total molecular lipophilicity. [46] In our studies, logP(o/w) is calculated from a linear atom type model by considering a pH value of the aqueous phase such that the predominant form of the chemical is un-ionized.

As anticipated, two-dimensional topological autocorrelations derive atom properties from the structure diagram, whereas spatial autocorrelation descriptors are based on the information encoded by the 3D molecular structure. We have to consider atom identities or a property (σ-electronegativity, π-electronegativity, σ-charge or π-charge) for their computation. In both cases, the function of autocorrelation was applied to derive the autocorrelation vectors.
Table 2.1: List of descriptors used in our analysis, arranged by class.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
<th>Ref(s).</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MW</td>
<td>molecular weight</td>
<td>[61]</td>
</tr>
<tr>
<td>2</td>
<td>HAccPot</td>
<td>highest hydrogen-bond acceptor potential</td>
<td>[61]</td>
</tr>
<tr>
<td>3</td>
<td>HDonPot</td>
<td>highest hydrogen-bond donor potential</td>
<td>[61]</td>
</tr>
<tr>
<td>4</td>
<td>HAcc</td>
<td>number of hydrogen-bonding acceptors derived from the sum of nitrogen and oxygen atoms in the molecule</td>
<td>[61]</td>
</tr>
<tr>
<td>5</td>
<td>HDon</td>
<td>number of hydrogen-bonding donors derived from the sum of NH and OH groups in the molecule</td>
<td>[61]</td>
</tr>
<tr>
<td>6</td>
<td>TPSA</td>
<td>topological polar surface area</td>
<td>[62]</td>
</tr>
<tr>
<td>7</td>
<td>ASA</td>
<td>approximate surface area</td>
<td>[63]</td>
</tr>
<tr>
<td>8</td>
<td>$\alpha$</td>
<td>mean polar polarizability</td>
<td>[64-67]</td>
</tr>
<tr>
<td>9</td>
<td>$\mu$</td>
<td>molecular dipole moment</td>
<td>[68]</td>
</tr>
<tr>
<td>10</td>
<td>$\log P(o/w)$</td>
<td>log of n-octanol/water partition coefficient</td>
<td></td>
</tr>
<tr>
<td>11-12</td>
<td>$\chi^0, \chi^1$</td>
<td>connectivity $\chi$ indices</td>
<td>[69]</td>
</tr>
<tr>
<td>13-14</td>
<td>$\kappa_1, \kappa_2$</td>
<td>$\kappa$ shape indices</td>
<td>[69]</td>
</tr>
<tr>
<td>15</td>
<td>W</td>
<td>Wiener path number</td>
<td>[70]</td>
</tr>
<tr>
<td>16</td>
<td>$\chi^R$</td>
<td>Randic index</td>
<td>[68]</td>
</tr>
<tr>
<td>17</td>
<td>$D_3$</td>
<td>diameter</td>
<td>[71]</td>
</tr>
<tr>
<td>18</td>
<td>$R_3$</td>
<td>radius</td>
<td>[70]</td>
</tr>
<tr>
<td>19</td>
<td>$I_3$</td>
<td>geometric shape coefficient</td>
<td>[71, 72]</td>
</tr>
<tr>
<td>20</td>
<td>$r_2$</td>
<td>radius perpendicular to $D_3$</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>$r_3$</td>
<td>radius perpendicular to $D_3$ and $r_2$</td>
<td></td>
</tr>
<tr>
<td>22-24</td>
<td>$\lambda_1, \lambda_2, \lambda_3$</td>
<td>principal moment of inertia</td>
<td>[68]</td>
</tr>
<tr>
<td>25</td>
<td>$r_{gyr}$</td>
<td>radius of gyration</td>
<td>[73, 74]</td>
</tr>
<tr>
<td>26</td>
<td>$r_{span}$</td>
<td>radius of the smallest sphere, centered at the center of the mass which completely encloses all atoms in the molecule</td>
<td>[74]</td>
</tr>
<tr>
<td>27</td>
<td>$\epsilon$</td>
<td>molecular eccentricity</td>
<td>[68]</td>
</tr>
<tr>
<td>28</td>
<td>$\Omega$</td>
<td>molecular asphericity</td>
<td>[68]</td>
</tr>
<tr>
<td>29</td>
<td>HOMO</td>
<td>energy (eV) of Highest Occupied Molecular Orbital</td>
<td>[75]</td>
</tr>
<tr>
<td>30</td>
<td>LUMO</td>
<td>energy (eV) of Lowest Unoccupied Molecular Orbital</td>
<td>[75]</td>
</tr>
<tr>
<td>31</td>
<td>LUMO-HOMO</td>
<td>difference between LUMO and HOMO</td>
<td>[75]</td>
</tr>
<tr>
<td>32</td>
<td>$n_{aliph_amino}$</td>
<td>number of aliphatic amino groups</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>$n_{aromatic_amino}$</td>
<td>number of aromatic amino groups</td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
2.3 Molecular descriptors

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
<th>Ref(s).</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>n\textsubscript{prim_amino}</td>
<td>number of primary aliphatic amino groups</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>n\textsubscript{sec_amino}</td>
<td>number of secondary aliphatic amino groups</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>n\textsubscript{tert_amino}</td>
<td>number of tertiary aliphatic amino groups</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>n\textsubscript{prim_sec_amino}</td>
<td>n\textsubscript{prim_amino}+n\textsubscript{sec_amino}</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>n\textsubscript{aro_hydroxy}</td>
<td>number of aromatic hydroxy groups</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>n\textsubscript{aliph_hydroxy}</td>
<td>number of aliphatic hydroxy groups</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>n\textsubscript{guanidine}</td>
<td>number of guanidine groups</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>n\textsubscript{basic_nitrogen}</td>
<td>number of basic, N-containing functional groups</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>n\textsubscript{acidic_groups}</td>
<td>number of acidic functional groups</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>n\textsubscript{acylsulfonamides}</td>
<td>number of sulfonamide-C=O groups</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>n\textsubscript{enolate_groups}</td>
<td>number of enolate groups</td>
<td></td>
</tr>
</tbody>
</table>

**Vectorial (topological and spatial)**

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
<th>Ref(s).</th>
</tr>
</thead>
<tbody>
<tr>
<td>45-55</td>
<td>2D-AC\textsubscript{χ\textsuperscript{LP}}</td>
<td>property: lone-pair electronegativity χ\textsubscript{LP}</td>
<td></td>
</tr>
<tr>
<td>56-66</td>
<td>2D-AC\textsubscript{χ\textsuperscript{σ}}</td>
<td>property: σ-electronegativity χ\textsubscript{σ}</td>
<td></td>
</tr>
<tr>
<td>67-77</td>
<td>2D-AC\textsubscript{χ\textsuperscript{π}}</td>
<td>property: π-electronegativity χ\textsubscript{π}</td>
<td></td>
</tr>
<tr>
<td>78-88</td>
<td>2D-AC\textsubscript{q\textsuperscript{σ}}</td>
<td>property: σ-charge q\textsubscript{σ}</td>
<td></td>
</tr>
<tr>
<td>89-99</td>
<td>2D-AC\textsubscript{q\textsuperscript{π}}</td>
<td>property: π-charge q\textsubscript{π}</td>
<td></td>
</tr>
<tr>
<td>100-110</td>
<td>2D-AC\textsubscript{q\textsuperscript{tot}}</td>
<td>property: total charge q\textsubscript{tot}</td>
<td></td>
</tr>
<tr>
<td>111-121</td>
<td>2D-AC\textsubscript{α}</td>
<td>property: polarizability α</td>
<td></td>
</tr>
<tr>
<td>122-249</td>
<td>3D-AC\textsubscript{identity}</td>
<td>property: identity</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>χ\textsubscript{σ_1} = \sum x_i^2</td>
<td>property: σ-electronegativity χ\textsubscript{σ}</td>
<td></td>
</tr>
<tr>
<td>251</td>
<td>χ\textsubscript{π_1} = \sum y_i^2</td>
<td>property: π-electronegativity χ\textsubscript{π}</td>
<td></td>
</tr>
<tr>
<td>252</td>
<td>q\textsubscript{σ_1} = \sum p_i^2</td>
<td>property: σ-charge q\textsubscript{σ}</td>
<td></td>
</tr>
<tr>
<td>253</td>
<td>q\textsubscript{π_1} = \sum q_i^2</td>
<td>property: π-charge q\textsubscript{π}</td>
<td></td>
</tr>
<tr>
<td>254-265</td>
<td>SurfACorr_HBP</td>
<td>property: hydrogen-bonding potential</td>
<td></td>
</tr>
</tbody>
</table>

The hydrogen atoms were included before the computation of the vectorial descriptors. In more detail, each component $A(d)$ of the autocorrelation vector for the topological distance $d$ is calculated as:

$$A(d) = \frac{1}{L} \sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_j \delta(d_{i,j}, d)$$

\[ \delta = \begin{cases} 1 & \forall \ d_{i,j} = d \\ 0 & \forall \ d_{i,j} \neq d \end{cases} \] (2.3.5)

where $A(d)$ represents the autocorrelation coefficient referring to atom pairs $i,j$; $N$ is the number of atoms in the molecule; $p_i$ and $p_j$ are the properties of atoms $i$ and $j$, respectively; $d_{i,j}$ is the $i,j$ topological distance (i.e. the number of bonds corresponding to the shortest path in the structure diagram). A maximum distance $d = 10$ was selected, and 11 2D topological autocorrelation components per molecule were obtained.

The calculation of 3D autocorrelation vectors was performed as described
in the 2.3.2 section, by using $p_i = p_j = 1$ in eq. 2.3.3 for property identity vectors (descriptors 122-249 in Table 2.1). Default parameter values were $d_{\text{lower}} = 1$ Å and $d_{\text{upper}} = 13.8$ Å, with a resolution of 0.1 Å. Then, 128 components for this descriptor were obtained. The sums of the squares of the σ-electronegativity (descriptor 250 in Table 2.1), π-electronegativity (descriptor 251 in Table 2.1), σ-charge (descriptor 252 in Table 2.1), and π-charge (descriptor 253 in Table 2.1) were calculated to reflect the electronegativities and charge distributions in the aliphatic and conjugated systems. They can be obtained by calculating the first component of the autocorrelation vector while setting the distance to zero ($d_{\text{lower}} = 0$ Å, $d_{\text{upper}} = 0$ Å, and number of intervals = 1). For these descriptors, only one component ($d = 0$) of the autocorrelation coefficients has been considered. The parameters for the calculation of the hydrogen-bonding potential (descriptors 254-265 in Table 2.1) autocorrelation coefficients were as follows: $d_{\text{lower}} = 1$ Å, $d_{\text{upper}} = 13$ Å, point density = 10 points/Å$^2$, and 12 autocorrelation coefficients were obtained.

An extensive presentation of the remaining descriptors in Table 2.1 was reported in a previous work. [76]

### 2.4 Data autoscaling

Once the molecular descriptors have been computed, the data is ordered in a matrix form suitable to proceed with the statistical analysis. The independent $x$ variables (molecular descriptors) should be distinguished from the dependent $y$ variable (experimental property). However, the descriptor values might cover different intervals and show diverse distribution. Consequently, a variable with high variance might have a stronger influence than the other variables in the model development, but this effect should be avoided. Then, each variable needs to be subjected to both scaling and mean-centering procedures, in the autoscaling process, in order to return the data in an unique scale and make them homogeneous. (Figure 2.4.1). [47] The scaling of data is achieved by multiplying the elements of each variable by the corresponding standard deviation (unit variance scaling), while in the mean-centering the mean value of the corresponding variable data is dettracted from each data. Finally, the information is contained in the same interval for all variables.
2.5 Linear strategies

2.5.1 Single and multiple regression

The linear regression represents the simplest mathematical technique to derive QSAR models, when the independent variables (molecular descriptors) are correlated with the dependent ones (experimental properties) in a linear way, as reported for example in the following relationship:

\[ y = ax + c \]  

(2.5.1)

with a single dependent \( y \) variable described as function of one independent \( x \) variable. The best straight line achieved is able to approximate data distribution with the minimum root mean squared error (RMSE) by considering the predicted and experimental \( y \) values\(^4\).

If further independent \( x \) variables are introduced, the calculation becomes more complex and in this case we refer to as multiple linear regression (MLR). The data is structured into a two-matrix organization and the variables are differently weighted, according to the angular coefficient of each contribute to the new regression straight line [48]:

\[ y = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_ix_i + c \]

\[ y = \sum a_nx_n + c \]  

(2.5.2)

\(^4\)The root mean square of error is defined as: \( RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_p-y_e)^2}{n}} \) where \( y_p = y_{predicted}, y_e = y_{experimental} \) and \( n \) is the total number of observations.
The problems related to this approach are due to the same variables, that should be independent from each other as more as possible, without errors, since they influence the model performance. Finally, they should be selected according to the relevance for the considered $y$ variable. Moreover, five samples at least are needed for each selected descriptor in the analysis. Therefore, novel tools have been introduced to overcome these requirements in the regression-based linear QSAR strategies.

2.5.2 Principal Component Analysis

A widely used approach to reduce the dataset dimensionality, i.e. the number of independent variables, is the Principal Component Analysis (PCA). This technique allows to optimize the information contained in the data matrix, by finding out the most contributing independent variables and eliminating intercorrelations between molecular descriptors. [46-49] In fact, if two variables $x_1$ and $x_2$ are highly correlated, it would be redundant to consider both of them in the model generation. So, a new single variable is introduced to synthesize the information: the principal component ($PC$) is a linear combination of $x_1$ and $x_2$.

If we consider a multivariate analysis with multiple descriptors, the dependent variable can be described by new variables, the principal components, linear combinations of the input independent variables:

$$PC_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n + b$$

$$PC_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n + b$$

$$PC_i = a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{in}x_n + b = \sum a_{in}x_n + b \quad (2.5.3)$$

Each $PC$ in the $X$ space is represented by the straight line that crosses the origin and best approximates the data distribution, in order to minimize the sum of the squared distances from the straight line, i.e. when $\sum e^2$ tends to assume a value close to zero. (Figure 2.5.1). The first principal component $PC_1$ is the straight line that maximizes the variance in the data, then, the data shows the best distribution along $PC_1$. The variance that has to be further explained, is progressively justified by the remaining principal components. Then, the second component is derived orthogonally to the first one and the following $PC$s are carried out in a similar way, as far as to equal
the number of the independent variables themselves. After the analysis, the data will present a large distribution of the descriptor values, with a low degree of correlation between the newly calculated variables.

Several statistical measures with the corresponding graphical representations can be considered in PCA: variances, loadings, scores, residuals and leverage. In Figure 2.5.2 a typical trend of the explained variance by increasing the number of PCs is reported. Each PC enriches the information carried by the previous one in terms of percentage (%) of variance, consequently, the last PCs are less significant than the first ones. Moreover, the decreasing slope highlighted by the graphical representation in Figure 2.5.2, allows the detection of the number of significant principal components, which defines also the model complexity.

In the loading plot the coefficients of the descriptors for each PC are shown. In particular, each descriptor is weighted according to cosine of the
angle that each variable $x$ forms with the considered $PC$ straight line, giving the loading value $p$. Furthermore, the correlation between the descriptors and PCs might be positive or negative and have a different importance: for a $PC$, the loading value of $x$ variables can change both its sign and intensity with respect to the zero reference.

After the calculation of the principal components, new coordinates are identified for each sample by the orthogonal projection of each point on the considered $PC$. The intercepted value represents the score, according to the different principal components, which the projection refers to. The new coordinates, indicated as $t_n$, with $n$ corresponding to the considered $PC$, reflect the information of the original space in the new $PC_i$ space. The score values can be graphically reported in a 2D plane, with the axis represented by two diverse $PC$s (Figure 2.5.3).

![Figure 2.5.3: Score plot $PC_1/PC_2$, with the outlier indicated by the arrow.](image)

In this plot, the samples in the same quadrant present similar properties, moreover, according to the sample position in the projected space, we can evaluate which $PC$s are more descriptive for a particular data. The anomalous position of data in the score plot (external to the ellipse, delimiting the confidence interval, as indicated in Figure 2.5.3) might correspond to outliers.

The leverage represents the data influence on the model: a sample with high leverage value tends to carry the analysis in a particular direction. This effect is proportional to the distance of the data from the global center. By evaluating the leverage, we are able to distinguish the presence of strong outliers.

---

5Discordant data if compared with the remaining data. Their chemical structure and molecular descriptors should be investigated before their exclusion from the model.
The detection of moderate outliers is achieved by the analysis of the variance not explained by the PCs, with the meaning of Distance to the Model in the X space or $D_{ModX}$: it indicates the residual (i.e. data variation non captured by the considered PCs) for each data in the $X$ space. The data can be represented in the plane $D_{ModX}$/number of observations, as shown in the Figure 2.5.4.

![Figure 2.5.4: Geometrical interpretation of $D_{ModX}$.](image)

Finally, the principal components are further correlated to the $y$ dependent variable by deriving a new linear regression model, called Principal Component Regression (PCR), represented by the following equation:

$$y = aPC_1 + bPC_2 + \cdots + kPC_n + d$$

(2.5.4)

### 2.5.3 Projection to Latent Structures by means of Partial Least Square

In chemometrics Projection to Latent Structure by means of Partial Least Square (PLS) analysis is able to detect linear correlations between $x$ independent variables and $y$ dependent variables by introducing new variables, the principal components, defined as latent variables, since they hide the distribution of data in the input $XY$ space. [77] PLS technique is an extension of the PCA methodology, by considering the computation of the principal components both in $X$ and $Y$ spaces.

In Figure 2.5.5 a dataset constituted of $X$ matrix, with $K$ corresponding to the number of descriptors ($K = 3$), and $Y$ matrix, with $M$ referring to the number of experimental data ($M = 3$), is reported as example. The aim is to build a model for the experimental data, to find the relationship between two groups of variables. As previously described for the PCA technique, several principal components can be defined as combinations of the input
Figure 2.5.5: Meaning of the PLS analysis with respect to PCR and PCA.

variables. PCs are straight lines passing by the origin and orthogonal one to each other, able to approximate the data distribution in both X and Y spaces (Figure 2.5.6).

Figure 2.5.6: Calculation of PCs in PLS; $t_1$ and $u_1$ scores referred to $PC_1$.

Then, for each point one can identify new coordinates, obtained by the projection of each point on the principal components: $t_{i1}$ and $u_{i1}$ scores. The following inner relationship between the projections is defined to correlate the data distribution in the X and Y spaces:

$$u_{i1} = mt_{i1} + h_i$$ (2.5.5)
where \( u \) and \( t \) are the score values of each data in the \( X \) and \( Y \) matrices, respectively, and \( h \) the residuals. In this way, the information is condensed in a smaller space than the input space, by achieving the desired dimensionality reduction. The relationship between the descriptors and the \( y \) property requires the obtainment of a straight line in the plot \( u_1/t_1 \) (Figure 2.5.6); moreover, the ideal function correlating \( t \) with \( u \) is the bisector of the first and third quadrants \( (m = 1) \). To achieve this objective, a further step is represented by the oscillation of the straight line \( PC_i \) in the \( X \) space in a way that \( t \) values result as more as equal to \( u \) values, tending to the ideal linear correlation in the plot \( t/u \).

If one dependent variable is analyzed, the data can be represented by using a single dimension in the \( Y \) space with \( u \) corresponding to \( y \) (Figure 2.5.7).

![Figure 2.5.7: \( y \) and \( t_1 \) scores referred to the first PC with three independent and one dependent variables.](image)

By introducing further PCs the percentage (%) of variance explained increases, but the contribute of each one progressively decreases, as previously described for the Principal Component Analysis.

---

\(^6\)Data variation not explained by the PCs.
The considerations related to the PLS analysis are similar to the statistical definitions in the last section. In more detail, the plots corresponding to the relationship between $X$ and $Y$ spaces, and the data representation in the $Y$ space can be also considered, as for example the $X$ and $Y$ score plots, the $X$ and $Y$ loadings. The score plot $u/t$ (for the different components) reports the observations in the $X$ ($T$) and $Y$ ($U$) projected spaces and gives information on the correlation between $Y$ and $X$ spaces.

In the $y_{\text{experimental}}$ vs $y_{\text{predicted}}$ plot the prediction results are compared with the experimental data (Figure 2.5.8).

![Figure 2.5.8: $y_{\text{experimental}}$ vs $y_{\text{predicted}}$ plot.]

Ideally, the data should lay on a straight line with an angular coefficient equal to one and crossing the origin. We can refer to the score plots, the $u/t$ score plot, the leverage, which is reported for both $X$ and $Y$ spaces, and to $X$- or $Y$-variance residuals/leverage plots to determine the presence of outliers\(^7\). Moreover, we can detect the outliers by observing the $y_{\text{experimental}}$ vs $y_{\text{predicted}}$ plot, if there is data distant from the ideal distribution. In the scatter plot the loading weights are reported, i.e. the weights of each descriptor in both $X$ and $Y$ spaces for the PCs, to quantify the influence of each input $x$ and $y$ variables in the analysis. The VIP plot (Variable Importance in the Projection) shows the coefficients of each original variable in the model for all the PCs.

The final mathematical model is able to summarize the variance in the data by introducing a number of latent variables (the new principal components) lower than the input variables (the molecular descriptors), but maintaining the input information.

\(^7\)The outliers are data with high leverage and residual values.
2.6 Nonlinear strategies

2.6.1 Response Surface Analysis

Response Surface Analysis (RSA) is widely applied in the Design of Experiments (DOE) approach, useful to solve optimization problems in research and development. [78] A Design of Experiment is a structured methodology consisting of a sequence of experimental determinations to describe the relationship between parameters $x_n$ involved in a process and the corresponding response variables $y_n$, as illustrated in Figure 2.6.1.

![Diagram](attachment:image.png)

**Figure 2.6.1:** Simplified graphical representation of Design of Experiments (DOE) approach. $x$ parameters (experimental factors) are related to $y$ variables (observed responses).

The objective of DOE is to design or improve a process/product of interest. During the experiments all relevant parameters are changed systematically, in order to find the ideal conditions. For example, we can consider the study of a formulation, in which different ingredients are mixed together. After the analysis of the results (formulation parameters), the optimal conditions and the role of the input parameters (ingredients factors) to determine the experimental outcome can be identified.

Several steps are involved in the DOE approach: a) define the objective to the study; b) select the design parameters to be varied during the experiment and their intervals; c) identify the response variables that will be measured; d) perform the experiments and collect the data. Only a small set of experiments is needed to be performed for studying a process under various conditions, if the computational methods are introduced.
In particular, Response Surface Analysis (RSA) refers to a collection of mathematical and statistical techniques, applied for analyzing the influence of the independent variables (input parameters) on the response(s). [78] Figure 2.6.2 graphically represents how one can derive the generic relationships between the response variables $y_1, y_2$, and the input variables $x_1, x_2$. The response values, corresponding to different $(x_1, x_2)$ pairs, yield as a surface lying above the $X$ plane. Moreover, we can derive the best $x_1, x_2$ pairs corresponding to the maximum $y$ values to achieve the optimal compromise solution for both $y_1$ and $y_2$ responses.

![Flowchart of an optimum design experiment by applying Response Surface Analysis (RSA).](image)

**Figure 2.6.2**: Flowchart of an optimum design experiment by applying Response Surface Analysis (RSA). Starting from experimental samples, each response surface is then generated for $x_1, x_2$ pairs. The ideal conditions are separately identified for the responses, to finally search for the best compromise solutions.

In most problems the relationship $f()$ between the response variable and the independent variables is unknown. Thus, RSA technique is aimed at
approximating the function $y = f(x_1, x_2, x_3, \ldots, x_n)$, where $f(x)$ is a first-order or a second-order equation. The approximation usually employs a low-order polynomial in some region of the $X$ space. If the response is well-modeled by a linear function of the $Y$ variables, then the approximating function is a first-order model, as the following:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2.$$  \hfill (2.6.1)

If a curvature is present in the system or in the region of the optimum, a polynomial of higher degree, corresponding to a nonlinear model, should be used to approximate the response, which is analyzed to locate the optimum, i.e. the set of independent variables such that the partial derivatives of the model response with respect to the individual independent variables is equal to zero. For the second-order equation the addition of a parameter of interaction between the independent variables $x_1 x_2$ is required to introduce a curvature in the response function:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{11} x_1^2 + \alpha_{22} x_2^2 + \alpha_{12} x_1 x_2.$$  \hfill (2.6.2)

Then, a more general formulation of the second-order response function is:

$$y = \alpha_0 + \sum \alpha_i x_i + \sum \alpha_{ii} x_i^2 + \sum \sum \alpha_{ji} x_j x_i.$$  \hfill (2.6.3)

The eventual objective of RSA is to determine the optimum operating conditions for the system, or a region which satisfies the operating specifications. Almost all RSA problems utilize one or both of these approximating polynomials. In the present thesis RSA is based on a multivariate thin plate spline algorithm derived by the Green’s theorem [79]:

$$y = \sum_{i=1}^{n} \alpha_i g(d_i) + \sum_{j=1}^{p} c_j x_j$$ \hfill (2.6.4)

where $\alpha_i$ and $c_j$ are the weight coefficients, $p$ the number of independent variables $x$, $n$ the number of data points and $g(d_i)$ the Green’s function applied to the Euclidean distances between the data $i$ and any coordinate in $x$ axis. According to this algorithm, the response surface function is the result of an elastic beam displacement in the $x_n$ space, where the elastic beam has to bent to reach the data points in the $y$ space. [79]
The workflow for the model development is reported in Figure 2.6.3. Input values are regarded as points of force actions, while output values as displaced values. The surface response is the result of a smoothing procedure, that reduces the influence of the background noise carried by the input data.

**Figure 2.6.3:** Schematic procedure of the thin plate spline algorithm applied in Response Surface Analysis.

In RSA technique, the selection of the most informative independent variables is performed to reduce the dimensionality of the final model and improve the model predictivity. Linear stepwise regression and nonlinear cluster analysis have been applied, and we combined the results to select the most statistically relevant independent variables. [79] The stepwise regression combines both forward selection and backward elimination processes, that progressively adds/eliminates the independent variables that are best/worst correlated to the response, respectively. The cluster analysis is a nonlinear analysis able to divide the dataset into groups, according to a similarity criterion applied to the samples. Then, similar samples belong to the same cluster, while different clusters contain diverse samples.

### 2.6.2 Support Vector Machine

Support Vector Machines (SVMs) are supervised learning systems originated from Statistical Learning Theory, recently developed by Vapnik, and characterized by novel attractive features and optimal generalization performance. [80, 81] The theory of SVM has been described in several books, and here we briefly introduce some principles. [82-87]

A supervised learning problem requires the resolution of a function approximation problem (approximation of an unknown response function), where the available data set (training set) is represented as a set of pairs (examples), \(T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i)\}\), where \(x_i\) is an input data and \(y_i\) is the corresponding observed response value. Usually, \(x_i \in \mathbb{R}^n\),
while if \( y_i \in \{-1,+1\} \) the learning problem is a binary classification, if \( y_i \in \mathbb{R}^n \) the learning problem is a regression. In both cases, the aim of the learning system is to select a hypothesis \( f(x) \) that approximates the desired response \( y_i \) in an optimal fashion, i.e. by minimizing some risk functional \( R \). In particular, we would like the function \( f(x) \) to be a reasonable estimate of the functional relation between input-output pairs (prediction or generalization property). \( R \) weights the cost of the approximation, while the error of a hypothesis is given by a loss function \( L() \) that measures the distance between \( y_i \) and \( f(x) \). A common example of loss is given by the quadratic error function:

\[
L(f(x_i), y_i) = (y_i - f(x_i))^2. \tag{2.6.5}
\]

The average error over the training set is the expected risk \( R \). If we assume that a probability distribution \( P(x_i, y_i) \) exists and it is known to govern the data and the underlying function dependences, \( R \) can be expressed as

\[
R = \int L(f(x_i), y_i) dP(x_i, y_i). \tag{2.6.6}
\]

The learning process aim at selecting the hypothesis \( f^{opt}(x) \) to minimize \( R \).

The only available information to the learning system is the training set; how the learning system uses the training set in order to minimize \( R \) is called the inductive principle, that is a general prescription for obtaining the estimate of \( f^{opt}(x) \) from the training set. Given an inductive principle, the learning algorithm tells how to use the data to obtain the estimate. One of the most popular inductive principles is Empirical Risk Minimization, which search for \( f^{opt}(x) \) by minimizing the empirical error:

\[
R_{emp} = \frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i). \tag{2.6.7}
\]

The linear SVM is based on: a) linear hypotheses corresponding to separating hyperplanes in the \( \mathbb{R}^n \) space, i.e. \( f(x) = w \cdot x + b = \sum_{i=1}^{n} w_i x_i + b \) where \( \cdot \) is the dot product between vectors; b) the solution of a quadratic optimization problem that represents a trade-off between the minimization of the empirical error, i.e. the error over the training set, and the maximization of the smoothness of \( f(x) \). [80] Nonlinear versions of SVM can be obtained by the introduction of a kernel. [88]

We will discuss these issues both for classification and regression.
Support Vector Classification

In the last years, several classification problems have been solved by using SVM approach, such as the discrimination between active and non active compounds. [89-99] Binary classification is widely performed to discriminate a set of examples in two classes. Different formulations for SVM are possible according to the loss function $L(f(x_i), y_j)$ used. We adopt the standard formulation derived by using the Hinge loss function and slack variables $\xi_i$:

$$\min_{w,b} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$  \hspace{1cm} (2.6.8)

subject to: $\forall i \in \{1, \ldots, n\}, \quad y_i (w \cdot x + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$,

where we recall that $y_i \in \{-1, +1\}$, $w$ and $b$ are the parameters that control the function $f(x)$, and the constraints are satisfied with zero error when it is possible to find a function able to classify any positive example ($y_i = +1$) by returning a positive value that has some margin from zero, i.e. $f(x) \geq 1$, and any negative example ($y_i = -1$) returning a negative value that has some margin from zero, i.e. $f(x) \leq -1$. If such function does not exist, then errors need to be compensated by choosing non zero values for the corresponding slack variables $\xi_i$. The geometrical interpretation of Support Vector Classification is shown in Figure 2.6.4.

**Figure 2.6.4:** A binary classification problem. The optimal separating hyperplane is orthogonal to the shortest line connecting the convex hulls of the two classes, and intersects it half-way between the two classes.
The trade-off between the minimization of the norm of the weight vector and the empirical error is given by the constant $C$. The above quadratic constrained minimization problem (eq. 2.6.8) can be more easily solved by resorting to the corresponding dual problem:

\[
\max_{\alpha} = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{2.6.9}
\]

subject to: $\sum_{i=1}^{n} y_i \alpha_i = 0$, and $\forall \; i \in \{1, \ldots, n\} \quad 0 \leq \alpha_i \leq C$.

The optimal weight vector $w^o$ of the first formulation is linked to the optimal solution vector $\alpha^o$ of the dual problem ($\alpha_i$ are called dual variables) by the following relation:

\[
w^o = \sum_{i=1}^{n} \alpha_i^o y_i x_i. \tag{2.6.10}
\]

The input vectors $x_i$ for which the corresponding dual variables satisfy $\alpha_i^o > 0$ are referred to as support vectors. Finally, the decision rule is given by $\text{sgn}(f(x))$. The characteristic nonlinearity of the boundary separating positive from negative samples is achieved by projecting the input vectors into a higher dimensional feature space, i.e. $x \mapsto \Phi(x)$ (Figure 2.6.5).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{kernel_transformation.png}
\caption{Transformation operated by the kernel.}
\end{figure}

In this way the dot product $x_i \cdot x_j$ is replaced by a kernel function $k(x_i, x_j)$ representing the dot product in the transformed space, i.e. $k(x_i, x_j) = \Phi(x_i)\Phi(x_j)$. The decision function takes the final form:

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b \right). \tag{2.6.11}
\]
An example of a kernel function is the Gaussian RBF (radial basis function), \( k(x_i, x_j) = e^{-\gamma(x_i - x_j)^2} \), which has demonstrated its good performances by producing a closed decision boundary. Regarding this kernel, the \( \gamma \) parameter should be appropriately selected: very small values of \( \gamma \) correspond to complex models with a high number of support vectors and risk of overfitting, while large values of \( \gamma \) might lead to a separating hyperplane described with few support vectors and too smooth for an accurate classification.

**Support Vector Regression**

In the last decade Support Vector Regression (SVR) has been used as a non-linear methodology to derive quantitative structure-activity relationships for the prediction of different chemical and biological properties. [100-106]

As anticipated, in a regression problem \( y_i \in \mathbb{R}^n \), then the mathematical formulation has to consider the approximation errors. A "reasonable" approximation is defined by introducing the constraint that for each input \( x_i \) we should have \( |y_i - f(x_i)| \leq \varepsilon \), where \( \varepsilon \) is a small positive constant representing the tolerance we allow on approximation errors. This requirement can be described by two linear constraints, i.e. \((y_i - w \cdot x_i - b) \leq \varepsilon \) and \((w \cdot x_i + b - y_i) \leq \varepsilon \). Errors above the tolerance are typically linearly penalized by resorting to the linear \( \varepsilon \)-insensitive loss function, in the following:

\[
L^\varepsilon(x, y, f) = |y - f(x)|_\varepsilon = \max(0, |y - f(x)| - \varepsilon)
\]

Based on the above considerations, the standard SVR model is defined as:

\[
\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*),
\]

subject to: \( \forall i \in \{1, \ldots, n\} \)

\[
\begin{align*}
(w \cdot x_i) + b - y_i & \leq \varepsilon + \xi_i, \\
y_i - (w \cdot x_i) + b & \leq \varepsilon + \xi_i^*, \\
\xi_i, \xi_i^* & \geq 0,
\end{align*}
\]

where a set of slack variables \( \xi_i, \xi_i^* \) is added to quantify the violation of the imposed constraints: \( \xi_i \) accounts for the underestimation of the target values, while \( \xi_i^* \) accounts for the overestimation of the target values.
The solution of a linear regression problem results to be a tube with radius $\varepsilon$ which approximates the data distribution (Figure 2.6.6). [80, 81, 88, 107] Even for SVR a kernel can be used to introduce nonlinearity. Then, the kernel expansion of the decision function $f$ is:

$$f(x, \alpha_i^*, \alpha_i) = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) k(x_i, x) + b. \quad (2.6.14)$$

The final hypothesis regression function is a weighted sum of the kernel function evaluated at the support vectors, defined as the training points located on the border of the regression tube (Figure 2.6.6).

**Figure 2.6.6:** In Support Vector Regression a tube with radius $\varepsilon$ is fitted to the data. The trade-off between model complexity and points lying outside of the tube (with positive slack variables $\xi$) is determined by minimizing eq. 2.6.13.

### 2.6.3 Cross-training with Support Vector Machine

The concept of cross-training has been introduced by Boutell and collaborators. [108] They turned from the previously less performing attempted strategies for multilabel data to present a new classification method suitable for multiple and overlapping classes tasks, with samples simultaneously associated to more than one class (Figure 2.6.7).

In the cross-training approach the multilabel data are used more than once when training the classification model. Moreover, each sample is assigned a positive label for each actual class to which it belongs. [108] Cross-training with Support Vector Machine (ct-SVM) represents a novel applica-
Figure 2.6.7: a) Single-label classification: the samples that belong to two different classes are often difficult to separate; b) multilabel classification: the data marked with both symbols belongs simultaneously to both classes.

tion of SVM analysis, when classes overlap in the feature space. In ct-SVM technique the output real-valued scores of the trained binary classifiers for each class are transformed into the final labels according to different testing criteria, as previously reported. \[109\] In more detail, \( n \) binary classifiers, \( n \) corresponding to the number of the considered target responses, were built using a radial basis function (RBF) kernel. The parameters for each SVM classifier (\( C \) and \( \gamma \)) were automatically optimized on the training set during the learning process by using a 10-fold cross-validation and predicting a small validation set. \[108\] After applying the cross-training approach, the real-valued scores were obtained.

Recently, three different testing criteria (P, T and C) have been proposed. \[108\] The P-criterion assigns to the samples all labels corresponding to a positive SVM score. If none of the scores is positive, the sample is classified as "unknown". The T-criterion uses the Closed World Assumption (CWA), according to which all samples belong to at least one class: if all SVM scores for a particular sample are negative, the pattern is assigned to the class corresponding to the less negative score. The C-criterion considers SVM scores without any sign and the decision depends on the closeness between the top SVM scores. In our studies a validation set has been used to select the closeness between two scores. Once each binary classifier was optimized by predicting a validation set, the training and the validation sets were merged and the new model was computed by using the previously optimized parameters. The final model was applied in the prediction of a test set in order to evaluate its statistical robustness.
2.6.4 Artificial Neural Networks

Artificial Neural Networks (ANN) have been introduced as a more flexible class of modeling techniques naturally able to deal with complex nonlinear systems both in classification and regression problems. Their architecture is particularly suitable in the studies with a large number of observations. The innovative potentialities and applicability of the neural network methodology in drug discovery have been recently described. [110]

The neural networks algorithm is able to model the functionality of the brain. In the biological neuron, the dendrites are fibers connecting each neuron to the neighboring neurons (Figure 2.6.8a). [46] A neuron receives the new information from the neighboring neurons and converts it to the final single signal in the soma. The signal, if strong enough (higher than a particular threshold value), is transmitted to the axon. Then, the axon carries the information further to the other dendrites and the transmission takes place at the level of the synapse. The aim of the learning process is to build the synapse strength.

Figure 2.6.8: a) Schematic structure of a biological neuron; b) unit of an artificial neuron.

Artificial neural network is an interconnected feed-forward network modeling consisting of interrelated neurons, which interchange signals. The feed-forward neural network depends on the input layer, where the units receive the input data from the previous layer, so the information is processed unidirectionally. It is formed by neurons (units) that process the information and cooperate, simulating the connections of biological neurons. [46, 111]

In fact, an artificial neuron (unit) behaves similarly to the biological
neuron (Figure 2.6.8b): the neuron $j$ receives the input signals $x_i$, that are multiplied by $w_{ji}$ (weights) and summed, obtaining the global signal $Net_j$, as in the following equation:

$$Net_j = \sum_i w_{ji}x_i$$  \hspace{1cm} (2.6.15)

where $w_{ji}$ are the weights codified by a vector, and establish the connection strength. The final signal $Net_j$ is filtered and modified by a transfer function, deciding whether the signal can be transmitted as $out_j$ to other neurons. The most common activation functions are linear, such as the sigmoidal transfer function: it can assume zero or one values as indicated by the relation:

$$out_j = \frac{1}{1 + e^{-\alpha Net_j + \vartheta}}.$$  \hspace{1cm} (2.6.16)

In ANN the units form a network and the network is structured in different interconnected layers: input layer, one or more hidden layers, with not accessible output, and output layer. [46] The neurons in the same layer receive the signal from the layer above and simultaneously produce a unique set of outputs. A typical network is shown in Figure 2.6.9.

Two different learning approaches are available: unsupervised and supervised. In the unsupervised learning the network classifies the input vectors, according to their similarity and the property to be analyzed is not used in the training process. Consequently, the data in the same neuron or in topologically adjacent neurons are similar, since similar data tends to form clusters. An example of unsupervised neural network is the Kohonen net-
work. [46, 111] The *supervised* learning is aimed at assigning the response signal to the input data. In the learning process the input data are sent to the input layer, then to the nodes of the hidden layer, and finally a response is elaborated in the output layer. In more detail, the input vectors are introduced to characterize the objects, and the output vectors correspond to the property of the object to investigate.

During the learning process (training) the neural networks learn from examples, so the connection between neurons are adapted, i.e. the weights are adjusted. The input data \( x_i \) enters the network system to generate the output \( \text{out}_i \), as previously described, and \( \text{out}_i \) is compared with the target value, yielding the error \( \delta \). Then, the weights are corrected to reduce \( \delta \) during more learning cycles (epochs), in which input data are processed in the network, as summarized in Figure 2.6.10. The supervised learning is applied, for example, in counter-propagation neural networks, where the error is back-propagated from the output layer to the previous hidden layers.

\[ \begin{align*}
\text{System} & \quad \delta \\
\text{Comparison} & \quad \text{error} \\
\text{Out} & \quad \text{Target}
\end{align*} \]

*Figure 2.6.10: Schematic representation of the supervised learning.*

The Kohonen Network or self-organizing map (SOM) is an example of recurrent associative neural network, where each layer feeds the input units back into its own units and the information is transmitted dynamically. It is able to project the samples from a multidimensional space into a two-dimensional plane, retaining the input information (Figure 2.6.11)[8]. [46] In the network architecture each column in this two-dimensional system represents a neuron; each cuboide in a column corresponds to a dimension of the input data (molecular descriptor) and it is also associated to a weight. At the beginning of the training process the weights are random numbers;

---

when an input vector enters the network, the weights more similar to that neuron are determined according to the Euclidean distance \( \sum_{i=1}^{m} (x_i - w_{ji})^2 \) calculated between the input vector \( x_s \) and the weights \( w_j \). The winning neuron is associated to the minimum Euclidean distance, according to:

\[
\min ||X(t) - W_j|| = \min \left\{ \sum_{i=1}^{m} [x_i(t) - w_{ji}]^2 \right\} \quad (2.6.17)
\]

where the index \( j \) refers to a particular neuron, \( n \) is the total number of neurons, \( W_j \) are the weight vectors and \( X(t) \) are the input vectors. Based on their distance from the winning neuron, the weights of the other neurons are adapted. The process is repeated for all remaining input data, until a training epoch is completed.

We have applied the supervised learning in a classification task by using the counter-propagation neural network methodology. A counter-propagation neural network (CPG NN) is a well-known extension of Kohonen self organizing maps analysis, where some output layers (output block), corresponding to the classes (\( Y \) variables), are added to the Kohonen input layers, that represent the molecular descriptors (\( X \) variables) (Figure 2.6.12). Figure 2.6.12 refers to a classification problem with four classes; in the output layers, each vector component is one for positive examples or zero for negative examples, according to the assigned classes. During the learning process only the input layers are considered to determine the winning neuron. Then all the weights, including the output layers, are adapted and the trained network can be used to predict unknown property vectors. Different topologies (rectangular and toroidal) and map sizes were used in the CPG NN analysis.

---

2.7 Validation and statistical evaluation

For regulatory purposes some reference principles, introduced by Organization of Economic Cooperation and Development (OECD), have been recommended in the QSAR model development. [112] The guideline document has underlined the following requirements: 

a) a defined endpoint, 
b) an unambiguous algorithm, 
c) a defined domain of applicability, 
d) appropriate measures of goodness-of-fit, robustness and predictivity and 
e) a mechanistic interpretation, if possible. Some of these principles and their role in the regulatory context have been clarified. [113]

We have used several methods to evaluate the statistical reliability of our models. Their predictive power was verified by performing a validation procedure: the internal validation was applied by excluding the samples composing the training set (cross-validation), the external validation has considered the prediction of the samples not used in the model generation (test set prediction). In particular, LOO (leave-one-out), 10-, 5-, 3- and 2-fold cross-validation procedures were performed.

In n-fold cross-validation procedure, the data set is divided randomly into n subsets with a similar number of samples and class distribution (in the classification approach), according to a stratified methodology. In the first step, one partition n is considered as test set, while the others (n-1) partitions are used to fit the model, used then to predict the test set. The process is repeated n times, until all the partitions are considered as test set.
Concerning the regression models, the correlation coefficient $r$ is calculated to evaluate the quality of the fitting process (model calibration), as follows:

$$r = \frac{\sum_{i=1}^{n}(x_{Exp} - \bar{x})(y_{Pred} - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_{Exp} - \bar{x})^2} \sqrt{\sum_{i=1}^{n}(y_{Pred} - \bar{y})^2}}$$  \hspace{1cm} (2.7.1)

where $x_{Exp}$ represents the experimental data, $y_{Pred}$ the predicted values and $\bar{x}$, $\bar{y}$ the corresponding averages. The values of the correlation coefficient are included in the interval between zero and one; $r$ indicates the ideal correlation when it is equal to one. The same parameter has been calculated after LOO cross-validation procedure ($r_{cv}$). The following statistical requirements should be satisfied to achieve a good modeling performance:

$$r_{cv}^2 > 0.3 \text{ and } r^2 / r_{cv}^2 \sim 1$$

A further validation procedure in the regression analysis is the $y$ variable randomization, which should give bad performing models.

In the single-label classification approach, following the OECD principles, we have applied an extensive $n$-fold cross-validation to evaluate the predictivity of our models. The average, standard deviation, minimum and maximum rates were collected for each $n$-fold cross-validation method. Moreover, the confusion matrix is required for the model evaluation (Figure 2.7.1).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{confusion_matrix.png}
\caption{Confusion matrix. The samples classified by the model (rows) and the experimental classes (columns) are reported. TP (true positives) are correct positive predictions, FP (false positives) are incorrect positive predictions, FN (false negatives) are incorrect negative predictions and TN (true negatives) are correct negative predictions.}
\end{figure}

Then, the true positive (TP) rate, the false positive (FP) rate, the true negative (TN) rate, the false negative (FN) rate, accuracy, recall, or sensitivity, and precision, or specificity, were calculated for each binary classifier from the confusion matrix, as summarized in Table 2.2.
Table 2.2: Statistical parameters to evaluate the classification models.

<table>
<thead>
<tr>
<th>Name</th>
<th>Calculation details</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive rate</td>
<td>TP/(FN+TP)</td>
</tr>
<tr>
<td>False positive rate</td>
<td>FP/(TN+FP)</td>
</tr>
<tr>
<td>True negative rate</td>
<td>TN/(TN+FP)</td>
</tr>
<tr>
<td>False negative rate</td>
<td>FN/(FN+TP)</td>
</tr>
<tr>
<td>Recall</td>
<td>TP rate</td>
</tr>
<tr>
<td>Precision</td>
<td>TP/(FP+TP)</td>
</tr>
<tr>
<td>% correct predictions</td>
<td>(TP+TN)/Total number of compounds · 100</td>
</tr>
<tr>
<td>Matthews correlation</td>
<td>MCC = ( \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} )</td>
</tr>
<tr>
<td>coefficient</td>
<td>TP = number of true positives; FP = number of false positives; TN = number of true negatives; FN = number of false negatives</td>
</tr>
</tbody>
</table>

The values of the rates are included in the interval between zero and one. Low values of FP and FN rates, high values of TP and TN rates, high percentage (%) values of correct predictions, recall and precision correspond to good modeling performances.

Moreover, we calculated the *Matthews correlation coefficient* (MCC), that falls in the range \(-1 \leq MCC \leq 1\). A value of MCC = 1 indicates perfect agreement between predicted and experimental classes for each binary classifier, whereas a value MCC = -1 indicates the worst possible prediction.

The evaluation of a multilabel classification model performance is more complicated in comparison with the statistical quality of a single-label classification model. The confusion matrix was extracted from the predictions of the validation set and the internal test set to assess the robustness of our models. In this case, the accuracy is referred to the overall performance on the tested data set, while theTPs, FPs, TNS, FNs, the recall and precision are base-class measures, calculated for each class after the comparison between actual and predicted labels by our multilabel models. [108, 109] In ct-SVM analysis, the ranking process provides a function to order the labels for each sample and to assign scores to the samples. Several ranking-based performance measures have been mathematically defined. [108, 109] One-error represents the ratio of the number of not top-ranked labels to the total number of actual labels. It can take on values between zero and one and values close to zero indicate a good performance. Coverage measures how far one needs, on average, to go down the list of labels to cover all actual
labels. The coverage interval is between one and the number of the classes; then, the best performance corresponds to a value of zero. *Average precision*, that refers to the whole system, reflects the effectiveness of the label ranking and indicates the frequency of the top-ranking for the actual labels. The extreme values are zero and one and the best performance is achieved when the average precision is equal to one. [108]

## 2.8 Software

Most modeling studies were carried out on a 16 CPU (Intel CoreTM2 Quad CPU 2.40 GHz) Linux cluster running under openMosix architecture (*Paper I, II, IV-VI*). [114]

Molecular structure building and *autocorrelation* molecular electrostatic potential (*auto*MEP) descriptors, based on Connolly’s solvent accessible surfaces, have been carried out using ADRIANA (version 2.0) [115] and ADRIANA Code suite (version 2.2) (*Papers I-VI*). [51] The number of hydrogen bonding donors and acceptors, Topological Polar Surface Area (TPSA) descriptors have been carried out using ADRIANA Code software (version 2.2) (*Papers III, VI*). [51] Sterimol descriptors, logP(o/w), Approximate Surface Area (ASA), HOMO and LUMO energy descriptors have been calculated using Molecular Operating Environment (MOE, ver. 2008.10) (*Paper VI*). [116] The remaining molecular descriptors in Table 2.1 have been calculated by using ADRIANA Code software (*Paper III*). [51]

Partial Least Square (PLS) analysis has been performed using "The Unscrambler" statistical software (*Paper I*). [117]

Response Surface Analysis (RSA) has been performed using DataFOR- EST and DataNESIA softwares (*Paper I, II*). [118, 119]

Some SVM classification analysis and Support Vector Regression models have been performed by using SVMlight software (*Paper IV*). [120]

Most single-label classification models were built using Weka data mining software (*Papers III, VI*). [121]

Cross-training with SVM (ct-SVM) multilabel classification models were generated with the R software and package e1071 (*Papers III, V*). [122, 123]

The counter-propagation neural network (CPG NN) analysis was performed using SONNNIA software (*Paper III*). [124]
Chapter 3

Estimation of the aqueous solvation free energy

Several quantitative structure-property relationship (QSPR) approaches have been explored for the prediction of aqueous solubility or aqueous solvation free energy, $\Delta G_{\text{hyd}}$, as crucial parameter affecting the pharmacokinetic profile and toxicity of chemical compounds. It is mostly accepted that aqueous solvation free energies can be expressed quantitatively in terms of properties of the molecular surface electrostatic potentials of the solutes. In the present study we have introduced autocorrelation molecular electrostatic potential (autoMEP) vectors in combination with nonlinear Response Surface Analysis (RSA) as alternative 3D-QSPR strategy to evaluate the aqueous solvation free energy of organic compounds. A robust QSPR model ($r_{cv} = 0.93$) has been obtained by using a collection of 248 organic chemicals. An external test set based on 23 molecules confirmed the good predictivity of the autoMEP/RSA model suggesting its further applicability in the in silico prediction of water solubility of large organic compounds libraries.
3.1 Introduction

In the last decades a wide interest has been focused on the prediction of aqueous solubility as relevant property affecting the pharmacokinetic profile and toxicity of chemical compounds. [125] Especially in the early phase of drug discovery, molecular solubility represents an important determinant of drug-likeness, since it relates, for example, to drug bioavailability. [61]

In fact, solvation effects play one of the major contributors influencing the quantity of free drug for biological processes, and increased solubility might correspond to improved therapeutic effectiveness of potential new drugs.

From a thermodynamic point of view, the solvation free energy describes the effects of the solvent on the solute, when the solute is transferred in solution phase at constant temperature and it is surrounded by the solvent molecules. The solvation process is energetically favored if the new interactions between the molecules in solution lead to a more stable thermodynamic system with respect to not interacting solvent and solute. Then, an efficient solvation process is due to favorable interactions between the solute and the water (aqueous solvation free energies, $\Delta G_{\text{hyd}}$). Moreover, the formation of a receptor-ligand complex requires a trade-off between an unfavorable electrostatic desolvation penalty, occurring when the ligand binds the receptor in aqueous solution, and the generally favorable intermolecular interactions involved in the complex. [126] So, the solvation effects are responsible for the most probable binding mode of a receptor-ligand complex as well as for the binding affinity of organic compounds.

Well performing computer simulations of the solute-solvent system are commonly applied to calculate the free energy of solvation, but their high computational and time demands are not consistent with the study of a large number of compounds. However, a classical quantitative structure-property relationship (QSPR) approach is suitable for the evaluation of any solute-solvent system, as previously reported in several papers. [127-131] Unfortunately, the datasets that were used for the generation of these models are restricted to organic compounds.

Recently, it was reported that autocorrelation Molecular Electrostatic Potential (autoMEP) vectors in combination to Partial Least Square (PLS) and/or Response Surface Analysis (RSA) techniques can represent a powerful three-dimensional quantitative structure-activity relationship (3D-QSAR) approach. [60, 132-135] In fact, topological and electrostatic complemen-
tarities are considered two key concepts in molecular recognition processes. Gasteiger and collaborators investigated the MEP on the molecular surface as particularly useful method for rationalizing the interactions between molecules and molecular recognition processes. [52, 57, 58] The electrostatic forces are a fundamental component of the interactions between the solute and the solvent. Moreover, the major contribution to the solvation free energy of the solute is represented by the surface of the solute that is accessible to the solvent, and by the screening effect of the solvent. Therefore, MEP distribution on the molecular surface can be used as parameter to describe aqueous solvation/desolvation processes. To this aim, we have introduced autoMEP vectors in combination with Response Surface Analysis (RSA) as alternative 3D-QSPR strategy for the estimation of the aqueous solvation free energy of organic compounds, as shown in Figure 3.1.1.

Figure 3.1.1: Flowchart illustrating 3D-QSPR strategy combining autoMEP descriptors computation with Response Surface Analysis (RSA) to predict the aqueous solvation free energy of small organic molecules.
3.2 Results and discussion

The solvation free energy is a thermodynamic parameter to describe the effects of the solvent. The solvation effect is globally due to intermolecular interactions between the solute and the solvent, as well as a change in the intramolecular interactions of the solute and, a reorganization of the solvent because of the solute. Among these phenomena, the electrostatics represent the main contribution in the abovementioned interactions.

As anticipated, autoMEP descriptors encode into autocorrelation vectors the three-dimensional spatial distribution and the intensity of the electrostatic potential projected on the molecular surface. Then, we have used autoMEP descriptors in combination with a response surface analysis technique (autoMEP/RSA) to predict the solvation free energy of a set of 248 organic chemicals (training set). This training set is a collection of small organic molecules, that belong to different chemical classes (Table 3.1).

Table 3.1: Frequency of functional groups in the training set.

<table>
<thead>
<tr>
<th>No.</th>
<th>Functionalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Alkanes</td>
</tr>
<tr>
<td>15</td>
<td>Alkenes</td>
</tr>
<tr>
<td>7</td>
<td>Alkynes</td>
</tr>
<tr>
<td>57</td>
<td>Halogen derivatives</td>
</tr>
<tr>
<td>22</td>
<td>Aromatics and cycles</td>
</tr>
<tr>
<td>20</td>
<td>Aromatics and N containing compounds</td>
</tr>
<tr>
<td>7</td>
<td>Nitro derivatives and nitriles</td>
</tr>
<tr>
<td>10</td>
<td>Amines</td>
</tr>
<tr>
<td>30</td>
<td>Alcohols</td>
</tr>
<tr>
<td>7</td>
<td>Ketones</td>
</tr>
<tr>
<td>12</td>
<td>Aldehydes</td>
</tr>
<tr>
<td>16</td>
<td>Ethers</td>
</tr>
<tr>
<td>26</td>
<td>Esters</td>
</tr>
<tr>
<td>5</td>
<td>S containing compounds</td>
</tr>
</tbody>
</table>

The parameters for the calculation of autocorrelation coefficients are the following: $d_{\text{lower}} = 0$; $d_{\text{upper}} = 5$; $L = 12$; point density $= 20$ points/Å$^2$, according to eq. 2.3.4. The preliminary application of the stepwise regression and the cluster analysis on the original twelve autoMEP descriptors led to the selection of five independent variables into RSA model: autoMEP 1, 7, 8, 10 and 12. The calibration step, performed as described in 2.6.1 section,
has provided a very high correlation coefficient \((r = 0.99)\), confirming the good choice of the independent variables, as summarized in Figure 3.2.1 and Table 3.2.

**Figure 3.2.1:** AutoMEP/RSA model; experimental \(\Delta G_{\text{hyd}}\) values vs predicted \(\Delta G_{\text{hyd}}\) values after LOO cross-validation on the training set.

**Table 3.2:** Summary of the statistical parameters of autoMEP/RSA model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of molecules</td>
<td>248</td>
</tr>
<tr>
<td>X variables</td>
<td>5</td>
</tr>
<tr>
<td>(r)</td>
<td>0.99</td>
</tr>
<tr>
<td>(r_{cv})^a</td>
<td>0.93</td>
</tr>
<tr>
<td>Slope</td>
<td>0.87</td>
</tr>
<tr>
<td>Offset</td>
<td>-0.25</td>
</tr>
<tr>
<td>(q^b)</td>
<td>0.92</td>
</tr>
<tr>
<td>RMR^c</td>
<td>0.009</td>
</tr>
<tr>
<td>RSS^d</td>
<td>1.19</td>
</tr>
</tbody>
</table>

^aCross-validated \(r\) after LOO cross-validation procedure: \(r_{cv} = \frac{\text{SXY}}{\sqrt{\text{SXX} \times \text{SYY}}}\), \(\text{SXY} = \sum(X - X_{\text{mean}})(Y - Y_{\text{mean}})\), \(\text{SXX} = \sum(X - X_{\text{mean}})^2\) and \(\text{SYY} = \sum(Y - Y_{\text{mean}})^2\) with \(X = Y_{\text{Experimental}}\) and \(Y = Y_{\text{predicted}}\).
^b\(r\) of the internal test set.
^cRoot mean square of residuals: RMR.
^dResidual sum of squares: RSS.

A LOO cross-validation technique has been applied for validating the final autoMEP/RSA model to statistically confirm its robustness \((r_{cv} = 0.93)\). Interestingly, autoMEPs 1 and 7 seem to play a major role in describing the complexity of the final response surface. A representation of solvation free energy as function of autoMEP 1 and autoMEP 7 is shown in Paper I.

By analyzing Table 3.1, the predictivity of autoMEP/RSA model does not present any particular dependence from the chemical structure of the
considered organic compounds. The residuals of 248 derivatives of the training set overcome 1 kcal/mol, and it happens especially whether chloride and fluorine atoms are present or for some aliphatic and aromatic alcohols and aromatic amines, as reported in Paper I. In most cases the solvation free energy of halogen derivatives is overestimated, while alcohols and aromatic amines are generally underestimated, if compared to the respective experimental values. However, it is interesting to note that as all other 3D-QSPR approaches, also autoMEP/RSA model is able to discriminate among stereoisomers, improving the limits of some models that have utilized, for example, atomic constants as molecular descriptors (Paper I). [128]

A test set of 23 molecules with a different chemical structure and solvation free energy values has been selected to further validate our autoMEP/RSA model. The experimental vs predicted solvation free energies values are collected in Figure 3.2.2 and Table 3.3.

The predicted and experimental solvation free energy values are very similar. A very good correlation coefficient calculated on the test set ($q = 0.92$) is an additional evidence about the good predictivity of the autoMEP/RSA model, as reported in Table 3.2 (Paper I). The predicted solvation free energies result very close to the experimental values, as shown in Figure 3.2.2.

\[ \Delta G_{\text{hyd}} \]

\[ \text{Molecules} \]

**Figure 3.2.2:** Comparison between experimental (▲) and predicted (■) $\Delta G_{\text{hyd}}$ values by autoMEP/RSA model for the test set.

52
Table 3.3: Experimental and predicted solvation free energies ($\Delta G_{\text{hyd}}$ in kcal/mol) by our autoMEP/RSA model for the test set of 23 molecules.

<table>
<thead>
<tr>
<th>No.</th>
<th>Molecule name</th>
<th>Exp. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
<th>Pred. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
<th>Residuals $^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2-methylpentane</td>
<td>2.56</td>
<td>2.53</td>
<td>-0.03</td>
</tr>
<tr>
<td>2</td>
<td>cis-1,2-dimethylcyclohexane</td>
<td>1.60</td>
<td>2.38</td>
<td>0.78</td>
</tr>
<tr>
<td>3</td>
<td>1-hexene</td>
<td>1.73</td>
<td>1.96</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>2,3-dimethyl-1,3-butadiene</td>
<td>0.40</td>
<td>0.33</td>
<td>-0.07</td>
</tr>
<tr>
<td>5</td>
<td>toluene</td>
<td>-0.77</td>
<td>0.28</td>
<td>1.05</td>
</tr>
<tr>
<td>6</td>
<td>tert-butylbenzene</td>
<td>-0.44</td>
<td>-0.30</td>
<td>0.14</td>
</tr>
<tr>
<td>7</td>
<td>dichloromethane</td>
<td>-1.42</td>
<td>-1.82</td>
<td>-0.40</td>
</tr>
<tr>
<td>8</td>
<td>1,3-dibromopropane</td>
<td>-1.99</td>
<td>-1.60</td>
<td>0.39</td>
</tr>
<tr>
<td>9</td>
<td>chloroethylene</td>
<td>0.50</td>
<td>-0.76</td>
<td>-1.26</td>
</tr>
<tr>
<td>10</td>
<td>1,4-dichlorobenzene</td>
<td>-1.02</td>
<td>-1.75</td>
<td>-0.73</td>
</tr>
<tr>
<td>11</td>
<td>diethyl sulfide</td>
<td>-1.45</td>
<td>-0.80</td>
<td>0.65</td>
</tr>
<tr>
<td>12</td>
<td>diisopropyl ether</td>
<td>-0.54</td>
<td>-0.96</td>
<td>-0.42</td>
</tr>
<tr>
<td>13</td>
<td>ethane thiol</td>
<td>-4.08</td>
<td>-2.98</td>
<td>1.10</td>
</tr>
<tr>
<td>14</td>
<td>3-hexanol</td>
<td>-3.73</td>
<td>-3.8</td>
<td>-0.07</td>
</tr>
<tr>
<td>15</td>
<td>hexanal</td>
<td>-2.85</td>
<td>-2.54</td>
<td>0.31</td>
</tr>
<tr>
<td>16</td>
<td>2-butanone</td>
<td>-3.76</td>
<td>-2.63</td>
<td>1.13</td>
</tr>
<tr>
<td>17</td>
<td>methylformate</td>
<td>-2.82</td>
<td>-4.34</td>
<td>-1.52</td>
</tr>
<tr>
<td>18</td>
<td>ethylpropionate</td>
<td>-2.83</td>
<td>-2.75</td>
<td>0.08</td>
</tr>
<tr>
<td>19</td>
<td>isoamylacetate</td>
<td>-2.24</td>
<td>-2.54</td>
<td>-0.30</td>
</tr>
<tr>
<td>20</td>
<td>propylamine</td>
<td>-4.56</td>
<td>-3.88</td>
<td>0.68</td>
</tr>
<tr>
<td>21</td>
<td>dibutylamine</td>
<td>-3.38</td>
<td>-1.46</td>
<td>1.92</td>
</tr>
<tr>
<td>22</td>
<td>1-nitropropane</td>
<td>-3.38</td>
<td>-4.18</td>
<td>-0.80</td>
</tr>
<tr>
<td>23</td>
<td>2-isobutylypyrazine</td>
<td>-5.11</td>
<td>-3.84</td>
<td>1.27</td>
</tr>
</tbody>
</table>

$^*$ Predicted $\Delta G_{\text{hyd}}$ (kcal/mol) - Experimental $\Delta G_{\text{hyd}}$ (kcal/mol).

Indeed, less accurate estimation are again reported for halogen derivatives and amines (see for example, molecules 9, 21 and 23 in Figure 3.2.2 and Table 3.3). These molecules have a chemical structure in common with the worst predicted compounds in the training set, such as chloroethylene, dibutylamine and 2-isobutylypyrazine, for which autoMEP/RSA deviations from the corresponding experimental values are higher than 1 kcal/mol. Overall, we can consider the combination of autocorrelation MEP vectors with a response surface analysis an alternative tool to evaluate the aqueous solvation free energy of organic compounds.
3.3 Final remarks

The solvent environment of molecules plays a very important role in their structure and function. Consequently, it is important to consider solvation effects accurately and efficiently in the prediction and simulation of the molecular properties.

In this work, we present an alternative 3D-QSPR approach combining autoMEP molecular descriptors with Response Surface Analysis (RSA) technique to evaluate the aqueous solvation free energy of organic compounds. Considering our results, autoMEP vectors can be considered an interesting electrostatic fingerprint able to describe the solvation effects, crucial in both pharmacodynamic and pharmacokinetic processes.
Parallel application of linear and nonlinear QSAR methodologies

The autocorrelated descriptors encoding for Molecular Electrostatic Potential (autoMEP) in combination with both linear (Partial Least Square, PLS) and nonlinear (Response Surface Analysis, RSA) strategies was demonstrated to be a reliable tool to quantitatively predict the binding affinity of human adenosine receptor antagonists. In this work, a collection of 127 known human hA$_{2A}$ antagonists has been utilized to generate two 3D-QSAR models (autoMEP/PLS and autoMEP/RSA). PLS analysis is able to describe linear correlations, whether RSA detects the possible nonlinearity in the relationships between the molecular descriptors and the target property. However, we show that the parallel approach by using both techniques can lead to a more robust consensus in the prediction results. To validate our approach we have used our strategy to predict the binding affinity of five new human hA$_{2A}$ pyrazolo-triazolo-pyrimidine antagonists.
4.1 Introduction

Ligand-based approaches are widely and successfully used to develop quantitative models able to correlate, and predict, the biological activities based on various molecular descriptors, especially when the bioactive conformation of the ligand is unknown, as in the case of some G protein-coupled receptors (GPCRs). The bioactive conformation represents the starting point of all 3D-QSAR strategies such as Comparative Molecular Field Analysis (CoMFA) or 3D-Pharmacophore search. [137, 138]

As anticipated, 3D-QSAR methods require the knowledge of the conformational properties of the molecules in order to calculate their structural or property descriptors. It was demonstrated that the autoMEP vectors in combination with Partial Least Square (PLS) analysis can represent an alternative 3D-QSAR tool to CoMFA. [132-134] However, both CoMFA and autoMEP/PLS methodologies can be classified as linear QSAR methods considering the mathematical relationship among molecular descriptors and the chemical/biological response space. Very recently, a nonlinear method based on a response surface analysis (RSA) application in tandem with the autoMEP descriptors (autoMEP/RSA) was also presented as an alternative 3D-QSAR method. [60, 135]

As case study we have considered the prediction of the pharmacodynamic profile of human adenosine A$_{2A}$ receptor antagonists. More specifically, we would like to show how the applicability in parallel of both linear and nonlinear 3D-QSAR methods (autoMEP/PLS and autoMEP/RSA) can help to predict the binding affinity data of a new set of human adenosine A$_{2A}$ receptor antagonists.

4.2 Human A$_{2A}$R antagonists dataset

Briefly, the adenosine A$_{2A}$ receptors are classified in the adenosine receptor (AR) family of GPCRs, which includes A$_1$, A$_{2A}$, A$_{2B}$ and A$_3$ different subtypes, abundantly expressed in diverse areas of human body and potentially in the same cellular types$^1$. [139] Being heptahelical transmembrane GPCRs, they are involved in several signal transduction pathways, as shown in Figure 4.2.1.

---

$^1$Further details on the four human AR subtypes are reported in section 7.2.
Moreover, they are codified by distinct genes and they have been cloned from various mammalian species, where they seem to differentiate for their pharmacological profile. [139] In particular, the human adenosine A$_2A$ receptor (hA$_{2A}$R) has been discovered to be crucial in some neurological disorders, which involve also other neurotransmission systems, above all the dopamine D$_2$ receptor, antagonistically associated to this adenosine receptor subtype. [140] It has been demonstrated that the activation of the human A$_2A$R causes the inhibition of platelet aggregation, attenuates the inflammatory responses mediated by cytokines, involves the regulation of the immune cells functions, while adenosine A$_2A$ receptor antagonists show a neuroprotective activity during ischemic processes. [141] The inhibition of A$_2A$ receptor, blocking the effects of adenosine, has been suggested as key strategy for the treatment of diverse pathologies. [142] In more detail, one of the main potential therapeutic applications of A$_2A$ receptor antagonists is the promotion of cellular survival and the reduction of neuronal damage in Parkinson’s or Huntington’s diseases. [142-144]

In the last few years, several different potent and selective human adenosine A$_2A$ receptor antagonists have been discovered. [145] In particular, pyrazolo-triazolo-pyrimidine and triazolo-pyridine derivatives were described as promising hA$_{2A}$R antagonists. Their chemical structures are summarized in Figure 4.2.2. In the present study, a collection of 127 known human A$_{2A}$R antagonists has been utilized to derive a couple of 3D-QSAR models (autoMEP/PLS and autoMEP/RSA). The binding affinity of the compounds is expressed as $K_i$ (nM) after displacement experiments by using [3H]-NECA
4.2 Human A\textsubscript{2A}R antagonists dataset

**Figure 4.2.2:** Chemical structures of known pyrazolo-triazolo-pyrimidine and triazolo-pyridine human A\textsubscript{2A}R antagonists in the training set.

Binding at human A\textsubscript{2A} receptors expressed in CHO or HEK-293 cells (*Paper II*). To validate our in tandem approach, they have been utilized to predict the binding affinity of five new human A\textsubscript{2A}R pyrazolo-triazolo-pyrimidine antagonists, following the flowchart shown in Figure 4.2.3.

**Figure 4.2.3:** Partial Least Square (PLS) and Response Surface Analysis (RSA) approaches applied in tandem.
As summarized in Figure 4.2.3, a collection of pyrazolo-triazolo-pyrimidine and triazolo-pyridine analogues (molecules 1-127) was selected as training set in both linear and nonlinear QSAR models. [146-152] An internal test set of 10 training set analogues (molecules 128-137) was selected for the validation process of both PLS and RSA models. Finally, five new pyrazolo-triazolo-pyrimidine analogues (molecules 138-142) has been analyzed as additional validation set².

4.3 Results and discussion

Topological and electrostatic complementarities are considered two key aspects in the molecular recognition processes. Both potentialities and advantages of the use of autoMEP descriptors in different 3D-QSAR applications have been previously discussed. [60, 132-135] In this work, using the autoMEP vectors, mentioned in chapter 2, we have assessed the possibility to combine in parallel two different linear and nonlinear strategies, PLS and RSA, to find a consensus in the quantitative binding affinity predictions.

4.3.1 PLS and RSA models

Both 3D-QSAR models have been derived by using 96 pyrazolo-triazolo-pyrimidine and 31 triazolo-pyridine derivatives as training set of known A₂A receptor antagonists. Moreover, both models have been subjected to a validation process by using a test set of 10 molecules (defined as the internal test set), structurally related to those included into the training set. Twelve autoMEP vectors have been used as independent variables in both PLS and RSA analysis (calculated as described in chapter 2); a preliminary variable selection step has been introduced before deriving RSA model (Paper II).

Concerning PLS analysis, the resulting model has shown acceptable statistical quality in both calibration and internal validation steps as demonstrated by the $r$ and $r_{cv}$ values of 0.80 and 0.78, respectively, using only three latent variables (Figure 4.3.1 and Table 4.1). The robustness of the PLS model is also supported by the good value of the correlation coefficient calculated on the test set ($q = 0.85$), as reported in Figure 4.3.2, Table 4.1 and Table 4.2.

²Experimental binding affinity data kindly provided by the work coordinated by Prof. G. Spalluto (University of Trieste) for the synthesis and by Prof. K. N. Klotz (University of Würzburg) for the pharmacological characterization.
Figure 4.3.1: AutoMEP/PLS model; experimental pK\textsubscript{i} data plotted vs predicted pK\textsubscript{i} values (after LOO cross-validation) for the training set.

Table 4.1: Summary of the statistical parameters of autoMEP/PLS model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of molecules</td>
<td>127</td>
</tr>
<tr>
<td>Latent variables</td>
<td>3</td>
</tr>
<tr>
<td>(r)</td>
<td>0.80</td>
</tr>
<tr>
<td>(r_{cv})</td>
<td>0.78</td>
</tr>
<tr>
<td>Slope</td>
<td>0.62</td>
</tr>
<tr>
<td>Offset</td>
<td>-0.57</td>
</tr>
<tr>
<td>(q^b)</td>
<td>0.85</td>
</tr>
<tr>
<td>(RMR^c)</td>
<td>0.043</td>
</tr>
</tbody>
</table>

*a* Cross-validated \(r\) after LOO cross-validation procedure: \(r_{cv} = \frac{S_{XY}}{\sqrt{(S_{XX})(S_{YY})}}\), where \(S_{XY} = \sum(X-Y \text{ mean})(Y-Y \text{ mean})\), \(S_{XX} = \sum(X-X \text{ mean})^2\) and \(S_{YY} = \sum(Y-Y \text{ mean})^2\) with \(X = Y_{\text{Experimental}}\) and \(Y = Y_{\text{predicted}}\). *b* \(r\) of the internal test set; *c* root mean square of residuals: RMR.
Figure 4.3.2: Comparison of autoMEP/PLS (■) and autoMEP/RSA predicted pK$_i$ (●) with experimental pK$_i$ values (▲) of the internal test set.

Table 4.2: Experimental and predicted pK$_i$ for the internal test set. Differences between predicted and experimental pK$_i$ values for both models are reported; PLS = autoMEP/PLS model; RSA = autoMEP/RSA model.

<table>
<thead>
<tr>
<th>No.</th>
<th>Exp. pK$_i$</th>
<th>Pred. pK$_i$, PLS</th>
<th>Pred. pK$_i$, RSA</th>
<th>Δ pK$_i$</th>
<th>Δ pK$_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.89</td>
<td>-0.11</td>
<td>0.25</td>
<td>-1.00</td>
<td>-0.64</td>
</tr>
<tr>
<td>129</td>
<td>-0.64</td>
<td>-0.55</td>
<td>-0.65</td>
<td>0.09</td>
<td>-0.01</td>
</tr>
<tr>
<td>130</td>
<td>-0.43</td>
<td>-0.04</td>
<td>-0.20</td>
<td>0.39</td>
<td>0.23</td>
</tr>
<tr>
<td>131</td>
<td>-2.79</td>
<td>-2.36</td>
<td>-2.66</td>
<td>0.43</td>
<td>0.13</td>
</tr>
<tr>
<td>132</td>
<td>-2.70</td>
<td>-1.47</td>
<td>-1.65</td>
<td>1.23</td>
<td>1.05</td>
</tr>
<tr>
<td>133</td>
<td>-3.02</td>
<td>-2.44</td>
<td>-2.08</td>
<td>0.58</td>
<td>0.94</td>
</tr>
<tr>
<td>134</td>
<td>-0.29</td>
<td>-1.37</td>
<td>-1.39</td>
<td>-1.08</td>
<td>-1.10</td>
</tr>
<tr>
<td>135</td>
<td>-1.00</td>
<td>-0.82</td>
<td>-0.85</td>
<td>0.18</td>
<td>0.15</td>
</tr>
<tr>
<td>136</td>
<td>-1.43</td>
<td>-1.08</td>
<td>-1.14</td>
<td>0.35</td>
<td>0.29</td>
</tr>
<tr>
<td>137</td>
<td>-1.86</td>
<td>-1.13</td>
<td>-1.11</td>
<td>0.73</td>
<td>0.75</td>
</tr>
</tbody>
</table>

$^{a}$Predicted pK$_i$ - Experimental pK$_i$. 

61
In parallel, we have delivered a nonlinear RSA model using the same training and test sets. The stepwise regression analysis together with the cluster analysis on the original 12 molecular descriptors led us to select five of them as final combination to utilize as independent variables into the RSA model: autoMEP 2, 4, 6, 7 and 11 (Paper II). The statistical parameters and the final RSA model are collected in Figure 4.3.3 and Table 4.3.

![Figure 4.3.3: AutoMEP/RSA model; experimental pKᵢ data plotted vs predicted pKᵢ values (after LOO cross-validation) for the training set.](image)

| Table 4.3: Summary of the statistical parameters of autoMEP/RSA model. |
|--------------------------|----------|----------|-----------|----------|
| Number of molecules      | 127      | X variables | 5         |
| r                       | 0.98     | r_cvᵇ     | 0.82      |
| Slope                   | 0.68     | Offset    | -0.49     |
| qᵇ                      | 0.87     | RMR       | 0.043     |

⁷Cross-validated r after LOO cross-validation procedure: \( r_{cv} = \frac{\sum(X-Y) \times \sum(Y-Y_{mean})}{\sum(X-X_{mean}) \times \sum(Y-Y_{mean})} \)

We can observe a very high correlation coefficient (\( r = 0.98 \)) for the calibration step confirming the good choice of the independent variables.
The correlation coefficients after LOO cross-validation and calculated on the test set are also appreciable ($r_{cv} = 0.82$ and $q = 0.87$, respectively). These results represent additional evidences about the good predictivity of the autoMEP/RSA model, as shown in Figure 4.3.2 and Table 4.2.

Even if both methods are statistically acceptable, the autoMEP/RSA model presents higher predictivity than autoMEP/PLS model (Table 4.2). However, both methodologies are able to coherently discriminate between "more active" and "less active" analogues. This result is very interesting because ensemble, or consensus, approaches to classification and regression have been considered as attractive tools. [153] In fact, these methods have been shown to outperform a single predictor usability on a wide range of scientific tasks. [153]

### 4.3.2 External test set prediction

The application in parallel of different strategies may confirm the predictions achieved by using single models alone. In this study, we used both autoMEP/PLS and autoMEP/RSA models as an ensemble of binding affinity predictors to prioritize the synthesis of new human A$_2$A receptors.

Following these encouraging results, we have tested the real predictive capability of our PLS and RSA models on an external test set, which consisted in five new pyrazolo-triazolo-pyrimidine analogues. This is a preliminary proof of concept of our parallel PLS/RSA approach. As anticipated in the Introduction, we aim at simultaneously performing different 3D-QSAR approaches to create a more even balance between false positive and false negative performance rates than the use of a single method can achieve.

In our laboratories, we are still developing new potent and selective human A$_2$A antagonists decorating the pyrazolo-triazolo-pyrimidine scaffold using different strategies. [154] In this context, we analyzed a new class of N-substituted derivatives in which different series of benzyloxy-phenyl-acetyl substituents are present$^3$. Once autoMEP vectors have been computed for this new set of ligands, we have applied both PLS and RSA models for their binding affinity predictions (Figure 4.3.4 and Table 4.4).

---

$^3$Experimental binding affinity data kindly provided by the work coordinated by Prof. G. Spalluto (University of Trieste) for the synthesis and by Prof. K. N. Klotz (University of Würzburg) for the pharmacological characterization.
As shown in Table 4.4, both methods predicted all five derivatives active in the low nM range \((K_i \text{ values in between to 50 and 150 nM})\), with a better performance of the autoMEP/RSA model with respect to autoMEP/PLS one. Interestingly, even if both methods overestimated all binding affinities, new synthesized pyrazolo-triazolo-pyrimidine analogues are active in the nM range \((K_i \text{ values in between to 230 and 750 nM})\), as theoretically predicted (Figure 4.3.4).
4.4 Final remarks

In light of the consideration that GPCR ligands represent one of the major continuing source of novel potent therapeutic agents, and that 3D structures of GPCRs as determined by experimental techniques are still unavailable, ligand-based drug discovery methods remain the most feasible computational approaches to the analysis of growing data sets of developmental GPCR ligands.

We have proposed the application of a couple of complementary QSAR methodologies to evaluate the binding affinity data of a new series of human A$_{2A}$R antagonists. Two statistically meaningful models have been generated from a common training set, and their predictivity has been evaluated by using both internal and external test sets. We are continuously analyzing new series of ligands with the aim to improve the robustness and the predictivity of our QSAR models. The purpose is to perform in silico screening of real or virtual libraries to research for new potent and selective human A$_{2A}$R antagonists.
Chapter 5

Isoform specificity of cytochrome P450 substrates

Each drug can potentially be metabolized by different cytochrome P450 (CYP450) isoforms. In the development of new drugs, the prediction of the metabolic fate is important to prevent drug-drug interactions. The present study deeply analyzes a collection of 554 CYP450 substrates by applying multi- and single-label classification strategies, after the computation and the selection of suitable molecular descriptors. Cross-training with support vector machine and counter-propagation neural network modeling methods were used in the multilabel approach, which allows one to classify the compounds simultaneously in multiple classes. In the single-label models automatic variable selection was combined with various cross-validation experiments and modeling techniques. Moreover, the reliability of both multi- and single-label models was assessed by the prediction of an external test set. Finally, the predicted results of the best models were compared to show that, even if the models present similar performances, the multilabel approach more coherently reflects the real metabolism information.

1This work has been carried out at Molecular Networks, Erlangen (Germany), with the supervision of Prof. J. Gasteiger and the collaboration of Dr. L. Terlloth.
5.1 Cytochrome P450 in drug metabolism

The metabolic profile of a drug candidate is an important aspect to be considered in the selection of a potential new drug. Several problems related to stability, toxicity of xenobiotics and drug interactions might represent serious adverse effects. In fact, in the case of co-administration of drugs, the pharmacological profile of each drug might be modified by the presence of other drugs in the human body. [155, 156] If two drugs are co-administered and metabolized by the same enzyme, the competition for the binding site can result in the inhibition of the biotransformation of one or both drugs.

Focusing the attention on metabolism in the ADMET process, a crucial role is played by the cytochrome P450 (CYP450) class of hemoprotein enzymes. The CYP450 superfamily of enzymes is abundantly expressed in the liver and remarkably in the small intestine, where it is responsible for the detoxification of xenobiotics. [157, 158] In the Phase I metabolism cytochrome P450 isoforms chemically modify a large variety of substrates mainly through oxidation reactions to make them more water-soluble and to ease their elimination. [159] This detoxification system is highly complex, since it includes many different CYP450 isoforms characterized by multiple binding sites, polymorphism and enzyme induction or modulation phenomena. [160] These aspects are involved in drug-drug interactions, which might lead to unpredictable blood concentrations of one of the xenobiotics with consequent possible toxic effects or loss of activity. CYP450 enzymes are classified in several isoforms according to the similarity of their amino acidic sequences. We have investigated CYP450 1A2, 2C9, 2D6, 2E1 and 3A4 substrates, that cover almost all possible metabolism routes (Figure 5.1.1).

![Figure 5.1.1: Importance, as percentage (%), of the different CYP450 isoforms in the metabolic process; CYP1A2 (orange), CYP2C9 (blue), CYP2C19 (purple), CYP2D6 (yellow), CYP2E1 (green), CYP3A4 (light blue), other isoforms (red).](image-url)
In our analysis, we have excluded CYP2C8 and CYP2C19 isoforms, poorly represented, as they are not much involved in the metabolic process. The same analysis with seven CYP450 isoforms has been performed, as reported in Paper III. CYP450 1A2 metabolizes planar molecules characterized by moderate volume and basicity. [157] CYP450 2C9 substrates are acidic or neutral, and lipophilic molecules, in particular sulfonylureas and NSAIDs (non steroidal anti-inflammatory drugs) drug classes. [157] CYP450 2D6 shows polymorphism and about 25% of all drugs are at least partial substrates of this isoform. They show a hydrophilic character and have a basic nitrogen atom. [157] Mostly small and polar molecules, like volatile anesthetics, are substrates of CYP450 2E1 isoform which is involved in many drug interactions. [157] The CYP450 3A4 isoform, ubiquitously found, is responsible for the metabolism of high volume and lipophilic xenobiotics; almost 50% of all drugs are metabolized by this isoform. [157] For this reason its activity is largely affected by chemically different compounds and CYP3A4 represents the most populated class in the dataset investigated in this study. However, different isoforms might be responsible for the detoxification of the same drug.

5.2 Computational approaches to CYP450

The early detection of ADMET properties of drugs under in vivo conditions is experimentally time-consuming and expensive, therefore computational methods can profitably speed up the collection of new data. [161-163] A challenging problem in this field is the prediction of CYP450 isoform specificity. Different chemoinformatic tools have already been attempted for the prediction of CYP-related metabolism properties. [164, 165] In more detail, several ligand-based approaches were applied to classify CYP450 substrates according to their route of metabolism. [44, 45, 166-170] Anyway, most solutions consider local models for each CYP450 isoform and they do not approach the problem globally. Among these, the traditional single-label classification deals exclusively with non-overlapping classes. Following this approach, a classification model was developed to predict the isoform specificity for CYP3A4, CYP2D6 and CYP2C9 substrates considering non-overlapping classes, i.e. assuming each compound to be metabolized by a single, predominant CYP450 isoform. [76]
In the present study we would like to extend the abovementioned model to cover other CYP450 isoforms and to find a strategy to predict the substrates which are metabolized by more than one isoform. Such a multilabel classification analysis represents a different approach that can be applied whether our dataset comprises elements assigned simultaneously to more than one class. The prediction of the metabolism profile of CYP450 substrates represents a novel application of this methodology.

This work aims at the prediction of the isoform specificity from information on the substrates metabolism. The substrates were represented by different sets of structural and physicochemical descriptors. Then, various classification techniques - both multilabel and single-label approaches - were used to derive models for the prediction of the isoform(s) responsible for metabolism of CYP450 1A2, 2C9, 2D6, 2E1 and 3A4 substrates. The procedures we followed are separately illustrated in Figure 5.2.1.

**Figure 5.2.1:** Flowchart of the multilabel and single-label approaches for the prediction of CYP450 isoform specificity.
We applied a multilabel classification method to distinguish between substrates of five CYP450 isoforms. Then, only single-label compounds, i.e. compounds metabolized by a single isoform, were utilized to build a classifier. Variable selection and optimization of the models were performed in the present study. The single-label classification models were directly generated after an automatic variable selection process. The modeling results and the metabolism profiles of the CYP450 substrates of the test sets predicted by the different approaches were compared to verify the reliability of both multi- and single-label classification methods.

5.3 Dataset

A collection of 554 cytochrome P450 substrates with different chemical structures was used to derive our classification models. In particular, we considered the dataset compiled in the recently published paper by Block et al. [170] It includes 253 substrates metabolized by CYP1A2, CYP2C19, CYP2C8, CYP2C9, CYP2D6 and CYP3A4 isoforms. At least one isoform may be responsible for the metabolism of a single substrate, i.e. a compound might be metabolized by several isoforms and then belongs to several classes. Since we considered all possible routes of metabolism, our classification problem is multilabel. Further substrates were extracted from other published papers and publicly available lists. [171-174] Moreover, 267 additional compounds from the Metabolite reaction database were included. [175] When inconsistencies resulted in the comparison between the datasets, we considered more reliable the information about the compounds metabolic fate published by Block et al. In the final collection (554 compounds) 484 substrates are metabolized by one CYP450 isoform and the remaining 70 compounds are metabolized by several CYPs. Only 13% of all compounds in the dataset are multilabel. Considering the 484 compounds metabolized by one isoform, 46 are CYP1A2 substrates (9.5%), 50 are CYP2C9 substrates (10.3%), 106 are CYP2D6 substrates (21.9%), 49 are CYP2E1 substrates (10.1%) and 233 are CYP3A4 substrates (48.2%). Not all isoforms have the same relevance in the xenobiotic metabolism, consequently these classes are differently populated and our dataset is quite unbalanced.

For our modeling studies two different data sets were compiled (Data set 1 and Data set 2). Data set 1 was manually split into training, validation
and test set with a similar distribution of the considered classes in the entire data set and the subsets. Most of the compounds from the Metabolite data set were used as test set, even if some substrates were included in the training and the validation set (Paper III). Data set 2 was simply split into a training and a validation set, applying the same selection criterion (similar distribution of the substrates in the considered classes).

### 5.3.1 Data set 1

Data set 1 comprises the initial collection of 554 chemically different substrates, metabolized by five CYP450 isoforms (CYP1A2, CYP2C9, CYP2D6, CYP2E1 and CYP3A4 single- and multilabel substrates). Data set 1 was used to perform a multilabel classification analysis. The distribution of the classes in the training, validation and test sets is reported in Table 5.1.

<table>
<thead>
<tr>
<th>CYP450 Isoform</th>
<th>Training set 1</th>
<th>Validation set 1</th>
<th>Test set 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>M</td>
<td>S</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>26</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>31</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>46</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>30</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>110</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>Total (554)</td>
<td>243</td>
<td>40</td>
<td>50</td>
</tr>
</tbody>
</table>

*Single-label substrates occur only once in each class; multilabel substrates belong to more than one class. Consequently, the sum of multilabel substrates for all the classes within a column is higher than the number of multilabel substrates for Training set 1, Validation set 1 and Test set 1; S = single-label; M = multilabel.

### 5.3.2 Data set 2

Only the single-label substrates in Data set 1 were selected to perform a single-label classification analysis. Then, 484 compounds, a subset of Data set 1, was used as Data set 2. In the new training set (Training set 2) exactly the same single-label substrates collected in the Training set 1 and Validation set 1 (totally 293 compounds) were included, while in the Test set 2 only the single-label substrates in the Test set 1 (191 compounds) were considered.
Data set 2 was utilized to develop a single-label classification model. The results of the splitting process for Data set 2 can be inferred by considering the "S" columns for the training, validation and test sets in Table 5.1.

5.4 Results

The prediction of isoform specificity represents a multilabel classification problem, characterized by high complexity of the feature space. In this study, we built a model to simultaneously classify a collection of substrates metabolized by five CYP450 isoforms (CYP1A2, CYP2C9, CYP2D6, CYP2E1 and CYP3A4). Various descriptors and data analysis techniques were combined to predict the isoform specificity using two different data sets, which correspond to (1) multilabel classification models (Data set 1) and (2) single-label classification model (Data set 2). In more detail, cross-training with Support Vector Machine (ct-SVM) and counter-propagation neural networks (CPG NN) have been applied as multilabel classification techniques, while SVM was used to develop single-label classification models. Only the best-performing experiments are reported in this chapter.

5.4.1 Multilabel classification with Data set 1

Since the information about the reaction site of the substrates was not reported, several models were derived combining global, topological, shape and functional group counts descriptors with 2D topological or 3D spatial autocorrelation vector components (descriptors reported in section 2.3 and in Table 2.1). A manual descriptor selection process was combined with ct-SVM multilabel modeling method in the optimization procedure using the Training set 1. The best subset of global, topological, shape and functional groups count descriptors was selected according to the model performance after each single-descriptor addition step. In more detail, a descriptor was included in the best subset if the model predictivity results on the Validation set 1 improved. The best model corresponded to the final set of 27 descriptors, as reported in Table 5.2.

The information encoded by the autocorrelation molecular electrostatic potential descriptors improved the predictivity of the model in comparison to the model computed without these twelve variables (Paper III). It supports the concept that the distribution of electrostatic properties on the molecu-
Table 5.2: Twenty seven descriptors selected for the training set in multilabel classification models with Data set 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>1</td>
<td>HA ccPot</td>
<td>Hydrogen bond acceptor potential</td>
</tr>
<tr>
<td>1</td>
<td>TPSA</td>
<td>Topological polar surface area</td>
</tr>
<tr>
<td>1</td>
<td>ASA</td>
<td>Approximate surface area</td>
</tr>
<tr>
<td>1</td>
<td>$D_3$</td>
<td>Diameter</td>
</tr>
<tr>
<td>1</td>
<td>$R_3$</td>
<td>Radius</td>
</tr>
<tr>
<td>1</td>
<td>$I_3$</td>
<td>Geometric shape coefficient</td>
</tr>
<tr>
<td>1</td>
<td>$r^2$</td>
<td>Radius perpendicular to $D_3$</td>
</tr>
<tr>
<td>1</td>
<td>$r^3$</td>
<td>Radius perpendicular to $D_3$ and $R_2$</td>
</tr>
<tr>
<td>1</td>
<td>$n_{aro_amino}$</td>
<td>Number of aromatic amino groups</td>
</tr>
<tr>
<td>1</td>
<td>$n_{tert_amino}$</td>
<td>Number of tertiary aliphatic amino groups</td>
</tr>
<tr>
<td>1</td>
<td>$n_{prim_sec_amino}$</td>
<td>$n_{prim_amino} + n_{sec_amino}$</td>
</tr>
<tr>
<td>1</td>
<td>$n_{basic_nitrogen}$</td>
<td>Number of basic, N containing functional groups</td>
</tr>
<tr>
<td>1</td>
<td>$n_{acidic_groups}$</td>
<td>Number of acidic functional groups</td>
</tr>
<tr>
<td>1</td>
<td>$q_{\pi-1} = \sum q_{\pi}^2$</td>
<td>property: $\pi$-charge $q_{\pi}$</td>
</tr>
<tr>
<td>12</td>
<td>SurfACorr_ESP</td>
<td>Surface autocorrelation; property: molecular electrostatic potential</td>
</tr>
</tbody>
</table>

Lar surface represents an important determinant in the prediction of isoform specificity. In the ct-SVM modeling method, the best results were achieved using the T-criterion to transform the real-valued scores assigned by the corresponding classifier into labels. This descriptor set was used to derive a first ct-SVM classification model with Training set 1. As previously described, the model parameters were optimized by predicting the Validation set 1. Then, TP, FP, TN, FN rates and the percentage (%) of correct predictions were calculated from the confusion matrix (Table 5.3).

Table 5.3: Multi-classification ct-SVM model; the statistical parameters for each class after prediction on the Validation set 1 (62 substrates) are reported.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate</th>
<th>FP rate</th>
<th>TN rate</th>
<th>FN rate</th>
<th>Recall</th>
<th>Precision</th>
<th>% correct predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.64</td>
<td>0.04</td>
<td>0.96</td>
<td>0.36</td>
<td>0.64</td>
<td>0.78</td>
<td>90.3</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.82</td>
<td>0.05</td>
<td>0.95</td>
<td>0.18</td>
<td>0.82</td>
<td>0.75</td>
<td>95.2</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.87</td>
<td>0.00</td>
<td>1.00</td>
<td>0.13</td>
<td>0.87</td>
<td>1.00</td>
<td>93.5</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>0.88</td>
<td>0.00</td>
<td>1.00</td>
<td>0.12</td>
<td>0.88</td>
<td>1.00</td>
<td>98.4</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.87</td>
<td>0.25</td>
<td>0.75</td>
<td>0.13</td>
<td>0.87</td>
<td>0.75</td>
<td>87.1</td>
</tr>
</tbody>
</table>

As seen in Table 5.3, a good predictivity is achieved for almost all the
classes if we analyze the values of recall and precision. Training set 1 and the Validation set 1 were merged to derive a new classifier with the optimized parameters. The performance measures of the final ct-SVM model are shown in Table 5.4.

**Table 5.4:** Multi-classification ct-SVM model; performance measures after prediction on the Validation set 1 (62 substrates) and the Test set 1 (209 compounds) are reported.

<table>
<thead>
<tr>
<th>Model prediction</th>
<th>Accuracy_{ML}</th>
<th>One-error</th>
<th>Coverage</th>
<th>Average precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation set 1</td>
<td>0.84</td>
<td>0.10</td>
<td>1.53</td>
<td>0.93</td>
</tr>
<tr>
<td>Test set 1</td>
<td>0.70</td>
<td>0.25</td>
<td>1.52</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Satisfactory results for the multilabel approach were also achieved by the application of CPG NN technique with the set of 27 descriptors in Table 5.2. A graphical representation of our analysis is shown in Figure 5.4.1.

![Flowchart of the CPG NN analysis with five output layers (CYP450 isoforms) applied to our multilabel classification problem.](image)

**Figure 5.4.1:** Flowchart of the CPG NN analysis with five output layers (CYP450 isoforms) applied to our multilabel classification problem.

Different map sizes, topologies and number of epochs were selected for the learning process and the best model with a rectangular topology is discussed here. After LOO and 5-fold cross-validation procedures, the best CPG NN model was selected. The model performances are reported in Table 5.5. Good values of precision and recall were obtained for CYP2C9, CYP2D6, CYP2E1 and CYP3A4 classes, by applying the best CPG NN model. A low predictivity of CPG NN model can be observed for CYP1A2 and CYP2C9.
5.4 Results

Table 5.5: Multi-classification CPG NN; the statistical parameters for each class after LOO and 5-fold cross-validation (345 substrates) are reported.

<table>
<thead>
<tr>
<th>Classes</th>
<th>LOO TP rate</th>
<th>LOO FP rate</th>
<th>LOO TN rate</th>
<th>LOO FN rate</th>
<th>LOO Recall</th>
<th>LOO Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.33</td>
<td>0.10</td>
<td>0.90</td>
<td>0.06</td>
<td>0.33</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.10</td>
<td>0.90</td>
<td>0.06</td>
<td>0.48</td>
<td>0.47</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.60</td>
<td>0.07</td>
<td>0.93</td>
<td>0.04</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>0.58</td>
<td>0.07</td>
<td>0.93</td>
<td>0.04</td>
<td>0.58</td>
<td>0.60</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.79</td>
<td>0.09</td>
<td>0.91</td>
<td>0.02</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>0.79</td>
<td>0.06</td>
<td>0.91</td>
<td>0.02</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>0.87</td>
<td>0.01</td>
<td>0.99</td>
<td>0.01</td>
<td>0.87</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>0.03</td>
<td>0.99</td>
<td>0.01</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.77</td>
<td>0.24</td>
<td>0.76</td>
<td>0.22</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>0.77</td>
<td>0.21</td>
<td>0.76</td>
<td>0.22</td>
<td>0.77</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Classes in terms of lower TP rate or, alternatively, higher FN rate.

In general, a model with a minimum predictivity for one class might not be able to detect substrates metabolized by this particular isoenzyme, whether it was applied to an external test set. However, at least 75% of the compounds for each class were correctly classified (Paper III). In Figure 5.4.2 the five output layers of the CPG NN network are shown.

Figure 5.4.2: CPG NN model; projection of the CYP450 substrates into five maps, corresponding to CYP1A2, CYP2C9, CYP2D6, CYP2E1 and CYP3A4 classes. CYP1A2 substrates in orange, CYP2C9 substrates in yellow, CYP2D6 substrates in blue, CYP2E1 substrates in green, CYP3A4 substrates in light blue are indicated (according to the most frequent output). For each map, neurons containing substrates which do not belong to the corresponding class are red. In the maps below conflicting neurons in black are also shown. Black neurons contain compounds of different CYP450 classes. White squares represent empty neurons.

In each layer, the CYP450 substrates tend to cluster. This tendency is not particularly evident for CYP1A2 class, where the occupied neurons are spread out in the corresponding layer. Few conflict neurons are present in the
maps, and most of them are caused by CYP3A4 substrates, conflicting with the substrates metabolized by other classes. In fact, CYP3A4 represents the major and the most chemically heterogeneous class. In some cases, the descriptor set we selected is not able to correctly classify the substrates with similar structural features and different class memberships.

5.4.2 Single-label classification with Data set 2

Data set 2 includes only single-label substrates extracted from Data set 1 and it is therefore suitable for the following data analysis. In the single-label approach, a systematic variable selection procedure was performed, by considering global, topological, shape and functional group counts, 128 spatial autocorrelation vectors (totally 168 descriptors in Table 2.1). Finally, nineteen descriptors were automatically selected by using the Training set 2, as summarized in Table 5.6.

Table 5.6: Nineteen descriptors resulting from automatic variable selection for the training set in single-label classification models with Data set 2.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HDonPot</td>
<td>Hydrogen bond donor potential</td>
</tr>
<tr>
<td>1</td>
<td>TPSA</td>
<td>Topological polar surface area</td>
</tr>
<tr>
<td>1</td>
<td>ASA</td>
<td>Approximate surface area</td>
</tr>
<tr>
<td>1</td>
<td>$\mu$</td>
<td>Molecular dipole moment</td>
</tr>
<tr>
<td>1</td>
<td>$r^2$</td>
<td>Radius perpendicular to $D_3$</td>
</tr>
<tr>
<td>1</td>
<td>$r^3$</td>
<td>Radius perpendicular to $D_3$ and $R_2$</td>
</tr>
<tr>
<td>1</td>
<td>$n_{\text{aliph_amino}}$</td>
<td>Number of aliphatic amino groups</td>
</tr>
<tr>
<td>1</td>
<td>$n_{\text{prim_sec_amino}}$</td>
<td>$n_{\text{prim_amino}} + n_{\text{sec_amino}}$</td>
</tr>
<tr>
<td>1</td>
<td>$n_{\text{basic_nitrogen}}$</td>
<td>Number of basic, N containing functional groups</td>
</tr>
<tr>
<td>1</td>
<td>$n_{\text{acidic_groups}}$</td>
<td>Number of acidic functional groups</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [1.2-1.3 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [1.3-1.4 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [1.4-1.5 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [1.7-1.8 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [2.3-2.4 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [2.7-2.8 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [3.1-3.2 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [4.2-4.3 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
<tr>
<td>1</td>
<td>3D-AC$_{\text{identity}}$ [5.3-5.4 Å]</td>
<td>Spatial autocorrelation; property: identity</td>
</tr>
</tbody>
</table>
5.4 Results

The 3D autocorrelation identity descriptors reflect the distribution of the interatomic distances in the 3D molecular structure and complete the information given by the first selected subset. In more detail, the Best First automatic criterion implemented in Weka was applied to select the variables. [121] The descriptor space was explored in order to detect the subset that is likely to predict the classes best. The attribute evaluator CfsSubsetEval combined with Best First search method has been applied. The variable selection process was repeated for each fold during the model validation step. The selected nine components of 3D autocorrelation identity vectors correspond to particular atom distances and show an important contribution in the model building process.

We have generated a single-label classification model, by combining the automatic variable selection with Support Vector Machine. The standard parameters suggested in Weka software were selected for the computation. In the SVM model, a polynomial kernel with exponent equal to three was used. The results are reported in Table 5.7.

Table 5.7: Single-label classification SVM model; the statistical parameters using nineteen descriptors for the Training set 2.

<table>
<thead>
<tr>
<th>Partition</th>
<th>CV</th>
<th>No. of runs</th>
<th>Mean</th>
<th>StDev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set 2</td>
<td>1</td>
<td>85.7</td>
<td>-</td>
<td>85.7</td>
<td>85.7</td>
<td></td>
</tr>
<tr>
<td>Training set 2</td>
<td>LOO</td>
<td>75.8</td>
<td>-</td>
<td>75.8</td>
<td>75.8</td>
<td></td>
</tr>
<tr>
<td>10-fold</td>
<td>10</td>
<td>76.3</td>
<td>1.3</td>
<td>74.7</td>
<td>78.2</td>
<td></td>
</tr>
<tr>
<td>5-fold</td>
<td>20</td>
<td>76.0</td>
<td>1.2</td>
<td>73.3</td>
<td>78.2</td>
<td></td>
</tr>
<tr>
<td>3-fold</td>
<td>33</td>
<td>75.3</td>
<td>1.7</td>
<td>71.0</td>
<td>78.1</td>
<td></td>
</tr>
<tr>
<td>2-fold</td>
<td>50</td>
<td>75.1</td>
<td>2.2</td>
<td>69.3</td>
<td>80.2</td>
<td></td>
</tr>
<tr>
<td>Test set 2</td>
<td>1</td>
<td>78.0</td>
<td>-</td>
<td>78.0</td>
<td>78.0</td>
<td></td>
</tr>
</tbody>
</table>

The model is quite stable if we analyze the profile of the standard deviation values for each n-fold cross-validation and the predictivity in Table 5.7. In the LOO cross-validation a predictivity of 75.8% was obtained. The percentage of correct predictions is lower for the other n-fold cross-validation procedures, with a difference of 10.6% between the Training set 2 predictivity and the average prediction accuracy in 2-fold cross-validation. Test set 2 is predicted with an accuracy of 78%. Table 5.8 shows the TP, FP, TN and FN rates of the Training set 2 in LOO cross-validation for SVM model. Further details on the models derived are reported in Paper III.
Table 5.8: Single-label classification SVM model; predictivity results after LOO cross-validation for Training set 2.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate</th>
<th>FP rate</th>
<th>TN rate</th>
<th>FN rate</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.34</td>
<td>0.01</td>
<td>0.99</td>
<td>0.66</td>
<td>0.34</td>
<td>0.73</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.72</td>
<td>0.04</td>
<td>0.96</td>
<td>0.28</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.77</td>
<td>0.05</td>
<td>0.95</td>
<td>0.23</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>0.89</td>
<td>0.06</td>
<td>0.94</td>
<td>0.11</td>
<td>0.89</td>
<td>0.68</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.82</td>
<td>0.18</td>
<td>0.82</td>
<td>0.18</td>
<td>0.82</td>
<td>0.79</td>
</tr>
</tbody>
</table>

5.4.3 Validation of the models with an external test set

In our analysis different test sets were studied, according to the data set used to build up the classification models, as described in paragraph 5.2. The isoform specificity was predicted for each test set by applying the multilabel (Test set 1) or the single-label classification (Test set 2) models.

Test set 1. 209 substrates (Test set 1) were analyzed by both ct-SVM and CPG NN isoform predictors. The prediction results on Test set 1 by using ct-SVM and CPG NN techniques are summarized in Table 5.9 and Table 5.10, respectively.

Table 5.9: Multilabel classification ct-SVM: predicted results of the model for the Test set 1. The number of true positives is indicated in the second column for each class.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate (TP)</th>
<th>TN rate</th>
<th>Recall</th>
<th>Precision</th>
<th>% correct predictions</th>
<th>Matthews correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.65 (15/23)</td>
<td>0.90</td>
<td>0.65</td>
<td>0.44</td>
<td>87.1</td>
<td>0.47</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.44 (8/18)</td>
<td>0.96</td>
<td>0.44</td>
<td>0.53</td>
<td>91.9</td>
<td>0.44</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.39 (38/64)</td>
<td>0.92</td>
<td>0.39</td>
<td>0.78</td>
<td>82.3</td>
<td>0.38</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>0.75 (9/12)</td>
<td>0.98</td>
<td>0.75</td>
<td>0.69</td>
<td>96.6</td>
<td>0.70</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.84 (97/116)</td>
<td>0.70</td>
<td>0.84</td>
<td>0.78</td>
<td>77.5</td>
<td>0.56</td>
</tr>
</tbody>
</table>

In the ct-SVM model predictions the FN rate is remarkable for CYP2C9 class. Considering that the classifier is based on five classes, the model prediction capability is quite accurate.

Regarding CPG NN model, only for the CYP2E1 class surprisingly good results were achieved, with a recall of 0.92, as reported in Table 5.10. Statistically acceptable values of TP rate corresponded to CYP2D6 (0.70) and CYP3A4 (0.72) isoforms, while a low predictivity was found for the remaining classes.
5.4 Results

Table 5.10: Multilabel classification CPG NN model; predicted results for the Test set 1.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate</th>
<th>TN rate</th>
<th>Recall</th>
<th>Precision</th>
<th>% correct predictions</th>
<th>Matthews correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.52</td>
<td>0.85</td>
<td>0.52</td>
<td>0.30</td>
<td>81.3</td>
<td>0.29</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.61</td>
<td>0.93</td>
<td>0.61</td>
<td>0.44</td>
<td>89.9</td>
<td>0.46</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.70</td>
<td>0.88</td>
<td>0.70</td>
<td>0.72</td>
<td>82.8</td>
<td>0.58</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>0.92</td>
<td>0.97</td>
<td>0.92</td>
<td>0.69</td>
<td>97.1</td>
<td>0.78</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.72</td>
<td>0.79</td>
<td>0.72</td>
<td>0.81</td>
<td>75.6</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Test set 2. In the single-label model the Test set 2 (191 CYP450 substrates) was used to assess its predictivity. As seen in Table 5.7, the percentage of correct predictions for Test set 2 resulted in 78% for the SVM model. This value is slightly higher than the corresponding value of correct predictions after 2-fold cross-validation. In Table 5.11 the prediction rates of the same single-classification model for each class of the Test set 2 are reported.

Table 5.11: Single-label classification SVM model; predicted results for the Test set 2. The number of true positives is indicated in the second column for each class.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate (TP)</th>
<th>TN rate</th>
<th>Recall</th>
<th>Precision</th>
<th>Matthews correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.64 (9/14)</td>
<td>0.97</td>
<td>0.64</td>
<td>0.60</td>
<td>0.59</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.73 (8/11)</td>
<td>0.98</td>
<td>0.73</td>
<td>0.67</td>
<td>0.68</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.77 (41/53)</td>
<td>0.91</td>
<td>0.77</td>
<td>0.76</td>
<td>0.67</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>1.00 (11/11)</td>
<td>0.97</td>
<td>1.00</td>
<td>0.65</td>
<td>0.79</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.78 (80/102)</td>
<td>0.85</td>
<td>0.78</td>
<td>0.85</td>
<td>0.64</td>
</tr>
</tbody>
</table>

All CYP2E1 substrates are correctly predicted (the TP rate is equal to one) and the values of TP rate for the remaining classes are included in the interval 60-80%. Moreover, the SVM model was applied to predict the eighteen multilabel substrates in Test set 1, not included in Test set 2. The prediction results were considered in the final comparison with the other multilabel classification methods.
5.5 Discussion

5.5.1 Aspects related to the data set

The classification of multilabel data represents a complex problem. So far, many different strategies were explored to classify drugs metabolized by a single isoform. However, this approach does not reflect the real scenario, in which the route of metabolism might involve several enzymes for the biotransformation. In fact, the same molecular structure can be recognized by different isoforms, on the other hand the same CYP450 isoform might metabolize chemically diverse compounds and this is particularly true for CYP3A4 isoform. Therefore, CYP450 enzymes are not selective and in the metabolic process the CYPs show a different role and relevance. As a consequence, the data set is unbalanced, with few compounds classified as CYP1A2, CP2C19, CYP2C8, CYP2C9, CYP2E1 substrates and more represented by CYP2D6 and CYP3A4 classes.

Moreover, in our analysis we dealt with information coming from different sources and in some cases it was inconsistent or incomplete. Then, the uncertainty of information led us to choose a compromise by considering more reliable the most recent source.

5.5.2 Considerations on the selected descriptors

In the metabolic process a recognition mechanism is responsible for the complementary interaction between the substrates and cytochrome P450 isoforms. Therefore, the chemical nature of the substrates and especially the distribution of particular properties on their surface are involved in the determination of their metabolic fate. The specificity of the interactions is driven by several molecular properties. Moreover, the function of autocorrelation is a useful strategy to overcome the dependence on the spatial rotation and translation of the molecules. In fact, the autocorrelation concept is able to describe the distribution of a particular property on the molecular surface and to represent molecules of different size with a vector of fixed length.

In ct-SVM model the selected descriptors (Table 5.2) have been confirmed relevant by the prediction results for CYP1A2, CYP2C9 and CYP3A4 classes. The CPG NN model is based on the same descriptors set. The clearly distinguishable clusters in the maps corresponding to each layer/class further on support the good choice of the variables. The results demonstrate that
the molecular size and the presence of particular functional groups as well as the distribution of the electrostatic or charge properties positively affect the model predictivity. However, the autocorrelation molecular electrostatic potential descriptors are more understandable than the 3D autocorrelation identity components used in the single-label model, so they can easily substitute the vectorial properties used in the paper by Terfloth et al. [76]

Seven out of nineteen descriptors in the single-label classification SVM model are in common with the manually selected variables in the multilabel classification models (TPSA, ASA, $r_2$, $r_3$, $n_{prim\_sec\_amino}$, $n_{basic\_nitrogen}$ and $n_{acidic\_groups}$ descriptors). Various shape and size related descriptors were recognized as important in the single-label analysis, while specific acid-basic properties were selected in all models, confirming that these descriptors are crucial in the prediction of isoform specificity.

5.5.3 ct-SVM and CPG NN models

The prediction results for the Test set 1 were analyzed to compare the performances of the ct-SVM and CPG NN models. The predictivity of ct-SVM for CYP1A2 and CYP3A4 classes is higher than the CPG NN model results. On the other hand, if we consider the remaining isoforms, the values of recall underline the better performances of the CPG NN model. After the comparison of the percentages of correct predictions, the values are similar for the corresponding isoforms.

5.5.4 ct-SVM and SVM models

We compared the prediction results of the multilabel classification ct-SVM model on Test set 1 (209 substrates) and the predictivity of single-label classification SVM model for Test set 2 (191 substrates) to verify whether the multilabel approach might be a valid alternative to the single-label methodology, by applying the same algorithm as modeling method. In this analysis we had to consider that the test sets comprise a different number of compounds, since in Test set 1 eighteen compounds are multilabel.

On first analysis, the recall shows similar performances of the models for CYP1A2 class and an increase of predictivity for CYP3A4 class in the multilabel model. Regarding CYP2C9, CYP2D6 and CYP2E1 classes, the recall drops down in the multilabel classifier. On the other hand, the profile of the precision values reflects better performances of the ct-SVM model if
CYP2D6 and CYP2E1 classes are considered, with the precision values of 0.78 and 0.69, respectively. In the single-label model the values of recall are 0.60 (CYP1A2 class), 0.67 (CYP2C9 class) and 0.85 (CYP3A4 class), higher than the corresponding values in the multilabel classifier. If we analyze the number of TP in the single-label model, 9 out of 14 CYP1A2 substrates, 8 out of 11 CYP2C9 substrates, 41 out of 53 CYP2D6 substrates, 11 out of 11 CYP2E1 substrates and 80 out of 102 CYP3A4 substrates resulted. The number of correctly predicted compounds using the single-label approach (Table 5.11) is very close in comparison to the number of TP in the multilabel results reported in Table 5.9.

It seems clear that the single-label model is not able to give a complete picture of the metabolism information. In fact, the ct-SVM model performances result at least comparable to the SVM model ones. However, in the single-label approach, inevitably, we lose important details about isoenzyme specificity, since each substrate is implicitly supposed to be metabolized by an unique CYP450 isoform.

5.5.5 Analysis of some classified compounds

We compared the prediction results of the ct-SVM, CPG NN and SVM models on the Test set 1, including multi- and single-label substrates. Also the multilabel compounds in the Test set 1 were predicted by the single-label model. The CYP450 substrates in Test set 1 not extracted from the Metabolite database were analyzed. [175] The experimental and predicted classes for these compounds are reported in Table 5.12.

Nine out of 44 compounds (two of them are multilabel) are incorrectly predicted by both multilabel ct-SVM and CPG NN models. The single-label SVM model assigned a wrong class to ten compounds and all the multilabel compounds were correctly assigned to one of the experimental classes by ct-SVM and CPG NN models. Regarding multilabel compounds, most ct-SVM predictions are correct even if partial. Two examples are Clomipramine and Methadone. The drug Clomipramine (173) is metabolized by four different CYP450 isoforms and the ct-SVM model correctly predicted three of them (CYP1A2, CYP2D6 and CYP3A4), while both CPG NN and SVM classifiers have only assigned one class. Similarly, Methadone (189) is predicted by the multilabel models to be metabolized by CYP2D6 and CYP3A4 isoforms, showing a good correspondence with the experimental metabolic profile.
Table 5.12: Some experimental and predicted isoforms after applying ct-SVM, CPG NN and SVM models are summarized. Ct-SVM and CPG NN models are multilabel; SVM model was carried out by using the single-label approach. In the second column the multilabel substrates are bold; M = multilabel; S = single-label.; 1A2 = CYP1A2; 2C9 = CYP2C9; 2D6 = CYP2D6; 2E1 = CYP2E1; 3A4 = CYP3A4.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Exp. classes</th>
<th>Pred. ct-SVM (M)</th>
<th>Pred. CPG NN (M)</th>
<th>SVM (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>166</td>
<td>Acetaminophen</td>
<td>1A2</td>
<td>1A2</td>
<td>1A2</td>
<td>1A2</td>
</tr>
<tr>
<td>167</td>
<td>Alpidem</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>168</td>
<td>Amiflamine</td>
<td>2D6</td>
<td>1A2</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>169</td>
<td>Aripiprazole</td>
<td>2D6 3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>170</td>
<td>Azatadine</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>171</td>
<td>Bufuralol</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>172</td>
<td>Cinnarizine</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>173</td>
<td>Clomipramine</td>
<td>1A2 2C9 2D6 3A4</td>
<td>1A2 2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>174</td>
<td>Clopidogrel</td>
<td>1A2 3A4</td>
<td>2D6</td>
<td>2D6</td>
<td>3A4</td>
</tr>
<tr>
<td>175</td>
<td>Deprenyl</td>
<td>2D6</td>
<td>3A4</td>
<td>2D6 3A4</td>
<td>2D6</td>
</tr>
<tr>
<td>176</td>
<td>Desipramine</td>
<td>1A2 2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>177</td>
<td>Dihydrocodeine</td>
<td>2D6 3A4</td>
<td>2D6 3A4</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>178</td>
<td>Ebsastine</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>179</td>
<td>Enalapril</td>
<td>3A4</td>
<td>3A4</td>
<td>2C9</td>
<td>1A2</td>
</tr>
<tr>
<td>180</td>
<td>Fluconazole</td>
<td>3A4</td>
<td>2C9 3A4</td>
<td>3A4</td>
<td>1A2</td>
</tr>
<tr>
<td>181</td>
<td>Fluvarine</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>182</td>
<td>Formoterol</td>
<td>2C9 2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>183</td>
<td>Indomethacin</td>
<td>2C9</td>
<td>3A4</td>
<td>2C9</td>
<td>2C9</td>
</tr>
<tr>
<td>184</td>
<td>Levonorgestrel</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>185</td>
<td>Lidocaine</td>
<td>2D6 3A4</td>
<td>1A2</td>
<td>2D6 3A4</td>
<td>2D6</td>
</tr>
<tr>
<td>186</td>
<td>Lisuride</td>
<td>3A4</td>
<td>1A2 3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>187</td>
<td>Lobeline</td>
<td>2D6</td>
<td>2D6 3A4</td>
<td>1A2 2D6 3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>188</td>
<td>Loroxican</td>
<td>2C9</td>
<td>2C9</td>
<td>2C9</td>
<td>2C9</td>
</tr>
<tr>
<td>189</td>
<td>Methadone</td>
<td>1A2 2D6 3A4</td>
<td>2D6 3A4</td>
<td>2D6 3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>190</td>
<td>Methoxyphenamine</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>191</td>
<td>Mexiletine</td>
<td>1A2 2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>192</td>
<td>Montelukast</td>
<td>2C9 3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>193</td>
<td>Omeprazole</td>
<td>2C9 3A4</td>
<td>3A4</td>
<td>1A2</td>
<td>3A4</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Exp. classes</th>
<th>Pred. ct-SVM (M)</th>
<th>Pred. CPG NN (M)</th>
<th>SVM (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>194</td>
<td>Ondansetron</td>
<td>1A2 2D6 3A4</td>
<td>1A2</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>195</td>
<td>Phenylbutazone</td>
<td>2C9</td>
<td>1A2</td>
<td>1A2</td>
<td>3A4</td>
</tr>
<tr>
<td>196</td>
<td>Quercetin</td>
<td>3A4</td>
<td>1A2</td>
<td>1A2</td>
<td>1A2</td>
</tr>
<tr>
<td>197</td>
<td>Ramelteon</td>
<td>1A2 2C9 3A4</td>
<td>3A4</td>
<td>2C9</td>
<td>3A4</td>
</tr>
<tr>
<td>198</td>
<td>Remoxipride</td>
<td>2D6</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>199</td>
<td>Sertindole</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>200</td>
<td>Sparteine</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
<td>2D6</td>
</tr>
<tr>
<td>201</td>
<td>Sulindanine</td>
<td>2C9</td>
<td>2C9</td>
<td>2C9</td>
<td>3A4</td>
</tr>
<tr>
<td>202</td>
<td>Sulfasalazine</td>
<td>3A4</td>
<td>3A4</td>
<td>2C9</td>
<td>3A4</td>
</tr>
<tr>
<td>203</td>
<td>Tamsulosin</td>
<td>2D6 3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>204</td>
<td>Theophylline</td>
<td>1A2 2E1 3A4</td>
<td>1A2</td>
<td>3A4</td>
<td>1A2</td>
</tr>
<tr>
<td>205</td>
<td>Tolterodine</td>
<td>2D6 3A4</td>
<td>2D6 3A4</td>
<td>2D6 3A4</td>
<td>2D6</td>
</tr>
<tr>
<td>206</td>
<td>Trimethoprim</td>
<td>2C9</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
</tr>
<tr>
<td>207</td>
<td>Valdecoxib</td>
<td>2C9 3A4</td>
<td>2C9</td>
<td>2C9</td>
<td>3A4</td>
</tr>
<tr>
<td>208</td>
<td>Zidovudine</td>
<td>3A4</td>
<td>3A4</td>
<td>3A4</td>
<td>1A2</td>
</tr>
<tr>
<td>209</td>
<td>Zileuton</td>
<td>1A2 2C9 3A4</td>
<td>3A4</td>
<td>2E1</td>
<td>1A2</td>
</tr>
</tbody>
</table>

These examples confirm that the multilabel approach is able to perform an extensive investigation of the drug metabolism, while the single-label results are limited to the prediction of a single class. A deep analysis of each compound in Table 5.12 is reported in Paper III.

### 5.6 Final remarks

In the present study, we investigated several classification strategies to predict the isoform specificity of known CYP1A2, CYP2C9, CYP2D6, CYP2E1 and CYP3A4 substrates. The multilabel approach was applied to a data set including five classes, by using the ct-SVM and the CPG NN methods. The best model (ct-SVM) was derived after the selection of 27 descriptors and yielded 77.5-96.6% of correct predictions for the five classes of the corresponding test set. Similarly, the CPG NN model achieved 75.6-97.1% of correct predictions. A five-classes data set was used to perform an extensive single-label classification analysis, in combination with automatic variable selection. The highest predictivity on the corresponding test set, achieved by using the SVM technique based on nineteen descriptors, was 78% of cor-
rect predictions. All the presented models show acceptable performances, however the multilabel prediction results reflect more coherently the real metabolic fate of drugs.

In conclusion, our results underline the high complexity of this classification problem and suggest the application of the multilabel approach to predict CYP450 isoform specificity. The advantage of the CPG NN technique is the graphical visualization of the results. Both ct-SVM and the CPG NN strategies might be extended to quantitative data. The multilabel methodology can be used to explore the metabolic profile of new chemical entities and the prediction capability might be improved by collecting other multilabel substrates in the dataset.
The selectivity is an important aspect of drug discovery, and in the development of G protein-coupled receptors (GPCRs) ligands to distinguish between related receptor subtypes is often the key to therapeutic success. Nowadays, the prediction of receptor subtype selectivity represents a very challenging task. In the present study, we present an alternative application of Support Vector Machine (SVM) and Support Vector Regression (SVR) methodologies to simultaneously describe both A2A versus A3 subtypes selectivity profile and the corresponding receptor binding affinities. We have implemented an integrated application of SVM-SVR approach, based on the use of the autocorrelated molecular descriptors encoding for the Molecular Electrostatic Potential (autoMEP), to simultaneously discriminate A2A versus A3 antagonists and to predict the binding affinity to the corresponding receptor subtype of a large dataset of known pyrazolo-triazolo-pyrimidine analogs. To validate our approach, we have synthesized 51 new pyrazolo-triazolo-pyrimidine derivatives anticipating both A2A/A3R subtypes selectivity and receptor binding affinity profiles.
6.1 Introduction

G protein-coupled receptors (GPCRs) represent the target for many drugs under development. The efficacy problems and limiting side-effects of some candidates are due to the lack of differentiation between receptor subtypes. There is thus considerable interest in attaining therapeutic selectivity by identifying the single receptor subtype that affects a particular physiology. The goal is to reduce, as more as possible, the side-effects, while retaining the desired function. To date, very few valuable computational tools are available for the prediction of receptor subtype selectivity, which is still considered a complex problem. Conversely, different in silico approaches are accessible to estimate the distinct receptor-ligand affinity, in particular QSAR is the commonly used approach in this field. [176, 177]

In the last few years, the possibility to discover new potent and selective adenosine receptors (ARs) antagonists has been intensively explored. Briefly, the adenosine receptor (AR) family belongs to GPCR family A, including four different subtypes, referred to as A\(_1\), A\(_2A\), A\(_2B\) and A\(_3\), which are widely but differentially distributed throughout the body\(^1\). [139, 142] In this study we will focus on A\(_2A\)R and A\(_3\)R subtypes and selective ligands to these ARs are becoming increasingly attractive drugs due to their potential role of this receptor in several physiopathological processes. [140, 141, 144, 178, 179] In particular, A\(_2A\)R antagonists seem to play a role in the reduction of neuronal damage in Parkinson’s or Huntington’s diseases, while A\(_3\)R antagonists have a potential application in the tumor growth inhibition and in the treatment of glaucoma. [142-144]

Consequently, several receptor-based and ligand-based drug design approaches have been carried out with the aim to improve potency and selectivity of different molecular scaffolds and, in particular, the pyrazolo-triazolo-pyrimidine scaffold has been extensively studied. Moreover, it has been demonstrated that proper substitutions at the N\(^5\) and N\(^8\) positions drive the antagonist selectivity to the human A\(_3\)R subtype. [149] On the other hand, the substitution at the position N\(^7\) shifts the selectivity profile to the human A\(_2A\)R subtype. [146] However, this very empirical rule based on experimental evidences does not provide a criterion to assign the correct pharmacological A\(_2A\) and A\(_3\) receptors profiles of novel pyrazolo-triazolo-pyrimidine derivatives.

\(^1\)Further details on the four human AR subtypes are reported in section 7.2.
Here, we describe an alternative application of the Support Vector Machine (SVM) and Support Vector Regression (SVR) methodologies to predict both A2AR versus A3R subtypes selectivity profile and the corresponding receptor binding affinities. As anticipated, SVM is very utilized to solve both classification and regression problems. [82, 85] In this study, we have implemented an integrated application of SVM-SVR approach, based on the use of autoMEP descriptors, to simultaneously discriminate A2AR versus A3R antagonists and to predict the binding affinity to the corresponding receptor subtype of a large dataset of known pyrazolo-triazolo-pyrimidine analogs. To validate our approach, we have newly synthesized 51 pyrazolo-triazolo-pyrimidine derivatives anticipating both A2AR/A3R subtypes selectivity and receptor binding affinity profiles.

6.2 Dataset

A collection of 104 selective N7- and N8-substituted pyrazolo-triazolo-pyrimidine analogues (molecules 1-104) has been selected as training set in the first SVM classification (SVMc) model. [146-151, 180]

In the SVR regression (SVR) model 104 N8-substituted pyrazolo-triazolo-pyrimidine derivatives (molecules 1-71, 105-137), selective and not selective, have been used as training set of both human A2AR and A3R nonlinear SVR models. [146, 149, 150, 180, 181]

Finally, a test set of 51 N8-substituted pyrazolo-triazolo-pyrimidine analogues (molecules 138-188) has been selected to validate both SVMc and SVR models (Table 6.1).

---

2Experimental binding affinity data kindly provided by the work coordinated by Prof. G. Spalluto (University of Trieste) for the synthesis and by Prof. K. N. Klotz (University of Würzburg) for the pharmacological characterization.
Table 6.1: Biological profile at the hA$_2$AR and hA$_3$R subtypes of the test set compounds.

<table>
<thead>
<tr>
<th>No.</th>
<th>R</th>
<th>R1</th>
<th>hA$_2$AR ($K_i$, nM)</th>
<th>hA$_3$R ($K_i$, nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>138</td>
<td>CH$_2$-α-naphthyl</td>
<td>CH$_3$CH$_2$CH$_3$Ph</td>
<td>305</td>
<td>343</td>
</tr>
<tr>
<td>139</td>
<td>CHPht$_2$</td>
<td>CH$_3$</td>
<td>216</td>
<td>0.25</td>
</tr>
<tr>
<td>140</td>
<td>CH$_2$-Ph-Ph</td>
<td>CH$_3$</td>
<td>193</td>
<td>11.2</td>
</tr>
<tr>
<td>141</td>
<td>CHPht$_2$</td>
<td>CH$_2$CH$_2$CH(CH$_3$)$_2$</td>
<td>159</td>
<td>5.86</td>
</tr>
<tr>
<td>142</td>
<td>CH$_2$-Ph-Ph</td>
<td>CH$_2$CH$_2$CH(CH$_3$)$_2$</td>
<td>9.18</td>
<td>268</td>
</tr>
<tr>
<td>143</td>
<td>CHPht$_2$</td>
<td>CH$_3$CH$_2$Ph</td>
<td>53.1</td>
<td>6.49</td>
</tr>
<tr>
<td>144</td>
<td>CH$_2$-Ph-Ph</td>
<td>CH$_3$CH$_2$Ph</td>
<td>46.9</td>
<td>125</td>
</tr>
<tr>
<td>145</td>
<td>CHPht$_2$</td>
<td>CH$_3$CH$_2$CH$_3$</td>
<td>114</td>
<td>1.20</td>
</tr>
<tr>
<td>146</td>
<td>CH$_2$-Ph-Ph</td>
<td>CH$_3$CH$_2$CH$_3$</td>
<td>30.6</td>
<td>189</td>
</tr>
<tr>
<td>147</td>
<td>CH$_2$-3-Cl-Ph</td>
<td>CH$_3$</td>
<td>30.5</td>
<td>1.94</td>
</tr>
<tr>
<td>148</td>
<td>CHPht$_2$</td>
<td>CH$_3$CH$_2$CH$_3$</td>
<td>46.5</td>
<td>0.93</td>
</tr>
<tr>
<td>149</td>
<td>CH$_2$-Ph-Ph</td>
<td>CH$_3$CH$_2$CH$_3$</td>
<td>52.5</td>
<td>65.4</td>
</tr>
<tr>
<td>150</td>
<td>CH$_2$-4-Cl-Ph</td>
<td>CH$_3$</td>
<td>15.6</td>
<td>1.27</td>
</tr>
<tr>
<td>151</td>
<td>CH$_2$-Ph-Ph</td>
<td>CH$_2$CH$_3$</td>
<td>70.9</td>
<td>14.2</td>
</tr>
<tr>
<td>152</td>
<td>CH$_2$-4-OCH$_3$-Ph</td>
<td>CH$_3$</td>
<td>62.5</td>
<td>0.95</td>
</tr>
<tr>
<td>153</td>
<td>CH$_2$-β-naphthyl</td>
<td>CH$_2$CH$_2$CH(CH$_3$)$_2$</td>
<td>15.7</td>
<td>409</td>
</tr>
<tr>
<td>154</td>
<td>CH$_2$-β-naphthyl</td>
<td>CH$_3$CH$_2$Ph</td>
<td>42.1</td>
<td>180</td>
</tr>
<tr>
<td>155</td>
<td>CH$_2$-2-thienyl</td>
<td>CH$_3$CH$_2$Ph</td>
<td>8.8</td>
<td>330</td>
</tr>
<tr>
<td>156</td>
<td>CH$_2$-3-thienyl</td>
<td>CH$_3$CH$_2$Ph</td>
<td>12.9</td>
<td>726</td>
</tr>
<tr>
<td>157</td>
<td>CH$_2$-α-naphthyl</td>
<td>CH$_3$CH$_2$Ph</td>
<td>74.6</td>
<td>3.05</td>
</tr>
<tr>
<td>158</td>
<td>CH$_2$-β-naphthyl</td>
<td>CH$_3$CH$_2$</td>
<td>18.4</td>
<td>36.3</td>
</tr>
<tr>
<td>159</td>
<td>CH$_2$-3-thienyl</td>
<td>CH$_3$CH$_3$</td>
<td>23</td>
<td>765</td>
</tr>
<tr>
<td>160</td>
<td>CH$_2$-2-thienyl</td>
<td>CH$_3$CH$_3$</td>
<td>15.9</td>
<td>196</td>
</tr>
<tr>
<td>161</td>
<td>CH$_2$-2-thienyl</td>
<td>CH$_3$</td>
<td>56</td>
<td>5.26</td>
</tr>
<tr>
<td>162</td>
<td>CH$_2$-3-thienyl</td>
<td>CH$_3$</td>
<td>31.3</td>
<td>1.25</td>
</tr>
<tr>
<td>163</td>
<td>CH$_2$-β-naphthyl</td>
<td>CH$_3$</td>
<td>77.5</td>
<td>14.5</td>
</tr>
<tr>
<td>164</td>
<td>CH$_2$-α-naphthyl</td>
<td>CH$_3$</td>
<td>80.5</td>
<td>3.47</td>
</tr>
<tr>
<td>165</td>
<td>CH$_2$-α-naphthyl</td>
<td>CH$_3$CH$_2$CH$_3$</td>
<td>38.7</td>
<td>17.3</td>
</tr>
</tbody>
</table>

*a*Displacement of specific [3H]-NECA binding at human A$_2$A receptors expressed in CHO cells;

*b* Displacement of specific [3H]-NECA binding at human A$_3$ receptors expressed in CHO cells.

(continued on next page)
6.3 Results and discussion

Support Vector Machine represents a group of supervised learning techniques, which find now diverse applications in classification and regression problems. SVM has been originally developed for classification, then the introduction of a suitable ε-insensitive loss function together with the advantages of the kernel representation have enabled its application in the regression analysis, as reported in chapter 2.6. Recently, the application of SVM and SVR approaches has helped to solve several classification problems, as for example active and non-active compounds discrimination, and to derive QSARs for the prediction of different chemical and biological properties. [89-92, 100, 103] SVR seems to be a promising tool, with good generalization performance and increased robustness compared with the neural networks.

---

<table>
<thead>
<tr>
<th>No.</th>
<th>R</th>
<th>R1</th>
<th>hA2,4 (K, nM)(^a)</th>
<th>hA3 (K, nM)(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>166</td>
<td>CH(_2)β-naphthyl</td>
<td>CH(_2)CH(_2)CH(_3)</td>
<td>7.99</td>
<td>95.9</td>
</tr>
<tr>
<td>167</td>
<td>CH(_2)α-naphthyl</td>
<td>CH(_2)CH(_2)CH(_2)CH(_3)</td>
<td>11.7</td>
<td>100</td>
</tr>
<tr>
<td>168</td>
<td>CH(_2)4-CF(_3)-Ph</td>
<td>CH(_3)</td>
<td>7.59</td>
<td>1.22</td>
</tr>
<tr>
<td>169</td>
<td>CH(_2)2-thienyl</td>
<td>CH(_2)CH(_2)CH(_2)CH(_3)</td>
<td>4.15</td>
<td>99.8</td>
</tr>
<tr>
<td>170</td>
<td>CH(_2)3-thienyl</td>
<td>CH(_2)CH(_2)CH(_2)CH(_3)</td>
<td>3.13</td>
<td>189</td>
</tr>
<tr>
<td>171</td>
<td>CH(_2)O-Ph-4-Cl</td>
<td>CH(_3)</td>
<td>39.3</td>
<td>223</td>
</tr>
<tr>
<td>172</td>
<td>CH(_2)3-Cl-Ph</td>
<td>CH(_2)CH(_2)CH(CH(_3))(_2)</td>
<td>1.86</td>
<td>273</td>
</tr>
<tr>
<td>173</td>
<td>CH(_2)4-Cl-Ph</td>
<td>CH(_2)CH(_2)CH(CH(_3))(_2)</td>
<td>2.75</td>
<td>56.5</td>
</tr>
<tr>
<td>174</td>
<td>CH(_2)3-Cl-Ph</td>
<td>CH(_2)CH(_2)Ph</td>
<td>5.75</td>
<td>273</td>
</tr>
<tr>
<td>175</td>
<td>CH(_2)4-CF(_3)-Ph</td>
<td>CH(_2)CH(_2)CH(CH(_3))(_2)</td>
<td>5.43</td>
<td>266</td>
</tr>
<tr>
<td>176</td>
<td>CH(_2)4-F-Ph</td>
<td>CH(_2)CH(_2)CH(CH(_3))(_2)</td>
<td>3.69</td>
<td>116</td>
</tr>
<tr>
<td>177</td>
<td>CH(_2)4-F-Ph</td>
<td>CH(_3)</td>
<td>54.1</td>
<td>0.97</td>
</tr>
<tr>
<td>178</td>
<td>CH(_2)2,6-Cl-Ph</td>
<td>CH(_2)CH(_2)CH(CH(_3))(_2)</td>
<td>18.7</td>
<td>207</td>
</tr>
<tr>
<td>179</td>
<td>CH(_2)2,6-Cl-Ph</td>
<td>CH(_3)</td>
<td>45.2</td>
<td>44.4</td>
</tr>
<tr>
<td>180</td>
<td>CH(_2)4-F-Ph</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>211</td>
<td>58.5</td>
</tr>
<tr>
<td>181</td>
<td>CH-Ph(_2)</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>326</td>
<td>12.6</td>
</tr>
<tr>
<td>182</td>
<td>CH(_2)β-naphthyl</td>
<td>CH(_2)CH(_2)CH(_3)Ph</td>
<td>73.6</td>
<td>71.7</td>
</tr>
<tr>
<td>183</td>
<td>CH(_3)Ph</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>43.9</td>
<td>5.49</td>
</tr>
<tr>
<td>184</td>
<td>CH(_2)3-Cl-Ph</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>182</td>
<td>110</td>
</tr>
<tr>
<td>185</td>
<td>CH(_2)4-Cl-Ph</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>89.9</td>
<td>30.5</td>
</tr>
<tr>
<td>186</td>
<td>CH(_2)O-Ph-1-Cl</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>27.2</td>
<td>400</td>
</tr>
<tr>
<td>187</td>
<td>CH(_2)2,6-Cl-Ph</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>186</td>
<td>601</td>
</tr>
<tr>
<td>188</td>
<td>CH(_2)Ph-Ph</td>
<td>CH(_2)CH(_2)CH(_2)Ph</td>
<td>256</td>
<td>410</td>
</tr>
</tbody>
</table>

\(^a\)Displacement of specific [3H]-NECA binding at human A\(_2,4\) receptors expressed in CHO cells;  
\(^b\)displacement of specific [3H]-NECA binding at human A\(_3\) receptors expressed in CHO cells.
We consider both topological and electrostatic complementarities extremely crucial in describing the receptor subtypes selectivity. Basing on the motivations underlined in section 4.3, we believe that the autoMEP vectors can be used as interesting molecular descriptors. We have also reported that pyrazolo-triazolo-pyrimidine is a versatile scaffold to cover a large spectrum of the adenosine receptor selectivity. As anticipated, pyrazolo-triazolo-pyrimidines bearing specific substitutions at the N^5 and N^8 positions have been described as highly potent and selective human A_3R antagonists, while the position N^7 shifts the selectivity profile to the human A_2A R subtype. [146, 149] However, the observation of the scaffold and its substitutions is not an unfailing strategy to correctly assign the selectivity profile of new pyrazolo-triazolo-pyrimidine antagonists.

In this chapter we present an integrated approach based on the introduction of two distinct support vector machine tools both using as input matrix the autoMEP vectors. The first tool is a SVM-driven selectivity classifier and the second one is a couple of SVR-driven receptor-affinity predictors. The workflow of the abovementioned procedure is summarized in Figure 6.3.1.

\textbf{Figure 6.3.1:} Flowchart of the in series autoMEP/SVM\textit{class} and autoMEP/SVR approach for the selection of new selective and potent human A_2AR and A_3R antagonists.
We have introduced the autoMEP of each pyrazolo-triazolo-pyrimidine antagonist to optimize the experimental $A_2A/R/A_3R$ subtypes selectivity profile using a SVM classifier (autoMEP/SVMclass). Then, the two different $A_2A/R$ and $A_3R$ receptor-affinity predictors can be generated using the autoMEP vectors as input values. The application of the autoMEP/SVMclass model ahead of the two receptor-affinity predictors can refine the prediction of both $A_2A/R/A_3R$ subtypes selectivity profile and $A_2A/R/A_3R$ binding affinity values of new pyrazolo-triazolo-pyrimidine derivatives.

### 6.3.1 SVM classification model

To build our SVM-driven selectivity classifier, we have selected 104 "selective" pyrazolo-triazolo-pyrimidines derivatives (molecules 1-104). In more detail, 48.1% of the SVMclass model training set include h$A_2A/R$ antagonists (50 compounds) and the remaining percentage (51.9%) is composed of h$A_3R$ antagonists (54 compounds). The definition of "selectivity" used for our classification approach is based on a simple binary criteria: the selectivity index is set to "+1" if it is referred to a selective h$A_2A/R$ antagonist, and is "-1" for a selective h$A_3R$ antagonists. Moreover, we have considered as selectivity threshold a difference of at least 2 orders of magnitude between the corresponding $K_i$ values.

Information encoded by twelve autoMEP vectors (calculated by selecting default parameters, as described in chapter 2) has been used as input matrix for our SVMclass model. The best classifier was obtained by using a Gaussian radial basis function kernel ($C = 150; \gamma = 0.01$) and this model has been subjected to an extensive $n$-fold cross-validation procedure (Table 6.2).

<table>
<thead>
<tr>
<th>Partition</th>
<th>CV</th>
<th>No. of runs</th>
<th>Mean</th>
<th>StDev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>LOO</td>
<td>1</td>
<td>93.3</td>
<td>-</td>
<td>93.3</td>
<td>93.3</td>
</tr>
<tr>
<td></td>
<td>10-fold</td>
<td>10</td>
<td>91.4</td>
<td>2.5</td>
<td>86.5</td>
<td>95.2</td>
</tr>
<tr>
<td></td>
<td>5-fold</td>
<td>10</td>
<td>91.7</td>
<td>2.1</td>
<td>88.5</td>
<td>94.2</td>
</tr>
<tr>
<td>Test set</td>
<td>1</td>
<td>78.4</td>
<td>-</td>
<td>78.4</td>
<td>78.4</td>
<td>78.4</td>
</tr>
</tbody>
</table>

Table 6.2: AutoMEP/SVMclass model; the statistical parameters after the cross-validation procedure on the selected classifier are collected.
The percentages obtained after repeated 10-fold and 5-fold cross-validation processes confirm the statistical reliability of this model. Interestingly, it yielded 93.3% correct predictions after LOO cross-validation. The statistical robustness is also confirmed by the percentage (%) values of sensitivity (92.0%) and specificity (94.4%). Therefore, we decided to select this model as SVM-driven selectivity classifier for the final validation step.

6.3.2 SVM regression models

A different collection of 104 pyrazolo-triazolo-pyrimidine analogs (molecules 1-71, 105-137) has been selected as training set of known hA$_{2A}$R and hA$_{3R}$ antagonists to derive our autoMEP/SVR models. The selection of all training set candidates was not performed according to a selectivity criterion, due to the fact that the principal aim of our regression model is to accurately predict the receptor binding affinity. Indeed we have utilized 17 hA$_{2A}$R selective antagonists (16.4%), 54 hA$_{3R}$ selective antagonists (51.9%) and 33 non selective antagonists (31.7%). Also in this step, information encoded by twelve autoMEP vectors of all training set antagonists has been used as input matrix. For the generation of both regression models, we have utilized a Gaussian radial basis function kernel.

An acceptable hA$_{2A}$R SVR model ($C = 200$, $\varepsilon = 0.0005$, $\gamma = 0.0006$) was obtained as indicated by the LOO cross-validated correlation coefficient ($r_{cv}$) of 0.78 and a root mean square of residuals (RMSR) of 0.050, as reported in Figure 6.3.2 and Table 6.3.
Chapter 6  6.3 Results and Discussion

**Figure 6.3.2:** AutoMEP/SVR hA2AR model; experimental pK$_i$ values vs predicted pK$_i$ values after LOO cross-validation on the training set.

**Table 6.3:** Statistical parameters of the autoMEP/SVR hA2AR model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of molecules</td>
<td>104</td>
</tr>
<tr>
<td>$r$</td>
<td>0.83</td>
</tr>
<tr>
<td>$r_{cv}$</td>
<td>0.78</td>
</tr>
<tr>
<td>Slope</td>
<td>0.62</td>
</tr>
<tr>
<td>Offset</td>
<td>-0.71</td>
</tr>
<tr>
<td>$q^b$</td>
<td>0.82</td>
</tr>
<tr>
<td>RMSR$^c$</td>
<td>0.050</td>
</tr>
</tbody>
</table>

$^a$Cross-validated $r$ after LOO cross-validation procedure: $r_{cv} = \sqrt{\frac{S_{XY}^2}{S_{XX}S_{YY}}}$, where $S_{XX} = \sum (X - X_{\text{mean}})^2$, $S_{YY} = \sum (Y - Y_{\text{mean}})^2$, $S_{XY} = \sum (X - X_{\text{mean}})(Y - Y_{\text{mean}})$.

$^b$Root mean square of residuals; RMSR.

On the other hand, the best hA$_3$R SVR model ($C = 150$, $\varepsilon = 0.3$, $\gamma = 0.005$) was derived as described by the LOO cross-validated correlation coefficient ($r_{cv}$) of 0.85 and a root mean square of residuals (RMSR) of 0.046 (Figure 6.3.3 and Table 6.4). The results of SVR analysis are noteworthy considering that the same training set was used to generate two different robust models. The validation of SVR models is discussed in the following paragraph.
6.3 Results and discussion

Figure 6.3.3: AutoMEP/SVR hA3R model; experimental pKi values vs predicted pKi values after LOO cross-validation on the training set.

Table 6.4: Statistical parameters of the autoMEP/SVR hA3R model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of molecules</td>
<td>104</td>
</tr>
<tr>
<td>$r$</td>
<td>0.95</td>
</tr>
<tr>
<td>$r_{cv}$</td>
<td>0.85</td>
</tr>
<tr>
<td>Slope</td>
<td>0.75</td>
</tr>
<tr>
<td>Offset</td>
<td>-0.11</td>
</tr>
<tr>
<td>$q$</td>
<td>0.85</td>
</tr>
<tr>
<td>RMSR</td>
<td>0.046</td>
</tr>
</tbody>
</table>

*a Cross-validated $r$ after LOO cross-validation procedure: $r_{cv} = \sqrt{\frac{SXY}{SXX} \frac{SYY}{SYY}}$, with $SXY = \sum (X - X_{mean})(Y - Y_{mean})$, $SXX = \sum (X - X_{mean})^2$, $SYY = \sum (Y - Y_{mean})^2$, and $X = Y_{Experimental}$ and $Y = Y_{predicted}$.

6.3.3 Validation of in series SVMclass and SVR models

As anticipated, the principal aim of the present work has been to evaluate the robustness of the tandem autoMEP/SVMclass and autoMEP/SVR models in the prediction of both A2AR/A3R subtypes selectivity and receptor binding affinity profiles of new pyrazolo-triazolo-pyrimidine derivatives, and in particular 51 analogs (molecules 138-188) were considered (Table 6.5).
Table 6.5: Predicted and experimental $hA_2AR$ and $hA_3R$ both $pK_i$ and $K_i$ for the test set. Differences between predicted and experimental $pK_i$ values for both SVR models ($hA_2AR$ and $hA_3R$) are reported. In the middle column $hA_2AR$ selective antagonists (red) and $hA_3R$ selective antagonists (blue) are highlighted.

<table>
<thead>
<tr>
<th>Pred. $pK_i$ (nM)</th>
<th>Exp. $pK_i$ (nM)</th>
<th>Δ $pK_i$</th>
<th>Pred. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>No.</th>
<th>Pred. $pK_i$ (nM)</th>
<th>Exp. $pK_i$ (nM)</th>
<th>Δ $pK_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$hA_2AR$</td>
<td>$hA_2AR$</td>
<td></td>
<td>$hA_2AR$</td>
<td>$hA_2AR$</td>
<td></td>
<td>$hA_2AR$</td>
<td>$hA_2AR$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.25</td>
<td>-2.48</td>
<td>0.23</td>
<td>177.83</td>
<td>305</td>
<td>138</td>
<td>343</td>
<td>107.15</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>-1.79</td>
<td>0.54</td>
<td>61.66</td>
<td>216</td>
<td>139</td>
<td>0.25</td>
<td>0.47</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>-2.04</td>
<td>0.25</td>
<td>109.65</td>
<td>193</td>
<td>140</td>
<td>11.2</td>
<td>35.48</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>-2.07</td>
<td>0.13</td>
<td>117.49</td>
<td>159</td>
<td>141</td>
<td>5.86</td>
<td>2.88</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>-2.05</td>
<td>-0.09</td>
<td>11.22</td>
<td>9.18</td>
<td>142</td>
<td>268</td>
<td>144.54</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>-1.47</td>
<td>0.26</td>
<td>29.51</td>
<td>53.1</td>
<td>143</td>
<td>6.49</td>
<td>1.70</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>-1.18</td>
<td>0.49</td>
<td>15.14</td>
<td>46.9</td>
<td>144</td>
<td>1.25</td>
<td>70.79</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>-1.85</td>
<td>0.21</td>
<td>70.79</td>
<td>114</td>
<td>145</td>
<td>1.20</td>
<td>0.56</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>-1.21</td>
<td>0.39</td>
<td>16.22</td>
<td>39.6</td>
<td>146</td>
<td>1.89</td>
<td>51.29</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>-1.59</td>
<td>-0.11</td>
<td>38.90</td>
<td>30.5</td>
<td>147</td>
<td>1.94</td>
<td>2.24</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>-1.87</td>
<td>-0.20</td>
<td>74.13</td>
<td>46.5</td>
<td>148</td>
<td>0.93</td>
<td>0.60</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>-1.15</td>
<td>0.57</td>
<td>14.13</td>
<td>52.5</td>
<td>149</td>
<td>65.4</td>
<td>44.67</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>-1.67</td>
<td>0.52</td>
<td>46.77</td>
<td>156</td>
<td>150</td>
<td>12.7</td>
<td>6.92</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>-1.41</td>
<td>0.44</td>
<td>25.70</td>
<td>70.9</td>
<td>151</td>
<td>14.2</td>
<td>72.44</td>
<td>-0.71</td>
</tr>
<tr>
<td></td>
<td>-1.78</td>
<td>0.02</td>
<td>60.36</td>
<td>62.5</td>
<td>152</td>
<td>0.95</td>
<td>3.30</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>-1.32</td>
<td>-0.12</td>
<td>20.80</td>
<td>15.7</td>
<td>153</td>
<td>409</td>
<td>85.11</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>-1.12</td>
<td>0.50</td>
<td>13.18</td>
<td>42.1</td>
<td>154</td>
<td>180</td>
<td>46.77</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>-1.41</td>
<td>-0.47</td>
<td>25.70</td>
<td>8.8</td>
<td>155</td>
<td>330</td>
<td>269.15</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>-1.58</td>
<td>-0.47</td>
<td>38.02</td>
<td>12.9</td>
<td>156</td>
<td>7.26</td>
<td>70.95</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>-1.65</td>
<td>0.22</td>
<td>44.67</td>
<td>74.6</td>
<td>157</td>
<td>3.05</td>
<td>11.48</td>
<td>-0.58</td>
</tr>
<tr>
<td></td>
<td>-1.52</td>
<td>-0.26</td>
<td>33.11</td>
<td>18.4</td>
<td>158</td>
<td>36.36</td>
<td>23.99</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>-1.73</td>
<td>-0.37</td>
<td>53.70</td>
<td>23</td>
<td>159</td>
<td>765</td>
<td>489.7</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>-1.52</td>
<td>-0.32</td>
<td>33.11</td>
<td>15.9</td>
<td>160</td>
<td>196</td>
<td>186.21</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>-1.64</td>
<td>0.11</td>
<td>43.65</td>
<td>56</td>
<td>161</td>
<td>5.26</td>
<td>1.12</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>-1.71</td>
<td>-0.21</td>
<td>51.29</td>
<td>31.3</td>
<td>162</td>
<td>1.25</td>
<td>3.55</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td>-1.63</td>
<td>0.26</td>
<td>42.66</td>
<td>77.5</td>
<td>163</td>
<td>14.5</td>
<td>3.55</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>-1.79</td>
<td>0.12</td>
<td>61.66</td>
<td>80.5</td>
<td>164</td>
<td>3.47</td>
<td>2.51</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>-1.69</td>
<td>-0.10</td>
<td>48.98</td>
<td>38.7</td>
<td>165</td>
<td>17.3</td>
<td>67.61</td>
<td>-0.39</td>
</tr>
<tr>
<td></td>
<td>-0.99</td>
<td>-0.09</td>
<td>9.77</td>
<td>7.99</td>
<td>166</td>
<td>95.9</td>
<td>57.54</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>-1.48</td>
<td>-0.41</td>
<td>30.20</td>
<td>11.7</td>
<td>167</td>
<td>100</td>
<td>61.66</td>
<td>0.21</td>
</tr>
</tbody>
</table>

*Predicted $pK_i$, • Experimental $pK_i$,.

(continued on next page)
6.3 Results and discussion

<table>
<thead>
<tr>
<th>Pred. $pK_i$ (nM)</th>
<th>Exp. $pK_i$ (nM)</th>
<th>$\Delta pK_i$</th>
<th>Pred. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>$\Delta K_i$</th>
<th>Pred. $pK_i$ (nM)</th>
<th>Exp. $pK_i$ (nM)</th>
<th>$\Delta pK_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hA2AR</td>
<td>hA2AR</td>
<td></td>
<td>hA2AR</td>
<td>hA2AR</td>
<td></td>
<td>hA2AR</td>
<td>hA2AR</td>
<td></td>
</tr>
<tr>
<td>-1.76</td>
<td>-1.88</td>
<td>0.12</td>
<td>57.54</td>
<td>75.9</td>
<td>168</td>
<td>1.22</td>
<td>18.30</td>
<td>-1.17</td>
</tr>
<tr>
<td>-1.41</td>
<td>-0.62</td>
<td>-0.79</td>
<td>25.70</td>
<td>4.15</td>
<td>169</td>
<td>99.8</td>
<td>173.78</td>
<td>-0.24</td>
</tr>
<tr>
<td>-0.76</td>
<td>-0.50</td>
<td>-0.26</td>
<td>5.75</td>
<td>3.13</td>
<td>170</td>
<td>189</td>
<td>22.30</td>
<td>0.93</td>
</tr>
<tr>
<td>-1.06</td>
<td>-1.59</td>
<td>0.53</td>
<td>11.48</td>
<td>39.3</td>
<td>171</td>
<td>223</td>
<td>52.48</td>
<td>0.63</td>
</tr>
<tr>
<td>-0.23</td>
<td>-0.27</td>
<td>0.04</td>
<td>1.70</td>
<td>1.86</td>
<td>172</td>
<td>273</td>
<td>97.72</td>
<td>0.45</td>
</tr>
<tr>
<td>0.11</td>
<td>-0.44</td>
<td>0.55</td>
<td>0.78</td>
<td>2.75</td>
<td>173</td>
<td>56.5</td>
<td>125.89</td>
<td>-0.35</td>
</tr>
<tr>
<td>-1.23</td>
<td>-0.76</td>
<td>-0.47</td>
<td>16.98</td>
<td>5.75</td>
<td>174</td>
<td>273</td>
<td>169.82</td>
<td>0.21</td>
</tr>
<tr>
<td>-0.86</td>
<td>-0.73</td>
<td>-0.13</td>
<td>7.24</td>
<td>5.43</td>
<td>175</td>
<td>266</td>
<td>89.13</td>
<td>0.47</td>
</tr>
<tr>
<td>-0.14</td>
<td>-0.57</td>
<td>0.43</td>
<td>1.38</td>
<td>3.69</td>
<td>176</td>
<td>116</td>
<td>173.78</td>
<td>-0.18</td>
</tr>
<tr>
<td>-1.93</td>
<td>-1.73</td>
<td>-0.20</td>
<td>85.11</td>
<td>54.1</td>
<td>177</td>
<td>0.97</td>
<td>14.45</td>
<td>-1.17</td>
</tr>
<tr>
<td>-0.85</td>
<td>-1.27</td>
<td>0.42</td>
<td>7.08</td>
<td>18.7</td>
<td>178</td>
<td>207</td>
<td>112.20</td>
<td>-0.27</td>
</tr>
<tr>
<td>-1.32</td>
<td>-1.66</td>
<td>0.34</td>
<td>20.89</td>
<td>45.2</td>
<td>179</td>
<td>44.4</td>
<td>8.51</td>
<td>0.72</td>
</tr>
<tr>
<td>-1.96</td>
<td>-2.32</td>
<td>0.36</td>
<td>91.20</td>
<td>21.1</td>
<td>180</td>
<td>58.5</td>
<td>223.87</td>
<td>-0.58</td>
</tr>
<tr>
<td>-2.34</td>
<td>-2.51</td>
<td>0.17</td>
<td>218.78</td>
<td>326</td>
<td>181</td>
<td>12.6</td>
<td>282</td>
<td>0.65</td>
</tr>
<tr>
<td>-1.82</td>
<td>-1.87</td>
<td>0.05</td>
<td>66.07</td>
<td>73.6</td>
<td>182</td>
<td>717</td>
<td>524.81</td>
<td>0.14</td>
</tr>
<tr>
<td>-1.33</td>
<td>-1.64</td>
<td>0.31</td>
<td>21.38</td>
<td>43.9</td>
<td>183</td>
<td>5.49</td>
<td>14.79</td>
<td>-0.43</td>
</tr>
<tr>
<td>-2.07</td>
<td>-2.26</td>
<td>0.19</td>
<td>117.49</td>
<td>182</td>
<td>184</td>
<td>110</td>
<td>245.47</td>
<td>-0.35</td>
</tr>
<tr>
<td>-1.26</td>
<td>-1.95</td>
<td>0.60</td>
<td>18.20</td>
<td>89.9</td>
<td>185</td>
<td>30.5</td>
<td>213.80</td>
<td>-0.85</td>
</tr>
<tr>
<td>-1.29</td>
<td>-1.43</td>
<td>0.14</td>
<td>19.50</td>
<td>27.2</td>
<td>186</td>
<td>400</td>
<td>602.56</td>
<td>-0.18</td>
</tr>
<tr>
<td>-2.48</td>
<td>-2.27</td>
<td>-0.21</td>
<td>302.00</td>
<td>186</td>
<td>187</td>
<td>601</td>
<td>1256.93</td>
<td>-0.32</td>
</tr>
<tr>
<td>-2.18</td>
<td>-2.41</td>
<td>0.23</td>
<td>151.36</td>
<td>256</td>
<td>188</td>
<td>410</td>
<td>48.98</td>
<td>0.92</td>
</tr>
</tbody>
</table>

*Predicted $pK_i$ - Experimental $pK_i$.

The experimental A2AR and A3R binding affinities are collected in Table 6.1. Moreover, 19.6% of the compounds are selective hA2AR (molecules 142, 153, 155, 156, 159, 170, 172, 174-176 in Table 6.5) and 9.8% are selective hA3R antagonists (molecules 139, 141, 145, 152, 181 in Table 6.5) in the test set. Following the workflow reported in Figure 6.3.1, the autoMEP vectors of these new 51 antagonists have been used as input matrix for the previously generated autoMEP/SVMclass model. Our classification model was able to correctly assign the 78.4% of the compounds in the collected test set to their class (Table 6.2). Almost all selective compounds of our test set are correctly classified and only 11 of them (22%) are erroneously recognized. In Paper IV additional information on autoMEP/SVMclass model predictions are reported.

After passing the selectivity filtering process, each of the hA2AR and hA3R antagonists has been analyzed by the corresponding SVR binding affin-
ity predictor. The comparison of all the experimental with the predicted $pK_i$ values by the abovementioned hA$_{2\alpha}$R and hA$_3$R SVR models on the test set again support the quality of the predictors, as underlined by the good values of the correlation coefficient ($q = 0.82$ and $q = 0.85$, respectively) (Table 6.3 and Table 6.4).

In Figure 6.3.4 and Figure 6.3.5 only the hA$_{2\alpha}$ classified antagonists predicted by the hA$_{2\alpha}$R SVR model and only the hA$_3$R classified antagonists predicted by the hA$_3$R SVR model, respectively, have been separately considered.

**Figure 6.3.4:** Test set prediction by autoMEP/SVR hA$_{2\alpha}$R model. a) Experimental $pK_i$ activity data plotted vs predicted $pK_i$ values; b) experimental $pK_i$ activity data (□) of the classified selective hA$_{2\alpha}$R antagonists in the test compared to the $pK_i$ values predicted by autoMEP/SVR hA$_{2\alpha}$R model (○).

**Figure 6.3.5:** Test set prediction by autoMEP/SVR hA$_3$R model. a) Experimental $pK_i$ activity data plotted vs predicted $pK_i$ values; b) experimental $pK_i$ activity data (□) of the classified selective hA$_3$R antagonists in the test compared to the $pK_i$ values predicted by autoMEP/SVR hA$_3$R model (○).
The prediction accuracies, as demonstrated by the differences between the experimental and the predicted pK\textsubscript{i} values, are statistically satisfactory for both hA\textsubscript{2A}R and hA\textsubscript{3}R SVR models, with few exceptions in particular regarding the hA\textsubscript{3}R binding affinity prediction, such as molecules 168, 170, 185 and 188 (see Table 6.5, Figure 6.3.4b and Figure 6.3.5b). In *Paper IV* additional considerations on the external test set potency profiles are reported.

### 6.4 Final remarks

To show the useful application of the machine learning in solving a pharmacodynamic selectivity problem, we have presented a combination of Support Vector Machine tools able to predict both A\textsubscript{2A}R versus A\textsubscript{3}R subtypes selectivity profile and the corresponding receptor binding affinities of a large dataset of known pyrazolo-triazolo-pyrimidine analogs. The preliminary results based on a new set of 51 pyrazolo-triazolo-pyrimidines are very encouraging. To further validate our integrated SVM approach, we are extending the applicability of this method to other classes of hAR antagonists and, at the same time, we are exploring the possibility to describe, using a multi-classifier, the full adenosine receptor selectivity spectrum, as we propose in the following chapter.
Nowadays, in medicinal chemistry adenosine receptors (ARs) represent some of the most studied targets, and there is growing interest on the different AR subtypes. The AR subtypes selectivity is highly desired in the development of potent ligands to achieve the therapeutic success. So far, very few ligand-based strategies have been investigated to predict the receptor subtypes selectivity. We have carried out a novel application of the multilabel classification approach by combining the autocorrelated molecular descriptors encoding for the Molecular Electrostatic Potential (autoMEP) with Support Vector Machines (SVMs). Three valuable models, based on decreasing thresholds of potency, have been generated as in series quantitative sieves for the simultaneous prediction of the hA1R, hA2AR, hA2BR and hA3R subtypes potency profile and selectivity of a large collection, more than 500, of known antagonists such as xanthine and pyrazolo-triazolo-pyrimidine analogs. The robustness and reliability of our multilabel classification models were assessed by predicting an internal test set. Finally, we have applied our strategy to 13 newly synthesized pyrazolo-triazolo-pyrimidine derivatives inferring their full adenosine receptor potency spectrum and hAR subtypes selectivity profile.
7.1 Introduction

Adenosine receptors (ARs) are widely considered interesting and promising therapeutic targets. In the last decade, the growing knowledge about the different adenosine receptor subtypes has inspired the development of potent and selective ligands. [142, 182] During the optimization step of drug discovery process, the general aim is to design drugs more effective in the therapeutic treatment, but with minimum side-effects. If compounds do not differentiate between receptor subtypes, their therapeutic application might be accompanied by efficacy problems or side-effects. Therefore, after identifying the single receptor subtype that is responsible for a particular function, the drug candidates may be sifted out according to criteria of high potency profile and subtype selectivity.

The detection of selective compounds by using in silico tools represents a difficult task and to date, not many examples of selectivity prediction have been described in the literature. [183, 184] Only few pioneer studies suggest an integration of both traditional classification and regression analysis as useful filtering strategy to select potent and selective ligands, as previously described in chapter 6.

We would like to demonstrate how a novel application of the multilabel classification approach by combining the autocorrelated molecular descriptors encoding for the Molecular Electrostatic Potential (autoMEP) vectors with Support Vector Machine (SVM) analysis can represent a very powerful tool to simultaneously describe the hA\textsubscript{1}R, hA\textsubscript{2\alpha}R, hA\textsubscript{2\beta}R and hA\textsubscript{3}R potency profiles and identify the possible subtype selectivity for hAR antagonists.

In the previous chapter, we have developed an integrated SVM-SVR method by using the autoMEP molecular descriptors to discriminate A\textsubscript{2\alpha}R versus A\textsubscript{3}R antagonists and to predict the binding affinity to the corresponding receptor subtype. However, in the traditional single-label classification classes are considered mutually exclusive. In our classification task some samples belong to multiple classes, since the hAR antagonists may present a good potency profile for more subtypes. The multilabel classification analysis seems to be appropriate, whereas our dataset deals with non-mutually exclusive and overlapping classes. In this field a novel multilabel classification technique, cross-training with Support Vector Machines (ct-SVM), has been recently proposed. [108, 109]
In the present study, the combination of autoMEP vectors with ct-SVM analysis (autoMEP/ct-SVM) represents a novel strategy for the prediction of the complete hARs potency profiles and infer hAR subtypes selectivity of known xanthine and pyrazolo-triazolo-pyrimidine analogs. Interestingly, our autoMEP/ct-SVM approach has been extended to all four hAR subtypes. In more detail, a large collection of hAR antagonists has been utilized to carry out and validate three autoMEP/ct-SVM models. They have been applied in series as quantitative sieves, based on decreasing thresholds of potency (500 nM, 250 nM and 100 nM), corresponding to different binding affinity $K_i$ values. For the further validation of our strategy, we have synthesized 13 new pyrazolo-triazolo-pyrimidine derivatives to inspect their $A_1$R, $A_2A$R, $A_2B$R and $A_3$R potency profiles. The workflow applied in our analysis is represented in Figure 7.1.1.

**Figure 7.1.1:** Workflow for the generation of autoMEP/ct-SVM multilabel classification models and the prediction of the hAR antagonists potency profiles.

---

1 Experimental binding affinity data kindly provided by the work coordinated by Prof. G. Spalluto (University of Trieste) for the synthesis and by Prof. K. N. Klotz (University of Würzburg) for the pharmacological characterization.
7.2 Adenosine receptor antagonists

In the last few years an intensive exploration of the chemical space has been pursued to discover new highly potent and selective adenosine receptors (ARs) antagonists. As anticipated, the adenosine receptor family belongs to GPCR (G protein-coupled receptors) family A, including four different subtypes, referred to as A<sub>1</sub>, A<sub>2A</sub>, A<sub>2B</sub> and A<sub>3</sub>, which are widely but differentially distributed in the tissues (Figure 7.2.1)<sup>2</sup>. Moreover, they have been cloned from various mammalian species, where they differentiate for both their pharmacological profile and effector coupling.<sup>[139]</sup>

![Figure 7.2.1: Signal transduction pathways involved in the activation of adenosine receptors. A<sub>1</sub> and A<sub>3</sub> activation inhibits adenylyl cyclase through G<sub>i</sub> family of G proteins, whereas A<sub>2A</sub> and A<sub>2B</sub> receptors activate G<sub>s</sub> family, that stimulates the adenylyl cyclase activity. Furthermore, A<sub>1</sub>R subtype may lead to the activation of G<sub>0</sub> family that increases the conductance of K<sup>+</sup> ions (efflux from inside to outside the cell) and influences some protein kinase activities. Finally, A<sub>3</sub> and A<sub>2B</sub> receptors activation can involve G<sub>q</sub> proteins, with the resulting stimulation of phospholipase C (PLC). Diacylglycerol (DAG) and inositol (1,4,5)-trisphosphate (IP<sub>3</sub>) are implicated in the regulation of protein kinase C (PKC) activity and the intracellular concentration of Ca<sup>2+</sup> ions, respectively. The adenosine receptors activation may involve other G proteins, affecting further cellular pathways.

Diverse potent and selective ligands for each subtype have demonstrated the potential therapeutic role of the adenosine receptor in several physiopathological processes. [140, 141, 143, 144, 178, 179, 185-187] In particular, A<sub>1</sub>R selective antagonists have shown anxiolytic effects and they have been

---

reported as promising candidates for the treatment of cognitive disorders, such as dementia. The antagonism selectivity for \( \text{A}_1 \)R is also the proposed mechanism for some diuretic agents, which are considered effective in congestive heart failure and in edema. [142, 185] \( \text{A}_2A \)R antagonists have a neuroprotective activity during ischemic processes and seem to play a role in the reduction of neuronal damage in Parkinson’s or Huntington’s diseases. [141-144] A potential therapeutic activity in the asthma disease has been discovered for \( \text{A}_2B \)R selective antagonists or mixed antagonists to \( \text{A}_2B \)Rs and \( \text{A}_3 \)Rs. [142] \( \text{A}_2B \)R antagonists are also studied as hypoglycaemic agents in diabetes, while \( \text{A}_3 \)R antagonists have a potential application in tumor growth inhibition and in the treatment of glaucoma. [142, 178, 179, 187]

Basing on different molecular scaffolds, diverse drug design approaches have been applied for the discovery of more potent and selective human AR (hAR) antagonists. To this aim, the xanthine and pyrazolo-triazolo-pyrimidine scaffolds have been properly modified to introduce novelty in the chemical space of known adenosine receptors antagonists. [146, 149, 188] In particular, improving selectivity for the hA\(_1\)R subtype has been obtained by decorating the classical xanthine scaffold with 8-aryl or 8-cycloalkyl substituents. [189] On the other hand, hA\(_2B\)R selective antagonists have been developed with different substitutions at the N\(^1\), N\(^3\) and 8 positions in the same xanthine scaffold. [190-192] Regarding the pyrazolo-triazolo-pyrimidine derivatives, the position N\(^7\) has been suggested to be crucial for the selectivity to the hA\(_2A\)R subtype. [146] Conversely, proper substituents at N\(^5\) and N\(^8\) positions shift the antagonism towards the hA\(_3\)R subtype. [149] Furthermore, an alternative imidazo[2,1-f]purinone scaffold has been discovered to improve potency and selectivity for the human A\(_3\)R antagonists. [193] This structural information is summarized in Table 7.1.

### 7.3 Dataset

Unfortunately, only a limited number of known AR antagonists has been synthesized and tested on all four human AR subtypes. In the past, most of the literature partially reported the binding affinity to some hAR subtypes or the dataset was obtained by using ARs cloned from other mammalian species. We have collected 514 hAR antagonists, synthesized and tested on all four hAR subtypes, to derive and validate our three autoMEP/ct-SVM multilabel
Table 7.1: Examples of potent and selective AR antagonists to the four $A_1$, $A_{2A}$, $A_{2B}$, and $A_3$ subtypes.

<table>
<thead>
<tr>
<th>Structures</th>
<th>Exp. binding affinity</th>
<th>Ref(s).</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$ adenosine receptor antagonists</td>
<td>$hA_1 R K_i = 7.1$ nM</td>
<td>[190]</td>
</tr>
<tr>
<td></td>
<td>$hA_2A R K_i = 1200$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_2B R K_i = 625$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_3 R K_i = 395$ nM</td>
<td></td>
</tr>
<tr>
<td>$A_{2A}$ adenosine receptor antagonists</td>
<td>$hA_1 R K_i = 2160$ nM</td>
<td>[146]</td>
</tr>
<tr>
<td></td>
<td>$hA_{2A} R K_i = 0.22$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_2B R K_i &gt; 10,000$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_3 R K_i &gt; 10,000$ nM</td>
<td></td>
</tr>
<tr>
<td>$A_{2B}$ adenosine receptor antagonists</td>
<td>$hA_1 R K_i = 566$ nM</td>
<td>[146]</td>
</tr>
<tr>
<td></td>
<td>$hA_{2A} R K_i &gt; 1,000$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_{2B} R K_i = 18$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_3 R K_i &gt; 1,000$ nM</td>
<td></td>
</tr>
<tr>
<td>$A_3$ adenosine receptor antagonists</td>
<td>$hA_1 R K_i &gt; 1,000$ nM</td>
<td>[192]</td>
</tr>
<tr>
<td></td>
<td>$hA_{2A} R K_i &gt; 1,000$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_{2B} R K_i &gt; 1,000$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_3 R K_i = 0.81$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_1 R K_i &gt; 1,000$ nM</td>
<td>[193]</td>
</tr>
<tr>
<td></td>
<td>$hA_{2A} R K_i &gt; 1,000$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_{2B} R K_i &gt; 1,000$ nM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$hA_3 R K_i = 0.8$ nM</td>
<td></td>
</tr>
</tbody>
</table>
classification models. [134, 146, 147, 149-151, 180, 181, 188-195] They are xanthine derivatives, N$^7$ and N$^8$ pyrazolo-triazolo-pyrimidine analogs. This large dataset was split into training set (318 compounds), validation set (65 compounds) and internal test set (131 compounds). In Figure 7.3.1 the potency and selectivity spectrum for each hAR subtype, considering only our collected homogeneous data on human ARs, is summarized.

![Figure 7.3.1: Distribution representation of the experimental $K_i$ binding affinity data of hAR antagonists in our dataset (514 molecules), including training set, validation set and internal test set. The classes are completely overlapping.](image.png)

The number of hAR antagonists are reported in different ranges of binding affinity for each subtype: $K_i \leq 100$ nM (●), $100$ nM < $K_i \leq 250$ nM (▲), $250$ nM < $K_i \leq 500$ nM (▲), $K_i > 500$ nM (▼). A similar distribution of the corresponding potency profiles intervals is present for hA$_2$A$_R$, hA$_2$B$_R$ and hA$_3$R subtypes, with a noteworthy number of potent antagonists having a binding affinity $K_i$ value lower or equal to 100 nM to the distinct subtypes (Figure 7.3.1). Concerning the hA$_1$R subtype, our collection is short of potent compounds and many hA$_1$R antagonists have a binding affinity $K_i$ value higher than 250 nM. In the final collection, 209 hAR antagonists (41%) are selective for one hAR subtype, with the corresponding $K_i$ values lower than 100 nM. Considering these 209 compounds, 5 are selective for hA$_1$R (2%), 31 are selective for hA$_2$A$_R$ (15%), 79 are hA$_2$B$_R$ selective antagonists (38%) and the remaining 94 are hA$_3$R selective antagonists (45%). Considering the limited information on hA$_1$R subtype and on human AR antagonists, it is very difficult to correctly predict the complete hAR potency profiles and infer the selectivity of novel xanthine and pyrazolo-triazolo-pyrimidine derivatives.

Finally, 13 newly synthesized pyrazolo-triazolo-pyrimidine analogues have been selected as external test set.
7.4 Results and discussion

In the last years, the Support Vector Machine method has shown good generalization performance and high accuracy as supervised learning technique in many classification tasks.\[82, 85\] Unfortunately, the traditional classification models do not give any quantitative information about the biological affinity of the compounds to the corresponding receptor. Moreover, in chemoinformatics the classical single-label classification considers mutually exclusive classes, while in some cases compounds are simultaneously labeled in multiple classes. In the present work, hAR antagonists may present multiple molecular properties of multiple hAR subtypes. Since our dataset deals with non-mutually exclusive and overlapping classes, the alternative ct-SVM multilabel classification provides an appropriate approach.\[108, 109\]

Both topological and electrostatic complementarities are crucial in describing the receptor subtypes selectivity, and the investigation of the MEP on the molecular surface is a useful strategy for rationalizing the interactions involved in the molecular recognition processes.\[52, 57, 58\] In this case, the distribution of the MEP on the molecular surface is able to discriminate the selectivity for different receptor subtypes. The introduction of the autocorrelation vector allows then for overcoming the MEP information inconvenience to be reliant on the spatial rotation and translation of the molecule.

As anticipated, the pyrazolo-triazolo-pyrimidine scaffold can be properly modified to obtain hA\(_2\)A\(_2\)R or hA\(_3\)R selectivity (chapter 6). On the other hand, the xanthine scaffold has been investigated to develop highly potent and selective hA\(_1\)R or A\(_2\)B\(_2\)R antagonists. However, the observation of the scaffold and its substitutions is not an unfailing strategy to correctly assign the potency profile covering all hAR subtypes of new human pyrazolo-triazolo-pyrimidine and xanthine AR antagonists.

Our "sieve system" is composed of three in series ct-SVM multilabel classification models using as input matrix our \textit{auto}MEP vectors (Figure 7.1.1). We aim at introducing \textit{auto}MEP descriptors of each antagonist to best approximate the experimental hA\(_1\)R, hA\(_2\)A\(_2\)R, hA\(_2\)B\(_2\)R and hA\(_3\)R subtypes potency profile applying three independent \textit{auto}MEP/ct-SVM classification models. In particular, our \textit{auto}MEP/ct-SVM models have been derived after the selection of different thresholds of binding affinity, corresponding to three diverse \(K_i\) values: 500 nM for MODEL 1, 250 nM in MODEL 2 and 100 nM for MODEL 3 (Figure 7.1.1). Interestingly, our models can provide at the
same time a quantitative information about the binding affinity $K_i$ values to all hAR subtypes. In fact, starting from the calculation of autoMEP vectors of novel pyrazolo-triazolo-pyrimidine and xanthine analogs, our MODELs 1, 2 and 3 are able to detect potent and selective hAR antagonists.

### 7.4.1 autoMEP/ct-SVM classification models

Each model is characterized by $n$ binary classifiers, with $n$ corresponding to the number of AR subtypes ($n=4$). The score values have been transformed in the predicted classes according to the $C$ criterion. [108] To derive our autoMEP/ct-SVM models we selected a collection of 318 pyrazolo-triazolo-pyrimidine and xanthine derivatives (molecules 1-318), and we have defined them as our training set. Furthermore, we have considered 65 additional pyrazolo-triazolo-pyrimidine and xanthine analogs (molecules 319-383) as validation set for each model to find the optimal parameters of the four binary classifiers. As previously described, in our autoMEP/ct-SVM models the actual labels (experimental classes) are assigned by selecting a different binding affinity $K_i$ value as threshold. In particular, a binary criterion assign "1" if the hAR subtype binding affinity $K_i$ value is lower than the selected threshold (500 nM, 250 nM or 100 nM according to the MODELs 1, 2 or 3, respectively). Conversely, "0" is assigned if the hAR antagonists are less potent than the threshold, i.e. the corresponding hAR subtype binding affinity $K_i$ value is higher than 500 nM, 250 nM or 100 nM in the corresponding MODELs 1, 2 and 3. The selected thresholds act as meshes of our "sieve system", able to filter hAR antagonists with increasing potency.

We have considered as selectivity criterion a difference of at least 2 orders of magnitude between the corresponding hAR subtypes $K_i$ values, with the lower receptor subtype $K_i$ value $\leq 100$ nM. To carry out the final autoMEP/ct-SVM classification models the corresponding training set and validation set were merged in a collection of 383 compounds for the training of the new classifiers with the ct-SVM optimized parameters\(^3\). Finally, an internal test set (molecules 384-514) has been selected to validate our MODELs 1, 2 and 3.

\(^3\)MODEL 1 classifiers: hA \(_{1R}\) ($C = 4, \gamma = 2$); hA \(_{2A R}\) ($C = 4, \gamma = 0.5$); hA \(_{2B R}\) ($C = 16, \gamma = 0.5$); hA \(_{3R}\) ($C = 16, \gamma = 0.5$). MODEL 2 classifiers: hA \(_{1R}\) ($C = 4, \gamma = 2$); hA \(_{2A R}\) ($C = 8, \gamma = 0.5$); hA \(_{2B R}\) ($C = 4, \gamma = 0.5$); hA \(_{3R}\) ($C = 16, \gamma = 1$). MODEL 3 classifiers: hA \(_{1R}\) ($C = 4, \gamma = 0.5$); hA \(_{2A R}\) ($C = 4, \gamma = 0.5$); hA \(_{2B R}\) ($C = 4, \gamma = 0.5$); hA \(_{3R}\) ($C = 4, \gamma = 0.5$).
As anticipated, our *autoMEP*/ct-SVM multilabel classification models have been evaluated in their in series applicability as "sieve system" for new xanthine and pyrazolo-triazolo-pyrimidine derivatives. In this validation process each of the hA1R, hA2AR, hA2BR and hA3R antagonists has been analyzed by MODELS 1, 2 and 3. In Figure 7.4.1 the potency profiles of five structurally different and correctly predicted hAR antagonists have been considered as examples for the predictions interpretation. Differently colored columns are used for each antagonist to show the predicted subtypes/labels by MODELS 1, 2 and 3. The predicted value "1" by the corresponding hAR binary classifier is represented with a color for each model to indicate the relative percentage (%) of potency profile. In Figure 7.4.1 the colors have been assigned in this way: hA1R (●), hA2AR (●), hA2BR (●) and hA3R (●).

Thus, the prediction "1"/"0" refers to the corresponding hAR subtype $K_i$ value, resulting to be lower/higher than the thresholds 500 nM, 250 nM or 100 nM for the MODELS 1, 2 or 3, respectively. A predicted relative percentage (%) of potency profile lower than 100% for at least one class means that the compound is more potent than the selected threshold $K_i$. 

**Figure 7.4.1:** Flowchart of the correctly predicted profiles of five hAR antagonists by applying our *autoMEP*/ct-SVM multilabel classification models on the internal test set.
for multiple hAR subtypes. Moreover, the selectivity can be inferred by one-color columns in all three models. Then, by observing Figure 7.4.1, the first N7-substituted pyrazolo-triazolo-pyrimidine analog (compound 388) resulted to have 100% hA2AR potency profile for all three models, with totally red columns. This corresponds to a $K_i$ value lower than 100 nM only for the hA2AR subtype and a $K_i$ value higher than 500 nM for the remaining hAR subtypes. Consequently, the compound 388 is hA2AR selective.

The compound 400, a xanthine derivative, is predicted to have a complete hA2BAR potency profile by the MODELS 1, 2 and 3. By observing the resulting columns, both compounds 429 and 480 show a mixed potency profile. They are N8-substituted pyrazolo-triazolo-pyrimidine analogs and each autoMEP/ct-SVM model is able to inform about the experimental $K_i$ values for all hAR subtypes. Finally, we have predicted a negative potency profile for all hAR subtypes in the last example (compound 491), characterized by a different chemical structure. In fact, no colored columns are present in the autoMEP/ct-SVM models prediction.

In our analysis the internal test set has been sifted out by three quantitative filters. We have compared the experimental with the corresponding predicted relative percentages (%) of potency profile by MODELS 1, 2 and 3 for the internal test set, as illustrated in Figure 7.4.2, 7.4.3 and 7.4.4, respectively. The colored columns indicate the positive potency profile for hA1R (●), hA2AR (●), hA2BAR (●) and hA3R (●) profiles.

**Figure 7.4.2:** Graphical representation of the (a) experimental and (b) predicted by MODEL 1 (threshold = 500 nM) potency profiles for the internal test set. The colored columns show whether the $K_i$ values to the relative hAR subtype are lower than 500 nM.
Good values of accuracy were obtained for our autoMEP/ct-SVM multi-label classification models in the prediction of the collected internal test set (86%, 81% and 78% for MODELs 1, 2 and 3, respectively). The satisfying prediction accuracies can be also inferred by the graphical comparison of the experimental (Figure 7.4.2a, 7.4.3a and 7.4.4a) with the predicted (Figure 7.4.2b, 7.4.3b and 7.4.4b) potency profiles for each model. The comparison of the experimental with the corresponding predicted graphical representations has shown high similarity in the color distribution.
Moreover, the satisfactory prediction results are highlighted by the good values of the statistical base-class parameters (Table 7.2).

**Table 7.2:** Statistical parameters of our models after prediction of the internal test set.

<table>
<thead>
<tr>
<th>Classes</th>
<th>MODEL 1</th>
<th></th>
<th>MODEL 2</th>
<th></th>
<th>MODEL 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>hA₁R</td>
<td>0.61</td>
<td>0.79</td>
<td>0.34</td>
<td>0.65</td>
<td>0.39</td>
<td>0.70</td>
</tr>
<tr>
<td>hA₂AᵣR</td>
<td>0.92</td>
<td>0.84</td>
<td>0.95</td>
<td>0.82</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td>hA₂BᵣR</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.89</td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td>hA₃R</td>
<td>0.94</td>
<td>0.97</td>
<td>0.80</td>
<td>0.92</td>
<td>0.92</td>
<td>0.98</td>
</tr>
</tbody>
</table>

The analysis of the experimental and the predicted hAR potency profiles of the internal test set by the abovementioned models, with the exception of the hA₁R subtype, again support the quality of our filtering strategy. After considering in series our **auto**MEP/ct-SVM models, 26/49 selective hAR antagonists have been perfectly predicted (molecules 385, 388, 390, 392, 395, 396, 400, 404-406, 408, 425, 431, 450, 454-459, 466, 468, 471, 474, 487 and 507) and we are able to infer hAR subtype selectivity of 10 compounds by analyzing their partially correct predictions (molecules 384, 386, 389, 394, 398, 403, 407, 452, 461 and 513). However, MODEL 3 has correctly assigned the potency profile of other 10 out of 49 selective hAR antagonists (molecules 412, 424, 427, 432, 434, 440, 441, 463, 472 and 473), as reported in Figure 7.2.2, 7.2.3 and 7.2.4. Concerning the prediction of potent ($K_i \leq 100$ nM) hAR antagonists, MODEL 3 has detected 39% potent hA₁R antagonists, 79% potent hA₂AᵣR antagonists, 78% potent hA₂BᵣR antagonists and 96% potent hA₃R antagonists.

### 7.4.3 External validation

In the optimization step of drug discovery, the principal application of the **auto**MEP/ct-SVM models described in the present work is the prediction of the complete hAR binding affinity profile and hAR subtypes selectivity of new potential antagonists. To evaluate the prediction capability of our **auto**MEP/ct-SVM strategy, we have considered as potential new hAR antagonists 13 novel N₈-substituted pyrazolo-triazolo-pyrimidine analogs (external test set, compounds 515-527), which have been synthetized and tested on all four hAR subtypes.
The experimental human A₁R, A₂₄R, A₂₅R and A₃R binding affinities are collected in Table 7.3.⁴

![Chemical structures](image)

### Table 7.3: Biological profile at the four hAR subtypes of the external test set (515-527).

<table>
<thead>
<tr>
<th>Mol.</th>
<th>R₁</th>
<th>R₂</th>
<th>hA₁ (Ki) nM</th>
<th>hA₂₄R (Ki) nM</th>
<th>hA₂₅R (Ki) nM</th>
<th>hA₃R (Ki) nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>515</td>
<td>-</td>
<td>-</td>
<td>1.250</td>
<td>87.3</td>
<td>3.700</td>
<td>1.537</td>
</tr>
<tr>
<td>516</td>
<td>-</td>
<td>-</td>
<td>1.710</td>
<td>2.520</td>
<td>&gt;10,000</td>
<td>25.5</td>
</tr>
<tr>
<td>517</td>
<td>CH₃</td>
<td>CH₂-O-Ph</td>
<td>42</td>
<td>15.1</td>
<td>&gt;10,000</td>
<td>692</td>
</tr>
<tr>
<td>518</td>
<td>CH₂CH₂</td>
<td>CHPh₂</td>
<td>156</td>
<td>131</td>
<td>248</td>
<td>0.98</td>
</tr>
<tr>
<td>519</td>
<td>CH₂CH₂</td>
<td>CH₂-2-thienyl</td>
<td>228</td>
<td>6.21</td>
<td>2,853</td>
<td>255</td>
</tr>
<tr>
<td>520</td>
<td>CH₂CH₂</td>
<td>CH₂-3-thienyl</td>
<td>170</td>
<td>4.51</td>
<td>2,820</td>
<td>3.14</td>
</tr>
<tr>
<td>521</td>
<td>CH₂CH₂</td>
<td>CH₂-β-naphthyl</td>
<td>113</td>
<td>7.94</td>
<td>1,200</td>
<td>174</td>
</tr>
<tr>
<td>522</td>
<td>CH₂CH₂</td>
<td>CH₂-4-OCH₃-Ph</td>
<td>22.6</td>
<td>1.95</td>
<td>460</td>
<td>359</td>
</tr>
<tr>
<td>523</td>
<td>CH₂CH₂</td>
<td>CH₂-2-thienyl</td>
<td>20.5</td>
<td>2.03</td>
<td>2,030</td>
<td>308</td>
</tr>
<tr>
<td>524</td>
<td>CH₂CH₂</td>
<td>CH₂-3-thienyl</td>
<td>31.5</td>
<td>3.88</td>
<td>2,487</td>
<td>540</td>
</tr>
<tr>
<td>525</td>
<td>CH₂CH₂</td>
<td>CH₂-O-Ph-4-Cl</td>
<td>155</td>
<td>16.1</td>
<td>&gt;10,000</td>
<td>2,306</td>
</tr>
<tr>
<td>526</td>
<td>CH₂CH₂</td>
<td>CH₂-4-OCH₃-Ph-4-Cl</td>
<td>199</td>
<td>31.7</td>
<td>&gt;30,000</td>
<td>3,251</td>
</tr>
<tr>
<td>527</td>
<td>CH₂CH₂</td>
<td>CH₂-4-OCH₃-Ph</td>
<td>44.8</td>
<td>8.93</td>
<td>2,690</td>
<td>120</td>
</tr>
</tbody>
</table>

⁴Displacement of specific [3H]-CCPA binding at human A₁ receptors expressed in CHO cells, (n=3-6);
⁵displacement of specific [3H]-NECA binding at human A₂₄ receptors expressed in CHO cells;
⁶Ki values of the inhibition of NECA-stimulated adenyl cyclase activity in CHO cells expressing hA₂₅ receptors;
⁷displacement of specific [3H]-NECA binding at human A₃ receptors expressed in CHO cells. Data are expressed as geometric means, with 95% confidence limits.

These derivatives are the result of various substitution and homologation experiments in two different positions to increase the selectivity of the pyrazolo-triazolo-pyrimidine scaffold to hA₂₄R and hA₃R subtypes. autoMEP vectors of these new 13 hAR antagonists have been used as input matrix for the previously generated autoMEP/ct-SVM multilabel classification MODELS 1, 2 and 3.

⁴Experimental binding affinity data kindly provided by the work coordinated by Prof. G. Spalluto (University of Trieste) for the synthesis and by Prof. K. N. Klotz (University of Würzburg) for the pharmacological characterization.
In Figure 7.4.5 we have reported the predicted potency profiles by MODELS 1, 2 and 3, where the relative percentages (%) indicate positive potency profile for hA1R (●), hA2AR (●), hA2BR (●) and hA3R (●) subtypes. In most cases our models are able to assign the potency profile with at least the 75% of accuracy for each compound in the collected external test set (Figure 7.4.5).

Our methodology has correctly classified most of the selective compounds in the external test set. In more detail, the compounds 515 and 518, hA2AR and hA3R selective antagonists, respectively, are perfectly predicted. Considering their similar structure, our autoMEP/ct-SVM models are able at least to select almost all potent hA2AR antagonists (molecules 519-527), without missing out any potent hA3R antagonist (molecule 520) in the filtering procedure (see Figure 7.4.5c). Surprisingly, MODEL 3 has not misattributed any compound to hA2BR class.

Figure 7.4.5: External test set predictions by a) MODEL 1 - 500 nM, b) MODEL 2 - 250 nM, c) MODEL 3 - 100 nM. The graphical representation indicates the percentage (%) of accuracy for each compound and the predicted relative percentages (%) of the potency profiles for hAR subtypes.
7.5 Final remarks

A novel multilabel classification approach combining autoMEP molecular descriptors with Support Vector Machine (autoMEP/ct-SVM) has been presented as powerful tool to predict hA\textsubscript{1}R, hA\textsubscript{2A}R, hA\textsubscript{2B}R and hA\textsubscript{3}R subtypes potency profile and infer the potential selectivity of known xanthine and pyrazolo-triazolo-pyrimidine derivatives. Three statistically meaningful models have been generated from the same training set by using different binding affinity $K_i$ values as thresholds for hAR classifiers and very positive results were achieved in the validation procedure.

The independent application of each of our models can be used to select with high accuracy hAR antagonists having a binding affinity $K_i$ value lower than 500 nM, 250 nM or 100 nM, according to the aim of the filtering process. To further improve the predictivity of our dynamic autoMEP/ct-SVM strategy, we aim at integrating new information on hAR antagonists in our dataset, especially regarding the hA\textsubscript{1}R subtype.
Quantitative structure-activity relationship (QSAR) analysis has been frequently utilized as computational tool for the prediction of several ecotoxicological parameters including the acute aquatic toxicity. In this chapter we describe a novel integrated strategy to predict the acute aquatic toxicity through the combination of both toxicokinetic and toxicodynamic behaviors of chemicals. In particular, a robust classification model (TOX\text{class}) has been derived by combining Support Vector Machine (SVM) analysis with three classes of toxicokinetic-like molecular descriptors: the autocorrelation molecular electrostatic potential (\textit{autoMEP}) vectors, Sterimol topological descriptors and logP(o/w) property values. TOX\text{class} model is able to assign chemicals to different levels of acute aquatic toxicity, providing an appropriate answer to the new regulatory requirements. Moreover, we have extended the abovementioned toxicokinetic-like descriptor set with more toxicodynamic-like descriptors, as for example HOMO and LUMO energies, to generate a valuable SVM classifier (MOA\text{class}) for the prediction of the mode of action (MOA) of toxic chemicals. As preliminary validation of our approach, the toxicokinetic (TOX\text{class}) and the toxicodynamic (MOA\text{class}) models have been applied in series to inspect both aquatic toxicity hazard and mode of action of 296 chemical substances with unknown or uncertain toxicodynamic information to assess their potential ecological risk and toxic mechanism.
8.1 Introduction

The need for various toxicological data of chemicals in limited time and animal experiments requires the application of alternative computational solutions. As anticipated in chapter 1, the attempt of REACH regulation is the improvement of the toxicity assessment process, by identifying the most hazardous properties of chemicals. \cite{24, 25} Regardless animal and \textit{in vivo} testing strategies are still reliable methods for the human and environmental toxicology risk assessment, the computational toxicology offers a valuable tool to speed up the costly screening of high numbers of compounds. As a consequence, several Intelligent or Integrated Testing Strategies (ITS) have been proposed as rapid, efficient approaches to obtain exposure and effects data and identify different modes of toxic action. \cite{28, 29} In more detail, \textit{in vitro} or computational methods, optimized \textit{in vivo} studies, chemical categories, read-across analysis and thresholds of toxicological concern (TTCs) are admitted non-testing strategies to replace missing data or endpoints, and profitably reduce costly animal experiments. \cite{27} So far, powerful computational toxicology prediction systems have been developed for the exposure and hazard assessment to satisfy the new regulatory requests. \cite{30} In drug discovery the \textit{in silico} approaches for the toxicity prediction of safety-relevant endpoints are precious contributions to early discovery of adverse drug reactions. \cite{31, 35}

In the last years several "data driven systems" and "expert systems" have become available for the prediction of toxicological endpoints. Data driven based programs generate statistically valuable structure-activity relationships (SARs) by processing large groups of unrelated chemicals, without user bias or prior organization, to find associations based on similar chemical structures, known as structural alerts, that most probably correspond to the same toxicological mechanism. Examples for data driven softwares able to predict toxicity endpoints are TOPKAT, MCASE and Lazar. \cite{196-198} Unfortunately, the ease prediction in these techniques is penalized by the accurate statistical validation needed. For this reason, they are better suggested to detect general alerting properties. On the other hand, the expert systems embody a series of knowledge based rules, considering small classes of similar-acting chemicals or groups of compounds with similar structure to build classes of potential toxicity. Even if their application is more limited in comparison with the data driven systems, the expert systems offer
more easily interpretable results. [31] In particular, various expert systems can provide the prediction of aquatic toxicity endpoints, such as ECOSAR, DEREK, HazardExpert and OASIS. [199-206]

In toxicology quantitative structure-activity relationships are widely used approaches to infer the toxicological properties of compounds from their molecular structure. The traditional linear QSAR methodologies are still the most applied strategies in this field. Several studies have focused on the prediction of the aquatic toxicity of chemical substances as basic information in the hazard and environmental risk assessment for species living in the water. [37-41] In this context, two main QSAR strategies, chemical class-based and toxicological-based, have been carried out. In the class-based approach the aquatic toxicity is modeled on small series of homologous chemical substances, according to the concept that similar compounds should behave with a similar toxic mode. [207-209] A problem in this classification scheme is represented by the difficult treatment of complex molecular structures. Alternatively, the toxicological-based QSAR models are developed for compounds supposed to act with the same toxic mechanism. [210-212] However, the toxicological-based approach is closely related to the toxicodynamic information and considers toxicity dependent on the mode of action (MOA). Moreover, in both methods the class assignment is not unambiguous, if more functional groups are involved in the same compound/mechanism of action.

In a previous work, Colombo has highlighted the advantages of the application of simple constitutional and quantum chemical descriptors for the classification of chemicals into structure-related subsets to derive QSAR local models, independently from the mode of action. [213] As proposed so far, most of the QSAR local models require the chemical grouping into structural or toxicodynamic classes first.

The typical endpoint for the assessment of acute toxicity is the concentration lethal to 50% of the organisms (LC$_{50}$), produced by chemicals with different mechanisms, and well-defined thresholds for acute toxicity have been established by UNECE in the Globally Harmonized System of Classification and Labeling of Chemicals (GHS). [214] The toxic effect may involve different types of biochemical molecular interaction between the chemicals and the biological target. Concerning aquatic toxicology, the classification scheme published by Verhaar and included in Toxtree is one of the first strategies to assign chemicals to mechanisms of action. [215, 216]
Russom and collaborators have utilized the information on the toxicodynamic profiles in the *fathead minnow* specie to develop an expert system based on substructural fragments for the classification of chemical substances in different modes of action. [217] In the recent years, alternative reliable classification models based on various molecular descriptors have been introduced. [218-222] The classification approach has recently been proposed in the prediction of genotoxicity, a property deeply involved in the toxicological profile of compounds. In such application, a preprocessing of experimental toxicity data by selecting a threshold was required for the classes definition. [223] Very recently, the well-known EPA Fathead Minnow Acute Toxicity (EPAFHM) database, reporting 96-h LC$_{50}$ values for fathead minnow of diverse industrial chemicals, has been selected to derive a local support vector regression model. [224]

In the present work we describe a novel classification approach able to assign a large number of compounds in the publicly available EPAFHM dataset to different classes of acute aquatic toxicity and modes of toxic action. The workflow of our approach is illustrated in Figure 8.1.1.

*Figure 8.1.1:* Flowchart of the in series TOXclass and MOAclass approach for the prediction of the toxicokinetic and toxicodynamic risk profiles of new chemicals.
Our first attempt is to provide an easily interpretable answer to the regulatory requirements by defining two classes of environmental hazard based on a 96-h LC$_{50}$ threshold value. We have carried out a couple of novel classification models combining Support Vector Machine with two different sets of molecular descriptors (TOX$_{class}$ and MOA$_{class}$ models). In particular, a statistically appreciable classification model (TOX$_{class}$) has been derived by combining SVM analysis with three classes of molecular descriptors: the autocorrelation molecular electrostatic potential (autoMEP) vectors, Sterimol topological descriptors and logP(o/w) property values. Once developed a model to predict the level of acute aquatic toxicity, we have extended the previous descriptor set by introducing further properties, that influence the toxicodynamic profile. Based on this descriptor set, we have generated a robust SVM classifier (MOA$_{class}$) for the prediction of the multiple MOA of toxic chemicals. The toxicokinetic and the toxicodynamic models have been applied in series to identify both aquatic toxicity classes and modes of action of 296 chemicals with unknown MOA or moderate confidence MOA assignment to assess both potential ecological risk and mechanism of toxicity.

### 8.2 Dataset

A collection of 559 industrial chemicals in the original EPAFHLM database has been selected to derive and validate our TOX$_{class}$ and MOA$_{class}$ models. [217] EPAFHLM dataset provides for each compound the chemical structure, 96-h LC$_{50}$ values in mg/L and mmol/L and specifies the mode of action with the corresponding confidence. The toxicity profiles distribution of the dataset considered in the present work is summarized in Table 8.1 and reported in Figure 8.2.1.

**Table 8.1:** The classification of substances in mg/L according to the GHS legislation and the corresponding LC$_{50}$ (mmol/L) intervals for our dataset are reported; tox = toxicity.

<table>
<thead>
<tr>
<th>Minor classes</th>
<th>Values LC$_{50}$ (mg/L)</th>
<th>Values LC$_{50}$ (mmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Tox 1 ($AT_1$)</td>
<td>$AT_1 \leq 1$ mg/L</td>
<td>$AT_1 \leq 0.00848$ mmol/L</td>
</tr>
<tr>
<td>Acute Tox 2 ($AT_2$)</td>
<td>$1 \text{ mg/L} &lt; AT_2 \leq 10$ mg/L</td>
<td>$0.001 \leq AT_2 \leq 0.131$ mmol/L</td>
</tr>
<tr>
<td>Acute Tox 3 ($AT_3$)</td>
<td>$10 \text{ mg/L} &lt; AT_3 \leq 100$ mg/L</td>
<td>$0.028 \leq AT_3 \leq 1.285$ mmol/L</td>
</tr>
<tr>
<td>No Acute Tox ($nAT$)</td>
<td>$nAT &gt; 100$ mg/L</td>
<td>$nAT \geq 0.360$ mmol/L</td>
</tr>
</tbody>
</table>

121
8.2 Dataset

Figure 8.2.1: Graphical representation of 559 chemicals in our dataset. a) Data distribution indicating the acute toxicity classes according to the GHS classification scheme for substances hazardous to the aquatic environment: \( LC_{50} \leq 1 \text{ mg/L} \) (red), \( 1 \text{ mg/L} < LC_{50} \leq 10 \text{ mg/L} \) (orange), \( 10 \text{ mg/L} < LC_{50} \leq 100 \text{ mg/L} \) (yellow), \( LC_{50} > 100 \text{ mg/L} \) (green); b) data distribution indicating the intervals of \( pLC_{50} \) (mmol/L) values corresponding to the classes reported in (a); the threshold for acute toxicity (\( pLC_{50} = 0.3 \)) is highlighted.

In the present study, a training set of 554 compounds has been used to carry out our TOX\textit{class} model (TOX\textit{class} training set). Moreover, 263 chemicals have been selected as training set of our MOA\textit{class} model (MOA\textit{class} training set). They show a high or high-moderate level of confidence in the assigned nine modes of action and this collection is partially overlapping to the abovementioned TOX\textit{class} training set. Finally, a preliminary test set comprising 296 compounds has been considered in both TOX\textit{class} and MOA\textit{class} models. These chemicals display a moderate or low level of confidence in the classification by MOA and they are partially included in the TOX\textit{class} training set. In particular, 187 chemicals in the test set (subset 1) have a low or moderate MOA confidence, while for the remaining 109 compounds (subset 2) the mode of action is unknown. The mechanistic information should be experimentally determined to verify the goodness of our predictions.

Acute toxicity data are expressed as \( LC_{50} \) values (in mmol/L) to quantify the concentration lethal to 50% juvenile fathead minnows (\textit{Pimephales promelas}) in 96-flow-through exposure tests. [217] In our analysis \( LC_{50} \) values in mmol/L are transformed in classes of aquatic toxicity, as described in section 8.3. The mode of action of chemicals was assigned based on joint toxic action studies, involving the analysis of behavioral responses and dose-response relationships after interpreting 96 h \( LC_{50} \) experiments. [217]
8.3 Results and discussion

Computer-based identification of molecular structure properties qualitatively (SAR) or quantitatively (QSAR) related to biological activity represents the main useful application in predictive toxicology. [30, 31, 35] In fact, the unknown toxicological properties of new chemicals can be inferred by considering the available information on toxic molecular structures or fragments.

Even if SAR and QSAR methodologies are more and more rising importance in this field, their applicability requires good quality and homogeneous experimental data. Firstly, LC$_{50}$ precise value of acute aquatic toxicity is not needed in the case of general evaluation of the environmental risk of a chemical. Just one unit of difference between experimental and predicted pLC$_{50}$ corresponds to ten times the same difference expressed in LC$_{50}$ (mmol/L). So the claim for the prediction of exactly the LC$_{50}$ value by using a classical QSAR local model is very challenging. Moreover, if the mechanism of action is unknown or uncertain, and LC$_{50}$ values are inaccurate, the possibility to derive reliable QSAR local models for each MOA subset is excluded. So, these methods are better tailored for a posteriori identification of alerting classes from the predicted values. Therefore, a classification approach independent from MOA, by selecting a particular threshold of aquatic toxicity, seems to be more appropriate.

To date, the classical models for acute aquatic toxicity prediction by chemical class or mode of action have approached toxicity as a property depending on the presence of particular scaffolds/functional groups and on the available toxicodynamic information. Some chemicals with different structure might be very toxic by acting with the same mechanism (Paper VI). Then, we can hypothesize that local models based on chemical classes are not able to correctly assess the level of toxicity of new chemically different compounds. Moreover, the different molecular intrinsic properties might be somehow responsible for a similar toxicokinetic. In EPAFHM dataset the toxic compounds displaying a different mode of action, but analogous molecular structures, are present (Paper VI). Even in this case, a local model by MOA might not recognize as toxic new compounds with similar structure but different toxicodynamic profile.

Our in series TOXclass/MOAclass strategy has been developed to overcome these problems in the toxicity prediction. Interestingly, we have considered the general toxicity as a property well described by molecular de-
8.3 Results and Discussion

Scriptors, rather than a priori influenced by the chemical structure or toxic mechanism. As synthetically represented in Figure 8.3.1, we have derived two robust TOXclass and MOAclass models to predict the level of acute aquatic toxicity and the mode of toxic action, respectively. Our procedure has analyzed toxicity as a property independent from the mechanism of action, since differently acting chemicals can be equally toxic. Moreover, MOA information is not needed to predict the potential toxicological risk. Consequently, we suggest to apply TOXclass and MOAclass models in series for the prediction of the toxicokinetic and toxicodynamic profiles of new chemicals.

Figure 8.3.1: Synthesis of the in series TOXclass and MOAclass approach.

8.3.1 TOXclass model

We have used 554 structurally different chemicals in the same EPAFHM dataset as our training set to carry out a first binary classification model (TOXclass). By considering data distribution of LC50 values in mmol/L according to GHS legislation, reported in Table 8.1 and Figure 8.2.1, we have selected 0.5 mmol/L as LC50 threshold value for high acute toxicity (pLC50=0.3). Then, we have divided our training set into two classes: high acute aquatic toxicity (LC50 ≤ 0.5 mmol/L), that comprises 66% of the training set, and low acute aquatic toxicity (LC50 > 0.5 mmol/L), including the remaining 34%. As anticipated, we have combined three classes of molecular descriptors with SVM analysis: autoMEP vectors (calculated by using default parameters indicated in chapter 2), Sterimol topological descriptors and logP(o/w) property values, as reported in Table 8.2.
Chapter 8  
8.3 Results and discussion

Table 8.2: List of the eighteen descriptors selected for the binary classification model to predict the level of acute toxicity.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>auto MEP vectors</td>
<td>Spatial autocorrelation; property: molecular electrostatic potential</td>
</tr>
<tr>
<td>5</td>
<td>Sterimol</td>
<td>Topological</td>
</tr>
<tr>
<td>1</td>
<td>logP(o/w)</td>
<td>Log octanol/water partition coefficient</td>
</tr>
</tbody>
</table>

In this study, the autocorrelation allows in particular to easily compare surface properties of structurally unrelated compounds of different size. Sterimol descriptors have been introduced to represent topological properties, as in previous papers. [60, 225] Finally, logP, or log of the n-octanol/water partition coefficient, describes the partitioning equilibria and is parameter commonly used as molecular descriptor in the evaluation of pharmacokinetic and pharmacodynamic properties and in modeling of toxicity. [212, 219-222] We have selected these molecular properties to represent properly the chemical influence to the toxicological profile of each compound.

A statistically appreciable TOXclass model has been carried out by using a Gaussian radial basis function kernel setting the \( C \) parameter value to 32 and the \( \gamma \) parameter value to 1. The results are shown in Table 8.3 and Table 8.4.

Table 8.3: Statistical parameters of TOXclass model after cross-validation procedure.

<table>
<thead>
<tr>
<th>Partition</th>
<th>CV</th>
<th>No. of runs</th>
<th>Mean</th>
<th>StDev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>1</td>
<td>88.6</td>
<td>-</td>
<td>88.6</td>
<td>88.6</td>
<td></td>
</tr>
<tr>
<td>Training set</td>
<td>LOO</td>
<td>1</td>
<td>84.3</td>
<td>-</td>
<td>84.3</td>
<td></td>
</tr>
<tr>
<td>10-fold</td>
<td>10</td>
<td>82.4</td>
<td>4.5</td>
<td>70.9</td>
<td>92.9</td>
<td></td>
</tr>
<tr>
<td>5-fold</td>
<td>20</td>
<td>82.0</td>
<td>3.4</td>
<td>70.3</td>
<td>89.2</td>
<td></td>
</tr>
<tr>
<td>3-fold</td>
<td>33</td>
<td>81.3</td>
<td>2.2</td>
<td>76.2</td>
<td>85.9</td>
<td></td>
</tr>
<tr>
<td>2-fold</td>
<td>50</td>
<td>80.5</td>
<td>1.8</td>
<td>76.2</td>
<td>85.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.4: Statistical parameters of TOXclass model after LOO cross-validation.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP (TP rate)</th>
<th>FP (FP rate)</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>high AT</td>
<td>323 (0.88)</td>
<td>43 (0.23)</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>low AT</td>
<td>144 (0.77)</td>
<td>44 (0.12)</td>
<td>0.77</td>
<td>0.77</td>
</tr>
</tbody>
</table>
8.3 Results and Discussion

The percentages (%) of correct predictions obtained after the extensive n-fold cross-validation procedures are higher than 80% (80.5% is the minimum average predictivity for the 2-fold cross-validation), with 84.3% correctly classified chemicals after LOO cross-validation procedure. The robustness of TOXclass model is confirmed by the values of recall and precision for both classes. These results demonstrate the good capability of TOXclass model to infer the toxicokinetic profile of chemicals prior any consideration on the toxicodynamic mechanism. Therefore, we have decided to apply this model as toxicity classifier, as described in 8.3.3 paragraph.

8.3.2 MOA class model

Regarding the MOA training set, 263 chemicals with high or high-moderate MOA confidence in the EPAFHM database have been assigned to nine different MOA and their distribution is given in Figure 8.3.2.

![Figure 8.3.2: MOA class model. Data distribution indicating the percentage (%) of the single MOA classes: baseline narcosis (green), polar narcosis (red), arylate and ester narcosis (light blue), electrophile or proelectrophile phosphorylation (yellow), neurodepressant (purple), uncoupler of oxidative phosphorylation (violet), central nervous system seizure mechanisms (blue), AchE inhibition (orange) and respiratory blocker or inhibition mechanisms (black). The percentages correspond to 142 baseline narcosis, 32 polar narcosis, 11 arylate and ester narcosis, 44 electrophile or proelectrophile phosphorylation, 6 neurodepressant, 11 uncoupler of oxidative phosphorylation, 9 central nervous system seizure mechanisms, 7 AchE inhibition and 1 respiratory blocker or inhibition acting chemicals, respectively.](image)

This training set is characterized by very unbalanced classes, with few chemicals acting with neurodepressant, uncoupler of oxidative phosphorylation, central nervous system seizure, AchE inhibition and respiratory blocker or inhibition mechanisms. However, this distribution reflects the actual well-established knowledge on the toxicodynamic profile of chemicals.

126
The descriptor set used in the MOA\textit{class} model is an extension of the family of descriptors reported above for the TOX\textit{class} model, as listed in Table 8.5.

\textbf{Table 8.5: List of the twenty four descriptors selected for the MOA\textit{class} model.}

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>\textit{auto} MEP vectors</td>
<td>Spatial autocorrelation; property: molecular electrostatic potential</td>
</tr>
<tr>
<td>5</td>
<td>Sterimol</td>
<td>Topological</td>
</tr>
<tr>
<td>1</td>
<td>logP(o/w)</td>
<td>Log octanol/water partition coefficient</td>
</tr>
<tr>
<td>1</td>
<td>HDon</td>
<td>Number of hydrogen bonding donors derived from the sum of NH and OH groups in the molecule</td>
</tr>
<tr>
<td>1</td>
<td>HAcc</td>
<td>Number of hydrogen bonding acceptors derived from the sum of nitrogen and oxygen atoms in the molecule</td>
</tr>
<tr>
<td>1</td>
<td>TPSA</td>
<td>Topological polar surface area</td>
</tr>
<tr>
<td>1</td>
<td>ASA</td>
<td>Approximate surface area</td>
</tr>
<tr>
<td>1</td>
<td>HOMO</td>
<td>The energy (eV) of the Highest Occupied Molecular Orbital</td>
</tr>
<tr>
<td>1</td>
<td>LUMO-HOMO</td>
<td>Difference between the energy (eV) of the Lowest Unoccupied Molecular Orbital and the energy (eV) of the Highest Occupied Molecular Orbital</td>
</tr>
</tbody>
</table>

Besides \textit{auto} MEP vectors, Sterimol and logP(o/w) descriptors, further quantum-physical-chemical properties were introduced to more accurately describe chemistry related the toxicodynamic profile (the number of hydrogen bonding donors and acceptors, the topological polar and approximate surface areas, HOMO and the difference LUMO-HOMO energy descriptors).

The SVM methodology has been utilized in combination with the above-mentioned descriptors to derive a robust MOA\textit{class} classifier, by using Gaussian radial basis function kernel ($C = 32$, $\gamma = 0.9$). MOA\textit{class} model statistical parameters are summarized in Table 8.6 and Table 8.7.

An acceptable MOA\textit{class} model has been obtained as indicated by the percentage (%) of correct predictions after LOO cross-validation procedure and the recall and precision values higher than 70% for the A, B, E, F and H MOA classes. Interestingly, 77% correct predictions, yielded after LOO cross-validation, confirm the reliability of this model. Regardless the high $n$-fold standard deviations, MOA\textit{class} analysis gave noteworthy results considering that MOA classes are highly unbalanced. Moreover, the inclusion of new molecular descriptors considerably improved the model predictivity for
Table 8.6: Statistical parameters of the MOA class model after cross-validation.

<table>
<thead>
<tr>
<th>Partition</th>
<th>CV</th>
<th>No. of runs</th>
<th>Mean</th>
<th>StdDev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>1</td>
<td>97.3</td>
<td>-</td>
<td>97.3</td>
<td>97.3</td>
<td>97.3</td>
</tr>
<tr>
<td>Training set</td>
<td>LOO</td>
<td>1</td>
<td>77.2</td>
<td>-</td>
<td>77.2</td>
<td>77.2</td>
</tr>
<tr>
<td>10-fold</td>
<td>10</td>
<td>75.1</td>
<td>8.0</td>
<td>57.7</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>5-fold</td>
<td>20</td>
<td>73.3</td>
<td>5.5</td>
<td>58.5</td>
<td>84.9</td>
<td></td>
</tr>
<tr>
<td>3-fold</td>
<td>33</td>
<td>69.7</td>
<td>4.3</td>
<td>55.7</td>
<td>78.2</td>
<td></td>
</tr>
<tr>
<td>2-fold</td>
<td>50</td>
<td>65.8</td>
<td>4.0</td>
<td>57.2</td>
<td>74.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.7: MOA class model. The statistical parameters after LOO cross-validation procedure. In the first column baseline narcosis (green), polar narcosis (red), ary late and ester narcosis (light blue), electrophile or proelectrophile phosphorylation (yellow), neurodepressant (purple), uncoupler of oxidative phosphorylation (violet), central nervous system seizure mechanisms (blue), AchE inhibition (orange) and respiratory blocker or inhibition (black) mode-of-action classes are highlighted.

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP (TP rate)</th>
<th>FP (FP rate)</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>126 (0.89)</td>
<td>30 (0.25)</td>
<td>0.89</td>
<td>0.81</td>
</tr>
<tr>
<td>(B)</td>
<td>23 (0.72)</td>
<td>10 (0.04)</td>
<td>0.72</td>
<td>0.70</td>
</tr>
<tr>
<td>(C)</td>
<td>5 (0.45)</td>
<td>4 (0.02)</td>
<td>0.45</td>
<td>0.56</td>
</tr>
<tr>
<td>(D)</td>
<td>25 (0.57)</td>
<td>11 (0.05)</td>
<td>0.57</td>
<td>0.70</td>
</tr>
<tr>
<td>(E)</td>
<td>5 (0.83)</td>
<td>0 (0.00)</td>
<td>0.83</td>
<td>1.00</td>
</tr>
<tr>
<td>(F)</td>
<td>9 (0.82)</td>
<td>2 (0.01)</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>(G)</td>
<td>5 (0.56)</td>
<td>3 (0.01)</td>
<td>0.56</td>
<td>0.62</td>
</tr>
<tr>
<td>(H)</td>
<td>5 (0.71)</td>
<td>0 (0.00)</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td>(I)</td>
<td>0 (0.00)</td>
<td>0 (0.00)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

B, C, D and G classes (polar narcosis, ary late and ester narcosis, electrophile or proelectrophile phosphorylation and central nervous system seizure mechanisms, respectively). This MOA class predictor has been selected to evaluate the test set, as described in the following paragraph.

8.3.3 Applicability of TOX class and MOA class models

As preliminary proof, TOX class and MOA class models have been applied in series to predict both toxicokinetic and toxicodynamic profiles of 296 chemicals (test set) with experimentally uncertain mechanisms of toxicity. In particular, our attempt is the evaluation of the potential ecological risk and,
in a second step, the mode of toxic action.

Our classification approach is able to intrinsically predict the toxicological classes instead of the numerical values of LC$_{50}$, with a low probability of significant errors with respect to classical QSAR regression methods. In more detail, our TOXclass model can directly assign chemicals to "high acute aquatic toxicity" or "low acute aquatic toxicity" classes, corresponding to LC$_{50}$ values in mmol/L lower or higher than 0.5, respectively. Following the workflow reported in Figure 8.3.1, the eighteen descriptors reported in Table 8.2 of all chemicals in the test set have been used as input matrix for TOXclass model and the prediction results are reported in Table 8.8.

**Table 8.8: Statistical parameters after the test set prediction by TOXclass model.**

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP (TP rate)</th>
<th>FP (FP rate)</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>high AT</td>
<td>183 (0.93)</td>
<td>18 (0.18)</td>
<td>0.93</td>
<td>0.91</td>
</tr>
<tr>
<td>low AT</td>
<td>81 (0.82)</td>
<td>14 (0.07)</td>
<td>0.82</td>
<td>0.85</td>
</tr>
</tbody>
</table>

TOXclass model yielded a percentage (%) of 89.2 correct predictions for the test set, with recall and precision higher than 80% for both aquatic toxicity classes. Only 14 high acute toxic aquatic chemicals have been erroneously recognized as low acute toxic.

After predicting the toxicokinetic profile, the ideal procedure should be the exclusive prediction of MOA for compounds with hazard toxicity (LC$_{50}$ ≤ 0.5 mmol/L). In this work, we have applied MOAclass model to the whole test set and the results have been analyzed separately for the subsets 1 and 2. In this case, the descriptor set reported in Table 8.5 was used to represent the compounds in the test set for the prediction by MOAclass model. The experimental (low and moderate confidence) and predicted MOA classes for subset 1 are summarized in Table 8.9 and Table 8.10, respectively.

Only 49% predicted classes corresponds to the experimental toxicodynamic profiles. The comparison of the experimental with the predicted MOA classes for the subset 1 supports the debatable quality of the experimental data. However, a good correspondence in the distribution of chemicals within the toxicological classes has been found, as underlined by the comparison of the graphical representations in Table 8.9 and Table 8.10.
Table 8.9: Graphical representation of the experimental toxicodynamic classes for chemicals with low or moderate MOA confidence in the test set (subset 1). The percentages (%) and the corresponding colors indicating the distribution of chemicals within the classes are reported in the table below.

<table>
<thead>
<tr>
<th>Experimental MOA classes</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) - Baseline narcosis</td>
<td>53</td>
</tr>
<tr>
<td>(B) - Polar narcosis</td>
<td>4</td>
</tr>
<tr>
<td>(C) - Arylate and ester narcosis</td>
<td>8</td>
</tr>
<tr>
<td>(D) - Electrophile or proelectrophile phosphorylation</td>
<td>27</td>
</tr>
<tr>
<td>(E) - Neurodepressant</td>
<td>0</td>
</tr>
<tr>
<td>(F) - Uncoupler of oxidative phosphorylation</td>
<td>1</td>
</tr>
<tr>
<td>(G) - Central nervous system seizure mechanisms</td>
<td>0</td>
</tr>
<tr>
<td>(H) - AchE inhibition</td>
<td>0</td>
</tr>
<tr>
<td>(I) - Respiratory blocker or inhibition</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8.10: Graphical representation of the predicted toxicodynamic classes for chemicals with low or moderate MOA confidence in the test set (subset 1). The percentages (%) and the corresponding colors indicating the distribution of chemicals within the classes are reported in the table below.

<table>
<thead>
<tr>
<th>Predicted MOA classes</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) - Baseline narcosis</td>
<td>61</td>
</tr>
<tr>
<td>(B) - Polar narcosis</td>
<td>12</td>
</tr>
<tr>
<td>(C) - Arylate and ester narcosis</td>
<td>2</td>
</tr>
<tr>
<td>(D) - Electrophile or proelectrophile phosphorylation</td>
<td>15</td>
</tr>
<tr>
<td>(E) - Neurodepressant</td>
<td>1</td>
</tr>
<tr>
<td>(F) - Uncoupler of oxidative phosphorylation</td>
<td>4</td>
</tr>
<tr>
<td>(G) - Central nervous system seizure mechanisms</td>
<td>3</td>
</tr>
<tr>
<td>(H) - AchE inhibition</td>
<td>2</td>
</tr>
<tr>
<td>(I) - Respiratory blocker or inhibition</td>
<td>0</td>
</tr>
</tbody>
</table>
In Table 8.11 the compounds in subset 2, classified by MOA class model, have been reported.

**Table 8.11:** Graphical representation of the predicted toxicodynamic classes by MOA class model for the chemicals with unknown MOA in the test set (subset 2). The number of compounds for each class are indicated, and the corresponding percentages (%) with the colors showing the distribution of chemicals within the classes are reported in the table below.

<table>
<thead>
<tr>
<th>Predicted MOA classes</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Baseline narcosis</td>
<td>67</td>
</tr>
<tr>
<td>(B) Polar narcosis</td>
<td>9</td>
</tr>
<tr>
<td>(C) Arylate and ester narcosis</td>
<td>4</td>
</tr>
<tr>
<td>(D) Electrophile or proelectrophile phosphorylation</td>
<td>15</td>
</tr>
<tr>
<td>(E) Neurodepressant</td>
<td>0</td>
</tr>
<tr>
<td>(F) Uncoupler of oxidative phosphorylation</td>
<td>1</td>
</tr>
<tr>
<td>(G) Central nervous system seizure mechanisms</td>
<td>3</td>
</tr>
<tr>
<td>(H) AchE inhibition</td>
<td>1</td>
</tr>
<tr>
<td>(I) Respiratory blocker or inhibition</td>
<td>0</td>
</tr>
</tbody>
</table>

Most chemicals have been predicted as baseline narcosis (74 compounds) and electrophile or proelectrophile phosphorylation (16 compounds) acting (Table 8.11). The following step should be to experimentally verify our subset 2 predictions and extend the applicability of our approach to new chemicals.

### 8.4 Final remarks

The toxicokinetic and toxicodynamic aspects of aquatic toxicity have been investigated to provide useful tools in agreement with the recent regulatory system for the evaluation of the environmental hazards. We have applied
8.4 Final remarks

novel powerful classification methods to discriminate chemicals into classes of toxicity and mode of toxic action. The novelty of the strategy is represented by a first approach to toxicity not related to toxicodynamic and chemical information. Moreover, our models have provided an easily interpretable answer to the regulatory requirements by defining two classes of acute aquatic toxicity. In particular, we have presented a couple of robust SVM classification models in combination with two families of molecular descriptors. The first TOX\textit{class} classifier has been applied to a test set to predict the level of aquatic toxicity, while the second MOA\textit{class} model has been used to infer the toxic mechanism.

The experimental evaluation of the acute aquatic toxicity and the mode of toxic action would represent an effective validation of our TOX\textit{class}/MOA\textit{class} approach. Finally, we aim at incorporating new compounds in the training sets to extend the TOX\textit{class}/MOA\textit{class} strategy to other chemicals for the aquatic toxicity prediction.
Chapter 9

General conclusions

The present thesis has focused on the development of nonlinear QSAR models, mainly by using machine learning methods, as an attractive and helpful strategy in drug discovery. Our first intent is to demonstrate the wide applicability of nonlinear methodologies in pharmaceutical research tasks. The promising prediction results underlined by the six reported case studies in the field of pharmacodynamic, pharmacokinetics and toxicology have been obtained by combining several families of molecular descriptors with nonlinear techniques, to properly describe the relationship between the molecular properties and the desired endpoint.

Response Surface Analysis in combination with autoMEP vectors, encoding for 3D molecular structures, resulted as a robust methodology in the evaluation of both aqueous solvation free energy of organic compounds and binding affinity $K_i$ values of human $A_2A$R antagonists, as described in chapters 3 and 4. Moreover, the case study discussed in chapter 4 has emphasized the potential parallel use of nonlinear methods to support the predictivity of linear strategies.

Especially Support Vector Machine has been suggested for its good predictive power and the generalization capability in a large number of regression and classification studies. The classification-based approaches aim at predicting classes of activity (high and low) or mechanisms of activity, while the regression-based methods offer the precise prediction of activity or property data. In chapter 5 we have reported interesting results about the application of nonlinear classification approaches to predict the cytochrome P450 metabolism undergone by xenobiotics in humans. In the pharmacodynamic
field, novel powerful SVM methodologies have been proposed to predict potency and selectivity of human AR antagonists, first focusing on $A_{2A}$R and $A_{3}$R subtypes (chapter 6) to finally extend our task to all hAR subtypes (chapter 7). In more detail, we have generated single-label and multilabel classification models, respectively, trying to develop a filtering strategy to detect potent and selective hAR antagonists. Further algorithms and screening approaches are being developing to improve the results in the selection process of drug candidates.

Regarding the prediction of the toxicological profile of chemicals, in silico approaches are more and more applied as alternative methods to animal testing in order to refine and reduce the experiments. Consequently, in computational toxicology investigations the QSAR models should be in agreement with the recent regulatory system for the evaluation of the environmental hazards. In chapter 8 we have discussed a novel classification strategy to evaluate the toxicokinetic and toxicodynamic aspects of aquatic toxicity. Interestingly, we introduced a new approach for studying toxicity, which is not strictly dependent on the toxicodynamic profile of chemicals. Then, the efforts should be directed to the interpretation of other toxicological endpoints, in order to built a complete system supporting costly animal testing protocols.
Bibliography


BIBLIOGRAPHY


[51] *ADRIANA.Code*; version 2.2; Molecular Networks GmbH: Erlangen, Germany, 2008.


[60] Bacilieri, M.; Varano, F.; Defflorian, F.; Marini, M.; Catarzi, D.; Colotta, V.; Filacchioni, G.; Galli, A.; Costagli, C.; Kaseda, C.; Moro, S. Tandem 3D-QSARs approach as valuable tool to predict binding affinity data: design of


[100] Liu, H. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D. QSAR study of ethyl 2-[(3-methyl-2,5-dioxo(3-pyrrolinyl)]amino]-4-(trifluoromethyl)pyrimi-

142


BIBLIOGRAPHY


[114] OpenMosix, version 2.4.26; Moshe Bar, Tel Aviv University: Israel, 2004.


144


Bibliography


148


Bibliography


Supporting Information

In the electronic version of this thesis you find attached files containing the supplementary information material.

Appendix A includes the training set, the test set with the relative references used to develop the autoMEP/RSA model for aqueous solvation free energy prediction, as described in chapter 3 (Paper I).

Appendix B reports the binding affinity data for the training set, both internal and external test sets, with the corresponding references, utilized for the generation of linear autoMEP/PLS and nonlinear autoMEP/RSA models derived in chapter 4 (Paper II).

Appendix C collects the whole dataset of cytochrome P450 substrates, with the relative references, as described in chapter 5 (Paper III). Additional tables reporting the results of a multi-classification model by considering 7 classes, together with further parameters corresponding to the models derived by using 5 classes are included.

Appendix D reports both training sets used in the SVM\textit{class} and SVR models, as summarized in chapter 6 (Paper IV). Moreover, the binding affinity data for the test set is reported.

Appendix E includes the training set, validation set and internal test set used in the multi-classification models to predict potency and selectivity of adenosine receptor antagonists, as described in chapter 7 (Paper V), together with further tables reporting additional information on the validation process and the predictions on the external test set.

Appendix F reports both training sets utilized in chapter 8 for the generation of TOX\textit{class} and MOA\textit{class} classifiers, with the predicted toxicodynamic and toxicokinetic information about the test set (Paper VI).
Appendix A

Paper I

Supplementary Information

A.1 Training set.
A.2 Test set.
A.3 References.
## A.1 Training set

**1-248** [1]

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>Exp. $\triangle G_{\text{hyd}}$ (kcal/mol)</th>
<th>Pred. $\triangle G_{\text{hyd}}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>methane</td>
<td>1.98</td>
<td>2.55</td>
</tr>
<tr>
<td>2</td>
<td>ethane</td>
<td>1.81</td>
<td>2.22</td>
</tr>
<tr>
<td>3</td>
<td>propane</td>
<td>2.02</td>
<td>2.08</td>
</tr>
<tr>
<td>4</td>
<td>n-butane</td>
<td>2.18</td>
<td>2.26</td>
</tr>
<tr>
<td>5</td>
<td>2-methylpropane</td>
<td>2.32</td>
<td>2.29</td>
</tr>
<tr>
<td>6</td>
<td>n-pentane</td>
<td>2.36</td>
<td>2.25</td>
</tr>
<tr>
<td>7</td>
<td>2,2-dimethylpropane</td>
<td>2.69</td>
<td>2.65</td>
</tr>
<tr>
<td>8</td>
<td>cyclopentane</td>
<td>1.22</td>
<td>1.87</td>
</tr>
<tr>
<td>9</td>
<td>n-hexane</td>
<td>2.58</td>
<td>2.07</td>
</tr>
<tr>
<td>10</td>
<td>3-methylpentane</td>
<td>2.54</td>
<td>2.57</td>
</tr>
<tr>
<td>11</td>
<td>cyclohexane</td>
<td>1.24</td>
<td>2.08</td>
</tr>
<tr>
<td>12</td>
<td>methylocyclopentane</td>
<td>1.62</td>
<td>1.87</td>
</tr>
<tr>
<td>13</td>
<td>2,2-dimethylbutane</td>
<td>2.63</td>
<td>2.63</td>
</tr>
<tr>
<td>14</td>
<td>n-heptane</td>
<td>2.65</td>
<td>2.49</td>
</tr>
<tr>
<td>15</td>
<td>2,4-dimethylpentane</td>
<td>2.92</td>
<td>2.48</td>
</tr>
<tr>
<td>16</td>
<td>methylocyclohexane</td>
<td>1.73</td>
<td>1.92</td>
</tr>
<tr>
<td>17</td>
<td>n-octane</td>
<td>2.93</td>
<td>2.43</td>
</tr>
<tr>
<td>18</td>
<td>2,2,4-trimethylpentane</td>
<td>2.89</td>
<td>2.47</td>
</tr>
<tr>
<td>19</td>
<td>ethylene</td>
<td>1.30</td>
<td>0.32</td>
</tr>
<tr>
<td>20</td>
<td>propylene</td>
<td>1.28</td>
<td>1.23</td>
</tr>
<tr>
<td>21</td>
<td>1-butene</td>
<td>1.40</td>
<td>1.51</td>
</tr>
<tr>
<td>22</td>
<td>2-methylpropene</td>
<td>1.30</td>
<td>1.06</td>
</tr>
<tr>
<td>23</td>
<td>1-pentene</td>
<td>1.69</td>
<td>1.80</td>
</tr>
<tr>
<td>24</td>
<td>trans-2-pentene</td>
<td>1.35</td>
<td>1.09</td>
</tr>
<tr>
<td>25</td>
<td>cyclopentene</td>
<td>0.57</td>
<td>1.29</td>
</tr>
<tr>
<td>26</td>
<td>2-methyl-2-buten</td>
<td>1.33</td>
<td>0.78</td>
</tr>
<tr>
<td>27</td>
<td>3-methyl-1-butene</td>
<td>1.85</td>
<td>1.58</td>
</tr>
<tr>
<td>28</td>
<td>cyclohexene</td>
<td>0.37</td>
<td>1.30</td>
</tr>
<tr>
<td>29</td>
<td>4-methyl-1-pentene</td>
<td>1.93</td>
<td>1.72</td>
</tr>
<tr>
<td>30</td>
<td>trans-2-heptene</td>
<td>1.69</td>
<td>1.61</td>
</tr>
<tr>
<td>31</td>
<td>1-methylocyclohexene</td>
<td>0.68</td>
<td>0.38</td>
</tr>
<tr>
<td>32</td>
<td>1-octene</td>
<td>2.20</td>
<td>1.85</td>
</tr>
<tr>
<td>33</td>
<td>1,3-butadiene</td>
<td>0.57</td>
<td>0.46</td>
</tr>
<tr>
<td>34</td>
<td>1,4-pentadiene</td>
<td>0.95</td>
<td>1.21</td>
</tr>
<tr>
<td>35</td>
<td>2-methyl-1,3-butadiene</td>
<td>0.69</td>
<td>0.44</td>
</tr>
<tr>
<td>36</td>
<td>1,5-hexadiene</td>
<td>1.02</td>
<td>1.39</td>
</tr>
<tr>
<td>37</td>
<td>propane</td>
<td>-0.48</td>
<td>-1.42</td>
</tr>
<tr>
<td>38</td>
<td>butyne</td>
<td>-0.17</td>
<td>-1.67</td>
</tr>
<tr>
<td>39</td>
<td>1-pentyne</td>
<td>0.01</td>
<td>-0.32</td>
</tr>
<tr>
<td>40</td>
<td>1-hexene</td>
<td>0.29</td>
<td>-0.37</td>
</tr>
</tbody>
</table>
## Training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>Exp. $\Delta G_{hyd}$ (kcal/mol)</th>
<th>Pred. $\Delta G_{hyd}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>1-heptyne</td>
<td>0.61</td>
<td>0.25</td>
</tr>
<tr>
<td>42</td>
<td>1-octyne</td>
<td>0.72</td>
<td>0.55</td>
</tr>
<tr>
<td>43</td>
<td>1-nonyne</td>
<td>1.06</td>
<td>0.37</td>
</tr>
<tr>
<td>44</td>
<td>benzene</td>
<td>-0.90</td>
<td>-0.30</td>
</tr>
<tr>
<td>45</td>
<td>ethylbenzene</td>
<td>-0.62</td>
<td>0.01</td>
</tr>
<tr>
<td>46</td>
<td>o-xylene</td>
<td>-0.91</td>
<td>-0.50</td>
</tr>
<tr>
<td>47</td>
<td>m-xylene</td>
<td>-0.82</td>
<td>-0.41</td>
</tr>
<tr>
<td>48</td>
<td>p-xylene</td>
<td>-0.82</td>
<td>-0.53</td>
</tr>
<tr>
<td>49</td>
<td>propylbenzene</td>
<td>-0.54</td>
<td>-0.10</td>
</tr>
<tr>
<td>50</td>
<td>2-propylbenzene</td>
<td>-0.30</td>
<td>-0.61</td>
</tr>
<tr>
<td>51</td>
<td>1,2,4-trimethylbenzene</td>
<td>-0.87</td>
<td>-0.15</td>
</tr>
<tr>
<td>52</td>
<td>butylbenzene</td>
<td>-0.40</td>
<td>-0.07</td>
</tr>
<tr>
<td>53</td>
<td>2-butylnbenzene</td>
<td>-0.46</td>
<td>-0.29</td>
</tr>
<tr>
<td>54</td>
<td>tert-amylbenzene</td>
<td>-0.18</td>
<td>-0.30</td>
</tr>
<tr>
<td>55</td>
<td>naphthalene</td>
<td>-2.45</td>
<td>-0.48</td>
</tr>
<tr>
<td>56</td>
<td>acenaphthene</td>
<td>-3.44</td>
<td>-2.07</td>
</tr>
<tr>
<td>57</td>
<td>anthracene</td>
<td>-4.34</td>
<td>-2.97</td>
</tr>
<tr>
<td>58</td>
<td>phenanthrene</td>
<td>-4.12</td>
<td>-3.98</td>
</tr>
<tr>
<td>59</td>
<td>chloromethane</td>
<td>-0.54</td>
<td>-2.30</td>
</tr>
<tr>
<td>60</td>
<td>trichloromethane</td>
<td>-1.04</td>
<td>-2.14</td>
</tr>
<tr>
<td>61</td>
<td>tetrachloromethane</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>62</td>
<td>bromomethane</td>
<td>-0.80</td>
<td>-1.45</td>
</tr>
<tr>
<td>63</td>
<td>dibromomethane</td>
<td>-1.99</td>
<td>-1.17</td>
</tr>
<tr>
<td>64</td>
<td>tribromomethane</td>
<td>-2.16</td>
<td>-1.91</td>
</tr>
<tr>
<td>65</td>
<td>iodomethane</td>
<td>-0.90</td>
<td>-0.85</td>
</tr>
<tr>
<td>66</td>
<td>chlorofluoromethane</td>
<td>-0.79</td>
<td>-2.47</td>
</tr>
<tr>
<td>67</td>
<td>chlorotrifluoromethane</td>
<td>2.56</td>
<td>-0.56</td>
</tr>
<tr>
<td>68</td>
<td>dichlorodifluoromethane</td>
<td>1.71</td>
<td>-0.11</td>
</tr>
<tr>
<td>69</td>
<td>chloroethane</td>
<td>-0.64</td>
<td>-1.03</td>
</tr>
<tr>
<td>70</td>
<td>bromoethane</td>
<td>-0.70</td>
<td>-0.12</td>
</tr>
<tr>
<td>71</td>
<td>iodoethane</td>
<td>-0.73</td>
<td>-0.22</td>
</tr>
<tr>
<td>72</td>
<td>1,1-dichloroethane</td>
<td>-0.86</td>
<td>-2.12</td>
</tr>
<tr>
<td>73</td>
<td>1,2-dichloroethane</td>
<td>-1.75</td>
<td>-2.34</td>
</tr>
<tr>
<td>74</td>
<td>1,2-dibromoethane</td>
<td>-2.13</td>
<td>-2.99</td>
</tr>
<tr>
<td>75</td>
<td>1-chloro-2-bromoethane</td>
<td>-1.98</td>
<td>-2.88</td>
</tr>
<tr>
<td>76</td>
<td>1,1,1-trichloroethane</td>
<td>-0.25</td>
<td>-0.76</td>
</tr>
<tr>
<td>77</td>
<td>1,1,2-trichloroethane</td>
<td>-1.98</td>
<td>-2.31</td>
</tr>
<tr>
<td>78</td>
<td>pentachloroethane</td>
<td>-1.38</td>
<td>-3.21</td>
</tr>
<tr>
<td>79</td>
<td>hexachloroethane</td>
<td>-1.42</td>
<td>0.38</td>
</tr>
<tr>
<td>80</td>
<td>chloropentafluoroethane</td>
<td>2.90</td>
<td>-0.08</td>
</tr>
<tr>
<td>81</td>
<td>1,1,2,2-tetrachloro-difluoroethane</td>
<td>0.83</td>
<td>1.02</td>
</tr>
<tr>
<td>82</td>
<td>1,1,2-trichloro-trifluoroethane</td>
<td>1.80</td>
<td>1.25</td>
</tr>
<tr>
<td>83</td>
<td>1,1-dichloro-tetrafluoroethane</td>
<td>2.54</td>
<td>0.68</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>Exp. $\Delta G_{\text{hyd}}$ (kcal/mol)</td>
<td>Pred. $\Delta G_{\text{hyd}}$ (kcal/mol)</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------</td>
<td>-----------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>84</td>
<td>1,2-dichloro-tetrafluor ethane</td>
<td>2.35</td>
<td>1.44</td>
</tr>
<tr>
<td>85</td>
<td>1-chloropropane</td>
<td>-0.36</td>
<td>-0.71</td>
</tr>
<tr>
<td>86</td>
<td>2-chloropropane</td>
<td>-0.25</td>
<td>-0.87</td>
</tr>
<tr>
<td>87</td>
<td>1-bromopropane</td>
<td>-0.57</td>
<td>-0.35</td>
</tr>
<tr>
<td>88</td>
<td>2-bromopropane</td>
<td>-0.48</td>
<td>-0.29</td>
</tr>
<tr>
<td>89</td>
<td>1-iodopropane</td>
<td>-0.62</td>
<td>-0.32</td>
</tr>
<tr>
<td>90</td>
<td>2-iodopropane</td>
<td>-0.47</td>
<td>-0.49</td>
</tr>
<tr>
<td>91</td>
<td>1,2-dichloropropane</td>
<td>-1.27</td>
<td>-2.88</td>
</tr>
<tr>
<td>92</td>
<td>1,3-dichloropropane</td>
<td>-1.92</td>
<td>-3.28</td>
</tr>
<tr>
<td>93</td>
<td>1,2-dibromopropane</td>
<td>-1.96</td>
<td>-2.10</td>
</tr>
<tr>
<td>94</td>
<td>1-chlorobutane</td>
<td>-0.14</td>
<td>-0.24</td>
</tr>
<tr>
<td>95</td>
<td>1-bromobutane</td>
<td>-0.41</td>
<td>-0.35</td>
</tr>
<tr>
<td>96</td>
<td>1-bromo-2-methylpropane</td>
<td>-0.03</td>
<td>-0.33</td>
</tr>
<tr>
<td>97</td>
<td>1-iodobutane</td>
<td>-0.26</td>
<td>-0.29</td>
</tr>
<tr>
<td>98</td>
<td>1,1-dichlorobutane</td>
<td>-0.70</td>
<td>-0.95</td>
</tr>
<tr>
<td>99</td>
<td>1-chloropentane</td>
<td>-0.07</td>
<td>-0.87</td>
</tr>
<tr>
<td>100</td>
<td>2-chloropentane</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>101</td>
<td>3-chloropentane</td>
<td>0.04</td>
<td>-0.13</td>
</tr>
<tr>
<td>102</td>
<td>1-bromo-3-methylbutane</td>
<td>0.21</td>
<td>-1.30</td>
</tr>
<tr>
<td>103</td>
<td>cis-1,2-dichloroethylene</td>
<td>-1.19</td>
<td>-1.37</td>
</tr>
<tr>
<td>104</td>
<td>trans-1,2-dichloroethylene</td>
<td>-0.77</td>
<td>-0.82</td>
</tr>
<tr>
<td>105</td>
<td>1,2,3-trichloroethylene</td>
<td>-0.44</td>
<td>-0.82</td>
</tr>
<tr>
<td>106</td>
<td>tetrachloroethylene</td>
<td>0.06</td>
<td>-0.76</td>
</tr>
<tr>
<td>107</td>
<td>3-chloropropene</td>
<td>-0.58</td>
<td>-0.70</td>
</tr>
<tr>
<td>108</td>
<td>chlorobenzene</td>
<td>-1.02</td>
<td>-1.93</td>
</tr>
<tr>
<td>109</td>
<td>bromobenzene</td>
<td>-1.48</td>
<td>-2.15</td>
</tr>
<tr>
<td>110</td>
<td>1,2-dichlorobenzene</td>
<td>-1.38</td>
<td>-1.30</td>
</tr>
<tr>
<td>111</td>
<td>1,3-dichlorobenzene</td>
<td>-0.99</td>
<td>-1.46</td>
</tr>
<tr>
<td>112</td>
<td>1,4-dichlorobenzene</td>
<td>-2.33</td>
<td>-2.46</td>
</tr>
<tr>
<td>113</td>
<td>$p$-bromotoluene</td>
<td>-1.41</td>
<td>-1.29</td>
</tr>
<tr>
<td>114</td>
<td>1-bromo-2-ethylbenzene</td>
<td>-1.30</td>
<td>-1.02</td>
</tr>
<tr>
<td>115</td>
<td>o-bromocumene</td>
<td>-0.86</td>
<td>-0.80</td>
</tr>
<tr>
<td>116</td>
<td>dimethyl ether</td>
<td>-1.92</td>
<td>-1.88</td>
</tr>
<tr>
<td>117</td>
<td>dimethyl sulfide</td>
<td>-1.56</td>
<td>-1.09</td>
</tr>
<tr>
<td>118</td>
<td>1,3-dioxolane</td>
<td>-4.14</td>
<td>-5.60</td>
</tr>
<tr>
<td>119</td>
<td>diethyl ether</td>
<td>-1.77</td>
<td>-2.00</td>
</tr>
<tr>
<td>120</td>
<td>methylpropyl ether</td>
<td>-1.69</td>
<td>-2.34</td>
</tr>
<tr>
<td>121</td>
<td>methyl isopropyl ether</td>
<td>-2.03</td>
<td>-2.00</td>
</tr>
<tr>
<td>122</td>
<td>tetrahydrofuran</td>
<td>-3.51</td>
<td>-3.20</td>
</tr>
<tr>
<td>123</td>
<td>dioxane</td>
<td>-5.11</td>
<td>-3.67</td>
</tr>
<tr>
<td>124</td>
<td>ethylpropyl ether</td>
<td>-1.84</td>
<td>-1.23</td>
</tr>
<tr>
<td>125</td>
<td>methyl tert-butyl ether</td>
<td>-2.24</td>
<td>-2.69</td>
</tr>
<tr>
<td>126</td>
<td>2-methyl tetrahydrofuran</td>
<td>-3.34</td>
<td>-3.11</td>
</tr>
</tbody>
</table>
### Appendix A

**Training set: continued**

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>Exp. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
<th>Pred. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>127</td>
<td>tetrahydropyran</td>
<td>-3.16</td>
<td>-2.32</td>
</tr>
<tr>
<td>128</td>
<td>dipropyl ether</td>
<td>-1.17</td>
<td>-1.10</td>
</tr>
<tr>
<td>129</td>
<td>1,2-diethoxythane</td>
<td>-3.30</td>
<td>-1.94</td>
</tr>
<tr>
<td>130</td>
<td>1,1-diethoxythane</td>
<td>-3.32</td>
<td>-3.43</td>
</tr>
<tr>
<td>131</td>
<td>di-$n$-butyl ether</td>
<td>-0.84</td>
<td>-0.38</td>
</tr>
<tr>
<td>132</td>
<td>anisole</td>
<td>-1.05</td>
<td>-3.11</td>
</tr>
<tr>
<td>133</td>
<td>thioanisole</td>
<td>-2.76</td>
<td>-1.21</td>
</tr>
<tr>
<td>134</td>
<td>2,2'-dichlorodiethyl sulfide</td>
<td>-3.97</td>
<td>-1.68</td>
</tr>
<tr>
<td>135</td>
<td>methanol</td>
<td>-5.14</td>
<td>-6.09</td>
</tr>
<tr>
<td>136</td>
<td>methanethiol</td>
<td>-1.26</td>
<td>-2.17</td>
</tr>
<tr>
<td>137</td>
<td>ethanol</td>
<td>-4.96</td>
<td>-4.66</td>
</tr>
<tr>
<td>138</td>
<td>2,2,2-trifluoroethanol</td>
<td>-4.35</td>
<td>-4.27</td>
</tr>
<tr>
<td>139</td>
<td>1-propanol</td>
<td>-4.92</td>
<td>-4.24</td>
</tr>
<tr>
<td>140</td>
<td>2-propanol</td>
<td>-4.81</td>
<td>-3.81</td>
</tr>
<tr>
<td>141</td>
<td>allyl alcohol</td>
<td>-5.10</td>
<td>-5.78</td>
</tr>
<tr>
<td>142</td>
<td>1,1,1-trifluoro-2-propanol</td>
<td>-4.21</td>
<td>-6.09</td>
</tr>
<tr>
<td>143</td>
<td>2,2,3,3-tetrafluoropropanol</td>
<td>-4.96</td>
<td>-4.77</td>
</tr>
<tr>
<td>144</td>
<td>2,2,3,3,3-pentafluoropropanol</td>
<td>-4.30</td>
<td>-4.95</td>
</tr>
<tr>
<td>145</td>
<td>1-butanol</td>
<td>-4.78</td>
<td>-4.85</td>
</tr>
<tr>
<td>146</td>
<td>2-butanol</td>
<td>-4.67</td>
<td>-3.80</td>
</tr>
<tr>
<td>147</td>
<td>tert-butyl alcohol</td>
<td>-4.57</td>
<td>-2.86</td>
</tr>
<tr>
<td>148</td>
<td>2-methyl-1-propanol</td>
<td>-4.57</td>
<td>-3.34</td>
</tr>
<tr>
<td>149</td>
<td>1-pentanol</td>
<td>-4.55</td>
<td>-4.46</td>
</tr>
<tr>
<td>150</td>
<td>2-pentanol</td>
<td>-4.45</td>
<td>-4.18</td>
</tr>
<tr>
<td>151</td>
<td>2-methyl-1-butanol</td>
<td>-4.48</td>
<td>-4.31</td>
</tr>
<tr>
<td>152</td>
<td>2-methyl-2-butanol</td>
<td>-4.49</td>
<td>-3.42</td>
</tr>
<tr>
<td>153</td>
<td>1-hexanol</td>
<td>-4.42</td>
<td>-4.15</td>
</tr>
<tr>
<td>154</td>
<td>cyclohexanol</td>
<td>-5.02</td>
<td>-4.74</td>
</tr>
<tr>
<td>155</td>
<td>2,3-dimethylbutanol</td>
<td>-3.97</td>
<td>-3.62</td>
</tr>
<tr>
<td>156</td>
<td>2-methyl-3-pentanol</td>
<td>-3.94</td>
<td>-4.33</td>
</tr>
<tr>
<td>157</td>
<td>4-methyl-2-pentanol</td>
<td>-3.79</td>
<td>-4.11</td>
</tr>
<tr>
<td>158</td>
<td>2-methyl-2-pentanol</td>
<td>-3.98</td>
<td>-3.09</td>
</tr>
<tr>
<td>159</td>
<td>1-heptanol</td>
<td>-4.31</td>
<td>-3.94</td>
</tr>
<tr>
<td>160</td>
<td>1-octanol</td>
<td>-4.16</td>
<td>-4.39</td>
</tr>
<tr>
<td>161</td>
<td>phenol</td>
<td>-6.62</td>
<td>-7.13</td>
</tr>
<tr>
<td>162</td>
<td>4-bromophenol</td>
<td>-7.20</td>
<td>-5.30</td>
</tr>
<tr>
<td>163</td>
<td>thiophenol</td>
<td>-2.58</td>
<td>-4.39</td>
</tr>
<tr>
<td>164</td>
<td>2-cresol</td>
<td>-5.94</td>
<td>-6.29</td>
</tr>
<tr>
<td>165</td>
<td>4-nitrophenol</td>
<td>-6.20</td>
<td>-6.36</td>
</tr>
<tr>
<td>166</td>
<td>4-tert-butylphenol</td>
<td>-6.00</td>
<td>-4.86</td>
</tr>
<tr>
<td>167</td>
<td>acetaldehyde</td>
<td>-3.55</td>
<td>-3.00</td>
</tr>
<tr>
<td>168</td>
<td>propional</td>
<td>-3.48</td>
<td>-3.30</td>
</tr>
<tr>
<td>169</td>
<td>butanal</td>
<td>-3.22</td>
<td>-3.56</td>
</tr>
</tbody>
</table>
### Training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>Exp. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
<th>Pred. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>170</td>
<td>pentanal</td>
<td>-3.07</td>
<td>-2.86</td>
</tr>
<tr>
<td>171</td>
<td>heptanal</td>
<td>-2.71</td>
<td>-2.24</td>
</tr>
<tr>
<td>172</td>
<td>octanal</td>
<td>-2.32</td>
<td>-2.45</td>
</tr>
<tr>
<td>173</td>
<td>nonanal</td>
<td>-2.10</td>
<td>-3.23</td>
</tr>
<tr>
<td>174</td>
<td>trans-2-butenal</td>
<td>-4.28</td>
<td>-4.03</td>
</tr>
<tr>
<td>175</td>
<td>trans-2-hexenal</td>
<td>-3.73</td>
<td>-4.26</td>
</tr>
<tr>
<td>176</td>
<td>trans-2-octenal</td>
<td>-3.48</td>
<td>-3.44</td>
</tr>
<tr>
<td>177</td>
<td>trans-trans-2,4-hexadienal</td>
<td>-4.70</td>
<td>-3.21</td>
</tr>
<tr>
<td>178</td>
<td>benzaldehyde</td>
<td>-4.08</td>
<td>-3.43</td>
</tr>
<tr>
<td>179</td>
<td>acetone</td>
<td>-3.85</td>
<td>-2.22</td>
</tr>
<tr>
<td>180</td>
<td>2-pentanone</td>
<td>-3.56</td>
<td>-2.33</td>
</tr>
<tr>
<td>181</td>
<td>2-heptanone</td>
<td>-3.11</td>
<td>-2.75</td>
</tr>
<tr>
<td>182</td>
<td>2-octanone</td>
<td>-2.92</td>
<td>-2.22</td>
</tr>
<tr>
<td>183</td>
<td>2-nonanone</td>
<td>-2.51</td>
<td>-1.56</td>
</tr>
<tr>
<td>184</td>
<td>2-undecanone</td>
<td>-2.18</td>
<td>-1.16</td>
</tr>
<tr>
<td>185</td>
<td>acetaldehyde</td>
<td>-4.64</td>
<td>-4.32</td>
</tr>
<tr>
<td>186</td>
<td>ethylformate</td>
<td>-2.68</td>
<td>-2.39</td>
</tr>
<tr>
<td>187</td>
<td>methylacetate</td>
<td>-3.36</td>
<td>-4.03</td>
</tr>
<tr>
<td>188</td>
<td>propylformate</td>
<td>-2.51</td>
<td>-2.71</td>
</tr>
<tr>
<td>189</td>
<td>isopropylformate</td>
<td>-2.04</td>
<td>-2.28</td>
</tr>
<tr>
<td>190</td>
<td>ethylacetate</td>
<td>-3.12</td>
<td>-3.08</td>
</tr>
<tr>
<td>191</td>
<td>methylpropionate</td>
<td>-3.01</td>
<td>-3.53</td>
</tr>
<tr>
<td>192</td>
<td>isobutylformate</td>
<td>-2.25</td>
<td>-2.44</td>
</tr>
<tr>
<td>193</td>
<td>propylacetate</td>
<td>-2.89</td>
<td>-2.86</td>
</tr>
<tr>
<td>194</td>
<td>isopropylacetate</td>
<td>-2.68</td>
<td>-2.95</td>
</tr>
<tr>
<td>195</td>
<td>methylbutyrate</td>
<td>-2.87</td>
<td>-2.93</td>
</tr>
<tr>
<td>196</td>
<td>isamyrylformate</td>
<td>-2.16</td>
<td>-2.70</td>
</tr>
<tr>
<td>197</td>
<td>butylacetate</td>
<td>-2.58</td>
<td>-2.45</td>
</tr>
<tr>
<td>198</td>
<td>isobutylacetate</td>
<td>-2.39</td>
<td>-2.48</td>
</tr>
<tr>
<td>199</td>
<td>propylpropionate</td>
<td>-2.49</td>
<td>-2.38</td>
</tr>
<tr>
<td>200</td>
<td>isopropylpropionate</td>
<td>-2.25</td>
<td>-2.61</td>
</tr>
<tr>
<td>201</td>
<td>ethylbutyrate</td>
<td>-2.53</td>
<td>-2.47</td>
</tr>
<tr>
<td>202</td>
<td>methylpentanoate</td>
<td>-2.57</td>
<td>-2.66</td>
</tr>
<tr>
<td>203</td>
<td>amylacetate</td>
<td>-2.49</td>
<td>-2.53</td>
</tr>
<tr>
<td>204</td>
<td>propylbutyrate</td>
<td>-2.31</td>
<td>-2.51</td>
</tr>
<tr>
<td>205</td>
<td>ethylpentanoate</td>
<td>-2.56</td>
<td>-2.09</td>
</tr>
<tr>
<td>206</td>
<td>methylhexanoate</td>
<td>-2.51</td>
<td>-2.87</td>
</tr>
<tr>
<td>207</td>
<td>hexylacetate</td>
<td>-2.29</td>
<td>-3.13</td>
</tr>
<tr>
<td>208</td>
<td>amylpropionate</td>
<td>-2.02</td>
<td>-2.38</td>
</tr>
<tr>
<td>209</td>
<td>methyloctanoate</td>
<td>-2.07</td>
<td>-3.73</td>
</tr>
<tr>
<td>210</td>
<td>ethylheptanoate</td>
<td>-2.33</td>
<td>-2.10</td>
</tr>
<tr>
<td>211</td>
<td>methyldioctanoate</td>
<td>-4.34</td>
<td>-4.76</td>
</tr>
<tr>
<td>212</td>
<td>ethylamine</td>
<td>-4.67</td>
<td>-3.76</td>
</tr>
</tbody>
</table>
### Training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>Exp. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
<th>Pred. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>213</td>
<td>butylamine</td>
<td>-4.43</td>
<td>-3.81</td>
</tr>
<tr>
<td>214</td>
<td>pentyamine</td>
<td>-4.14</td>
<td>-4.15</td>
</tr>
<tr>
<td>215</td>
<td>hexylamine</td>
<td>-4.09</td>
<td>-2.43</td>
</tr>
<tr>
<td>216</td>
<td>dimethylamine</td>
<td>-4.34</td>
<td>-3.83</td>
</tr>
<tr>
<td>217</td>
<td>diethylamine</td>
<td>-4.12</td>
<td>-2.39</td>
</tr>
<tr>
<td>218</td>
<td>pyrrolidine</td>
<td>-5.54</td>
<td>-3.71</td>
</tr>
<tr>
<td>219</td>
<td>piperidine</td>
<td>-5.17</td>
<td>-3.74</td>
</tr>
<tr>
<td>220</td>
<td>dipropylamine</td>
<td>-3.70</td>
<td>-1.16</td>
</tr>
<tr>
<td>221</td>
<td>hexamethyleneimine</td>
<td>-4.97</td>
<td>-3.42</td>
</tr>
<tr>
<td>222</td>
<td>trimethylamine</td>
<td>-3.27</td>
<td>-4.15</td>
</tr>
<tr>
<td>223</td>
<td>triethylamine</td>
<td>-3.07</td>
<td>-3.61</td>
</tr>
<tr>
<td>224</td>
<td>n-methylpyrrolidine</td>
<td>-4.02</td>
<td>-5.29</td>
</tr>
<tr>
<td>225</td>
<td>n-methylpiperidine</td>
<td>-3.94</td>
<td>-5.04</td>
</tr>
<tr>
<td>226</td>
<td>propionitrile</td>
<td>-3.90</td>
<td>-4.34</td>
</tr>
<tr>
<td>227</td>
<td>butyronitrile</td>
<td>-3.69</td>
<td>-5.07</td>
</tr>
<tr>
<td>228</td>
<td>nitroethane</td>
<td>-3.76</td>
<td>-4.51</td>
</tr>
<tr>
<td>229</td>
<td>2-nitropropane</td>
<td>-3.18</td>
<td>-3.58</td>
</tr>
<tr>
<td>230</td>
<td>nitrobenzene</td>
<td>-4.17</td>
<td>-4.54</td>
</tr>
<tr>
<td>231</td>
<td>2-nitrotoluene</td>
<td>-3.63</td>
<td>-3.67</td>
</tr>
<tr>
<td>232</td>
<td>3-nitrotoluene</td>
<td>-3.50</td>
<td>-3.87</td>
</tr>
<tr>
<td>233</td>
<td>pyridine</td>
<td>-4.75</td>
<td>-4.26</td>
</tr>
<tr>
<td>234</td>
<td>2-methylpyridine</td>
<td>-4.68</td>
<td>-4.58</td>
</tr>
<tr>
<td>235</td>
<td>3-methylpyridine</td>
<td>-4.84</td>
<td>-4.70</td>
</tr>
<tr>
<td>236</td>
<td>4-methylpyridine</td>
<td>-4.99</td>
<td>-5.01</td>
</tr>
<tr>
<td>237</td>
<td>2-ethylpyridine</td>
<td>-4.38</td>
<td>-4.38</td>
</tr>
<tr>
<td>238</td>
<td>4-ethylpyridine</td>
<td>-4.78</td>
<td>-2.45</td>
</tr>
<tr>
<td>239</td>
<td>2,3-dimethylpyridine</td>
<td>-4.88</td>
<td>-5.02</td>
</tr>
<tr>
<td>240</td>
<td>2,4-dimethylpyridine</td>
<td>-4.92</td>
<td>-4.97</td>
</tr>
<tr>
<td>241</td>
<td>2,5-dimethylpyridine</td>
<td>-4.77</td>
<td>-4.55</td>
</tr>
<tr>
<td>242</td>
<td>2,6-dimethylpyridine</td>
<td>-4.66</td>
<td>-3.66</td>
</tr>
<tr>
<td>243</td>
<td>3,4-dimethylpyridine</td>
<td>-5.28</td>
<td>-4.85</td>
</tr>
<tr>
<td>244</td>
<td>3,5-dimethylpyridine</td>
<td>-4.90</td>
<td>-4.76</td>
</tr>
<tr>
<td>245</td>
<td>2-methylpyrazine</td>
<td>-5.58</td>
<td>-4.45</td>
</tr>
<tr>
<td>246</td>
<td>2-ethylpyrazine</td>
<td>-5.53</td>
<td>-4.23</td>
</tr>
<tr>
<td>247</td>
<td>2-ethyl-3-methoxypyrazine</td>
<td>-4.45</td>
<td>-4.19</td>
</tr>
<tr>
<td>248</td>
<td>2-isobutyl-3-methoxypirazine</td>
<td>-3.73</td>
<td>-4.56</td>
</tr>
</tbody>
</table>
A.2 Test set

1-23: [1]

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>Exp. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
<th>Pred. $\Delta G_{\text{hyd}}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2-methylpentane</td>
<td>2.56</td>
<td>2.53</td>
</tr>
<tr>
<td>2</td>
<td>cis-1,2-dimethylcyclohexane</td>
<td>1.60</td>
<td>2.38</td>
</tr>
<tr>
<td>3</td>
<td>1-hexene</td>
<td>1.73</td>
<td>1.96</td>
</tr>
<tr>
<td>4</td>
<td>2,3-dimethyl-1,3-butadiene</td>
<td>0.40</td>
<td>0.33</td>
</tr>
<tr>
<td>5</td>
<td>toluene</td>
<td>-0.77</td>
<td>0.28</td>
</tr>
<tr>
<td>6</td>
<td>tert-butylbenzene</td>
<td>-0.44</td>
<td>-0.3</td>
</tr>
<tr>
<td>7</td>
<td>dichloromethane</td>
<td>-1.42</td>
<td>-1.82</td>
</tr>
<tr>
<td>8</td>
<td>1,3-dibromopropane</td>
<td>-1.99</td>
<td>-1.6</td>
</tr>
<tr>
<td>9</td>
<td>chloroethylene</td>
<td>0.50</td>
<td>-0.76</td>
</tr>
<tr>
<td>10</td>
<td>1,4-dichlorobenzene</td>
<td>-1.02</td>
<td>-1.75</td>
</tr>
<tr>
<td>11</td>
<td>diethyl sulfide</td>
<td>-1.45</td>
<td>-0.8</td>
</tr>
<tr>
<td>12</td>
<td>diisopropyl ether</td>
<td>-0.54</td>
<td>-0.96</td>
</tr>
<tr>
<td>13</td>
<td>ethane thiol</td>
<td>-4.08</td>
<td>-2.98</td>
</tr>
<tr>
<td>14</td>
<td>3-hexanol</td>
<td>-3.73</td>
<td>-3.8</td>
</tr>
<tr>
<td>15</td>
<td>hexanal</td>
<td>-2.85</td>
<td>-2.54</td>
</tr>
<tr>
<td>16</td>
<td>2-butanol</td>
<td>-3.76</td>
<td>-2.63</td>
</tr>
<tr>
<td>17</td>
<td>methylformate</td>
<td>-2.82</td>
<td>-4.34</td>
</tr>
<tr>
<td>18</td>
<td>ethylpropionate</td>
<td>-2.83</td>
<td>-2.75</td>
</tr>
<tr>
<td>19</td>
<td>isooctylacetate</td>
<td>-2.24</td>
<td>-2.54</td>
</tr>
<tr>
<td>20</td>
<td>propylamine</td>
<td>-4.56</td>
<td>-3.88</td>
</tr>
<tr>
<td>21</td>
<td>dibutylamine</td>
<td>-3.38</td>
<td>-1.46</td>
</tr>
<tr>
<td>22</td>
<td>1-nitropropane</td>
<td>-3.38</td>
<td>-4.18</td>
</tr>
<tr>
<td>23</td>
<td>2-isobutylpyrazine</td>
<td>-5.11</td>
<td>-3.84</td>
</tr>
</tbody>
</table>

A.3 References

APPENDIX B

Paper II
Supplementary Information

B.1 Training set.
B.2 Internal test set.
B.3 External test set.
B.4 References.
B.1 Training set

1-15: [1]; 16: [2]; 17-18: [3]; 19-30: [1]; 31-34: [4]; 35-41: [5]; 42-47: [6]; 48-50: [7]; 51-53: [8]; 54-59: [6]; 60-63: [7]; 64-69: [5]; 70-73: [4]; 74-79: [1]; 80-85: [5]; 86-87: [4]; 88-90: [1]; 91-96: [6]; 97-108: [7]; 109-127: [7]
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁/Amîne*</th>
<th>R₂</th>
<th>Pred. pKᵢ (nM) (PLS)</th>
<th>Pred. pKᵢ (nM) (RSA)</th>
<th>Exp. pKᵢ (nM) hA₂Aᵢ R</th>
<th>Exp. pKᵢ (nM) hA₂Aᵢ R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>N(CH₂CH₂OH)₂</td>
<td>H</td>
<td>–</td>
<td>-0.49</td>
<td>0.14</td>
<td>0.92</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>NH₂</td>
<td>H</td>
<td>–</td>
<td>-0.24</td>
<td>0.11</td>
<td>0.66</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>NO₂</td>
<td>H</td>
<td>–</td>
<td>-0.14</td>
<td>-0.49</td>
<td>0.00</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>NHCOCH₃</td>
<td>H</td>
<td>–</td>
<td>-0.45</td>
<td>-0.38</td>
<td>-0.68</td>
<td>4.80</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>N(CH₂CH₂OH)₂</td>
<td>H</td>
<td>–</td>
<td>-0.52</td>
<td>0.16</td>
<td>-0.04</td>
<td>1.10</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>COOCH₂CH₃</td>
<td>H</td>
<td>–</td>
<td>-0.08</td>
<td>-0.52</td>
<td>-0.60</td>
<td>4.00</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>OCH₂COOCH₂CH₃</td>
<td>H</td>
<td>–</td>
<td>-0.27</td>
<td>-0.45</td>
<td>0.37</td>
<td>0.43</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>C[NOH]NH₂</td>
<td>H</td>
<td>–</td>
<td>-0.40</td>
<td>-0.54</td>
<td>-0.78</td>
<td>6.00</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>COOH</td>
<td>H</td>
<td>–</td>
<td>-0.34</td>
<td>-0.40</td>
<td>-0.67</td>
<td>4.63</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>SO₂NH₂</td>
<td>H</td>
<td>–</td>
<td>-1.02</td>
<td>-1.65</td>
<td>-0.12</td>
<td>1.31</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>SO₂N(CH₂CH₂OH)₂</td>
<td>H</td>
<td>–</td>
<td>-0.11</td>
<td>0.07</td>
<td>0.10</td>
<td>0.80</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>-SO₂N(CH₂CH₂NH₂)</td>
<td>H</td>
<td>–</td>
<td>-0.24</td>
<td>-0.38</td>
<td>-0.58</td>
<td>3.80</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>SO₂N(CH₃CH₂Cl)₂</td>
<td>H</td>
<td>–</td>
<td>-0.75</td>
<td>-0.88</td>
<td>0.23</td>
<td>0.59</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>SO₂NHCH₂COOH</td>
<td>H</td>
<td>–</td>
<td>-0.61</td>
<td>-0.23</td>
<td>-1.70</td>
<td>50.0</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>H</td>
<td>H</td>
<td>–</td>
<td>-0.20</td>
<td>-0.16</td>
<td>-0.08</td>
<td>1.20</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>H</td>
<td>H</td>
<td>–</td>
<td>-0.14</td>
<td>-0.08</td>
<td>-0.04</td>
<td>1.10</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>OCH₂O</td>
<td>–</td>
<td>–</td>
<td>-0.01</td>
<td>-0.42</td>
<td>0.52</td>
<td>3.30</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>OH</td>
<td>H</td>
<td>–</td>
<td>-0.12</td>
<td>-0.01</td>
<td>-0.18</td>
<td>1.50</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>CH₃OH</td>
<td>CH₃OH</td>
<td>H</td>
<td>-0.01</td>
<td>-0.73</td>
<td>0.72</td>
<td>0.19</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>COOCH₂CH₃</td>
<td>H</td>
<td>H</td>
<td>-0.14</td>
<td>-0.39</td>
<td>-0.74</td>
<td>5.48</td>
</tr>
<tr>
<td>21</td>
<td>–</td>
<td>CH₃</td>
<td>3,4-Cl₂-Ph-NH</td>
<td>–</td>
<td>-2.42</td>
<td>-2.38</td>
<td>-2.16</td>
<td>143</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>$R_1$/Amine*</td>
<td>$R_2$</td>
<td>Pred. $pK_i$ (nM) (PLS)</td>
<td>Pred. $pK_i$ (nM) (RSA)</td>
<td>Exp. $pK_i$ (nM) h$A_{2A}$R</td>
<td>Exp. $pK_i$ (nM) h$A_{2A}$R</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>---------------</td>
<td>--------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>22</td>
<td>–</td>
<td>CH$_3$</td>
<td>3,4-OCH$_2$O-Ph-NH</td>
<td>–</td>
<td>-2.38</td>
<td>-2.67</td>
<td>-2.83</td>
<td>680</td>
</tr>
<tr>
<td>23</td>
<td>–</td>
<td>CH$_3$</td>
<td>4-NO$_2$Ph-NH</td>
<td>–</td>
<td>-2.81</td>
<td>-2.74</td>
<td>-2.84</td>
<td>695</td>
</tr>
<tr>
<td>24</td>
<td>–</td>
<td>CH$_3$</td>
<td>4-CH$_3$Ph-NH</td>
<td>–</td>
<td>-2.57</td>
<td>-2.40</td>
<td>-2.04</td>
<td>110</td>
</tr>
<tr>
<td>25</td>
<td>–</td>
<td>CH$_3$</td>
<td>4-Br-Ph-NH</td>
<td>–</td>
<td>-2.47</td>
<td>-2.13</td>
<td>-2.00</td>
<td>100</td>
</tr>
<tr>
<td>26</td>
<td>–</td>
<td>CH$_3$</td>
<td>4-F-Ph-NH</td>
<td>–</td>
<td>-2.75</td>
<td>-2.53</td>
<td>-2.08</td>
<td>120</td>
</tr>
<tr>
<td>27</td>
<td>–</td>
<td>CH$_3$</td>
<td>4-CF$_3$Ph-NH</td>
<td>–</td>
<td>-2.37</td>
<td>-2.14</td>
<td>-2.15</td>
<td>140</td>
</tr>
<tr>
<td>28</td>
<td>–</td>
<td>CH$_3$</td>
<td>2-OCH$_3$Ph-NH</td>
<td>–</td>
<td>-2.44</td>
<td>-2.28</td>
<td>-2.26</td>
<td>180</td>
</tr>
<tr>
<td>29</td>
<td>–</td>
<td>CH$_3$</td>
<td>3-OCH$_3$Ph-NH</td>
<td>–</td>
<td>-2.35</td>
<td>-2.74</td>
<td>-2.20</td>
<td>160</td>
</tr>
<tr>
<td>30</td>
<td>–</td>
<td>CH$_3$</td>
<td>2-Cl-Ph-NH</td>
<td>–</td>
<td>-2.48</td>
<td>-1.88</td>
<td>-2.30</td>
<td>200</td>
</tr>
<tr>
<td>31</td>
<td>–</td>
<td>CH$_3$</td>
<td>4-Cl-Ph-NH</td>
<td>–</td>
<td>-2.56</td>
<td>-2.16</td>
<td>-2.26</td>
<td>180</td>
</tr>
<tr>
<td>32</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>3,4-C$_2$H$_5$Ph-NH</td>
<td>–</td>
<td>-1.79</td>
<td>-2.30</td>
<td>-2.55</td>
<td>352</td>
</tr>
<tr>
<td>33</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>3,4-C$_2$H$_5$O-Ph-NH</td>
<td>–</td>
<td>-2.05</td>
<td>-2.36</td>
<td>-2.76</td>
<td>576</td>
</tr>
<tr>
<td>34</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>4-CH$_3$-Ph-NH</td>
<td>–</td>
<td>-2.15</td>
<td>-2.13</td>
<td>-1.48</td>
<td>30</td>
</tr>
<tr>
<td>35</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>4-Br-Ph-NH</td>
<td>–</td>
<td>-2.12</td>
<td>-2.05</td>
<td>-1.60</td>
<td>40</td>
</tr>
<tr>
<td>36</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>4-F-Ph-NH</td>
<td>–</td>
<td>-2.30</td>
<td>-2.08</td>
<td>-1.78</td>
<td>60</td>
</tr>
<tr>
<td>37</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>4-CF$_3$-Ph-NH</td>
<td>–</td>
<td>-1.90</td>
<td>-1.93</td>
<td>-1.72</td>
<td>53</td>
</tr>
<tr>
<td>38</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>2-OCH$_3$-Ph-NH</td>
<td>–</td>
<td>-2.18</td>
<td>-2.09</td>
<td>-2.12</td>
<td>133</td>
</tr>
<tr>
<td>39</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>3-OCH$_3$-Ph-NH</td>
<td>–</td>
<td>-2.03</td>
<td>-2.26</td>
<td>-2.15</td>
<td>140</td>
</tr>
<tr>
<td>40</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>2-Cl-Ph-NH</td>
<td>–</td>
<td>-2.07</td>
<td>-1.91</td>
<td>-2.18</td>
<td>150</td>
</tr>
<tr>
<td>41</td>
<td>–</td>
<td>C$_2$H$_5$</td>
<td>4-Cl-Ph-NH</td>
<td>–</td>
<td>-2.16</td>
<td>-1.86</td>
<td>-2.20</td>
<td>160</td>
</tr>
<tr>
<td>42</td>
<td>–</td>
<td>n-C$_3$H$_7$</td>
<td>3,4-C$_3$H$_7$-Ph-NH</td>
<td>–</td>
<td>-1.59</td>
<td>-2.01</td>
<td>-2.60</td>
<td>401</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁/Amine*</td>
<td>R₂</td>
<td>Pred. pKᵢ (nM) (PLS)</td>
<td>Pred. pKᵢ (nM) (RSA)</td>
<td>Exp. pKᵢ (nM) hₐ₂ₐₐR</td>
<td>Exp. pKᵢ (nM) hₐ₂ₐₐR</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>43</td>
<td>n-C₃H₇</td>
<td>3,4-OCH₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.85</td>
<td>-2.23</td>
<td>-2.82</td>
<td>667</td>
</tr>
<tr>
<td>44</td>
<td>n-C₃H₇</td>
<td>4-NO₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-2.06</td>
<td>-2.44</td>
<td>-3.05</td>
<td>1115</td>
</tr>
<tr>
<td>45</td>
<td>n-C₃H₇</td>
<td>4-CH₃Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.92</td>
<td>-2.08</td>
<td>-1.08</td>
<td>12</td>
</tr>
<tr>
<td>46</td>
<td>n-C₃H₇</td>
<td>4-BrPh-NH</td>
<td>-</td>
<td></td>
<td>-1.88</td>
<td>-1.87</td>
<td>-1.70</td>
<td>50</td>
</tr>
<tr>
<td>47</td>
<td>n-C₃H₇</td>
<td>4-FPh-NH</td>
<td>-</td>
<td></td>
<td>-2.06</td>
<td>-2.01</td>
<td>-1.62</td>
<td>42</td>
</tr>
<tr>
<td>48</td>
<td>n-C₃H₇</td>
<td>4-CF₃Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.68</td>
<td>-1.79</td>
<td>-1.53</td>
<td>34</td>
</tr>
<tr>
<td>49</td>
<td>n-C₃H₇</td>
<td>2-OCH₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.99</td>
<td>-1.92</td>
<td>-2.00</td>
<td>100</td>
</tr>
<tr>
<td>50</td>
<td>n-C₃H₇</td>
<td>3-OCH₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.89</td>
<td>-2.14</td>
<td>-2.05</td>
<td>113</td>
</tr>
<tr>
<td>51</td>
<td>n-C₃H₇</td>
<td>2-ClPh-NH</td>
<td>-</td>
<td></td>
<td>-1.83</td>
<td>-1.97</td>
<td>-2.08</td>
<td>121</td>
</tr>
<tr>
<td>52</td>
<td>n-C₃H₇</td>
<td>4-ClPh-NH</td>
<td>-</td>
<td></td>
<td>-1.93</td>
<td>-1.77</td>
<td>-2.15</td>
<td>140</td>
</tr>
<tr>
<td>53</td>
<td>n-C₄H₉</td>
<td>3,4-OCH₂O-Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.75</td>
<td>-1.83</td>
<td>-2.58</td>
<td>376</td>
</tr>
<tr>
<td>54</td>
<td>n-C₄H₉</td>
<td>4-NO₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.90</td>
<td>-2.36</td>
<td>-2.70</td>
<td>503</td>
</tr>
<tr>
<td>55</td>
<td>n-C₄H₉</td>
<td>4-CH₃Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.82</td>
<td>-2.19</td>
<td>-1.38</td>
<td>24</td>
</tr>
<tr>
<td>56</td>
<td>n-C₄H₉</td>
<td>4-BrPh-NH</td>
<td>-</td>
<td></td>
<td>-1.77</td>
<td>-1.99</td>
<td>-1.82</td>
<td>66</td>
</tr>
<tr>
<td>57</td>
<td>n-C₄H₉</td>
<td>4-FPh-NH</td>
<td>-</td>
<td></td>
<td>-1.91</td>
<td>-1.80</td>
<td>-1.70</td>
<td>50</td>
</tr>
<tr>
<td>58</td>
<td>n-C₄H₉</td>
<td>4-CF₃Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.60</td>
<td>-1.79</td>
<td>-1.49</td>
<td>31</td>
</tr>
<tr>
<td>59</td>
<td>n-C₄H₉</td>
<td>2-OCH₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.90</td>
<td>-1.76</td>
<td>-1.96</td>
<td>91</td>
</tr>
<tr>
<td>60</td>
<td>n-C₄H₉</td>
<td>3-OCH₂Ph-NH</td>
<td>-</td>
<td></td>
<td>-1.78</td>
<td>-2.19</td>
<td>-1.98</td>
<td>95</td>
</tr>
<tr>
<td>61</td>
<td>n-C₄H₉</td>
<td>2-ClPh-NH</td>
<td>-</td>
<td></td>
<td>-1.76</td>
<td>-2.01</td>
<td>-2.00</td>
<td>100</td>
</tr>
<tr>
<td>62</td>
<td>n-C₄H₉</td>
<td>4-ClPh-NH</td>
<td>-</td>
<td></td>
<td>-1.80</td>
<td>-1.75</td>
<td>-2.05</td>
<td>111</td>
</tr>
<tr>
<td>63</td>
<td>CH₃</td>
<td>Ph-CH₂</td>
<td>-</td>
<td></td>
<td>-2.30</td>
<td>-1.96</td>
<td>-2.63</td>
<td>423</td>
</tr>
</tbody>
</table>
### Training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁/Amine*</th>
<th>R₂</th>
<th>Pred. pK₁ (nM) (PLS)</th>
<th>Pred. pK₁ (nM) (RSA)</th>
<th>Exp. pK₁ (nM) hA₂hR</th>
<th>Exp. pK₁ (nM) hA₂hR</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td></td>
<td>C₂H₅</td>
<td>Ph-CH₂</td>
<td>–</td>
<td>–.98</td>
<td>–.99</td>
<td>–.89</td>
<td>335</td>
</tr>
<tr>
<td>65</td>
<td></td>
<td>n-C₃H₇</td>
<td>Ph-CH₂</td>
<td>–</td>
<td>–.68</td>
<td>–.96</td>
<td>–.24</td>
<td>395</td>
</tr>
<tr>
<td>66</td>
<td></td>
<td>n-C₄H₉</td>
<td>Ph-CH₂</td>
<td>–</td>
<td>–.65</td>
<td>–.92</td>
<td>–.60</td>
<td>400</td>
</tr>
<tr>
<td>67</td>
<td></td>
<td>CH₃</td>
<td>Ph-NH</td>
<td>–</td>
<td>–.74</td>
<td>–.60</td>
<td>–.58</td>
<td>381</td>
</tr>
<tr>
<td>68</td>
<td></td>
<td>C₂H₅</td>
<td>Ph-NH</td>
<td>–</td>
<td>–.28</td>
<td>–.27</td>
<td>–.09</td>
<td>40</td>
</tr>
<tr>
<td>69</td>
<td></td>
<td>n-C₃H₇</td>
<td>Ph-NH</td>
<td>–</td>
<td>–.04</td>
<td>–.86</td>
<td>–.79</td>
<td>62</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>n-C₄H₉</td>
<td>Ph-NH</td>
<td>–</td>
<td>–.94</td>
<td>–.78</td>
<td>–.21</td>
<td>148</td>
</tr>
<tr>
<td>71</td>
<td></td>
<td>CH₃</td>
<td>4-SO₃Ph-NH</td>
<td>–</td>
<td>–.88</td>
<td>–.65</td>
<td>–.27</td>
<td>594</td>
</tr>
<tr>
<td>72</td>
<td></td>
<td>C₂H₅</td>
<td>4-SO₃Ph-NH</td>
<td>–</td>
<td>–.37</td>
<td>–.62</td>
<td>–.40</td>
<td>249</td>
</tr>
<tr>
<td>73</td>
<td></td>
<td>n-C₃H₇</td>
<td>4-SO₃Ph-NH</td>
<td>–</td>
<td>–.05</td>
<td>–.62</td>
<td>–.28</td>
<td>305</td>
</tr>
<tr>
<td>74</td>
<td></td>
<td>n-C₄H₉</td>
<td>4-SO₃Ph-NH</td>
<td>–</td>
<td>–.97</td>
<td>–.46</td>
<td>–.55</td>
<td>352</td>
</tr>
<tr>
<td>75</td>
<td></td>
<td>CH₃</td>
<td>4-OCH₃Ph-NH</td>
<td>–</td>
<td>–.77</td>
<td>–.43</td>
<td>–.14</td>
<td>1300</td>
</tr>
<tr>
<td>76</td>
<td></td>
<td>CH₃</td>
<td>3-ClPh-NH</td>
<td>–</td>
<td>–.19</td>
<td>–.45</td>
<td>–.08</td>
<td>1195</td>
</tr>
<tr>
<td>77</td>
<td></td>
<td>C₂H₅</td>
<td>3-ClPh-NH</td>
<td>–</td>
<td>–.98</td>
<td>–.44</td>
<td>–.26</td>
<td>180</td>
</tr>
<tr>
<td>78</td>
<td></td>
<td>n-C₃H₇</td>
<td>4-OCH₃Ph-NH</td>
<td>–</td>
<td>–.23</td>
<td>–.01</td>
<td>–.15</td>
<td>140</td>
</tr>
<tr>
<td>79</td>
<td></td>
<td>n-C₄H₉</td>
<td>3-ClPh-NH</td>
<td>–</td>
<td>–.73</td>
<td>–.27</td>
<td>–.09</td>
<td>1220</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>n-C₄H₉</td>
<td>4-OCH₃Ph-NH</td>
<td>–</td>
<td>–.20</td>
<td>–.22</td>
<td>–.19</td>
<td>80</td>
</tr>
<tr>
<td>81</td>
<td></td>
<td>n-C₄H₉</td>
<td>3-ClPh-NH</td>
<td>–</td>
<td>–.64</td>
<td>–.01</td>
<td>–.98</td>
<td>95</td>
</tr>
<tr>
<td>82</td>
<td></td>
<td>Ph-CH₂CH₂</td>
<td>4-OCH₃Ph-NH</td>
<td>–</td>
<td>–.99</td>
<td>–.02</td>
<td>–.08</td>
<td>120</td>
</tr>
<tr>
<td>83</td>
<td></td>
<td>Ph-CH₂CH₂</td>
<td>3-ClPh-NH</td>
<td>–</td>
<td>–.84</td>
<td>–.62</td>
<td>–.02</td>
<td>105</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R$_1$/Amine*</td>
<td>R$_2$</td>
<td>Pred. pK$_i$ (nM) (PLS)</td>
<td>Pred. pK$_i$ (nM) (RSA)</td>
<td>Exp. pK$_i$ (nM)</td>
<td>Exp. pK$_i$ (nM)</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
<td>--------------</td>
<td>--------------</td>
<td>-------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>84</td>
<td>-</td>
<td>Ph-CH$_2$CH$_2$CH$_2$</td>
<td>3-Cl-Ph-NH</td>
<td>-</td>
<td>-1.51</td>
<td>-1.36</td>
<td>-2.30</td>
<td>200</td>
</tr>
<tr>
<td>85</td>
<td>-</td>
<td>H</td>
<td>4-0CH$_3$-Ph-NH</td>
<td>-</td>
<td>-3.41</td>
<td>-2.87</td>
<td>-2.72</td>
<td>520</td>
</tr>
<tr>
<td>86</td>
<td>-</td>
<td>-</td>
<td>4-0CH$_3$-Ph-NH</td>
<td>-</td>
<td>-1.80</td>
<td>-2.07</td>
<td>-0.97</td>
<td>9.40</td>
</tr>
<tr>
<td>87</td>
<td>-</td>
<td>-</td>
<td>3-Cl-Ph-NH</td>
<td>-</td>
<td>-1.46</td>
<td>-1.72</td>
<td>-0.98</td>
<td>9.50</td>
</tr>
<tr>
<td>88</td>
<td>-</td>
<td>n-C$_3$H$_7$</td>
<td>-</td>
<td>-</td>
<td>-1.16</td>
<td>-1.36</td>
<td>-0.40</td>
<td>2.51</td>
</tr>
<tr>
<td>89</td>
<td>-</td>
<td>Ph-CH$_2$CH$_2$</td>
<td>-</td>
<td>-</td>
<td>-0.96</td>
<td>-0.22</td>
<td>0.47</td>
<td>0.34</td>
</tr>
<tr>
<td>90</td>
<td>-</td>
<td>Ph-CH$_2$CH$_2$CH$_2$</td>
<td>-</td>
<td>-</td>
<td>-0.77</td>
<td>-0.16</td>
<td>0.80</td>
<td>0.16</td>
</tr>
<tr>
<td>91</td>
<td>-</td>
<td>CH$_3$</td>
<td>SCH$_3$</td>
<td>H</td>
<td>-0.60</td>
<td>-0.81</td>
<td>-0.08</td>
<td>1.2</td>
</tr>
<tr>
<td>92</td>
<td>-</td>
<td>CH$_3$</td>
<td>Si(CH$_3$)$_2$CH$_3$</td>
<td>H</td>
<td>-0.26</td>
<td>0.08</td>
<td>-0.32</td>
<td>2.1</td>
</tr>
<tr>
<td>93</td>
<td>-</td>
<td>CH$_3$</td>
<td>NHCH$_2$CH$_3$</td>
<td>CO$_2$-NH-Ph-4-OCH$_3$</td>
<td>-2.11</td>
<td>-2.26</td>
<td>-1.32</td>
<td>21</td>
</tr>
<tr>
<td>94</td>
<td>-</td>
<td>CH$_3$</td>
<td>SCH$_3$</td>
<td>COCH$_2$Ph-4-OCH$_3$</td>
<td>-1.70</td>
<td>-2.16</td>
<td>-1.18</td>
<td>15</td>
</tr>
<tr>
<td>95</td>
<td>-</td>
<td>CH$_3$</td>
<td>SCH$_3$</td>
<td>COCH$_2$Ph-4-isobutyl</td>
<td>-1.40</td>
<td>-1.32</td>
<td>-1.70</td>
<td>50</td>
</tr>
<tr>
<td>96</td>
<td>-</td>
<td>CH$_3$</td>
<td>NHCH$_2$CH$_3$</td>
<td>COCH$_2$Ph-3,4-medioxy</td>
<td>-1.68</td>
<td>-1.59</td>
<td>1.79</td>
<td>61</td>
</tr>
<tr>
<td>97</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-1.31</td>
<td>-1.09</td>
<td>-1.54</td>
<td>35</td>
</tr>
<tr>
<td>98</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-1.06</td>
<td>-0.90</td>
<td>-1.43</td>
<td>27</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁/Amine</td>
<td>R₂</td>
<td>Pred. pKᵢ (nM) (PLS)</td>
<td>Pred. pKᵢ (nM) (RSA)</td>
<td>Exp. pKᵢ (nM) hA₂ΔR</td>
<td>Exp. pKᵢ (nM) hA₂ΔR</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>---------</td>
<td>----------</td>
<td>----</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>99</td>
<td></td>
<td><img src="image1" alt="Structure 1" /></td>
<td></td>
<td></td>
<td>-0.84</td>
<td>-0.58</td>
<td>-1.60</td>
<td>40</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td><img src="image2" alt="Structure 2" /></td>
<td></td>
<td></td>
<td>-1.09</td>
<td>-1.67</td>
<td>-1.20</td>
<td>16</td>
</tr>
<tr>
<td>101</td>
<td></td>
<td><img src="image3" alt="Structure 3" /></td>
<td></td>
<td></td>
<td>-1.10</td>
<td>-1.18</td>
<td>-1.79</td>
<td>61</td>
</tr>
<tr>
<td>102</td>
<td></td>
<td><img src="image4" alt="Structure 4" /></td>
<td></td>
<td></td>
<td>-2.01</td>
<td>-1.77</td>
<td>-1.36</td>
<td>23</td>
</tr>
<tr>
<td>103</td>
<td></td>
<td><img src="image5" alt="Structure 5" /></td>
<td></td>
<td></td>
<td>-1.02</td>
<td>-0.91</td>
<td>-1.18</td>
<td>15</td>
</tr>
<tr>
<td>104</td>
<td></td>
<td><img src="image6" alt="Structure 6" /></td>
<td></td>
<td></td>
<td>-0.95</td>
<td>-1.08</td>
<td>-1.88</td>
<td>76</td>
</tr>
<tr>
<td>105</td>
<td></td>
<td><img src="image7" alt="Structure 7" /></td>
<td></td>
<td></td>
<td>-1.44</td>
<td>-1.37</td>
<td>-2.28</td>
<td>192</td>
</tr>
</tbody>
</table>
## Training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁/Amine*</th>
<th>R₂</th>
<th>Pred. $pK_i$ (nM) (PLS)</th>
<th>Pred. $pK_i$ (nM) (RSA)</th>
<th>Exp. $pK_i$ (nM) hA₂AR</th>
<th>Exp. $K_i$ (nM) hA₂AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>106</td>
<td></td>
<td><img src="image" alt="Molecule 106" /></td>
<td><img src="image" alt="Molecule 106" /></td>
<td>-</td>
<td>-1.42</td>
<td>-1.25</td>
<td>-1.76</td>
<td>57</td>
</tr>
<tr>
<td>107</td>
<td></td>
<td><img src="image" alt="Molecule 107" /></td>
<td><img src="image" alt="Molecule 107" /></td>
<td>-</td>
<td>-1.16</td>
<td>-1.35</td>
<td>-1.88</td>
<td>76</td>
</tr>
<tr>
<td>108</td>
<td></td>
<td><img src="image" alt="Molecule 108" /></td>
<td><img src="image" alt="Molecule 108" /></td>
<td>-</td>
<td>-0.91</td>
<td>-0.67</td>
<td>-1.76</td>
<td>57</td>
</tr>
<tr>
<td>109</td>
<td></td>
<td><img src="image" alt="Molecule 109" /></td>
<td><img src="image" alt="Molecule 109" /></td>
<td>-</td>
<td>-0.55</td>
<td>-0.53</td>
<td>-0.48</td>
<td>3</td>
</tr>
<tr>
<td>110</td>
<td></td>
<td><img src="image" alt="Molecule 110" /></td>
<td><img src="image" alt="Molecule 110" /></td>
<td>-</td>
<td>-1.49</td>
<td>-1.21</td>
<td>-0.95</td>
<td>9</td>
</tr>
<tr>
<td>111</td>
<td></td>
<td><img src="image" alt="Molecule 111" /></td>
<td><img src="image" alt="Molecule 111" /></td>
<td>-</td>
<td>-1.39</td>
<td>-1.07</td>
<td>-0.60</td>
<td>4</td>
</tr>
<tr>
<td>112</td>
<td></td>
<td><img src="image" alt="Molecule 112" /></td>
<td><img src="image" alt="Molecule 112" /></td>
<td>-</td>
<td>-1.30</td>
<td>-1.09</td>
<td>-0.78</td>
<td>6</td>
</tr>
<tr>
<td>113</td>
<td></td>
<td><img src="image" alt="Molecule 113" /></td>
<td><img src="image" alt="Molecule 113" /></td>
<td>-</td>
<td>-0.64</td>
<td>-0.71</td>
<td>-0.90</td>
<td>8</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁/Amine*</td>
<td>R₂</td>
<td>Pred. pKᵢ (nM) (PLS)</td>
<td>Pred. pKᵢ (nM) (RSA)</td>
<td>Exp. pKᵢ (nM) hA₂₅₆R</td>
<td>Exp. pKᵢ (nM) hA₂₅₆R</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>---------</td>
<td>-----------</td>
<td>----</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>114</td>
<td></td>
<td><img src="image1" alt="Image" /></td>
<td></td>
<td></td>
<td>-0.49</td>
<td>-0.60</td>
<td>-0.30</td>
<td>2</td>
</tr>
<tr>
<td>115</td>
<td></td>
<td><img src="image2" alt="Image" /></td>
<td></td>
<td></td>
<td>-1.20</td>
<td>-0.88</td>
<td>-1.08</td>
<td>12</td>
</tr>
<tr>
<td>116</td>
<td></td>
<td><img src="image3" alt="Image" /></td>
<td></td>
<td></td>
<td>-1.20</td>
<td>-1.27</td>
<td>-0.30</td>
<td>2</td>
</tr>
<tr>
<td>117</td>
<td></td>
<td><img src="image4" alt="Image" /></td>
<td></td>
<td></td>
<td>-1.18</td>
<td>-1.14</td>
<td>-0.48</td>
<td>3</td>
</tr>
<tr>
<td>118</td>
<td></td>
<td><img src="image5" alt="Image" /></td>
<td></td>
<td></td>
<td>-0.23</td>
<td>-0.27</td>
<td>-0.78</td>
<td>6</td>
</tr>
<tr>
<td>119</td>
<td></td>
<td><img src="image6" alt="Image" /></td>
<td></td>
<td></td>
<td>-0.81</td>
<td>-0.80</td>
<td>-0.60</td>
<td>4</td>
</tr>
<tr>
<td>120</td>
<td></td>
<td><img src="image7" alt="Image" /></td>
<td></td>
<td></td>
<td>-1.36</td>
<td>-0.96</td>
<td>-0.95</td>
<td>9</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁/Amine⁺</td>
<td>R₂</td>
<td>Pred. pKᵢ (nM) (PLS)</td>
<td>Pred. pKᵢ (nM) (RSA)</td>
<td>Exp. pKᵢ (nM)</td>
<td>hₐ₂ₐ₁R</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.66</td>
<td>-0.54</td>
<td>-0.70</td>
<td>5</td>
</tr>
<tr>
<td>122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.58</td>
<td>-1.32</td>
<td>-1.76</td>
<td>57</td>
</tr>
<tr>
<td>123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.63</td>
<td>-1.70</td>
<td>-1.04</td>
<td>11</td>
</tr>
<tr>
<td>124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.43</td>
<td>-1.27</td>
<td>-1.43</td>
<td>27</td>
</tr>
<tr>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.25</td>
<td>-1.36</td>
<td>-2.02</td>
<td>104</td>
</tr>
<tr>
<td>126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.35</td>
<td>-1.62</td>
<td>-1.23</td>
<td>17</td>
</tr>
<tr>
<td>127</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.17</td>
<td>-1.39</td>
<td>-1.15</td>
<td>14</td>
</tr>
</tbody>
</table>
B.2 Internal test set

128-130

131-133

134

135

136-137

128-129: [1]; 130: [3]; 131-133: [4]; 134: [5]; 135: [6]; 136-137: [7]
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R₁</th>
<th>R₂</th>
<th>R₃</th>
<th>pKᵢ (nM)</th>
<th>pKᵢ (mM)</th>
<th>Exp. pKᵢ (nM)</th>
<th>Pred. pKᵢ (mM)</th>
<th>Exp. pKᵢ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-0.11</td>
<td>0.25</td>
<td>0.89</td>
<td>4.40</td>
<td>2.70</td>
</tr>
<tr>
<td>129</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-0.55</td>
<td>-0.05</td>
<td>-0.13</td>
<td>6.14</td>
<td>6.14</td>
</tr>
<tr>
<td>130</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-2.36</td>
<td>-2.06</td>
<td>-2.79</td>
<td>3.70</td>
<td>6.14</td>
</tr>
<tr>
<td>131</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-1.47</td>
<td>-2.08</td>
<td>-3.02</td>
<td>4.40</td>
<td>4.40</td>
</tr>
<tr>
<td>132</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-1.37</td>
<td>-1.39</td>
<td>-1.00</td>
<td>10.00</td>
<td>10.00</td>
</tr>
<tr>
<td>133</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-0.82</td>
<td>-0.35</td>
<td>-0.08</td>
<td>1.95</td>
<td>1.95</td>
</tr>
<tr>
<td>134</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-1.14</td>
<td>1.68</td>
<td>-1.13</td>
<td>6.04</td>
<td>6.04</td>
</tr>
<tr>
<td>135</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-1.11</td>
<td>-1.11</td>
<td>-1.86</td>
<td>7.50</td>
<td>7.50</td>
</tr>
<tr>
<td>136</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-1.11</td>
<td>-1.11</td>
<td>-1.86</td>
<td>7.50</td>
<td>7.50</td>
</tr>
<tr>
<td>137</td>
<td>CH₂NH₂</td>
<td></td>
<td></td>
<td>-1.11</td>
<td>-1.11</td>
<td>-1.86</td>
<td>7.50</td>
<td>7.50</td>
</tr>
</tbody>
</table>
### B.3 External test set

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>Predicted $pK_i$ (nM)</th>
<th>Predicted $pK_i$ (nM)</th>
<th>Experimental $h\alpha_{2\beta}R$</th>
<th>Experimental $h\alpha_{2\beta}R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>138</td>
<td>H</td>
<td>-1.96</td>
<td>-2.14</td>
<td>-2.37</td>
<td>233</td>
</tr>
<tr>
<td>139</td>
<td>2-CH$_3$</td>
<td>-1.87</td>
<td>-1.85</td>
<td>-2.52</td>
<td>333</td>
</tr>
<tr>
<td>140</td>
<td>2,6-Cl</td>
<td>-1.83</td>
<td>-1.94</td>
<td>-2.45</td>
<td>281</td>
</tr>
<tr>
<td>141</td>
<td>3-Cl</td>
<td>-1.67</td>
<td>-1.98</td>
<td>-2.44</td>
<td>278</td>
</tr>
<tr>
<td>142</td>
<td>4-CH$_3$</td>
<td>-1.82</td>
<td>-1.97</td>
<td>-2.88</td>
<td>751</td>
</tr>
</tbody>
</table>

138-142: [8]
B.4 References


APPENDIX C

Paper III
Supplementary Information

C.1 Dataset.
C.2 References.

Table C.1 Descriptors selected for the training set in Model 3 in the previous publication to classify CYP450 substrates in three classes. The numbers in the first column refer to the publication by Terfloh et al.

Table C.2 Multi-classification model by applying the ct-SVM technique to Data set 1 after the computation of the twelve selected descriptors in Model 3 published in Terfloh et al. paper: the statistical parameters for each class after prediction on the Validation set 1 (67 substrates) are reported.

Table C.3 Multi-classification model by applying the ct-SVM technique to Data set 1 after the computation of the twelve selected descriptors in Model 3 published in Terfloh et al. paper: performance measures after prediction on the Validation set 1 (67 substrates) and on the Test set (217 substrates) are summarized.

Table C.4 Multi-label classification ct-SVM/7classes model using the same descriptors of Model 3 in ISOCYP paper: predicted results of the model for the Test set 1.

Table C.5 ct-SVM/7classes model parameters.

Table C.6 ct-SVM/5classes model parameters.

Table C.7 CPG-NN/5classes model parameters.

Table C.8 SVM/5classes model parameters.

Table C.9 CPG-NN/5classes percentage (%) of correct predictions after LOO validation procedure.
### C.1 Dataset

<table>
<thead>
<tr>
<th>Mol</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>322</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>2</td>
<td>859</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>3</td>
<td>864</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>4</td>
<td>868</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>5</td>
<td>2689</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>6</td>
<td>3088</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>7</td>
<td>3179</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>8</td>
<td>3326</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>9</td>
<td>3676</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>10</td>
<td>3908</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>11</td>
<td>5018</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>12</td>
<td>5354</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>13</td>
<td>5476</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>14</td>
<td>5491</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>15</td>
<td>5516</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>16</td>
<td>5712</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>17</td>
<td>5726</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>18</td>
<td>6095</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>19</td>
<td>6228</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>20</td>
<td>6389</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>21</td>
<td>6609</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>22</td>
<td>6610</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>MolId</td>
<td>Molécule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>23</td>
<td>6634</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>24</td>
<td>6635</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>25</td>
<td>8606</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>26</td>
<td>8616</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>27</td>
<td>8641</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>28</td>
<td>9063</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>29</td>
<td>9348</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>30</td>
<td>9862</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>31</td>
<td>9948</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>32</td>
<td>10011</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>33</td>
<td>10089</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>34</td>
<td>11120</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>35</td>
<td>11124</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>36</td>
<td>11295</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>37</td>
<td>11300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>38</td>
<td>11441</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>39</td>
<td>11443</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>40</td>
<td>11447</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>41</td>
<td>11449</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>42</td>
<td>11750</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>43</td>
<td>12345</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>44</td>
<td>12451</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>45</td>
<td>13212</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>46</td>
<td>13414</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>47</td>
<td>13499</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>48</td>
<td>14334</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>49</td>
<td>14343</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>50</td>
<td>14467</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>51</td>
<td>14628</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>52</td>
<td>15061</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>53</td>
<td>15508</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>54</td>
<td>15852</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>55</td>
<td>16165</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>56</td>
<td>16306</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>57</td>
<td>16430</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>58</td>
<td>16457</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>59</td>
<td>16460</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>60</td>
<td>16462</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>61</td>
<td>16608</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>62</td>
<td>16611</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>63</td>
<td>16781</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>64</td>
<td>17178</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>65</td>
<td>17304</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>66</td>
<td>17308</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
</tbody>
</table>
## Dataset: continued

<table>
<thead>
<tr>
<th>MolId</th>
<th>Molécula náme</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>18423</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td></td>
<td>[1]</td>
</tr>
<tr>
<td>68</td>
<td>19936</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>19943</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>19960</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>20159</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>20140</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>21115</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>21125</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>21138</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>21342</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>21344</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>21491</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>22139</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>22636</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>23276</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>23281</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>23669</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>23677</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>24195</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>25165</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>25307</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>25318</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>25319</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>90</td>
<td>25414</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>91</td>
<td>25647</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>92</td>
<td>26299</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>93</td>
<td>27253</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>94</td>
<td>27906</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>95</td>
<td>27907</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>96</td>
<td>27909</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>97</td>
<td>29127</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>98</td>
<td>29870</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>99</td>
<td>29889</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>100</td>
<td>30029</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>101</td>
<td>30031</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>102</td>
<td>30034</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>103</td>
<td>30085</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>104</td>
<td>30095</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>105</td>
<td>30529</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>106</td>
<td>30531</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>107</td>
<td>31064</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>108</td>
<td>31357</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>109</td>
<td>31543</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>110</td>
<td>31809</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
</tbody>
</table>
## Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>34120</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>112</td>
<td>34454</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>113</td>
<td>34455</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>114</td>
<td>34457</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>115</td>
<td>34458</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>116</td>
<td>34475</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>117</td>
<td>34579</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>118</td>
<td>34600</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>119</td>
<td>35597</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>120</td>
<td>35600</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>121</td>
<td>36005</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>122</td>
<td>36731</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>123</td>
<td>37213</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>124</td>
<td>37474</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>125</td>
<td>37599</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>126</td>
<td>38395</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>127</td>
<td>38398</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>128</td>
<td>38402</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>129</td>
<td>38404</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>130</td>
<td>38407</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>131</td>
<td>38416</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>132</td>
<td>38528</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>133</td>
<td>38680</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>134</td>
<td>38744</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>135</td>
<td>39279</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>136</td>
<td>40334</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>137</td>
<td>40472</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>40931</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>40994</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>140</td>
<td>41354</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>41405</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>142</td>
<td>41788</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>41972</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>44100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>44405</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>44972</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>44978</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>148</td>
<td>45411</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>46098</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>46231</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>46232</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>46346</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>46350</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
<td></td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molécule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>156</td>
<td>47019</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>157</td>
<td>47020</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>158</td>
<td>47041</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>159</td>
<td>47755</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>160</td>
<td>48285</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>161</td>
<td>48336</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>162</td>
<td>48337</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>163</td>
<td>48566</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>164</td>
<td>48657</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>166</td>
<td>48685</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>167</td>
<td>49000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>168</td>
<td>504113</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>169</td>
<td>506888</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>170</td>
<td>508356</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>171</td>
<td>51201</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>172</td>
<td>51488</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>173</td>
<td>1727</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>174</td>
<td>51728</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>175</td>
<td>51800</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>176</td>
<td>51831</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>177</td>
<td>52799</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>178</td>
<td>52801</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>179</td>
<td>53365</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>180</td>
<td>53367</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>181</td>
<td>55175</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>182</td>
<td>55955</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>183</td>
<td>57002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>184</td>
<td>57004</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>185</td>
<td>57197</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>186</td>
<td>58162</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>187</td>
<td>58164</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>188</td>
<td>58173</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>189</td>
<td>58176</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>190</td>
<td>598883</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>191</td>
<td>598887</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>192</td>
<td>598889</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>193</td>
<td>60153</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>194</td>
<td>60598</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>195</td>
<td>60616</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>196</td>
<td>60621</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>197</td>
<td>60622</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>198</td>
<td>60623</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>199</td>
<td>60824</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
<tr>
<td>200</td>
<td>61327</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No [1]</td>
</tr>
</tbody>
</table>
## Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molécule n°</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>61469</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>202</td>
<td>61470</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>203</td>
<td>61472</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>204</td>
<td>61550</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>205</td>
<td>61815</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>206</td>
<td>61816</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>207</td>
<td>620027</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>208</td>
<td>62126</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>209</td>
<td>62006</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>210</td>
<td>62744</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>211</td>
<td>63097</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>212</td>
<td>63098</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>213</td>
<td>63099</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>214</td>
<td>63100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>215</td>
<td>63101</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>216</td>
<td>63102</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>217</td>
<td>63104</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>218</td>
<td>63115</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>219</td>
<td>63333</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>220</td>
<td>63924</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>221</td>
<td>63929</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>222</td>
<td>64618</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>223</td>
<td>64627</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>224</td>
<td>64819</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>225</td>
<td>64580</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>226</td>
<td>64929</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>227</td>
<td>65245</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>228</td>
<td>66235</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>229</td>
<td>66260</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>230</td>
<td>66337</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>231</td>
<td>66427</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>232</td>
<td>66478</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>233</td>
<td>67107</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>234</td>
<td>67111</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>235</td>
<td>67517</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>236</td>
<td>67596</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>237</td>
<td>68665</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>238</td>
<td>68590</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>239</td>
<td>68791</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>240</td>
<td>69572</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>241</td>
<td>69573</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>242</td>
<td>69575</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>243</td>
<td>69578</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
</tbody>
</table>
## Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molécule n°</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>244</td>
<td>69579</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>245</td>
<td>69585</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>246</td>
<td>69734</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>247</td>
<td>69807</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>248</td>
<td>70101</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>249</td>
<td>70678</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>250</td>
<td>70714</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>251</td>
<td>71062</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>252</td>
<td>71072</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>253</td>
<td>71098</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>254</td>
<td>71099</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>255</td>
<td>71410</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>256</td>
<td>72142</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>257</td>
<td>72588</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>258</td>
<td>72922</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>259</td>
<td>73006</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>260</td>
<td>73066</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>261</td>
<td>73076</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>262</td>
<td>73184</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>263</td>
<td>73186</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>264</td>
<td>73201</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>265</td>
<td>73211</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>266</td>
<td>73212</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>267</td>
<td>73333</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>268</td>
<td>Acetofenac</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[1]</td>
</tr>
<tr>
<td>269</td>
<td>Acenocoumarol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[2]</td>
</tr>
<tr>
<td>270</td>
<td>Acetaminophen</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>271</td>
<td>Alfentanil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>272</td>
<td>Alfuzosin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>273</td>
<td>Almotriptan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>274</td>
<td>Alpiden</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>275</td>
<td>Alprazolam</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>276</td>
<td>Alprostadil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[5]</td>
</tr>
<tr>
<td>277</td>
<td>Amitriptyline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>278</td>
<td>Amiodarone</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>279</td>
<td>Amisulpride</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>280</td>
<td>Amlopidine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>281</td>
<td>Amphetamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>282</td>
<td>Amprenavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>283</td>
<td>Aprepitant</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>284</td>
<td>Arbidolol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[5]</td>
</tr>
<tr>
<td>285</td>
<td>Atenolol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>286</td>
<td>Atazanavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>287</td>
<td>Atomoxetine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>288</td>
<td>Atorvastatin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>289</td>
<td>Azatadin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>290</td>
<td>Beclomethasone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>291</td>
<td>Benzphetamine</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>292</td>
<td>Bepridil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>293</td>
<td>Benazotene</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>294</td>
<td>Bisoprolol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>295</td>
<td>Bromocriptine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>296</td>
<td>Budesonide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>297</td>
<td>Bufamidol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>298</td>
<td>Buspirone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>299</td>
<td>Busulfan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>300</td>
<td>Caffeine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>301</td>
<td>Carbamazepine</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>302</td>
<td>Carisoprodol</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>303</td>
<td>Carmustine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>304</td>
<td>Carvedilol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>305</td>
<td>Celecoxib</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>306</td>
<td>Celecoxib</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[5]</td>
</tr>
<tr>
<td>307</td>
<td>Cerivastatin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>308</td>
<td>Chlorbemepride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>309</td>
<td>Chlorpheniramine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>310</td>
<td>Chlorpromazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>311</td>
<td>Chlorpropamide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>312</td>
<td>Chlorzoxazone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>313</td>
<td>Clofazol</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>314</td>
<td>Cinacalcet</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>315</td>
<td>Cinnarizine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>316</td>
<td>Cisapride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>317</td>
<td>Citalopram</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>318</td>
<td>Clarithromycin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>319</td>
<td>Clindamycin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>320</td>
<td>Clomipramine</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>321</td>
<td>Clozapine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>322</td>
<td>Clopidogrel</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>323</td>
<td>Clozapine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>324</td>
<td>Cocaine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>325</td>
<td>Codeine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>326</td>
<td>Colchicine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>327</td>
<td>Cotrimoxazole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>328</td>
<td>Cyclophosphamide</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>329</td>
<td>Cyclophosphamide</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>330</td>
<td>Cyclosporin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>331</td>
<td>Dapsone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>332</td>
<td>Darifenacin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>333</td>
<td>Debrisoquine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[5]</td>
</tr>
<tr>
<td>334</td>
<td>Delavirdine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>335</td>
<td>Deprenyl</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>336</td>
<td>Desipramine</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>337</td>
<td>Desogestrel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>338</td>
<td>Decamethasone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>339</td>
<td>Desfluramine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>340</td>
<td>Dextromethorphan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>341</td>
<td>Diclofenac</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>342</td>
<td>Dilofencan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>343</td>
<td>Dihydrocodeine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[2]</td>
</tr>
<tr>
<td>344</td>
<td>Dihydroergotamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>345</td>
<td>Diltiazem</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>346</td>
<td>Disopyramide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>347</td>
<td>Docetaxel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>348</td>
<td>Dofetilide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>349</td>
<td>Dolasetron</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>350</td>
<td>Domperidone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>351</td>
<td>Donepezil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>352</td>
<td>Doxepin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>353</td>
<td>Doxorubicin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>354</td>
<td>Dronabinol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>355</td>
<td>Duloxetine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>356</td>
<td>Dutasteride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>357</td>
<td>Ebastine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>358</td>
<td>Efavirenz</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>359</td>
<td>Enalapril</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>360</td>
<td>Escinüide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>361</td>
<td>Enflurane</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>362</td>
<td>Eplerenone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>363</td>
<td>Ergotamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>364</td>
<td>Erlotinib</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>365</td>
<td>Erythromycin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>366</td>
<td>Estradiol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>367</td>
<td>Esliclociclov</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>368</td>
<td>Ethanol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>369</td>
<td>Ethynylestradiol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[5]</td>
</tr>
<tr>
<td>370</td>
<td>Ethoxanidamide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>371</td>
<td>Ethylmorphine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[2]</td>
</tr>
<tr>
<td>372</td>
<td>Etoposide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>373</td>
<td>Etoposide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>374</td>
<td>Exemestane</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>375</td>
<td>Felodipine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>376</td>
<td>Fenfluramine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>377</td>
<td>Fentanyl</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>378</td>
<td>Fesofenadine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>379</td>
<td>Finasteride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>380</td>
<td>Flecaïnide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>381</td>
<td>Flavoxate</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>382</td>
<td>Flunarizine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>383</td>
<td>Flunitrazepam</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>384</td>
<td>Flusoxetine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>385</td>
<td>Fluphenazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>386</td>
<td>Flurbiprofen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>387</td>
<td>Flutamide</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>388</td>
<td>Fluticasone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>389</td>
<td>Fluvastatin</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>390</td>
<td>Fluvastatin</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>391</td>
<td>Formoterol</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>392</td>
<td>Fulvestrant</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>393</td>
<td>Galantamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>394</td>
<td>Glimépiride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>395</td>
<td>Glipizide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>396</td>
<td>Glyburide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>397</td>
<td>Granisetron</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[2]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYPA4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>398</td>
<td>Haloperidol</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>399</td>
<td>Halothane</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>400</td>
<td>Hexobarbital</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>401</td>
<td>Hydrocodone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>402</td>
<td>Ibuprofen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>403</td>
<td>Ifosfamide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>404</td>
<td>Imatinib</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>405</td>
<td>Imipramine</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>406</td>
<td>Indinavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>407</td>
<td>Indomethacin</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>408</td>
<td>Irbesartan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>409</td>
<td>Irinotecan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>410</td>
<td>Isotretinoin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>411</td>
<td>Letrozole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>412</td>
<td>Lisinopril</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>413</td>
<td>Itraconazole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>414</td>
<td>Ketoconazole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>415</td>
<td>Lansoprazole</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>416</td>
<td>Lenovozepine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>417</td>
<td>Levodopa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>418</td>
<td>Levobupivacaine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>419</td>
<td>Levonorgestrel</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Mol name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>420</td>
<td>Lidocaine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>421</td>
<td>Lisuride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>422</td>
<td>Lobeline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>423</td>
<td>Lopinavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>424</td>
<td>Loratadine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>425</td>
<td>Lormoxacin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>426</td>
<td>Losartan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>427</td>
<td>Lovastatin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>428</td>
<td>Maprotiline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>429</td>
<td>Maraviroc</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>430</td>
<td>Medroxyprogesterone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>431</td>
<td>Mefenamic acid</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>432</td>
<td>Meloxicam</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>433</td>
<td>Meperidine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>434</td>
<td>Meprobamate</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>435</td>
<td>Methadone</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>436</td>
<td>Methamphetamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>437</td>
<td>Methoxyamphetamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>438</td>
<td>Methoxyflurane</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>439</td>
<td>Methoxyphenamine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>440</td>
<td>Methylprednisolone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>441</td>
<td>Metoclopramide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[5]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>442</td>
<td>Metoprolol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>443</td>
<td>Mexiletine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>444</td>
<td>Mianserin</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[2]</td>
</tr>
<tr>
<td>445</td>
<td>Miconazole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>446</td>
<td>Midazolam</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>447</td>
<td>Mifepristone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>448</td>
<td>Minaprine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>449</td>
<td>Mirtazapine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>450</td>
<td>Moclobemide</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>451</td>
<td>Modafinil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>452</td>
<td>Mometasone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>453</td>
<td>Montelukast</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>454</td>
<td>Morphine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>455</td>
<td>Naproxen</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>456</td>
<td>Nateglinide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>457</td>
<td>Nefarinavir</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>458</td>
<td>Nevirapine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>459</td>
<td>Nicardipine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>460</td>
<td>Nifedipine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>461</td>
<td>Nilotamide</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>462</td>
<td>Nimodipine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>464</td>
<td>Nisoldipine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>465</td>
<td>Nitrendipine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>466</td>
<td>Norethisterone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>467</td>
<td>Nortriptyline</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>468</td>
<td>Olanzapine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>469</td>
<td>Omeprazole</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>470</td>
<td>Ondansetron</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td></td>
<td>[3]</td>
</tr>
<tr>
<td>471</td>
<td>Oxybutynin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td></td>
<td>[3]</td>
</tr>
<tr>
<td>472</td>
<td>Oxycodone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>473</td>
<td>Paclitaxel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>474</td>
<td>Pantoprazole</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>475</td>
<td>Paroxetine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>476</td>
<td>Pentamidine</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td></td>
<td>[3]</td>
</tr>
<tr>
<td>477</td>
<td>Perhexiline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>478</td>
<td>Perphenazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>479</td>
<td>Phenformin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>480</td>
<td>Phenylbutazone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>481</td>
<td>Phenytol</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>482</td>
<td>Pimozone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>483</td>
<td>Pindolol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>484</td>
<td>Piroxicam</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>486</td>
<td>Prednisolone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>487</td>
<td>Prednisone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>488</td>
<td>Progesterone</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>489</td>
<td>Proguanil</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>490</td>
<td>Promethazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>491</td>
<td>Propafenone</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>492</td>
<td>Propoxyphene</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>493</td>
<td>Proparanol</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>494</td>
<td>Quercetin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>495</td>
<td>Quetiapine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>496</td>
<td>Quinidine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>497</td>
<td>Quinine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>498</td>
<td>Rabeproside</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>499</td>
<td>Ramelteon</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>500</td>
<td>Ranolazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>501</td>
<td>Remoxipride</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>502</td>
<td>Repaglinide</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>503</td>
<td>Retinoic acid</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>504</td>
<td>Retinol</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>505</td>
<td>Rifabutin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>506</td>
<td>Rifampicin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>507</td>
<td>Riluzole</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
</tbody>
</table>
## Dataset: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>508</td>
<td>Risperidone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>509</td>
<td>Ritonavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>510</td>
<td>Ropivacaine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>511</td>
<td>Ropinirole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>512</td>
<td>Rosiglitazone</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>513</td>
<td>Salmeterol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>514</td>
<td>Saquinavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>515</td>
<td>Sertindole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>516</td>
<td>Sertaline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>517</td>
<td>Serotonin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>518</td>
<td>Sitagliptin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>519</td>
<td>Sildenafil</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>520</td>
<td>Sildenafil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>521</td>
<td>Sirolimus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>522</td>
<td>Sotalol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>523</td>
<td>Sotalol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>524</td>
<td>Sotufenec</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>525</td>
<td>Sufentanil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>526</td>
<td>Sulfadiazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>527</td>
<td>Sulfamethoxazole</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>528</td>
<td>Sulfadiazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>529</td>
<td>Sunitinib</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
</tbody>
</table>
### Dataset: continued

<table>
<thead>
<tr>
<th>MolId</th>
<th>Molecule name</th>
<th>CYP1A2</th>
<th>CYP2C19</th>
<th>CYP2C8</th>
<th>CYP2C9</th>
<th>CYP2D6</th>
<th>CYP2E1</th>
<th>CYP3A4</th>
<th>Multilabel</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>530</td>
<td>Suprofen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[5]</td>
</tr>
<tr>
<td>531</td>
<td>Tacrine</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>532</td>
<td>Tacrolimus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>533</td>
<td>Tadalafil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>534</td>
<td>Tamoxifen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>535</td>
<td>Tamsulosin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[2]</td>
</tr>
<tr>
<td>536</td>
<td>Telithromycin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>537</td>
<td>Temazepam</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>538</td>
<td>Teniposide</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>539</td>
<td>Tenoxécum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>540</td>
<td>Terfenadine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>541</td>
<td>Testosterone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>542</td>
<td>Theophylline</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>543</td>
<td>Thiordanizine</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>544</td>
<td>Timolol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>545</td>
<td>Timidazine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>546</td>
<td>Tiropranavir</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>547</td>
<td>Tizanidine</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>548</td>
<td>Tolbutamidé</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>549</td>
<td>Tolterodine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>550</td>
<td>Torasemide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>551</td>
<td>Torsemide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>Mol Id</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>552</td>
<td>Tramadol</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>553</td>
<td>Trazodone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>554</td>
<td>Trimazolam</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>555</td>
<td>Trimethoprim</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>556</td>
<td>Trimethoprim</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>557</td>
<td>Trimipramine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>558</td>
<td>Trospium</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[2]</td>
</tr>
<tr>
<td>559</td>
<td>Valdecoxib</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>560</td>
<td>Valsartan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>561</td>
<td>Vardenafil</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>562</td>
<td>Venlafaxine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>563</td>
<td>Verapamil</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>564</td>
<td>Vinblastine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>565</td>
<td>Vincristine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>566</td>
<td>Vindesine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[2]</td>
</tr>
<tr>
<td>567</td>
<td>Vinorelbine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>568</td>
<td>Voriconazole</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>569</td>
<td>Warfarin-(R)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>570</td>
<td>Warfarin-(S)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>571</td>
<td>Zafirlukast</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>572</td>
<td>Zaleplon</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>573</td>
<td>Zidovudine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[4]</td>
</tr>
<tr>
<td>Mol</td>
<td>Molecule name</td>
<td>CYP1A2</td>
<td>CYP2C19</td>
<td>CYP2C8</td>
<td>CYP2C9</td>
<td>CYP2D6</td>
<td>CYP2E1</td>
<td>CYP3A4</td>
<td>Multilabel</td>
<td>References</td>
</tr>
<tr>
<td>-----</td>
<td>----------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>574</td>
<td>Zileuton</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td>[3]</td>
</tr>
<tr>
<td>575</td>
<td>Ziprasidone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>576</td>
<td>Zolmitriptan</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[5]</td>
</tr>
<tr>
<td>577</td>
<td>Zolpidem</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>578</td>
<td>Zonisamide</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>579</td>
<td>Zopiclone</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
<tr>
<td>580</td>
<td>Zopiclone</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>[3]</td>
</tr>
</tbody>
</table>
C.2 References


Table C.1

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>2D-AC_{identity} (5)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>145</td>
<td>2D-AC_{q\pi} (3)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>148</td>
<td>2D-AC_{q\pi} (6)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>126</td>
<td>2D-AC_{\chi\pi} (5)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>133</td>
<td>2D-AC_{q\sigma} (1)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>134</td>
<td>2D-AC_{q\sigma} (2)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>116</td>
<td>2D-AC_{\chi\sigma} (6)</td>
<td>Topological autocorrelation</td>
</tr>
<tr>
<td>223</td>
<td>3D-AC_{identity} ([5.8-5.9]Å)</td>
<td>Spatial autocorrelation</td>
</tr>
<tr>
<td>33</td>
<td>n_{acidic groups}</td>
<td>Number of acidic functional groups</td>
</tr>
<tr>
<td>27</td>
<td>n_{aliph_amino}</td>
<td>Number of aliphatic amino groups</td>
</tr>
<tr>
<td>32</td>
<td>n_{basic_nitrogen}</td>
<td>Number of basic, nitrogen containing functional groups</td>
</tr>
<tr>
<td>26</td>
<td>r_{3}</td>
<td>Radius perpendicular to $D_3$ and $R_2$</td>
</tr>
</tbody>
</table>
Table C.2

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate</th>
<th>FP rate</th>
<th>TN rate</th>
<th>FN rate</th>
<th>Recall</th>
<th>Precision</th>
<th>% correct predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.09</td>
<td>0.02</td>
<td>0.98</td>
<td>0.91</td>
<td>0.09</td>
<td>0.50</td>
<td>83.6</td>
</tr>
<tr>
<td>CYP2C19</td>
<td>0.25</td>
<td>0.02</td>
<td>0.98</td>
<td>0.75</td>
<td>0.25</td>
<td>0.50</td>
<td>94.0</td>
</tr>
<tr>
<td>CYP2C8</td>
<td>0.25</td>
<td>0.02</td>
<td>0.98</td>
<td>0.75</td>
<td>0.25</td>
<td>0.50</td>
<td>94.0</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.73</td>
<td>0.04</td>
<td>0.96</td>
<td>0.27</td>
<td>0.73</td>
<td>0.80</td>
<td>92.5</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.93</td>
<td>0.08</td>
<td>0.92</td>
<td>0.07</td>
<td>0.93</td>
<td>0.78</td>
<td>92.5</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>1.00</td>
<td>0.02</td>
<td>0.98</td>
<td>0.00</td>
<td>1.00</td>
<td>0.89</td>
<td>98.5</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.68</td>
<td>0.17</td>
<td>0.83</td>
<td>0.32</td>
<td>0.68</td>
<td>0.78</td>
<td>76.1</td>
</tr>
</tbody>
</table>
### Table C.3

<table>
<thead>
<tr>
<th>Model prediction</th>
<th>Accuracy $ML_{(n=1)}$</th>
<th>One-error</th>
<th>Coverage</th>
<th>Average precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation set 1</td>
<td>0.70</td>
<td>0.19</td>
<td>2.04</td>
<td>0.85</td>
</tr>
<tr>
<td>Test set 1</td>
<td>0.71</td>
<td>0.25</td>
<td>1.94</td>
<td>0.82</td>
</tr>
</tbody>
</table>
Table C.4

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP rate</th>
<th>FP rate</th>
<th>TN rate</th>
<th>FN rate</th>
<th>Recall</th>
<th>Precision</th>
<th>% correct predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>0.48</td>
<td>0.05</td>
<td>0.95</td>
<td>0.52</td>
<td>0.48</td>
<td>0.52</td>
<td>89.9</td>
</tr>
<tr>
<td>CYP2C19</td>
<td>0.10</td>
<td>0.01</td>
<td>0.99</td>
<td>0.90</td>
<td>0.10</td>
<td>0.25</td>
<td>94.5</td>
</tr>
<tr>
<td>CYP2C8</td>
<td>0.33</td>
<td>0.02</td>
<td>0.98</td>
<td>0.77</td>
<td>0.33</td>
<td>0.17</td>
<td>94.5</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>0.44</td>
<td>0.03</td>
<td>0.97</td>
<td>0.56</td>
<td>0.44</td>
<td>0.61</td>
<td>93.1</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>0.86</td>
<td>0.15</td>
<td>0.85</td>
<td>0.14</td>
<td>0.86</td>
<td>0.70</td>
<td>85.2</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>0.92</td>
<td>0.03</td>
<td>0.97</td>
<td>0.08</td>
<td>0.92</td>
<td>0.61</td>
<td>96.3</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>0.78</td>
<td>0.22</td>
<td>0.78</td>
<td>0.22</td>
<td>0.78</td>
<td>0.80</td>
<td>78.3</td>
</tr>
</tbody>
</table>
### Table C.5

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Classes CYP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1A2</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table C.6

<table>
<thead>
<tr>
<th>Parameters</th>
<th>1A2</th>
<th>2C9</th>
<th>2D6</th>
<th>2E1</th>
<th>3A4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>γ</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table C.7

<table>
<thead>
<tr>
<th>Parameter set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net dimensions</td>
</tr>
<tr>
<td>Initial span ((x, y))</td>
</tr>
<tr>
<td>Span step ((x, y))</td>
</tr>
<tr>
<td>Number of cycles</td>
</tr>
<tr>
<td>Topology</td>
</tr>
</tbody>
</table>
### Table C.8

<table>
<thead>
<tr>
<th>Parameter set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel polynomial, degree=3</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
### Table C.9

<table>
<thead>
<tr>
<th>Classes</th>
<th>% correct predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>81.4</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>87.8</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>87.8</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>98.0</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>76.5</td>
</tr>
</tbody>
</table>
Appendix D

Paper IV

Supplementary Information

D.1 SVM classification model, training set.
D.2 SVM regression model, training set.
D.3 Test set.
D.4 References.
D.1 SVM classification model, training set

1-42: [1]; 43: [2]; 44-56: [3]; 57-62: [4]; 63-71: [5]; 72-91: [2]; 92-93: [6]; 94-96: [7]; 97-101: [2]; 102-104: [8]
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Class model</th>
<th>Exp. Kᵢ (nM)</th>
<th>Exp. Kᵢ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>CH₃</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td>-</td>
<td>SVM model</td>
<td>143</td>
<td>3.40</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>CH₃</td>
<td>3,4-Cl₂-CH₃-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>680</td>
<td>0.24</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>CH₃</td>
<td>4-NO₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>695</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>CH₃</td>
<td>4-CH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>110</td>
<td>0.31</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>CH₃</td>
<td>4-Brc-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>100</td>
<td>0.46</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>CH₃</td>
<td>4-F-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>120</td>
<td>0.34</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>CH₃</td>
<td>4-CF₃-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>140</td>
<td>0.74</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>CH₃</td>
<td>2-OCH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>180</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>CH₃</td>
<td>3-OCH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>160</td>
<td>0.80</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>CH₃</td>
<td>2-C₃H₇-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>200</td>
<td>0.91</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>CH₃</td>
<td>4-C₃H₇-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>180</td>
<td>0.29</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>C₂H₅</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>352</td>
<td>3.00</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>C₂H₅</td>
<td>3,4-Cl₂-CH₃-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>576</td>
<td>0.27</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-NO₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>614</td>
<td>0.65</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>C₂H₅</td>
<td>2-OCH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>133</td>
<td>0.56</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>C₂H₅</td>
<td>3-OCH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>140</td>
<td>0.86</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>C₂H₅</td>
<td>2-C₃H₇-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>150</td>
<td>0.30</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-C₃H₇-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>160</td>
<td>0.20</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>401</td>
<td>2.50</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>3,4-Cl₂-CH₃-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>667</td>
<td>0.30</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-NO₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>1115</td>
<td>0.81</td>
</tr>
<tr>
<td>22</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>2-OCH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>100</td>
<td>0.34</td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>3-OCH₂-Ph-NH-CO</td>
<td>-</td>
<td>-1</td>
<td>113</td>
<td>0.40</td>
</tr>
</tbody>
</table>
### SVM classification model, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>R&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Class SVM model</th>
<th>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM) h&lt;sub&gt;A&lt;/sub&gt;&lt;sub&gt;2&lt;/sub&gt;R</th>
<th>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM) h&lt;sub&gt;A&lt;/sub&gt;&lt;sub&gt;3&lt;/sub&gt;R</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>121</td>
<td>0.71</td>
</tr>
<tr>
<td>25</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>140</td>
<td>0.34</td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>3,4-Cl&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>405</td>
<td>3.70</td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>3,4-OCH&lt;sub&gt;3&lt;/sub&gt;O-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>376</td>
<td>0.50</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>4-NO&lt;sub&gt;2&lt;/sub&gt;Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>503</td>
<td>0.55</td>
</tr>
<tr>
<td>29</td>
<td>-</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>0.86</td>
</tr>
<tr>
<td>30</td>
<td>-</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>111</td>
<td>0.43</td>
</tr>
<tr>
<td>31</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>-</td>
<td>423</td>
<td>0.81</td>
</tr>
<tr>
<td>32</td>
<td>-</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>-</td>
<td>335</td>
<td>1.03</td>
</tr>
<tr>
<td>33</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>-</td>
<td>306</td>
<td>1.01</td>
</tr>
<tr>
<td>34</td>
<td>-</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>-</td>
<td>400</td>
<td>1.11</td>
</tr>
<tr>
<td>35</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>381</td>
<td>0.16</td>
</tr>
<tr>
<td>36</td>
<td>-</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>148</td>
<td>0.21</td>
</tr>
<tr>
<td>37</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>1300</td>
<td>0.20</td>
</tr>
<tr>
<td>38</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>1195</td>
<td>0.40</td>
</tr>
<tr>
<td>39</td>
<td>-</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>1040</td>
<td>0.60</td>
</tr>
<tr>
<td>40</td>
<td>-</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>180</td>
<td>1.60</td>
</tr>
<tr>
<td>41</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>140</td>
<td>0.80</td>
</tr>
<tr>
<td>42</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>1220</td>
<td>0.91</td>
</tr>
<tr>
<td>43</td>
<td>-</td>
<td>H</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>520</td>
<td>0.14</td>
</tr>
<tr>
<td>44</td>
<td>-</td>
<td>H</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>-</td>
<td>248</td>
<td>0.50</td>
</tr>
</tbody>
</table>
### SVM classification model, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Class SVM model</th>
<th>Exp. (nM) hA₂₄R</th>
<th>Kᵢ (nM) hAᵢR</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td></td>
<td>CH₃</td>
<td>H</td>
<td></td>
<td>1</td>
<td>2.80</td>
<td>300</td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>CH₂=CH₂</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>176</td>
<td>0.48</td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>n-C₃H₇</td>
<td>H</td>
<td></td>
<td>1</td>
<td>1.60</td>
<td>600</td>
</tr>
<tr>
<td>48</td>
<td></td>
<td>t-C₄H₉</td>
<td>H</td>
<td></td>
<td>1</td>
<td>45</td>
<td>1149</td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>t-C₄H₉</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>545</td>
<td>0.80</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>t-C₄H₉</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>796</td>
<td>2.78</td>
</tr>
<tr>
<td>51</td>
<td></td>
<td>(CH₃)₂CH=CH₂</td>
<td>H</td>
<td></td>
<td>1</td>
<td>0.78</td>
<td>700</td>
</tr>
<tr>
<td>52</td>
<td></td>
<td>(CH₃)₂C=CH₂=CH₂</td>
<td>H</td>
<td></td>
<td>1</td>
<td>0.80</td>
<td>811</td>
</tr>
<tr>
<td>53</td>
<td></td>
<td>(CH₃)₂C=CH₂=CH₂</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>102.5</td>
<td>40</td>
</tr>
<tr>
<td>54</td>
<td></td>
<td>2,4,5-Br₃-Ph₂CH₂=CH₂</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>2030</td>
<td>25</td>
</tr>
<tr>
<td>55</td>
<td></td>
<td>2,4,5-Br₃-Ph₂CH₂=CH₂</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>3800</td>
<td>71</td>
</tr>
<tr>
<td>56</td>
<td></td>
<td>2-(α-naphthyl)ethyl</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>3360</td>
<td>16</td>
</tr>
<tr>
<td>57</td>
<td></td>
<td>C₂H₅</td>
<td>H</td>
<td></td>
<td>1</td>
<td>1.95</td>
<td>3799</td>
</tr>
<tr>
<td>58</td>
<td></td>
<td>n-C₅H₇</td>
<td>H</td>
<td></td>
<td>1</td>
<td>2.51</td>
<td>613</td>
</tr>
<tr>
<td>59</td>
<td></td>
<td>Ph-CH₂=CH₂</td>
<td>H</td>
<td></td>
<td>1</td>
<td>0.34</td>
<td>2785</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>Ph-CH₂=CH₂=CH₂</td>
<td>H</td>
<td></td>
<td>1</td>
<td>0.16</td>
<td>2666</td>
</tr>
<tr>
<td>61</td>
<td></td>
<td>Ph-CH₂=CH₂</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>120</td>
<td>1.47</td>
</tr>
<tr>
<td>62</td>
<td></td>
<td>Ph-CH₂=CH₂=CH₂</td>
<td>CH₂</td>
<td></td>
<td>-1</td>
<td>1010</td>
<td>19.81</td>
</tr>
<tr>
<td>63</td>
<td></td>
<td>CH₂CH₃</td>
<td>CO-CH₂Ph₂</td>
<td></td>
<td>-1</td>
<td>131</td>
<td>0.98</td>
</tr>
<tr>
<td>64</td>
<td></td>
<td>CH₂CH₂=CH(CH₃)₂</td>
<td>CO-CH₂-2-thienyl</td>
<td>-</td>
<td>1</td>
<td>2.03</td>
<td>308</td>
</tr>
</tbody>
</table>
**SVM classification model, training set: continued**

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>R&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Class SVM&lt;sub&gt;class&lt;/sub&gt; model</th>
<th>Exp. h&lt;sub&gt;A&lt;sub&gt;2&lt;/sub&gt;R&lt;/sub&gt; (nM)</th>
<th>K&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
<th>K&lt;sub&gt;i&lt;/sub&gt; (nM) h&lt;sub&gt;A&lt;sub&gt;3&lt;/sub&gt;R&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td></td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH(Ch&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-3-thienyl</td>
<td>-</td>
<td>1</td>
<td>3.88</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-β-naphthyl</td>
<td>-</td>
<td>1</td>
<td>7.94</td>
<td>174</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-O-Ph</td>
<td>-</td>
<td>1</td>
<td>15.1</td>
<td>692</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH(Ch&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>-</td>
<td>1</td>
<td>1.95</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH(Ch&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-O-Ph-4-Cl</td>
<td>-</td>
<td>1</td>
<td>16.1</td>
<td>2306</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-O-Ph-4-Cl</td>
<td>-</td>
<td>1</td>
<td>31.7</td>
<td>3321</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH=Ph</td>
<td>CO-CH&lt;sub&gt;2&lt;/sub&gt;-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>-</td>
<td>1</td>
<td>8.93</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>2</td>
<td>N(CH&lt;sub&gt;2&lt;/sub&gt;CH=OH)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.12</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>3</td>
<td>NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.22</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>3</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.13</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>2</td>
<td>SO&lt;sub&gt;2&lt;/sub&gt;H</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>10.0</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>3</td>
<td>NO&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.00</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>2</td>
<td>NHCOCH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>4.80</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>3</td>
<td>N(CH&lt;sub&gt;2&lt;/sub&gt;CH&lt;sub&gt;2&lt;/sub&gt;OH)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.10</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>3</td>
<td>CN</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>86.0</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>3</td>
<td>COOEt</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>4.00</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>2</td>
<td>OCH&lt;sub&gt;2&lt;/sub&gt;COOEt</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.43</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>3</td>
<td>C(NOH)NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>6.00</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>2</td>
<td>NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>5.5</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>3</td>
<td>C(NH)NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>4.4</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>3</td>
<td>COOH</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>4.63</td>
<td>&gt;10,000</td>
<td></td>
</tr>
</tbody>
</table>
### SVM classification model, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Class in SVM model</th>
<th>Exp. Kᵢ (nM) hA₂₄R</th>
<th>Exp. Kᵢ (nM) hA₃₄R</th>
</tr>
</thead>
<tbody>
<tr>
<td>86</td>
<td>2</td>
<td>SO₂NH₂</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.31</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>87</td>
<td>2</td>
<td>SO₂N(CH₂CH₂OH)₂</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.80</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>88</td>
<td>2</td>
<td>SO₂N(CH₃)₂</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>3.80</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>89</td>
<td>2</td>
<td>SO₂N(CH₂CH₂Cl)₂</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.59</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>90</td>
<td>2</td>
<td>SO₂NHCH₂COOH</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.80</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>91</td>
<td>3</td>
<td>SO₃H</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.20</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>92</td>
<td>3</td>
<td>H</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.10</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>93</td>
<td>3</td>
<td>H</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.10</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>94</td>
<td>3</td>
<td>OCH₂O</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.59</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>95</td>
<td>3</td>
<td>OCH₃</td>
<td>OCH₃</td>
<td>-</td>
<td>1</td>
<td>2.7</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>96</td>
<td>3</td>
<td>OH</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>1.50</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>97</td>
<td>3</td>
<td>CH₂OH</td>
<td>CH₂OH</td>
<td>H</td>
<td>1</td>
<td>0.19</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>98</td>
<td>3</td>
<td>H</td>
<td>H</td>
<td>SO₃H</td>
<td>1</td>
<td>75</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>99</td>
<td>3</td>
<td>COOEt</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>5.48</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>COOH</td>
<td>COOH</td>
<td>H</td>
<td>1</td>
<td>120</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>101</td>
<td>3</td>
<td>COOH</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>59</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>102</td>
<td>-</td>
<td>CH₃</td>
<td>S(CH₂)₂CH₃</td>
<td>H</td>
<td>1</td>
<td>2.1</td>
<td>224</td>
</tr>
<tr>
<td>103</td>
<td>-</td>
<td>CH₃</td>
<td>SCH₃</td>
<td>COCH₂Ph-4OCH₃</td>
<td>1</td>
<td>15</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>104</td>
<td>-</td>
<td>CH₃</td>
<td>SCH₃</td>
<td>COCH₂Ph-4isobutyl</td>
<td>1</td>
<td>50</td>
<td>&gt;1,000</td>
</tr>
</tbody>
</table>
D.2 SVM regression model, training set

1-71, 105-126, 129-137

127-128

1-42: [1]; 43: [2]; 44-56: [3]; 57-62: [4]; 63-71: [5]; 105-128: [1]; 129-133: [3]; 134-135: [4]; 136-137: [9]
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R1</th>
<th>Pred. pK$<em>i$ (nM) hA$</em>{2A}$R (SVM model)</th>
<th>pK$<em>i$ (nM) hA$</em>{2A}$R (SVM model)</th>
<th>Exp. pK$<em>i$ (nM) hA$</em>{2A}$R (SVM model)</th>
<th>Exp. pK$<em>i$ (nM) hA$</em>{2A}$R (SVM model)</th>
<th>Exp. pK$<em>i$ (nM) hA$</em>{2A}$R (SVM model)</th>
<th>Exp. pK$<em>i$ (nM) hA$</em>{2A}$R (SVM model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1$^a$</td>
<td>CH$_3$</td>
<td>3,4-Cl$_2$-Ph-NH-CO</td>
<td>-2.25</td>
<td>0.62</td>
<td>-2.16</td>
<td>-0.53</td>
<td>143</td>
<td>3.40</td>
</tr>
<tr>
<td>2$^a$</td>
<td>CH$_3$</td>
<td>3,4-OCH$_3$-O-Ph-NH-CO</td>
<td>-2.71</td>
<td>0.60</td>
<td>-2.83</td>
<td>0.62</td>
<td>680</td>
<td>0.24</td>
</tr>
<tr>
<td>3$^a$</td>
<td>CH$_3$</td>
<td>4-N$_2$-Ph-NH-CO</td>
<td>-2.94</td>
<td>0.35</td>
<td>-2.84</td>
<td>0.37</td>
<td>695</td>
<td>0.43</td>
</tr>
<tr>
<td>4$^a$</td>
<td>CH$_3$</td>
<td>4-CH$_3$-Ph-NH-CO</td>
<td>-2.14</td>
<td>1.22</td>
<td>-2.04</td>
<td>0.51</td>
<td>110</td>
<td>0.31</td>
</tr>
<tr>
<td>5$^a$</td>
<td>CH$_3$</td>
<td>4-Br-Ph-NH-CO</td>
<td>-2.04</td>
<td>0.33</td>
<td>-2.00</td>
<td>0.34</td>
<td>100</td>
<td>0.46</td>
</tr>
<tr>
<td>6$^a$</td>
<td>CH$_3$</td>
<td>4-F-Ph-NH-CO</td>
<td>-2.31</td>
<td>0.16</td>
<td>-2.08</td>
<td>0.47</td>
<td>120</td>
<td>0.34</td>
</tr>
<tr>
<td>7$^a$</td>
<td>CH$_3$</td>
<td>4-CF$_3$-Ph-NH-CO</td>
<td>-2.14</td>
<td>0.04</td>
<td>-2.15</td>
<td>0.13</td>
<td>140</td>
<td>0.74</td>
</tr>
<tr>
<td>8$^a$</td>
<td>CH$_3$</td>
<td>2-OCH$_3$-Ph-NH-CO</td>
<td>-2.84</td>
<td>0.14</td>
<td>-2.36</td>
<td>0.15</td>
<td>180</td>
<td>0.70</td>
</tr>
<tr>
<td>9$^a$</td>
<td>CH$_3$</td>
<td>3-OCH$_3$-Ph-NH-CO</td>
<td>-2.60</td>
<td>0.52</td>
<td>-2.20</td>
<td>0.10</td>
<td>160</td>
<td>0.80</td>
</tr>
<tr>
<td>10$^a$</td>
<td>CH$_3$</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-2.26</td>
<td>-0.15</td>
<td>-2.30</td>
<td>0.04</td>
<td>200</td>
<td>0.91</td>
</tr>
<tr>
<td>11$^a$</td>
<td>CH$_3$</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-2.02</td>
<td>0.24</td>
<td>-2.26</td>
<td>0.54</td>
<td>180</td>
<td>0.29</td>
</tr>
<tr>
<td>12$^a$</td>
<td>C$_2$H$_5$</td>
<td>3,4-Cl$_2$-Ph-NH-CO</td>
<td>-2.34</td>
<td>-0.10</td>
<td>-2.55</td>
<td>-0.48</td>
<td>352</td>
<td>3.00</td>
</tr>
<tr>
<td>13$^a$</td>
<td>C$_2$H$_5$</td>
<td>3,4-OCH$_3$-O-Ph-NH-CO</td>
<td>-2.45</td>
<td>0.74</td>
<td>-2.76</td>
<td>0.57</td>
<td>576</td>
<td>0.27</td>
</tr>
<tr>
<td>14$^a$</td>
<td>C$_2$H$_5$</td>
<td>4-N$_2$-Ph-NH-CO</td>
<td>-2.63</td>
<td>-0.34</td>
<td>-2.79</td>
<td>0.19</td>
<td>614</td>
<td>0.65</td>
</tr>
<tr>
<td>15$^a$</td>
<td>C$_2$H$_5$</td>
<td>2-OCH$_3$-Ph-NH-CO</td>
<td>-2.30</td>
<td>0.81</td>
<td>-2.12</td>
<td>0.25</td>
<td>133</td>
<td>0.56</td>
</tr>
<tr>
<td>16$^a$</td>
<td>C$_2$H$_5$</td>
<td>3-CH$_3$-Ph-NH-CO</td>
<td>-2.28</td>
<td>0.59</td>
<td>-2.15</td>
<td>0.07</td>
<td>140</td>
<td>0.86</td>
</tr>
<tr>
<td>17$^a$</td>
<td>C$_2$H$_5$</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-2.10</td>
<td>0.23</td>
<td>-2.18</td>
<td>0.52</td>
<td>150</td>
<td>0.30</td>
</tr>
<tr>
<td>18$^a$</td>
<td>C$_2$H$_5$</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-1.88</td>
<td>0.36</td>
<td>-2.20</td>
<td>0.70</td>
<td>160</td>
<td>0.20</td>
</tr>
<tr>
<td>19$^a$</td>
<td>n-C$_3$H$_7$</td>
<td>3,4-Cl$_2$-Ph-NH-CO</td>
<td>-2.65</td>
<td>-0.28</td>
<td>-2.60</td>
<td>-0.40</td>
<td>401</td>
<td>2.50</td>
</tr>
<tr>
<td>20$^a$</td>
<td>n-C$_3$H$_7$</td>
<td>3,4-OCH$_3$-O-Ph-NH-CO</td>
<td>-2.61</td>
<td>0.51</td>
<td>-2.82</td>
<td>0.52</td>
<td>667</td>
<td>0.30</td>
</tr>
</tbody>
</table>

$^a$ Molecules already introduced in the previous SVM classification model
### SVM regression model training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R&lt;sub&gt;i&lt;/sub&gt;</th>
<th>Pred. p&lt;sub&gt;Ki&lt;/sub&gt; (nM) hA&lt;sub&gt;2,3&lt;/sub&gt;R</th>
<th>Pred. p&lt;sub&gt;Ki&lt;/sub&gt; (nM) hA&lt;sub&gt;2,4&lt;/sub&gt;R</th>
<th>Exp. p&lt;sub&gt;Ki&lt;/sub&gt; (nM) hA&lt;sub&gt;2,3&lt;/sub&gt;R</th>
<th>Exp. p&lt;sub&gt;Ki&lt;/sub&gt; (nM) hA&lt;sub&gt;2,4&lt;/sub&gt;R</th>
<th>Exp. p&lt;sub&gt;Ki&lt;/sub&gt; (nM) hA&lt;sub&gt;2,3&lt;/sub&gt;R</th>
<th>Exp. p&lt;sub&gt;Ki&lt;/sub&gt; (nM) hA&lt;sub&gt;2,4&lt;/sub&gt;R</th>
</tr>
</thead>
<tbody>
<tr>
<td>21&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-NO&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO</td>
<td>-2.80</td>
<td>-0.15</td>
<td>-3.05</td>
<td>0.09</td>
<td>1115</td>
<td>0.81</td>
</tr>
<tr>
<td>22&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-1.77</td>
<td>-0.28</td>
<td>-2.00</td>
<td>0.47</td>
<td>100</td>
<td>0.34</td>
</tr>
<tr>
<td>23&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>3-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-2.10</td>
<td>-0.14</td>
<td>-2.05</td>
<td>0.40</td>
<td>113</td>
<td>0.40</td>
</tr>
<tr>
<td>24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-2.93</td>
<td>-0.23</td>
<td>-2.70</td>
<td>0.26</td>
<td>503</td>
<td>0.55</td>
</tr>
<tr>
<td>25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-1.85</td>
<td>-0.71</td>
<td>-2.00</td>
<td>0.07</td>
<td>100</td>
<td>0.86</td>
</tr>
<tr>
<td>26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>3,4-Cl&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO</td>
<td>-2.13</td>
<td>0.39</td>
<td>-2.05</td>
<td>0.37</td>
<td>111</td>
<td>0.43</td>
</tr>
<tr>
<td>27&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>3,4-OCH&lt;sub&gt;2&lt;/sub&gt;O-Ph-NH-CO</td>
<td>-1.43</td>
<td>0.25</td>
<td>-2.63</td>
<td>0.09</td>
<td>423</td>
<td>0.81</td>
</tr>
<tr>
<td>28&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>4-NO&lt;sub&gt;2&lt;/sub&gt;Ph-NH-CO</td>
<td>-1.28</td>
<td>0.62</td>
<td>-2.53</td>
<td>-0.01</td>
<td>335</td>
<td>1.03</td>
</tr>
<tr>
<td>29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-1.34</td>
<td>-0.86</td>
<td>-2.49</td>
<td>0.00</td>
<td>306</td>
<td>1.01</td>
</tr>
<tr>
<td>30&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-1.28</td>
<td>-1.75</td>
<td>-2.60</td>
<td>-0.05</td>
<td>400</td>
<td>1.11</td>
</tr>
<tr>
<td>31&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-1.78</td>
<td>0.13</td>
<td>-2.58</td>
<td>0.80</td>
<td>381</td>
<td>0.16</td>
</tr>
<tr>
<td>32&lt;sup&gt;a&lt;/sup&gt;</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-2.93</td>
<td>-0.23</td>
<td>-2.70</td>
<td>0.26</td>
<td>503</td>
<td>0.55</td>
</tr>
<tr>
<td>33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-1.85</td>
<td>-0.71</td>
<td>-2.00</td>
<td>0.07</td>
<td>100</td>
<td>0.86</td>
</tr>
<tr>
<td>34&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-2.13</td>
<td>0.39</td>
<td>-2.05</td>
<td>0.37</td>
<td>111</td>
<td>0.43</td>
</tr>
<tr>
<td>35&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Ph-NH-CO</td>
<td>-1.43</td>
<td>0.25</td>
<td>-2.63</td>
<td>0.09</td>
<td>423</td>
<td>0.81</td>
</tr>
<tr>
<td>36&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n-C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;9&lt;/sub&gt;</td>
<td>Ph-NH-CO</td>
<td>-1.75</td>
<td>0.37</td>
<td>-2.17</td>
<td>0.68</td>
<td>148</td>
<td>0.21</td>
</tr>
<tr>
<td>37&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-2.30</td>
<td>0.28</td>
<td>-3.14</td>
<td>0.70</td>
<td>1390</td>
<td>0.20</td>
</tr>
<tr>
<td>38&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-2.39</td>
<td>0.20</td>
<td>-3.08</td>
<td>0.40</td>
<td>1195</td>
<td>0.40</td>
</tr>
</tbody>
</table>

<sup>a</sup> Molecules already introduced in the previous SVM classification model
SVM regression model, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R_1</th>
<th>Pred. $pK_i$ (nM) h_{A2,R}</th>
<th>Pred. $pK_i$ (nM) h_{A3,R}</th>
<th>Exp. $pK_i$ (nM) h_{A2,R}</th>
<th>Exp. $pK_i$ (nM) h_{A3,R}</th>
<th>Exp. $K_i$ (nM) h_{A2,R}</th>
<th>Exp. $K_i$ (nM) h_{A3,R}</th>
</tr>
</thead>
<tbody>
<tr>
<td>39^a</td>
<td>C_2H_5</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-2.26</td>
<td>0.07</td>
<td>-3.02</td>
<td>0.22</td>
<td>1040</td>
<td>0.60</td>
</tr>
<tr>
<td>40^a</td>
<td>C_2H_5</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-2.06</td>
<td>-0.15</td>
<td>-2.26</td>
<td>-0.20</td>
<td>180</td>
<td>1.60</td>
</tr>
<tr>
<td>41^a</td>
<td>n-C_3H_7</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-2.06</td>
<td>-0.07</td>
<td>-2.15</td>
<td>0.10</td>
<td>140</td>
<td>0.80</td>
</tr>
<tr>
<td>42^a</td>
<td>n-C_3H_7</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-2.08</td>
<td>-0.70</td>
<td>-3.09</td>
<td>0.04</td>
<td>1220</td>
<td>0.91</td>
</tr>
<tr>
<td>43^a</td>
<td>H</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-2.42</td>
<td>0.09</td>
<td>-2.72</td>
<td>0.85</td>
<td>520</td>
<td>0.14</td>
</tr>
<tr>
<td>44^a</td>
<td>H</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-2.60</td>
<td>0.81</td>
<td>-2.39</td>
<td>0.30</td>
<td>248</td>
<td>0.50</td>
</tr>
<tr>
<td>45^a</td>
<td>CH_3</td>
<td>H</td>
<td>-0.50</td>
<td>-2.55</td>
<td>-0.45</td>
<td>-2.48</td>
<td>2.80</td>
<td>300</td>
</tr>
<tr>
<td>46^a</td>
<td>CH_2=CH=CH_2</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-2.36</td>
<td>-0.12</td>
<td>-2.25</td>
<td>-3.31</td>
<td>176</td>
<td>0.48</td>
</tr>
<tr>
<td>47^a</td>
<td>n-C_4H_9</td>
<td>H</td>
<td>-0.21</td>
<td>-3.41</td>
<td>-0.20</td>
<td>-2.78</td>
<td>1.60</td>
<td>600</td>
</tr>
<tr>
<td>48^a</td>
<td>t-C_4H_9</td>
<td>H</td>
<td>-1.47</td>
<td>-1.97</td>
<td>-1.65</td>
<td>-3.06</td>
<td>45</td>
<td>1149</td>
</tr>
<tr>
<td>49^a</td>
<td>t-C_4H_9</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-2.96</td>
<td>0.54</td>
<td>-2.74</td>
<td>0.10</td>
<td>545</td>
<td>0.80</td>
</tr>
<tr>
<td>50^a</td>
<td>t-C_4H_9</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-2.91</td>
<td>0.16</td>
<td>-2.90</td>
<td>-0.44</td>
<td>796</td>
<td>2.78</td>
</tr>
<tr>
<td>51^a</td>
<td>(CH_3)_2CH=CH_2=CH_2</td>
<td>H</td>
<td>0.08</td>
<td>-2.49</td>
<td>0.11</td>
<td>-2.85</td>
<td>0.78</td>
<td>700</td>
</tr>
<tr>
<td>52^a</td>
<td>(CH_3)_2C=CH_2=CH_2</td>
<td>H</td>
<td>-0.32</td>
<td>-2.57</td>
<td>0.10</td>
<td>-2.91</td>
<td>0.80</td>
<td>811</td>
</tr>
<tr>
<td>53^a</td>
<td>(CH_3)_2C=CH_2=CH_2</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-1.89</td>
<td>-0.08</td>
<td>-2.01</td>
<td>-1.60</td>
<td>1025</td>
<td>40</td>
</tr>
<tr>
<td>54^a</td>
<td>2,4,5-Br_3-Ph_2-CH_2=CH_2</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-2.68</td>
<td>-1.56</td>
<td>-3.31</td>
<td>-1.40</td>
<td>2030</td>
<td>25</td>
</tr>
<tr>
<td>55^a</td>
<td>2,4,5-Br_3-Ph_2-CH_2=CH_2</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-2.62</td>
<td>-0.27</td>
<td>-3.56</td>
<td>-1.85</td>
<td>3600</td>
<td>71</td>
</tr>
<tr>
<td>56^a</td>
<td>2-(α-naphthyl)ethyl</td>
<td>4-OC_6H_5-Ph-NH-CO</td>
<td>-1.83</td>
<td>-1.10</td>
<td>-3.51</td>
<td>-1.20</td>
<td>3200</td>
<td>16</td>
</tr>
</tbody>
</table>

^a Molecules already introduced in the previous SVM classification model
## SVM regression model training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R$_j$</th>
<th>Pred. $pK_i$ (nM) SVM hA$_2$R</th>
<th>Pred. $pK_i$ (nM) hA$_2$R</th>
<th>Exp. $pK_i$ (nM) hA$_3$R</th>
<th>Exp. $pK_i$ (nM) hA$_3$R</th>
</tr>
</thead>
<tbody>
<tr>
<td>57$^a$</td>
<td>C$_2$H$_5$</td>
<td>H</td>
<td>-0.33</td>
<td>-2.97</td>
<td>-0.29</td>
<td>-3.58</td>
</tr>
<tr>
<td>58$^a$</td>
<td>$a$-C$_3$H$_7$</td>
<td>H</td>
<td>-0.31</td>
<td>-2.88</td>
<td>-0.40</td>
<td>-2.79</td>
</tr>
<tr>
<td>59$^a$</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>H</td>
<td>0.51</td>
<td>-3.03</td>
<td>0.47</td>
<td>-3.44</td>
</tr>
<tr>
<td>60$^a$</td>
<td>Ph-CH$_2$-CH$_2$-CH$_2$</td>
<td>H</td>
<td>0.01</td>
<td>-3.10</td>
<td>0.80</td>
<td>-3.43</td>
</tr>
<tr>
<td>61$^a$</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>4-OC$_6$H$_4$-Ph-NH-CO</td>
<td>-2.06</td>
<td>-0.59</td>
<td>-2.08</td>
<td>-0.17</td>
</tr>
<tr>
<td>62$^a$</td>
<td>Ph-CH$_2$-CH$_2$-CH$_2$</td>
<td>4-OC$_6$H$_4$-Ph-NH-CO</td>
<td>-1.89</td>
<td>-0.91</td>
<td>-3.00</td>
<td>-1.30</td>
</tr>
<tr>
<td>63$^a$</td>
<td>CH$_2$-CH$_3$</td>
<td>CO-CHPh$_2$</td>
<td>-1.21</td>
<td>-0.54</td>
<td>-2.12</td>
<td>0.01</td>
</tr>
<tr>
<td>64$^a$</td>
<td>CH$_2$-CH$_2$-CH$_2$</td>
<td>CO-CH$_2$-2-thienyl</td>
<td>-1.30</td>
<td>-0.78</td>
<td>-0.31</td>
<td>-2.49</td>
</tr>
<tr>
<td>65$^a$</td>
<td>CH$_2$-CH$_2$-CH$_2$</td>
<td>CO-CH$_2$-3-thienyl</td>
<td>-1.62</td>
<td>-1.51</td>
<td>-0.50</td>
<td>-2.73</td>
</tr>
<tr>
<td>66$^a$</td>
<td>CH$_2$-CH$_2$-CH$_2$</td>
<td>CO-CH$_2$-3-naphthyl</td>
<td>-1.47</td>
<td>-1.90</td>
<td>-0.90</td>
<td>-2.24</td>
</tr>
<tr>
<td>67$^a$</td>
<td>CH$_3$</td>
<td>CO-CH$_2$-O-Ph</td>
<td>-1.88</td>
<td>-0.86</td>
<td>-1.18</td>
<td>-2.84</td>
</tr>
<tr>
<td>68$^a$</td>
<td>CH$_2$-CH$_2$-CH$_2$</td>
<td>CO-CH$_2$-4-OC$_6$H$_4$-Ph</td>
<td>-1.49</td>
<td>-1.75</td>
<td>-0.29</td>
<td>-2.56</td>
</tr>
<tr>
<td>69$^a$</td>
<td>CH$_2$-CH$_2$-CH$_2$</td>
<td>CO-CH$_2$-O-Ph-4-Cl</td>
<td>-1.23</td>
<td>-2.37</td>
<td>-1.21</td>
<td>-3.36</td>
</tr>
<tr>
<td>70$^a$</td>
<td>CH$_2$-CH$_2$-Ph</td>
<td>CO-CH$_2$-O-Ph-4-Cl</td>
<td>-1.24</td>
<td>-2.61</td>
<td>-1.50</td>
<td>-3.51</td>
</tr>
<tr>
<td>71$^a$</td>
<td>CH$_2$-CH$_2$-CH$_2$-Ph</td>
<td>CO-CH$_2$-4-OC$_6$H$_4$-Ph</td>
<td>-1.38</td>
<td>-3.80</td>
<td>-0.95</td>
<td>-2.08</td>
</tr>
</tbody>
</table>

$^a$ Molecules already introduced in the previous SVM classification model
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R$_i$</th>
<th>Pred. $pK_i$ (nM) hA$_{24}$R</th>
<th>Pred. $pK_i$ (nM) hA$_{24}$R</th>
<th>Exp. $pK_i$ (nM) hA$_{24}$R</th>
<th>Exp. $K_i$ (nM) hA$_{24}$R</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>C$_2$H$_5$</td>
<td>4-CH$_3$-Ph-NH-CO</td>
<td>-1.87</td>
<td>0.86</td>
<td>-1.48</td>
<td>0.85</td>
</tr>
<tr>
<td>106</td>
<td>C$_2$H$_5$</td>
<td>4-Br-Ph-NH-CO</td>
<td>-1.99</td>
<td>0.80</td>
<td>-1.60</td>
<td>0.43</td>
</tr>
<tr>
<td>107</td>
<td>C$_2$H$_5$</td>
<td>4-F-Ph-NH-CO</td>
<td>-2.17</td>
<td>0.18</td>
<td>-1.78</td>
<td>0.07</td>
</tr>
<tr>
<td>108</td>
<td>C$_2$H$_5$</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-2.10</td>
<td>0.33</td>
<td>-1.72</td>
<td>0.01</td>
</tr>
<tr>
<td>109</td>
<td>n-C$_3$H$_7$</td>
<td>4-CH$_3$-Ph-NH-CO</td>
<td>-1.70</td>
<td>0.61</td>
<td>-1.08</td>
<td>0.40</td>
</tr>
<tr>
<td>110</td>
<td>n-C$_3$H$_7$</td>
<td>4-Br-Ph-NH-CO</td>
<td>-1.96</td>
<td>0.60</td>
<td>-1.70</td>
<td>0.35</td>
</tr>
<tr>
<td>111</td>
<td>n-C$_3$H$_7$</td>
<td>4-F-Ph-NH-CO</td>
<td>-2.32</td>
<td>0.36</td>
<td>-1.62</td>
<td>0.54</td>
</tr>
<tr>
<td>112</td>
<td>n-C$_3$H$_7$</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-2.14</td>
<td>0.01</td>
<td>-1.53</td>
<td>0.29</td>
</tr>
<tr>
<td>113</td>
<td>n-C$_4$H$_9$</td>
<td>4-CH$_3$-Ph-NH-CO</td>
<td>-1.81</td>
<td>0.35</td>
<td>-1.38</td>
<td>0.68</td>
</tr>
<tr>
<td>114</td>
<td>n-C$_4$H$_9$</td>
<td>4-Br-Ph-NH-CO</td>
<td>-2.10</td>
<td>0.27</td>
<td>-1.82</td>
<td>0.04</td>
</tr>
<tr>
<td>115</td>
<td>n-C$_4$H$_9$</td>
<td>4-F-Ph-NH-CO</td>
<td>-2.38</td>
<td>0.58</td>
<td>-1.70</td>
<td>0.10</td>
</tr>
<tr>
<td>116</td>
<td>n-C$_4$H$_9$</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-2.26</td>
<td>0.40</td>
<td>-1.49</td>
<td>0.14</td>
</tr>
<tr>
<td>117</td>
<td>n-C$_4$H$_9$</td>
<td>2-OCH$_3$-Ph-NH-CO</td>
<td>-1.49</td>
<td>0.38</td>
<td>-1.96</td>
<td>0.24</td>
</tr>
<tr>
<td>118</td>
<td>n-C$_4$H$_9$</td>
<td>3-OCH$_3$-Ph-NH-CO</td>
<td>-1.96</td>
<td>0.10</td>
<td>-1.98</td>
<td>0.22</td>
</tr>
<tr>
<td>119</td>
<td>C$_2$H$_5$</td>
<td>Ph-NH-CO</td>
<td>-1.95</td>
<td>0.42</td>
<td>-1.60</td>
<td>0.74</td>
</tr>
<tr>
<td>120</td>
<td>n-C$_3$H$_7$</td>
<td>Ph-NH-CO</td>
<td>-1.71</td>
<td>0.28</td>
<td>-1.79</td>
<td>0.82</td>
</tr>
<tr>
<td>121</td>
<td>CH$_3$</td>
<td>4-SO$_2$-Ph-NH-CO</td>
<td>-2.52</td>
<td>-1.50</td>
<td>-2.77</td>
<td>-1.40</td>
</tr>
<tr>
<td>122</td>
<td>C$_2$H$_5$</td>
<td>4-SO$_2$-Ph-NH-CO</td>
<td>-2.96</td>
<td>-1.14</td>
<td>-2.40</td>
<td>-1.60</td>
</tr>
<tr>
<td>123</td>
<td>n-C$_3$H$_7$</td>
<td>4-SO$_2$-Ph-NH-CO</td>
<td>-2.49</td>
<td>-1.57</td>
<td>-2.48</td>
<td>-1.48</td>
</tr>
</tbody>
</table>
**SVM regression model, training set: continued**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>n-C₄H₉</td>
<td>4-SO₃Ph-NH-CO</td>
<td>-2.52</td>
<td>-1.30</td>
<td>-2.55</td>
<td>-1.67</td>
<td>352</td>
<td>47</td>
</tr>
<tr>
<td>125</td>
<td>n-C₄H₉</td>
<td>4-OCH₃Ph-NH-CO</td>
<td>-2.17</td>
<td>-0.14</td>
<td>-1.90</td>
<td>0.49</td>
<td>80</td>
<td>0.32</td>
</tr>
<tr>
<td>126</td>
<td>n-C₄H₉</td>
<td>3-ClPh-NH-CO</td>
<td>-2.38</td>
<td>-0.35</td>
<td>-1.98</td>
<td>0.22</td>
<td>95</td>
<td>0.60</td>
</tr>
<tr>
<td>127</td>
<td>-</td>
<td>4-OCH₃Ph-NH-CO</td>
<td>-0.60</td>
<td>-1.59</td>
<td>-0.97</td>
<td>0.85</td>
<td>9.40</td>
<td>0.14</td>
</tr>
<tr>
<td>128</td>
<td>-</td>
<td>3-ClPh-NH-CO</td>
<td>-2.19</td>
<td>-1.09</td>
<td>-0.98</td>
<td>0.72</td>
<td>9.50</td>
<td>0.19</td>
</tr>
<tr>
<td>129</td>
<td>H</td>
<td>H</td>
<td>-1.08</td>
<td>-3.02</td>
<td>-1.30</td>
<td>-2.54</td>
<td>20</td>
<td>348</td>
</tr>
<tr>
<td>130</td>
<td>(CH₃)₂CH-C₄H₂</td>
<td>4-OCH₃Ph-NH-CO</td>
<td>-2.07</td>
<td>-0.07</td>
<td>-1.81</td>
<td>-1.48</td>
<td>65</td>
<td>30</td>
</tr>
<tr>
<td>131</td>
<td>(CH₃)₂CH-C₄H₂</td>
<td>3-ClPh-NH-CO</td>
<td>-2.32</td>
<td>-0.28</td>
<td>-2.06</td>
<td>-1.10</td>
<td>115</td>
<td>40</td>
</tr>
<tr>
<td>132</td>
<td>(CH₃)₂C=CH₂-C₄H₂</td>
<td>3-ClPh-NH-CO</td>
<td>-2.26</td>
<td>-0.36</td>
<td>-2.04</td>
<td>-1.74</td>
<td>110</td>
<td>55</td>
</tr>
<tr>
<td>133</td>
<td>2,4,5-Br₃Ph-C₄H₂</td>
<td>H</td>
<td>-1.97</td>
<td>-3.52</td>
<td>-2.54</td>
<td>-3.65</td>
<td>348</td>
<td>4481</td>
</tr>
<tr>
<td>134</td>
<td>Ph-C₄H₂-Ch₂</td>
<td>3-ClPh-NH-CO</td>
<td>-2.08</td>
<td>-0.39</td>
<td>-2.02</td>
<td>-1.12</td>
<td>105</td>
<td>13.28</td>
</tr>
<tr>
<td>135</td>
<td>Ph-C₄H₂-Ch₂-Ch₂</td>
<td>3-ClPh-NH-CO</td>
<td>-1.85</td>
<td>-1.23</td>
<td>-2.30</td>
<td>-1.63</td>
<td>200</td>
<td>42.65</td>
</tr>
<tr>
<td>136</td>
<td>Ph-C₄H₂-Ch₂</td>
<td>(CH₃)₂COCONH(CH₂)₃-CO</td>
<td>-2.43</td>
<td>-3.77</td>
<td>-2.46</td>
<td>-3.17</td>
<td>288</td>
<td>1480</td>
</tr>
<tr>
<td>137</td>
<td>(CH₃)₂CH-C₄H₂</td>
<td>(CH₃)₂COCONH(CH₂)₃-CO</td>
<td>-2.45</td>
<td>-2.80</td>
<td>-2.47</td>
<td>-3.40</td>
<td>257</td>
<td>2488</td>
</tr>
</tbody>
</table>
D.3 Test set

138-188: [5]
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>Exp. class by SVM model</th>
<th>Pred. class by SVM model</th>
<th>Pred. pK&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
<th>Pred. pK&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
<th>Exp. pK&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
<th>Exp. pK&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
<th>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
<th>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>138</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-α-naphthyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>1</td>
<td>1</td>
<td>-1.63</td>
<td>2.76</td>
<td>-2.48</td>
<td>-2.54</td>
<td>305</td>
<td>343</td>
</tr>
<tr>
<td>139</td>
<td>CHPh&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-1.08</td>
<td>0.74</td>
<td>-2.33</td>
<td>0.60</td>
<td>216</td>
<td>0.25</td>
</tr>
<tr>
<td>140</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>1</td>
<td>-1.75</td>
<td>-1.46</td>
<td>-2.29</td>
<td>-1.05</td>
<td>193</td>
<td>11.2</td>
</tr>
<tr>
<td>141</td>
<td>CHPh&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;/sub&gt;&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-1.20</td>
<td>0.02</td>
<td>-2.20</td>
<td>-0.77</td>
<td>159</td>
<td>5.86</td>
</tr>
<tr>
<td>142</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Ph-Ph</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;/sub&gt;&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1</td>
<td>1</td>
<td>-1.05</td>
<td>-1.60</td>
<td>-0.96</td>
<td>-2.43</td>
<td>9.18</td>
<td>206</td>
</tr>
<tr>
<td>143</td>
<td>CHPh&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>-1</td>
<td>-1</td>
<td>-1.57</td>
<td>-0.75</td>
<td>-1.73</td>
<td>-0.81</td>
<td>53.1</td>
<td>6.49</td>
</tr>
<tr>
<td>144</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>1</td>
<td>1</td>
<td>-1.59</td>
<td>-2.45</td>
<td>-1.67</td>
<td>-2.10</td>
<td>46.9</td>
<td>125</td>
</tr>
<tr>
<td>145</td>
<td>CHPh&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-1.19</td>
<td>0.16</td>
<td>-2.06</td>
<td>-0.08</td>
<td>114</td>
<td>1.20</td>
</tr>
<tr>
<td>146</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1</td>
<td>1</td>
<td>-1.59</td>
<td>-1.47</td>
<td>-1.60</td>
<td>-2.28</td>
<td>30.6</td>
<td>189</td>
</tr>
<tr>
<td>147</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Cl</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-1.79</td>
<td>-0.31</td>
<td>-1.48</td>
<td>-0.29</td>
<td>30.5</td>
<td>1.94</td>
</tr>
<tr>
<td>148</td>
<td>CHPh&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-1.23</td>
<td>-0.04</td>
<td>-1.67</td>
<td>0.03</td>
<td>46.5</td>
<td>0.93</td>
</tr>
<tr>
<td>149</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1</td>
<td>1</td>
<td>-1.49</td>
<td>-1.37</td>
<td>-1.72</td>
<td>-1.82</td>
<td>52.5</td>
<td>65.4</td>
</tr>
<tr>
<td>150</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Cl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-1.36</td>
<td>-3.43</td>
<td>-2.19</td>
<td>-1.10</td>
<td>156</td>
<td>12.7</td>
</tr>
<tr>
<td>151</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>1</td>
<td>-1.84</td>
<td>-1.81</td>
<td>-1.85</td>
<td>-1.15</td>
<td>70.9</td>
<td>14.2</td>
</tr>
<tr>
<td>152</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;-Ph</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-2.06</td>
<td>-0.45</td>
<td>-1.80</td>
<td>0.02</td>
<td>62.5</td>
<td>0.95</td>
</tr>
<tr>
<td>153</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-β-naphthyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;/sub&gt;&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1</td>
<td>1</td>
<td>-1.59</td>
<td>-1.52</td>
<td>-1.20</td>
<td>-2.61</td>
<td>15.7</td>
<td>409</td>
</tr>
<tr>
<td>154</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-β-naphthyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>1</td>
<td>1</td>
<td>-1.61</td>
<td>-2.18</td>
<td>-1.62</td>
<td>-2.26</td>
<td>42.1</td>
<td>180</td>
</tr>
<tr>
<td>155</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-thienyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>1</td>
<td>1</td>
<td>-1.69</td>
<td>-1.45</td>
<td>-0.94</td>
<td>-2.52</td>
<td>8.8</td>
<td>330</td>
</tr>
<tr>
<td>156</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-thienyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-Ph</td>
<td>1</td>
<td>1</td>
<td>-1.73</td>
<td>-1.90</td>
<td>-1.11</td>
<td>-2.86</td>
<td>12.9</td>
<td>726</td>
</tr>
<tr>
<td>157</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-α-naphthyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-1</td>
<td>-1</td>
<td>-2.21</td>
<td>-0.43</td>
<td>-1.87</td>
<td>-0.48</td>
<td>74.6</td>
<td>3.05</td>
</tr>
<tr>
<td>158</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-β-naphthyl</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1</td>
<td>1</td>
<td>-1.42</td>
<td>-2.40</td>
<td>-1.26</td>
<td>-1.56</td>
<td>18.4</td>
<td>36.3</td>
</tr>
<tr>
<td>Mol Id</td>
<td>R</td>
<td>R₁</td>
<td>Exp. class in SVM model</td>
<td>Pred. class by SVM model</td>
<td>pKᵢ  (nM) hA₂₄ R</td>
<td>hA₂₄ R</td>
<td>hA₃ R</td>
<td>hA₃ R</td>
<td>Exp. Kᵢ  (nM)</td>
<td>Exp. Kᵢ  (nM)</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>-----</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>159</td>
<td>CH₂-3-thienyl</td>
<td>CH₃</td>
<td>1</td>
<td>-1</td>
<td>-2.01</td>
<td>0.11</td>
<td>-1.36</td>
<td>-2.88</td>
<td>23</td>
<td>765</td>
</tr>
<tr>
<td>160</td>
<td>CH₂-2-thienyl</td>
<td>CH₃</td>
<td>1</td>
<td>-1</td>
<td>-1.80</td>
<td>0.41</td>
<td>-1.20</td>
<td>-2.20</td>
<td>15.9</td>
<td>196</td>
</tr>
<tr>
<td>161</td>
<td>CH₂-2-thienyl</td>
<td>CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-1.79</td>
<td>0.24</td>
<td>-1.75</td>
<td>-0.72</td>
<td>56</td>
<td>5.26</td>
</tr>
<tr>
<td>162</td>
<td>CH₂-3-thienyl</td>
<td>CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-1.82</td>
<td>-0.13</td>
<td>-1.50</td>
<td>-0.10</td>
<td>31.3</td>
<td>1.25</td>
</tr>
<tr>
<td>163</td>
<td>CH₂-β-naphthyl</td>
<td>CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-1.91</td>
<td>-0.55</td>
<td>-1.89</td>
<td>-1.16</td>
<td>77.5</td>
<td>14.5</td>
</tr>
<tr>
<td>164</td>
<td>CH₂-α-naphthyl</td>
<td>CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-2.21</td>
<td>-0.13</td>
<td>-1.91</td>
<td>-0.54</td>
<td>80.5</td>
<td>3.47</td>
</tr>
<tr>
<td>165</td>
<td>CH₂-α-naphthyl</td>
<td>CH₂-CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-1.70</td>
<td>-1.32</td>
<td>-1.59</td>
<td>-1.24</td>
<td>38.7</td>
<td>17.3</td>
</tr>
<tr>
<td>166</td>
<td>CH₂-β-naphthyl</td>
<td>CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.83</td>
<td>-1.07</td>
<td>-0.90</td>
<td>-1.98</td>
<td>7.99</td>
<td>95.9</td>
</tr>
<tr>
<td>167</td>
<td>CH₂-α-naphthyl</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.81</td>
<td>-0.96</td>
<td>-1.07</td>
<td>-2.00</td>
<td>11.7</td>
<td>100</td>
</tr>
<tr>
<td>168</td>
<td>CH₂-LCF₃-Ph</td>
<td>CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-1.68</td>
<td>-1.81</td>
<td>-1.88</td>
<td>-0.09</td>
<td>78.9</td>
<td>1.22</td>
</tr>
<tr>
<td>169</td>
<td>CH₂-2-thienyl</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.58</td>
<td>-0.71</td>
<td>-0.62</td>
<td>-2.00</td>
<td>4.15</td>
<td>99.8</td>
</tr>
<tr>
<td>170</td>
<td>CH₂-3-thienyl</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.73</td>
<td>-0.77</td>
<td>-0.50</td>
<td>-2.28</td>
<td>3.13</td>
<td>189</td>
</tr>
<tr>
<td>171</td>
<td>CH₂-O-Ph-LCl</td>
<td>CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.65</td>
<td>-3.50</td>
<td>-1.59</td>
<td>-2.35</td>
<td>30.3</td>
<td>223</td>
</tr>
<tr>
<td>172</td>
<td>CH₂-3-CI-Ph</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.58</td>
<td>-1.43</td>
<td>-0.27</td>
<td>-2.44</td>
<td>1.86</td>
<td>273</td>
</tr>
<tr>
<td>173</td>
<td>CH₂-4-CI-Ph</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.58</td>
<td>-1.60</td>
<td>-0.44</td>
<td>-1.75</td>
<td>2.75</td>
<td>56.5</td>
</tr>
<tr>
<td>174</td>
<td>CH₂-2-CI-Ph</td>
<td>CH₂-CH₂-Ph</td>
<td>1</td>
<td>1</td>
<td>-1.71</td>
<td>-1.92</td>
<td>-0.76</td>
<td>-2.44</td>
<td>5.75</td>
<td>273</td>
</tr>
<tr>
<td>175</td>
<td>CH₂-4-CF₃-Ph</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-2.14</td>
<td>-2.27</td>
<td>-0.73</td>
<td>-2.42</td>
<td>5.43</td>
<td>206</td>
</tr>
<tr>
<td>176</td>
<td>CH₂-4-F-Ph</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.83</td>
<td>-1.63</td>
<td>-0.57</td>
<td>-2.06</td>
<td>3.09</td>
<td>116</td>
</tr>
<tr>
<td>177</td>
<td>CH₂-4-F-Ph</td>
<td>CH₃</td>
<td>-1</td>
<td>-1</td>
<td>-2.06</td>
<td>-0.64</td>
<td>-1.73</td>
<td>0.01</td>
<td>54.1</td>
<td>0.97</td>
</tr>
<tr>
<td>178</td>
<td>CH₂-2,6-Cl₂-Ph</td>
<td>CH₂-CH₂-CH₃</td>
<td>1</td>
<td>1</td>
<td>-1.12</td>
<td>-1.03</td>
<td>-1.27</td>
<td>-2.32</td>
<td>18.7</td>
<td>207</td>
</tr>
<tr>
<td>Mol Id</td>
<td>R</td>
<td>R₁</td>
<td>Exp. class in SVM class model</td>
<td>Pred. class by SVM model</td>
<td>pKᵢ</td>
<td>pKᵢ</td>
<td>Exp. pKᵢ</td>
<td>Exp. pKᵢ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>CH₂-2,6-Cl₂-Ph</td>
<td>CH₃</td>
<td>-1 -1 -1.65 -2.47 -1.65 45.2</td>
<td>44.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>CH₂-4-F-Ph</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>-1 1 -0.98 -3.24 -1.77 211</td>
<td>58.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>CHPh₂</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>-1 1 -1.29 -2.74 -1.10 326</td>
<td>12.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>182</td>
<td>CH₂-β-naphthyl</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>1 1 -1.38 -2.50 -1.87 73.6</td>
<td>717</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>CH₂-Ph</td>
<td>CH₂-CH₂-Ph</td>
<td>-1 1 -1.70 -1.64 -1.64 43.9</td>
<td>5.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>184</td>
<td>CH₂-3-Cl-Ph</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>-1 1 -1.54 -2.31 -2.26 182</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>CH₂-4-Cl-Ph</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>-1 1 -1.72 -1.97 -1.95 88.9</td>
<td>30.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>CH₂-O-Ph-4-Cl</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>1 1 -1.32 -2.45 -1.43 27.2</td>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>CH₂-2,6-Cl₂-Ph</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>-1 -1 -0.88 -2.14 -2.27 186</td>
<td>601</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>188</td>
<td>CH₂-Ph-Ph</td>
<td>CH₂-CH₂-CH₂-Ph</td>
<td>1 -1 -0.75 -2.06 -2.41 256</td>
<td>410</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D.4 References


APPENDIX E

Paper V
Supplementary Information

E.1 Models 1, 2, 3, training set.
E.2 Models 1, 2, 3, validation set.
E.3 Models 1, 2, 3, internal test set.
E.4 References.

Table E.1 Prediction of the validation set (MODEL 1).
Table E.2 Prediction of the validation set (MODEL 2).
Table E.3 Prediction of the validation set (MODEL 3).
Table E.4 Models 1, 2, 3: prediction of the external test set.
E.1 MODELS 1, 2, 3, training set.
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R1</th>
<th>R2</th>
<th>Exp. Kᵢ (nM) hA₁R or (IC₅₀ nM)</th>
<th>Exp. Kᵢ (nM) hA₂R or (IC₅₀ nM)</th>
<th>Exp. Kᵢ (nM) hA₃R or (IC₅₀ nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>N(CH₃CH₂OH)₂</td>
<td>H</td>
<td>-</td>
<td>123</td>
<td>0.12</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>NH₂</td>
<td>H</td>
<td>-</td>
<td>2,100</td>
<td>0.22</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>CH₃NH₂</td>
<td>H</td>
<td>-</td>
<td>576</td>
<td>0.13</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>NO₂</td>
<td>H</td>
<td>-</td>
<td>1,025</td>
<td>1.00</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>NHCOCH₃</td>
<td>H</td>
<td>-</td>
<td>419</td>
<td>4.80</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>N(CH₃CH₂OH)₂</td>
<td>H</td>
<td>-</td>
<td>558</td>
<td>1.10</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>OCH₂COOEt</td>
<td>H</td>
<td>-</td>
<td>4,197</td>
<td>0.43</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>NH₂</td>
<td>H</td>
<td>-</td>
<td>75</td>
<td>5.5</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>COOH</td>
<td>H</td>
<td>-</td>
<td>4,927</td>
<td>4.63</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>SO₃N(CH₃CH₂OH)₂</td>
<td>H</td>
<td>-</td>
<td>2,495</td>
<td>0.80</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>SO₃N(CH₃CH₂Cl)₂</td>
<td>H</td>
<td>-</td>
<td>309</td>
<td>3.80</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>SO₃H</td>
<td>H</td>
<td>-</td>
<td>5,207</td>
<td>0.70</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>SO₃H</td>
<td>H</td>
<td>-</td>
<td>139</td>
<td>1.40</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>H</td>
<td>H</td>
<td>SO₃H</td>
<td>6,382</td>
<td>7.5</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>COOEt</td>
<td>H</td>
<td>H</td>
<td>1,703</td>
<td>5.48</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>COOH</td>
<td>COOH</td>
<td>H</td>
<td>9,299</td>
<td>1.20</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>COOH</td>
<td>H</td>
<td>H</td>
<td>3,599</td>
<td>5.99</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>H</td>
<td>H</td>
<td>-</td>
<td>350</td>
<td>1.20</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>H</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>448</td>
<td>520</td>
<td>350</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. $K_i$ (nM)</td>
<td>Exp. $K_i$ (nM)</td>
<td>Exp. $K_i$ (nM)</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>CH₃</td>
<td>3,4-OCH₂O-Ph-NH-CO</td>
<td>-</td>
<td>1,015</td>
<td>680</td>
<td>142</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>CH₃</td>
<td>4-NO₂-Ph-NH-CO</td>
<td>-</td>
<td>1,115</td>
<td>695</td>
<td>180</td>
</tr>
<tr>
<td>22</td>
<td>-</td>
<td>CH₃</td>
<td>4-CH₃-Ph-NH-CO</td>
<td>-</td>
<td>731</td>
<td>110</td>
<td>302</td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>CH₃</td>
<td>4-Bz-Ph-NH-CO</td>
<td>-</td>
<td>600</td>
<td>100</td>
<td>181</td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>CH₃</td>
<td>2-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>450</td>
<td>180</td>
<td>223</td>
</tr>
<tr>
<td>25</td>
<td>-</td>
<td>CH₃</td>
<td>3-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>500</td>
<td>160</td>
<td>251</td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>CH₃</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-</td>
<td>400</td>
<td>200</td>
<td>101</td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>C₂H₅</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td>-</td>
<td>526</td>
<td>352</td>
<td>40</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>C₂H₅</td>
<td>3,4-OCH₂O-Ph-NH-CO</td>
<td>-</td>
<td>802</td>
<td>576</td>
<td>48</td>
</tr>
<tr>
<td>29</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-Bz-Ph-NH-CO</td>
<td>-</td>
<td>500</td>
<td>40</td>
<td>296</td>
</tr>
<tr>
<td>30</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-F-Ph-NH-CO</td>
<td>-</td>
<td>602</td>
<td>60</td>
<td>246</td>
</tr>
<tr>
<td>31</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-CF₃-Ph-NH-CO</td>
<td>-</td>
<td>450</td>
<td>53</td>
<td>235</td>
</tr>
<tr>
<td>32</td>
<td>-</td>
<td>C₂H₅</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-</td>
<td>348</td>
<td>150</td>
<td>97</td>
</tr>
<tr>
<td>33</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-</td>
<td>400</td>
<td>160</td>
<td>85</td>
</tr>
<tr>
<td>34</td>
<td>n-C₃H₇</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td>-</td>
<td>611</td>
<td>401</td>
<td>31</td>
<td>2.50</td>
</tr>
<tr>
<td>35</td>
<td>n-C₃H₇</td>
<td>3,4-OCH₂O-Ph-NH-CO</td>
<td>-</td>
<td>842</td>
<td>667</td>
<td>53</td>
<td>0.30</td>
</tr>
<tr>
<td>36</td>
<td>n-C₃H₇</td>
<td>4-NO₂-Ph-NH-CO</td>
<td>-</td>
<td>1,214</td>
<td>1,115</td>
<td>305</td>
<td>0.81</td>
</tr>
<tr>
<td>37</td>
<td>n-C₃H₇</td>
<td>2-Cl-Ph-NH-CO</td>
<td>-</td>
<td>300</td>
<td>121</td>
<td>61</td>
<td>0.71</td>
</tr>
<tr>
<td>38</td>
<td>n-C₃H₇</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-</td>
<td>325</td>
<td>140</td>
<td>76</td>
<td>0.34</td>
</tr>
</tbody>
</table>
### Table: Models 1, 2, 3 training set continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. Kᵢ (nM)</th>
<th>Exp. Kᵢᵢ (nM)</th>
<th>Exp. Kᵢᵢᵢ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td>-</td>
<td>708</td>
<td>495</td>
<td>74</td>
</tr>
<tr>
<td>40</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>3,4-OCH₂O-Ph-NH-CO</td>
<td>-</td>
<td>410</td>
<td>376</td>
<td>78</td>
</tr>
<tr>
<td>41</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>4-N₂O₂Ph-NH-CO</td>
<td>-</td>
<td>584</td>
<td>503</td>
<td>52</td>
</tr>
<tr>
<td>42</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>4-CH₃Ph-NH-CO</td>
<td>-</td>
<td>281</td>
<td>241</td>
<td>161</td>
</tr>
<tr>
<td>43</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>4-BrPh-NH-CO</td>
<td>-</td>
<td>300</td>
<td>66</td>
<td>121</td>
</tr>
<tr>
<td>44</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>4-FPh-NH-CO</td>
<td>-</td>
<td>318</td>
<td>50</td>
<td>136</td>
</tr>
<tr>
<td>45</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>4-Cl₂Ph-NH-CO</td>
<td>-</td>
<td>350</td>
<td>31</td>
<td>190</td>
</tr>
<tr>
<td>46</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>2-OCH₃Ph-NH-CO</td>
<td>-</td>
<td>200</td>
<td>91</td>
<td>125</td>
</tr>
<tr>
<td>47</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>3-OCH₃Ph-NH-CO</td>
<td>-</td>
<td>248</td>
<td>95</td>
<td>150</td>
</tr>
<tr>
<td>48</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>2-ClPh-NH-CO</td>
<td>-</td>
<td>280</td>
<td>100</td>
<td>78</td>
</tr>
<tr>
<td>49</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>Ph-CH₂-CO</td>
<td>-</td>
<td>702</td>
<td>423</td>
<td>165</td>
</tr>
<tr>
<td>50</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>Ph-CH₂-CO</td>
<td>-</td>
<td>714</td>
<td>335</td>
<td>161</td>
</tr>
<tr>
<td>51</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>Ph-CH₂-CO</td>
<td>-</td>
<td>351</td>
<td>306</td>
<td>143</td>
</tr>
<tr>
<td>52</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>Ph-NH-CO</td>
<td>-</td>
<td>176</td>
<td>62</td>
<td>241</td>
</tr>
<tr>
<td>53</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>Ph-NH-CO</td>
<td>-</td>
<td>409</td>
<td>148</td>
<td>50</td>
</tr>
<tr>
<td>54</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-SO₂HPh-NH-CO</td>
<td>-</td>
<td>&gt;10,000</td>
<td>764</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>55</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-SO₂HPh-NH-CO</td>
<td>-</td>
<td>&gt;10,000</td>
<td>705</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>56</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-SO₂HPh-NH-CO</td>
<td>-</td>
<td>&gt;10,000</td>
<td>352</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>57</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-CH₃Ph-NH-CO</td>
<td>-</td>
<td>1,097</td>
<td>1,300</td>
<td>261</td>
</tr>
</tbody>
</table>
### MODELS 1, 2, 3, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. $K_i$ (nM) hA₁R or (IC₅₀ nM)</th>
<th>Exp. $K_i$ (nM) hA₂₂R or (IC₅₀ nM)</th>
<th>Exp. $K_i$ (nM) hA₂₄R or (IC₅₀ nM)</th>
<th>Exp. $K_i$ (nM) hA₃₃R or (IC₅₀ nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>58</td>
<td></td>
<td>CH₃</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>350</td>
<td>1,195</td>
<td>205</td>
<td>0.40</td>
</tr>
<tr>
<td>59</td>
<td></td>
<td>C₂H₅</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>249</td>
<td>180</td>
<td>150</td>
<td>1.60</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>n-C₃H₇</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>1,197</td>
<td>140</td>
<td>2,056</td>
<td>0.80</td>
</tr>
<tr>
<td>61</td>
<td></td>
<td>n-C₄H₉</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>296</td>
<td>80</td>
<td>303</td>
<td>0.32</td>
</tr>
<tr>
<td>62</td>
<td></td>
<td>n-C₄H₉</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>245</td>
<td>95</td>
<td>70</td>
<td>0.60</td>
</tr>
<tr>
<td>63</td>
<td></td>
<td>-</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>7.60</td>
<td>9.40</td>
<td>22</td>
<td>0.14</td>
</tr>
<tr>
<td>64</td>
<td></td>
<td>-</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>8.10</td>
<td>9.50</td>
<td>30</td>
<td>0.19</td>
</tr>
<tr>
<td>65</td>
<td></td>
<td>Ph-CH₂₂-CH₂</td>
<td>H</td>
<td>-</td>
<td>1</td>
<td>0.34</td>
<td>5.1</td>
<td>2,785</td>
</tr>
<tr>
<td>66</td>
<td></td>
<td>Ph-CH₂₂-CH₂₂CH₂</td>
<td>H</td>
<td>-</td>
<td>3</td>
<td>0.16</td>
<td>49</td>
<td>2,666</td>
</tr>
<tr>
<td>67</td>
<td></td>
<td>Ph-CH₂₂-CH₂</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>251</td>
<td>105</td>
<td>90</td>
<td>13,28</td>
</tr>
<tr>
<td>68</td>
<td></td>
<td>Ph-CH₂₂-CH₂₂CH₂</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>1,500</td>
<td>200</td>
<td>226</td>
<td>42.65</td>
</tr>
<tr>
<td>69</td>
<td></td>
<td>(CH₃)₂CH-CH₂₂CH₂</td>
<td>H</td>
<td>-</td>
<td>2.00</td>
<td>0.78</td>
<td>9.1</td>
<td>700</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>2,4,5-Br₃-Ph-CH₂₂-CH₂</td>
<td>H</td>
<td>-</td>
<td>247</td>
<td>3.48</td>
<td>1,113</td>
<td>4,481</td>
</tr>
<tr>
<td>71</td>
<td></td>
<td>H</td>
<td>H</td>
<td>-</td>
<td>140</td>
<td>20</td>
<td>51</td>
<td>348</td>
</tr>
<tr>
<td>72</td>
<td></td>
<td>2-C₅H₉</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>1,015</td>
<td>706</td>
<td>450</td>
<td>2.78</td>
</tr>
<tr>
<td>73</td>
<td></td>
<td>(CH₃)₂CH-CH₂₂CH₂</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>100</td>
<td>115</td>
<td>61</td>
<td>40</td>
</tr>
<tr>
<td>74</td>
<td></td>
<td>(CH₃)₂CH=CH₂₂CH₂</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>300</td>
<td>1,025</td>
<td>58</td>
<td>40</td>
</tr>
<tr>
<td>75</td>
<td></td>
<td>(CH₃)₂CH=CH₂₂CH₂</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>332</td>
<td>110</td>
<td>81</td>
<td>55</td>
</tr>
<tr>
<td>76</td>
<td></td>
<td>2,4,5-Br₃-Ph-CH₂₂-CH₂</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>255</td>
<td>2,000</td>
<td>&gt;10,000</td>
<td>25</td>
</tr>
</tbody>
</table>
### Appendix E: Paper V

**Models 1, 2, 3, training set continued**

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R1</th>
<th>R2</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (IC$_{50}$ nM)</th>
<th>Exp. $K_i$ (IC$_{50}$ nM)</th>
<th>Exp. $K_i$ (IC$_{50}$ nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>-</td>
<td>2,4,5-Br$_3$-Ph-CH$_2$-CH$_2$</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>389</td>
<td>3,000</td>
<td>1,816</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>-</td>
<td>2-((naphthyl)ethyl)</td>
<td>4-OCH$_3$-Ph-NH-CO</td>
<td>-</td>
<td>2,106</td>
<td>3,200</td>
<td>&gt;10,000</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>-</td>
<td>2-((naphthyl)ethyl)</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>3,164</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>-</td>
<td>n-C$_4$H$_9$</td>
<td>H</td>
<td>-</td>
<td>14</td>
<td>1.60</td>
<td>33</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CO$_2$NH$_2$(CH$_3$)$_2$</td>
<td>-</td>
<td>60</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CO$_2$NH$_2$(CH$_3$)$_3$</td>
<td>-</td>
<td>35</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>-</td>
<td>(CH$_3$)$_2$CH-CH$_2$-CH$_2$</td>
<td>CO$_2$NH$_2$(CH$_3$)$_2$</td>
<td>-</td>
<td>389</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$-CH$_2$</td>
<td>CO$_2$NH$_2$(CH$_3$)$_2$</td>
<td>-</td>
<td>849</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CO((CH$_2$)$_3$)NH$_3$+Cl$^-$</td>
<td>-</td>
<td>1.6</td>
<td>54.5</td>
<td>27.6</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CO((CH$_2$)$_3$)NHCOOC(CH$_3$)$_3$</td>
<td>-</td>
<td>11.0</td>
<td>298</td>
<td>&gt;10,000</td>
<td>1,480</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>-</td>
<td>(CH$_3$)$_2$CH-CH$_2$-CH$_2$</td>
<td>CO((CH$_2$)$_3$)NH$_3$+Cl$^-$</td>
<td>-</td>
<td>9.9</td>
<td>241</td>
<td>52.8</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>-</td>
<td>(CH$_3$)$_2$CH-CH$_2$-CH$_2$</td>
<td>CO((CH$_2$)$_3$)NHCOOC(CH$_3$)$_3$</td>
<td>-</td>
<td>99</td>
<td>297</td>
<td>&gt;10,000</td>
<td>2,488</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CH$_2$-CH$_2$-OH</td>
<td>-</td>
<td>5,485</td>
<td>84.9</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CH$_2$-CH$_2$-OH</td>
<td>-</td>
<td>50.4</td>
<td>750</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>-</td>
<td>cC$<em>6$H$</em>{11}$-CH$_2$-CH$_2$</td>
<td>CH$_2$-CH$_2$-OH</td>
<td>-</td>
<td>714</td>
<td>51.3</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CO-CH$_2$-NH$_2$+Cl$^-$</td>
<td>-</td>
<td>4.5</td>
<td>182</td>
<td>70</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>-</td>
<td>Ph-CH$_2$-CH$_2$</td>
<td>CO((CH$_2$)$_2$)NH$_3$+Cl$^-$</td>
<td>-</td>
<td>2.5</td>
<td>50.0</td>
<td>37.6</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>-</td>
<td>CH$_3$</td>
<td>Ph$_2$-CH-CO</td>
<td>-</td>
<td>139</td>
<td>216</td>
<td>116</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>-</td>
<td>CH$_3$</td>
<td>Ph$_2$-Ph-CH$_2$-CO</td>
<td>-</td>
<td>1,400</td>
<td>193</td>
<td>(&gt;10,000)</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>--------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>CH₂-CH₂-CH(CH₃)₂</td>
<td>Ph₂-CH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>441</td>
<td>159</td>
<td>(&gt;10,000)</td>
<td>5.86</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>CH₂-CH₂-CH(CH₃)₂</td>
<td>Ph₂-CH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>84.3</td>
<td>9.18</td>
<td>(8,900)</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Ph-CH₂-CH₂</td>
<td>Ph₂-CH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>115</td>
<td>53.1</td>
<td>(5,840)</td>
<td>6.49</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Ph-CH₂-CH₂</td>
<td>Ph₂-CH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>84.3</td>
<td>46.9</td>
<td>(11,100)</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>C₂H₅</td>
<td>Ph₂-CH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>557</td>
<td>70.9</td>
<td>(&gt;50,000)</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>CH₃</td>
<td>4-OCH₃Ph₂-CH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>133</td>
<td>62.5</td>
<td>(&gt;10,000)</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>CH₂-CH₂-CH(CH₃)₂</td>
<td>β-naphthyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>140</td>
<td>15.7</td>
<td>(&gt;10,000)</td>
<td>409</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>Ph₂-CH₂-Ch₂</td>
<td>β-naphthyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>57.5</td>
<td>42.1</td>
<td>(&gt;50,000)</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>Ph₂-CH₂-Ch₂</td>
<td>2-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>18.5</td>
<td>8.8</td>
<td>(&gt;10,000)</td>
<td>330</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Ph₂-CH₂-Ch₂</td>
<td>3-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>22.8</td>
<td>12.9</td>
<td>(&gt;10,000)</td>
<td>726</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>C₂H₅</td>
<td>3-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>154</td>
<td>23</td>
<td>(9,370)</td>
<td>765</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>C₂H₅</td>
<td>2-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>148</td>
<td>15.9</td>
<td>(15,000)</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>CH₃</td>
<td>2-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>444</td>
<td>56</td>
<td>(&gt;10,000)</td>
<td>5.28</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>CH₃</td>
<td>3-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>461</td>
<td>31.3</td>
<td>(18,700)</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>CH₃</td>
<td>β-naphthyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>190</td>
<td>77.5</td>
<td>(&gt;10,000)</td>
<td>14.5</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>CH₃</td>
<td>α-naphthyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>872</td>
<td>80.5</td>
<td>(&gt;10,000)</td>
<td>3.47</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>n-C₃H₇</td>
<td>β-naphthyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>160</td>
<td>7.99</td>
<td>(5,370)</td>
<td>95.9</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>n-C₃H₇</td>
<td>3-thienyl-CH₂</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>295</td>
<td>3.13</td>
<td>290</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>CH₃</td>
<td>4-Cl-Ph-OCH₂-CO</td>
<td>-</td>
<td>hA₁R or (% displ. 10 μM)</td>
<td>963</td>
<td>39.3</td>
<td>(&gt;10,000)</td>
<td>223</td>
<td></td>
</tr>
</tbody>
</table>
### MODELS 1, 2, 3 training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. $K_i$ (nM) hₐ₁R or (% displ. 10 µM)</th>
<th>Exp. $K_i$ (nM) hₐ₂₄R</th>
<th>Exp. $K_i$ (nM) hₐ₂₅R or (IC₅₀ nM)</th>
<th>Exp. $K_i$ (nM) hₐ₃₅R or (% displ. 10 µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td>-</td>
<td>CH₂CH₂CH(CH₃)₂</td>
<td>4-Cl-PHEN-CH₂-CO</td>
<td>-</td>
<td>22.6 2.75 391 56.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>-</td>
<td>CH₂CH₂CH(CH₃)₂</td>
<td>4-CF₃-PHEN-CH₂-CO</td>
<td>-</td>
<td>48.2 5.43 (5,480) 266</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>-</td>
<td>CH₂CH₂CH(CH₃)₂</td>
<td>4-F-PHEN-CH₂-CO</td>
<td>-</td>
<td>51.8 3.69 (5,000) 116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>-</td>
<td>CH₂CH₂CH(CH₃)₂</td>
<td>2,6-Cl-PHEN-CH₂-CO</td>
<td>-</td>
<td>121 18.7 461 207</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>-</td>
<td>CH₃</td>
<td>2,6-Cl-PHEN-CH₂-CO</td>
<td>-</td>
<td>908 45.2 (6,450) 44.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>-</td>
<td>Ph-CH₂CH₂CH₂</td>
<td>4-F-PHEN-CH₂-CO</td>
<td>-</td>
<td>726 21.1 (&gt;50,000) 58.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>-</td>
<td>Ph-CH₂CH₂CH₂</td>
<td>Ph₂CH₂-CO</td>
<td>-</td>
<td>307 326 (&gt;10,000) 12.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>-</td>
<td>Ph-CH₂CH₂CH₂</td>
<td>3-Cl-PHEN-CH₂-CO</td>
<td>-</td>
<td>605 182 (50,000) 110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>-</td>
<td>Ph-CH₂CH₂CH₂</td>
<td>4-Cl-PHEN-CH₂-CO</td>
<td>-</td>
<td>256 88.0 (&gt;10,000) 30.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>-</td>
<td>Ph-CH₂CH₂CH₂</td>
<td>2,6-Cl-PHEN-CH₂-CO</td>
<td>-</td>
<td>786 186 (&gt;10,000) 60.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>-</td>
<td>Ph-CH₂CH₂CH₂</td>
<td>Ph₂-PHEN-CO</td>
<td>-</td>
<td>1,170 256 (&gt;100,000) 410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>-</td>
<td>CH₃</td>
<td>Ph₂-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>2,410 233 (12,900) 29.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>-</td>
<td>CH₃</td>
<td>2,4-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>2,940 333 (&gt;30,000) 6.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>-</td>
<td>CH₃</td>
<td>2,6-Cl-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>1,660 281 (17,800) 35.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>-</td>
<td>CH₃</td>
<td>3-Cl-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>1,980 278 (10,400) 9.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>-</td>
<td>CH₃</td>
<td>4-CH₃-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>6,760 751 (&gt;30,000) 24.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>-</td>
<td>CH₃</td>
<td>2,5-CH₃-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>5,730 609 (&gt;30,000) 11.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>133</td>
<td>-</td>
<td>CH₃</td>
<td>3,4-CH₃-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>5,250 397 (&gt;30,000) 10.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>134</td>
<td>-</td>
<td>CH₃</td>
<td>3,5-CH₃-PHEN₂O-PHEN₂-CO</td>
<td>-</td>
<td>7,200 357 (&gt;30,000) 11.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MODELs 1, 2, 3, training set: continued

| Mol Id | n   | R          | R1          | R2          | Exp. $K_i$ (nM) | Exp. $K_i$ (nM) | Exp. $K_i$ (nM) | Exp. $K_i$ (nM) |
|--------|-----|------------|-------------|-------------|----------------|----------------|----------------|----------------|----------------|
| 135    | -   | CH$_2$-CH$_2$-CH(=CH$_2$)$_2$ | 3,4-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 396            | 572            | 30,000         | 135            |
| 136    | -   | CH$_2$-CH$_2$-CH(=CH$_2$)$_2$ | 3,2-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 7,620          | 954            | 30,000         | 936            |
| 137    | -   | CH$_2$-CH$_2$-CH(=CH$_2$)$_2$ | 2,4,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 975            | 83.7           | >10,000         | 197            |
| 138    | -   | Ph-CH$_2$-CH$_2$ | 2,5-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 148            | 52.9           | >10,000         | 61.3           |
| 139    | -   | Ph-CH$_2$-CH$_2$ | 3,5-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 348            | 156            | >10,000         | 46.3           |
| 140    | -   | Ph-CH$_2$-CH$_2$ | 2,4,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 1,180          | 185            | >10,000         | 79.5           |
| 141    | -   | Ph-CH$_2$-CH$_2$ | 2,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 4,590          | 432            | >10,000         | 278            |
| 142    | -   | Ph-CH$_2$-CH$_2$ | 4,8,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 1,680          | 222            | >10,000         | 88.8           |
| 143    | -   | Ph-CH$_2$-CH$_2$ | 3,4,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 1,550          | 351            | >10,000         | 279            |
| 144    | -   | Ph-CH$_2$-CH$_2$ | 2,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 3,610          | 469            | >10,000         | 271            |
| 145    | -   | Ph-CH$_2$-CH$_2$ | 3,4,6-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 63.3           | 8.41           | >10,000         | 150            |
| 146    | -   | Ph-CH$_2$-CH$_2$ | 2,4,6,8-CH$_3$-Ph-CH$_2$O-Ph-CH$_2$-CO | -           | 2,190          | 759            | >10,000         | 587            |
| 147    | -   | H          | NH(=CH$_2$)$_2$Ph-OH | -           | 774            | 1.6            | 75             | 743            |
| 148    | -   | H          | OPh         | -           | 2,720          | 18.3           | (3,420)         | 489            |
| 149    | -   | H          | SCH$_3$ | -           | 1,730          | 96.5           | (14,900)        | 2,576           |
| 150    | -   | Ph-CH$_2$-CO | SCH$_3$ | -           | 1,420          | 429            | (>100,000)      | 4,200           |
| 151    | -   | 4-CH$_3$Ph-NH-CO | SCH$_3$ | -           | 3,440          | 742            | (16,800)        | 2,300           |
| 152    | -   | 4-CH$_3$(CH$_2$)$_4$CO | OPh | -           | 2,900          | 189            | (>100,000)      | 2,300           |
| 153    | -   | 4-CH$_3$Ph-NH-CO | OPh | -           | 35% inhib      | 214            | (20,000)        | 750            |
### Appendix E: Paper V

**MODELS 1, 2, 3; training set: continued**

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>R&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$hA_1 R$ or $(% \text{ displ.} \ 10 \mu M)$</td>
<td>$hA_2R$ or $(% \text{ displ.} \ 10 \mu M)$</td>
<td>$hA_2B R$ or $(% \text{ displ.} \ 10 \mu M)$</td>
<td>$hA_3 R$ or $(% \text{ displ.} \ 10 \mu M)$</td>
</tr>
<tr>
<td>154</td>
<td></td>
<td>Ph&lt;sub&gt;2&lt;/sub&gt;-CH-CO</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>951</td>
<td>4,090</td>
<td>(&gt;100,000)</td>
<td>473</td>
</tr>
<tr>
<td>155</td>
<td></td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>6,000</td>
<td>6,700</td>
<td>(&gt;100,000)</td>
<td>1,050</td>
</tr>
<tr>
<td>156</td>
<td></td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;(CH&lt;sub&gt;2&lt;/sub&gt;)&lt;sub&gt;4&lt;/sub&gt;CO</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>1,570</td>
<td>10,300</td>
<td>(&gt;100,000)</td>
<td>2,200</td>
</tr>
<tr>
<td>157</td>
<td></td>
<td>Ph-NH-CO</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>3,150</td>
<td>580</td>
<td>(&gt;100,000)</td>
<td>311</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td>4-CH&lt;sub&gt;3&lt;/sub&gt;Ph-NH-CO</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>2,720</td>
<td>2,720</td>
<td>(&gt;100,000)</td>
<td>9,600</td>
</tr>
<tr>
<td>159</td>
<td></td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
<td>423</td>
<td>(&gt;10,000)</td>
<td>(56.5)</td>
</tr>
<tr>
<td>160</td>
<td></td>
<td>NH(CH&lt;sub&gt;2&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;NH-Boc</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-NH-Boc</td>
<td>-</td>
<td>1,148</td>
<td>3,180</td>
<td>(&gt;100,000)</td>
<td>788</td>
</tr>
<tr>
<td>161</td>
<td></td>
<td>H</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-NH-Boc</td>
<td>-</td>
<td>1,410</td>
<td>463</td>
<td>(30,000)</td>
<td>696</td>
</tr>
<tr>
<td>162</td>
<td></td>
<td>H</td>
<td>NH(CH&lt;sub&gt;2&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-NH-Boc</td>
<td>-</td>
<td>408</td>
<td>65</td>
<td>(20,000)</td>
<td>(77.3)</td>
</tr>
<tr>
<td>163</td>
<td></td>
<td>H</td>
<td>N-N=CH&lt;sub&gt;2&lt;/sub&gt;Ph</td>
<td>-</td>
<td>141</td>
<td>23</td>
<td>(&gt;30,000)</td>
<td>(71.8)</td>
</tr>
<tr>
<td>164</td>
<td></td>
<td>H</td>
<td>N-N=CH&lt;sub&gt;2&lt;/sub&gt;F-Pht</td>
<td>-</td>
<td>5,630</td>
<td>20</td>
<td>(&gt;30,000)</td>
<td>(52.7)</td>
</tr>
<tr>
<td>165</td>
<td></td>
<td>H</td>
<td>N-N=CH&lt;sub&gt;2&lt;/sub&gt;F-Pht</td>
<td>-</td>
<td>787</td>
<td>174</td>
<td>(&gt;10,000)</td>
<td>(71.2)</td>
</tr>
<tr>
<td>166</td>
<td></td>
<td>H</td>
<td>N-N=CH&lt;sub&gt;2&lt;/sub&gt;4Br-Pht</td>
<td>-</td>
<td>255</td>
<td>9.5</td>
<td>(30,000)</td>
<td>(14.6)</td>
</tr>
<tr>
<td>167</td>
<td></td>
<td>H</td>
<td>N-N=CH&lt;sub&gt;2&lt;/sub&gt;N=CH&lt;sub&gt;2&lt;/sub&gt;CH=CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>303</td>
<td>252</td>
<td>(&gt;100,000)</td>
<td>(49.6)</td>
</tr>
<tr>
<td>168</td>
<td></td>
<td>H</td>
<td>N-N=CH&lt;sub&gt;2&lt;/sub&gt;OC&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>-</td>
<td>120</td>
<td>439</td>
<td>(&gt;100,000)</td>
<td>(33.5)</td>
</tr>
</tbody>
</table>
### TABLE 2.3.1

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. Kᵣ₃ (nM) hAᵣ₁R or (% displ. 10 µM)</th>
<th>Exp. Kᵣ₃ (nM) hAᵣ₂R or (IC₅₀ nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>169</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>415</td>
<td>1,670</td>
</tr>
<tr>
<td>170</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
<td>3,780</td>
</tr>
<tr>
<td>171</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>626</td>
<td>83</td>
</tr>
<tr>
<td>172</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
<td>578</td>
</tr>
<tr>
<td>173</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
<td>404</td>
</tr>
<tr>
<td>174</td>
<td>-</td>
<td>PhCO</td>
<td>OPh</td>
<td>-</td>
<td>&gt;10,000</td>
<td>87</td>
</tr>
<tr>
<td>175</td>
<td>-</td>
<td>PhCO</td>
<td>H</td>
<td>-</td>
<td>3,830</td>
<td>32</td>
</tr>
<tr>
<td>176</td>
<td>-</td>
<td>CH₃</td>
<td>H</td>
<td>NHCH₂CH₃</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>177</td>
<td>-</td>
<td>CH₃</td>
<td>H</td>
<td>4-OCH₃Ph-NH</td>
<td>260</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>178</td>
<td>-</td>
<td>CH₃</td>
<td>H</td>
<td>CH₃-N-piperazine</td>
<td>30</td>
<td>156</td>
</tr>
<tr>
<td>179</td>
<td>-</td>
<td>CH₃</td>
<td>H</td>
<td>S(CH₂)₃CH₃</td>
<td>9</td>
<td>2.1</td>
</tr>
<tr>
<td>180</td>
<td>-</td>
<td>CH₃</td>
<td>4-OCH₃Ph-NH-CO</td>
<td>CH₃-N-piperazine</td>
<td>316</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>181</td>
<td>-</td>
<td>SCH₃</td>
<td>4-OCH₃Ph-NH-CO</td>
<td>SCH₃</td>
<td>70</td>
<td>3.1</td>
</tr>
<tr>
<td>182</td>
<td>-</td>
<td>SCH₃</td>
<td>4-OCH₃Ph-CH₂-CO</td>
<td>SCH₃</td>
<td>80</td>
<td>1.5</td>
</tr>
<tr>
<td>183</td>
<td>-</td>
<td>SCH₃</td>
<td>4-isobutyl-Ph-CH₂-CO</td>
<td>SCH₃</td>
<td>780</td>
<td>50</td>
</tr>
</tbody>
</table>
### Models 1, 2, 3, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>R&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>184</td>
<td>-</td>
<td>S&lt;sub&gt;CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3,4-methoxy-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>SCH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>70</td>
<td>4.1</td>
<td>30</td>
<td>110</td>
</tr>
<tr>
<td>185</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3,4-methoxy-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>NHCH&lt;sub&gt;2&lt;/sub&gt;CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>136</td>
<td>61</td>
<td>65</td>
<td>183</td>
</tr>
<tr>
<td>186</td>
<td>-</td>
<td>N-methylpiperazine</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>187</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>4-OH-Ph-NH</td>
<td>666</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>308</td>
</tr>
<tr>
<td>188</td>
<td>-</td>
<td>2-OEt</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>348</td>
</tr>
<tr>
<td>189</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>235</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>190</td>
<td>-</td>
<td>4-OCH&lt;sub&gt;2&lt;/sub&gt;CO&lt;sub&gt;2&lt;/sub&gt;Et</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>191</td>
<td>-</td>
<td>4-OCH&lt;sub&gt;2&lt;/sub&gt;CO&lt;sub&gt;2&lt;/sub&gt;H</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>192</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-F-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>500</td>
<td>&gt;1,000</td>
<td>14</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>193</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-Cl-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>7</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>194</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-Br-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>35</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>195</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-i-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>1,000</td>
<td>&gt;1,000</td>
<td>28</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>196</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-nec-butyl-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>197</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>480</td>
<td>&gt;1,000</td>
<td>65</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>198</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3,4-dimethoxy-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>60</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>199</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1-naphthyl-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>200</td>
<td>-</td>
<td>Cl</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>222</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>201</td>
<td>-</td>
<td>Br</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>250</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>202</td>
<td>-</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. $K_i$ (nM) hA₁R or (% disp. 10 µM)</td>
<td>Exp. $K_i$ (nM) hA₂₄R</td>
<td>Exp. $K_i$ (nM) hA₂₆R or (IC₅₀ nM)</td>
<td>Exp. $K_i$ (nM) hA₃R or (% disp. 10 µM)</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>203</td>
<td></td>
<td>- propyl</td>
<td>- propyl</td>
<td>cyclopentyl</td>
<td>3</td>
<td>60</td>
<td>51</td>
<td>243</td>
</tr>
<tr>
<td>204</td>
<td></td>
<td>- propyl</td>
<td>(S)-1-phenylethyl</td>
<td>biphenyl</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>455</td>
</tr>
<tr>
<td>205</td>
<td></td>
<td>- butyl</td>
<td>3-hydroxypropyl</td>
<td>3-noradamantyl</td>
<td>0.7</td>
<td>980</td>
<td>187</td>
<td>2,300</td>
</tr>
<tr>
<td>206</td>
<td></td>
<td>- propyl</td>
<td>3-noradamantyl</td>
<td>-</td>
<td>13.8</td>
<td>25,000</td>
<td>22,300</td>
<td>188</td>
</tr>
<tr>
<td>207</td>
<td></td>
<td>- H</td>
<td>CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>0.8</td>
</tr>
<tr>
<td>208</td>
<td></td>
<td>- H</td>
<td>CH₃CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>15</td>
</tr>
<tr>
<td>209</td>
<td></td>
<td>- H</td>
<td>CH(CH₃)₂</td>
<td>-</td>
<td>460</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>31</td>
</tr>
<tr>
<td>210</td>
<td></td>
<td>- H</td>
<td>cyclohexyl</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>555</td>
</tr>
<tr>
<td>211</td>
<td></td>
<td>- CH₃</td>
<td>CH₂CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>60</td>
</tr>
<tr>
<td>212</td>
<td></td>
<td>- H</td>
<td>4-OCH₃-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>55</td>
</tr>
<tr>
<td>213</td>
<td></td>
<td>- H</td>
<td>4-Ph-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>214</td>
<td></td>
<td>- H</td>
<td>4-F-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>22</td>
</tr>
<tr>
<td>215</td>
<td></td>
<td>- H</td>
<td>CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>400</td>
<td>8</td>
</tr>
<tr>
<td>216</td>
<td></td>
<td>- H</td>
<td>CH₂CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>3.45</td>
</tr>
<tr>
<td>217</td>
<td></td>
<td>- CH₃</td>
<td>CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>80</td>
</tr>
<tr>
<td>218</td>
<td></td>
<td>- H</td>
<td>CH₂-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>148</td>
</tr>
<tr>
<td>219</td>
<td></td>
<td>- CH₂-CH₂-OH</td>
<td>CH₂-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>38</td>
</tr>
<tr>
<td>220</td>
<td></td>
<td>- CH₂-CH=CH₂</td>
<td>CH₂-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>5.13</td>
</tr>
</tbody>
</table>
### Models 1, 2, 3, training set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R1</th>
<th>R2</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>Exp. $K_i$ (IC$_{50}$ nM)</th>
<th>Exp. $K_i$ (10 µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>221</td>
<td>-</td>
<td>CH$_2$C≡CH</td>
<td>CH$_2$Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>15</td>
</tr>
<tr>
<td>222</td>
<td>-</td>
<td>CH$_2$CH$_2$N(CH$_3$)$_2$</td>
<td>CH$_2$Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>98</td>
</tr>
<tr>
<td>223</td>
<td>-</td>
<td>H</td>
<td>1-hydroxy-propyl-2-yl</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>224</td>
<td>-</td>
<td>H</td>
<td>CH$_2$CH$_3$</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>225</td>
<td>-</td>
<td>CH$_2$-CH$_2$CH$_3$</td>
<td>Ph</td>
<td>-</td>
<td>373</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>99</td>
</tr>
<tr>
<td>226</td>
<td>-</td>
<td>CH$_2$-CH(CH$_3$)$_2$</td>
<td>Ph</td>
<td>-</td>
<td>476</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>144</td>
</tr>
<tr>
<td>227</td>
<td>-</td>
<td>CH$_3$</td>
<td>Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>228</td>
<td>-</td>
<td>CH$_3$</td>
<td>CH$_3$</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>229</td>
<td>-</td>
<td>Ph</td>
<td>-</td>
<td>-</td>
<td>102</td>
<td>1,500</td>
<td>22</td>
<td>1,200</td>
</tr>
<tr>
<td>230</td>
<td>-</td>
<td>2-Cl-Ph</td>
<td>-</td>
<td>-</td>
<td>2,500</td>
<td>2,500</td>
<td>19</td>
<td>490</td>
</tr>
<tr>
<td>231</td>
<td>-</td>
<td>4-CN-Ph</td>
<td>-</td>
<td>-</td>
<td>1,200</td>
<td>&gt;5,000</td>
<td>18</td>
<td>950</td>
</tr>
<tr>
<td>233</td>
<td>-</td>
<td>3-Cl-Ph</td>
<td>-</td>
<td>-</td>
<td>652</td>
<td>&gt;5,000</td>
<td>82</td>
<td>490</td>
</tr>
<tr>
<td>234</td>
<td>-</td>
<td>3-CF$_3$-Ph</td>
<td>-</td>
<td>-</td>
<td>3,500</td>
<td>&gt;5,000</td>
<td>42</td>
<td>960</td>
</tr>
<tr>
<td>235</td>
<td>-</td>
<td>2-CF$_3$-Ph</td>
<td>-</td>
<td>-</td>
<td>530</td>
<td>&gt;5,000</td>
<td>74</td>
<td>340</td>
</tr>
<tr>
<td>236</td>
<td>-</td>
<td>2-OCH$_3$-Ph</td>
<td>-</td>
<td>-</td>
<td>1,900</td>
<td>1,600</td>
<td>82</td>
<td>410</td>
</tr>
<tr>
<td>237</td>
<td>-</td>
<td>3-OCH$_3$-Ph</td>
<td>-</td>
<td>-</td>
<td>2,200</td>
<td>&gt;5,000</td>
<td>82</td>
<td>600</td>
</tr>
<tr>
<td>238</td>
<td>-</td>
<td>4-CH$_3$Ph</td>
<td>-</td>
<td>-</td>
<td>2,600</td>
<td>3,000</td>
<td>54</td>
<td>8,700</td>
</tr>
<tr>
<td>239</td>
<td>-</td>
<td>CH$_2$-CH$_2$CH$_3$</td>
<td>CH$_2$-CH$_2$-CH$_3$</td>
<td>Ph</td>
<td>1,000</td>
<td>1,800</td>
<td>21</td>
<td>630</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>240</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-Cl-Ph</td>
<td>2,000</td>
<td>&gt;5,000</td>
<td>39</td>
<td>&gt;8,500</td>
</tr>
<tr>
<td>241</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-CF₃-Ph</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>21</td>
<td>1,300</td>
</tr>
<tr>
<td>242</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>3-Cl-Ph</td>
<td>3,000</td>
<td>&gt;5,000</td>
<td>134</td>
<td>3,000</td>
</tr>
<tr>
<td>243</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>3-CF₃-Ph</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>64</td>
<td>1,800</td>
</tr>
<tr>
<td>244</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-OCH₃-Ph</td>
<td>3,500</td>
<td>&gt;5,000</td>
<td>123</td>
<td>2,300</td>
</tr>
<tr>
<td>245</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>2-OCH₃-Ph</td>
<td>930</td>
<td>3,600</td>
<td>88</td>
<td>&gt;8,500</td>
</tr>
<tr>
<td>246</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>Ph</td>
<td>1,500</td>
<td>420</td>
<td>14</td>
<td>1,800</td>
</tr>
<tr>
<td>247</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-Cl-Ph</td>
<td>540</td>
<td>1,300</td>
<td>82</td>
<td>3,300</td>
</tr>
<tr>
<td>248</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-CF₃-Ph</td>
<td>2,300</td>
<td>1,700</td>
<td>48</td>
<td>1,200</td>
</tr>
<tr>
<td>249</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>3-OCH₃-Ph</td>
<td>3,500</td>
<td>&gt;5,000</td>
<td>131</td>
<td>&gt;5,000</td>
</tr>
<tr>
<td>250</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-OCH₃-Ph</td>
<td>2,000</td>
<td>&gt;5,000</td>
<td>63</td>
<td>3,000</td>
</tr>
<tr>
<td>251</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>2-OCH₃-Ph</td>
<td>3,300</td>
<td>&gt;5,000</td>
<td>195</td>
<td>4,900</td>
</tr>
<tr>
<td>252</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>76</td>
<td>290</td>
<td>11</td>
<td>170</td>
</tr>
<tr>
<td>253</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>F</td>
<td>160</td>
<td>400</td>
<td>14</td>
<td>140</td>
</tr>
<tr>
<td>254</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>H</td>
<td>H</td>
<td>5,600</td>
<td>3,800</td>
<td>37</td>
<td>1,500</td>
</tr>
<tr>
<td>255</td>
<td>-</td>
<td>n-C₅H₁¹</td>
<td>H</td>
<td>H</td>
<td>5,100</td>
<td>3,400</td>
<td>34</td>
<td>110</td>
</tr>
<tr>
<td>256</td>
<td>-</td>
<td>i-C₅H₁³</td>
<td>H</td>
<td>H</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>13</td>
<td>1,100</td>
</tr>
<tr>
<td>257</td>
<td>-</td>
<td>cyclopropyl methyl</td>
<td>H</td>
<td>H</td>
<td>3,250</td>
<td>2,700</td>
<td>6</td>
<td>180</td>
</tr>
<tr>
<td>258</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>H</td>
<td>F</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>73</td>
<td>2,117</td>
</tr>
</tbody>
</table>
**MODELS 1, 2, 3, training set: continued**

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. Kᵢ (nM)</th>
<th>Exp. Kᵢ (nM)</th>
<th>Exp. Kᵢ (IC₅₀ nM)</th>
<th>Exp. Kᵢ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>259</td>
<td>-</td>
<td>CH₂-CH₃</td>
<td>H</td>
<td>CF₃</td>
<td>6,000</td>
<td>&gt;5,000</td>
<td>6</td>
<td>9,000</td>
</tr>
<tr>
<td>260</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>CH₂</td>
<td>1,800</td>
<td>1,730</td>
<td>11</td>
<td>160</td>
</tr>
<tr>
<td>261</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>CF₃</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>8</td>
<td>700</td>
</tr>
<tr>
<td>262</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>H</td>
<td>F</td>
<td>4,200</td>
<td>&gt;5,000</td>
<td>18</td>
<td>270</td>
</tr>
<tr>
<td>263</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>H</td>
<td>CH₂</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>30</td>
<td>&gt;9,000</td>
</tr>
<tr>
<td>264</td>
<td>-</td>
<td>i-C₃H₉</td>
<td>H</td>
<td>CF₃</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>24</td>
<td>1,600</td>
</tr>
<tr>
<td>265</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>4-Cl-Ph</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>7</td>
<td>&gt;9,000</td>
</tr>
<tr>
<td>266</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>4-CF₃-Ph</td>
<td>1,400</td>
<td>&gt;5,000</td>
<td>15</td>
<td>&gt;9,000</td>
</tr>
<tr>
<td>267</td>
<td>-</td>
<td>cyclopropyl methyl</td>
<td>H</td>
<td>4-Cl-Ph</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>48</td>
<td>&gt;9,000</td>
</tr>
<tr>
<td>268</td>
<td>-</td>
<td>i-C₃H₉</td>
<td>H</td>
<td>Ph</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>12</td>
<td>&gt;9,000</td>
</tr>
<tr>
<td>269</td>
<td>-</td>
<td>cyclopropyl methyl</td>
<td>H</td>
<td>4-CF₃-Ph</td>
<td>&gt;6,000</td>
<td>&gt;5,000</td>
<td>15</td>
<td>&gt;9,000</td>
</tr>
<tr>
<td>270</td>
<td>-</td>
<td>CH₃</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>175</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>271</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>58</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>272</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>235</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>Exp. $K_i$ (nM)</td>
<td>Exp. $K_i$ (nM)</td>
<td>Exp. $K_i$ (nM)</td>
<td>Exp. $K_i$ (nM)</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>273</td>
<td></td>
<td>n-C₃H₇</td>
<td>H</td>
<td>65</td>
<td>&gt;1,000</td>
<td>9</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>274</td>
<td></td>
<td>n-C₃H₇</td>
<td>CH₃</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>516</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>275</td>
<td></td>
<td>n-C₃H₇</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>1,000</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>276</td>
<td></td>
<td>n-C₃H₇</td>
<td>H</td>
<td>548</td>
<td>&gt;1,000</td>
<td>2,065</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>277</td>
<td></td>
<td>n-C₃H₇</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>1,000</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td></td>
<td>n-C₃H₇</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>1,000</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>279</td>
<td></td>
<td>H</td>
<td>H</td>
<td>900</td>
<td>&gt;1,000</td>
<td>35</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>280</td>
<td></td>
<td>OCH₃</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>96</td>
<td>&gt;1,000</td>
<td></td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R&lt;sub&gt;1&lt;/sub&gt;</td>
<td>R&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM)</td>
<td>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM)</td>
<td>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM)</td>
<td>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM)</td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>281</td>
<td></td>
<td>H</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;CH(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>282</td>
<td></td>
<td>H</td>
<td>NO&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>78</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>283</td>
<td></td>
<td>H</td>
<td>OH</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>103</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>284</td>
<td>-</td>
<td>H</td>
<td>F</td>
<td>H</td>
<td>200</td>
<td>&gt;1,000</td>
<td>88</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>285</td>
<td>-</td>
<td>H</td>
<td>OCH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>850</td>
<td>&gt;1,000</td>
<td>100</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>286</td>
<td>-</td>
<td>Cl</td>
<td>H</td>
<td>H</td>
<td>4,481</td>
<td>&gt;1,000</td>
<td>160</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>287</td>
<td>-</td>
<td>F</td>
<td>H</td>
<td>H</td>
<td>3,227</td>
<td>&gt;1,000</td>
<td>50</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>288</td>
<td>-</td>
<td>H</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>1,628</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>289</td>
<td>-</td>
<td>H</td>
<td>Cl</td>
<td>H</td>
<td>520</td>
<td>&gt;1,000</td>
<td>28</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>290</td>
<td>-</td>
<td>H</td>
<td>OCH&lt;sub&gt;2&lt;/sub&gt;-m-CF&lt;sub&gt;3&lt;/sub&gt;-C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>H</td>
<td>100</td>
<td>&gt;1,000</td>
<td>90</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>291</td>
<td>-</td>
<td>H</td>
<td>OCH&lt;sub&gt;2&lt;/sub&gt;-p-NO&lt;sub&gt;2&lt;/sub&gt;-C&lt;sub&gt;6&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>H</td>
<td>163</td>
<td>&gt;1,000</td>
<td>111</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>292</td>
<td>-</td>
<td>H</td>
<td>CF&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>746</td>
<td>&gt;1,000</td>
<td>130</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>293</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>OCH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>OCH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>294</td>
<td>-</td>
<td>-</td>
<td>OCH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>H</td>
<td>566</td>
<td>&gt;1,000</td>
<td>18</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>295</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>569</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>296</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>619</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>297</td>
<td>-</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>OCH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-</td>
<td>1,725</td>
<td>&gt;1,000</td>
<td>95</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>298</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55</td>
<td>&gt;1,000</td>
<td>34</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>299</td>
<td>-</td>
<td>H</td>
<td>N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>2,410</td>
<td>&gt;1,000</td>
<td>50</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. $K_i$ (nM)</td>
<td>$hA_1$R or hA₂₄R (% displ.) 10 µM</td>
<td>Exp. $K_i$ (nM)</td>
<td>$hA_2$₂₄R or $IC_{50}$ nM</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>300</td>
<td>-</td>
<td>OCH₃</td>
<td>H</td>
<td>-</td>
<td>1,993</td>
<td>&gt;1,000</td>
<td>90</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>301</td>
<td>-</td>
<td>Cl</td>
<td>H</td>
<td>-</td>
<td>1,361</td>
<td>&gt;1,000</td>
<td>110</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>302</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-F-Ph</td>
<td>-</td>
<td>65</td>
<td>&gt;1,000</td>
<td>12</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>303</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-Br-Ph</td>
<td>-</td>
<td>150</td>
<td>&gt;1,000</td>
<td>20</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>304</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>3,4-[methyleneoxy]phenyl</td>
<td>-</td>
<td>200</td>
<td>&gt;1,000</td>
<td>55</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>305</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>5-methylpyridin-2-yl</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>1,012</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>306</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-&lt;sup&gt;i&lt;/sup&gt;-Bu-phenyl</td>
<td>-</td>
<td>1,005</td>
<td>&gt;1,000</td>
<td>74</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>307</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-tolyl</td>
<td>-</td>
<td>79</td>
<td>&gt;1,000</td>
<td>19</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>308</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>pyridin-4-yl</td>
<td>-</td>
<td>955</td>
<td>&gt;1,000</td>
<td>41</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>309</td>
<td>-</td>
<td>&lt;i&gt;i&lt;/i&gt;C₄H₉</td>
<td>4-F-Ph</td>
<td>-</td>
<td>467</td>
<td>&gt;1,000</td>
<td>303</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>310</td>
<td>-</td>
<td>&lt;i&gt;i&lt;/i&gt;C₄H₉</td>
<td>4-Br-Ph</td>
<td>-</td>
<td>2,427</td>
<td>&gt;1,000</td>
<td>132</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>311</td>
<td>-</td>
<td>n-C₆H₆</td>
<td>4-F-Ph</td>
<td>-</td>
<td>181</td>
<td>&gt;1,000</td>
<td>185</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>312</td>
<td>-</td>
<td>&lt;i&gt;i&lt;/i&gt;C₄H₉</td>
<td>4-Br-Ph</td>
<td>-</td>
<td>49</td>
<td>&gt;1,000</td>
<td>66</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>313</td>
<td>-</td>
<td>&lt;i&gt;i&lt;/i&gt;C₄H₉</td>
<td>4-F-Ph</td>
<td>-</td>
<td>72</td>
<td>&gt;1,000</td>
<td>207</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>314</td>
<td>-</td>
<td>n-C₅H₅</td>
<td>CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>122</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>315</td>
<td>-</td>
<td>H</td>
<td>F</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>70</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>316</td>
<td>-</td>
<td>H</td>
<td>OCH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>53</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>317</td>
<td>-</td>
<td>4-OCH₃-Ph</td>
<td>-</td>
<td>-</td>
<td>1,221</td>
<td>21.2</td>
<td>&gt;10,000</td>
<td>1,760</td>
</tr>
<tr>
<td>318</td>
<td>-</td>
<td>4-Cl-Ph</td>
<td>-</td>
<td>-</td>
<td>187</td>
<td>67.3</td>
<td>&gt;10,000</td>
<td>1,490</td>
</tr>
</tbody>
</table>
E.2 MODELS 1, 2, 3, validation set.
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. Ki (nM) hA₁R or (% disp.)</th>
<th>Exp. Ki (nM) hA₂R or (% disp.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>319</td>
<td>2</td>
<td>SO₃H</td>
<td>H</td>
<td>-</td>
<td>190</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>320</td>
<td>2</td>
<td>SO₃NH₂</td>
<td>H</td>
<td>-</td>
<td>2,630</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>321</td>
<td>2</td>
<td>SO₃NHCH₂COOH</td>
<td>H</td>
<td>-</td>
<td>9,320</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>322</td>
<td>-</td>
<td>CH₃</td>
<td>4-CF₃-Ph-NH-CO</td>
<td>-</td>
<td>750</td>
<td>286</td>
</tr>
<tr>
<td>323</td>
<td>-</td>
<td>CH₃</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-</td>
<td>430</td>
<td>128</td>
</tr>
<tr>
<td>324</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-NO₂-Ph-NH-CO</td>
<td>-</td>
<td>1,134</td>
<td>228</td>
</tr>
<tr>
<td>325</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-CH₃-Ph-NH-CO</td>
<td>-</td>
<td>262</td>
<td>132</td>
</tr>
<tr>
<td>326</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>4-CF₃-Ph-NH-CO</td>
<td>-</td>
<td>600</td>
<td>101</td>
</tr>
<tr>
<td>327</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>Ph-CH₂-CO</td>
<td>-</td>
<td>326</td>
<td>88</td>
</tr>
<tr>
<td>328</td>
<td>-</td>
<td>C₂H₅</td>
<td>Ph-NH-CO</td>
<td>-</td>
<td>1,006</td>
<td>245</td>
</tr>
<tr>
<td>329</td>
<td>-</td>
<td>C₂H₅</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>-</td>
<td>5,000</td>
<td>65</td>
</tr>
<tr>
<td>330</td>
<td>-</td>
<td>C₂H₅</td>
<td>H</td>
<td>-</td>
<td>5,000</td>
<td>65</td>
</tr>
<tr>
<td>331</td>
<td>-</td>
<td>2-(α-naphthyl)ethyl</td>
<td>H</td>
<td>-</td>
<td>348</td>
<td>4,294</td>
</tr>
<tr>
<td>332</td>
<td>-</td>
<td>4-C₄H₉</td>
<td>H</td>
<td>-</td>
<td>600</td>
<td>400</td>
</tr>
<tr>
<td>333</td>
<td>-</td>
<td>3-Cl-Ph-NH-CO</td>
<td>H</td>
<td>-</td>
<td>228</td>
<td>70</td>
</tr>
<tr>
<td>334</td>
<td>-</td>
<td>C₃H₁₁-C₆H₅-C₂H₂</td>
<td>H</td>
<td>-</td>
<td>510</td>
<td>4,294</td>
</tr>
<tr>
<td>335</td>
<td>-</td>
<td>CH₂-C₆H₅-C₂H₂</td>
<td>CO-NH-C(CH₃)₃</td>
<td>-</td>
<td>500</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>336</td>
<td>-</td>
<td>Ph-C₂H₅-C₂H₂-C₂H₂</td>
<td>CO-NH-C(CH₃)₃</td>
<td>-</td>
<td>875</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>337</td>
<td>-</td>
<td>CH₂-C₂H₅-C₂H₃</td>
<td>Ph-C₂H₅-CO</td>
<td>-</td>
<td>802</td>
<td>157</td>
</tr>
<tr>
<td>338</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>α-naphthyl-CH₂</td>
<td>-</td>
<td>598</td>
<td>74</td>
</tr>
<tr>
<td>339</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>β-naphthyl-CH₂</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>340</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>341</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>342</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>343</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>344</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>345</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>346</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>347</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>348</td>
<td>-</td>
<td>CH₂-C₂H₅</td>
<td>H</td>
<td>-</td>
<td>185</td>
<td>18.4</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. $K_i$ (nM)</td>
<td>Exp. $hA_{1}R$ or (% displ. 10 nM)</td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>--------------</td>
<td>-------------------</td>
<td>------</td>
<td>-----------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>340</td>
<td>CH₂-CH₂-CH₃</td>
<td>α-naphthyl-CH₂</td>
<td>-</td>
<td>388</td>
<td>38.7</td>
<td>(12,400)</td>
</tr>
<tr>
<td>341</td>
<td>Ph-CH₂-CH₂</td>
<td>3-Cl-Ph-CH₂-CO</td>
<td>-</td>
<td>13.2</td>
<td>5.75</td>
<td>(9,100)</td>
</tr>
<tr>
<td>342</td>
<td>CH₃</td>
<td>4-F-Ph-CH₂-CO</td>
<td>-</td>
<td>74.1</td>
<td>54.1</td>
<td>(13,000)</td>
</tr>
<tr>
<td>343</td>
<td>Ph-CH₂-CH₂-CH₂</td>
<td>4-Cl-Ph-CH₂-CO</td>
<td>-</td>
<td>71.1</td>
<td>27.2</td>
<td>(&gt;10,000)</td>
</tr>
<tr>
<td>344</td>
<td>CH₂-CH₂-CH(CH₃)₂</td>
<td>2-CH₃-Ph-CH₂O-Ph-CH₂-CO</td>
<td>-</td>
<td>9066</td>
<td>52.3</td>
<td>(&gt;30,000)</td>
</tr>
<tr>
<td>345</td>
<td>CH₂-CH₂-CH(CH₃)₂</td>
<td>3-CH₃-Ph-CH₂O-Ph-CH₂-CO</td>
<td>-</td>
<td>564</td>
<td>29.9</td>
<td>(&gt;30,000)</td>
</tr>
<tr>
<td>346</td>
<td>Ph-CH₂-CH₂</td>
<td>3,4-CH₃-Ph-CH₂O-Ph-CH₂-CO</td>
<td>-</td>
<td>260</td>
<td>100</td>
<td>(&gt;10,000)</td>
</tr>
<tr>
<td>347</td>
<td>Ph-CH₂-CH₂-CH₂</td>
<td>3-CH₃-Ph-CH₂O-Ph-CH₂-CO</td>
<td>-</td>
<td>582</td>
<td>97.3</td>
<td>(&gt;10,000)</td>
</tr>
<tr>
<td>348</td>
<td>Ph-CH₂-CH₂-CH₂</td>
<td>3,5-CH₃-Ph-CH₂O-Ph-CH₂-CO</td>
<td>-</td>
<td>2,190</td>
<td>248</td>
<td>(&gt;10,000)</td>
</tr>
<tr>
<td>349</td>
<td>CO-CH-Ph₂</td>
<td>OPh</td>
<td>-</td>
<td>&lt;10% inhibit</td>
<td>1,860</td>
<td>(&gt;100,000)</td>
</tr>
<tr>
<td>350</td>
<td>CO-CH₂-Ph</td>
<td>OPh</td>
<td>-</td>
<td>7,260</td>
<td>136</td>
<td>(&gt;100,000)</td>
</tr>
<tr>
<td>351</td>
<td>4-OCH₃-Ph-CH₂-CO</td>
<td>OPh</td>
<td>-</td>
<td>&lt;12% inhibit</td>
<td>882</td>
<td>(&gt;100,000)</td>
</tr>
<tr>
<td>352</td>
<td>CO-NH-Ph</td>
<td>OPh</td>
<td>-</td>
<td>&lt;17% inhibit</td>
<td>38.9</td>
<td>(8,870)</td>
</tr>
<tr>
<td>353</td>
<td>H</td>
<td>-</td>
<td>107</td>
<td>(&gt;100,000)</td>
<td>107</td>
<td>(100,000)</td>
</tr>
<tr>
<td>354</td>
<td>H</td>
<td>-</td>
<td>2,140</td>
<td>(&gt;100,000)</td>
<td>2,140</td>
<td>(100,000)</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. Kᵢ (nM) hₐ₁R or (% displ. 10 µM)</td>
<td>Exp. Kᵢ (nM) hₐ₂₄R</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>355</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
<td>192</td>
</tr>
<tr>
<td>356</td>
<td>-</td>
<td>Ph-CH₂-CH₂-CH₂</td>
<td>H</td>
<td>SCH₂</td>
<td>1.75</td>
<td>22</td>
</tr>
<tr>
<td>357</td>
<td>-</td>
<td>N-methylpipерazine</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>358</td>
<td>-</td>
<td>morpholine</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>359</td>
<td>-</td>
<td>CH₃</td>
<td>Ph-NH-CO-CH₂</td>
<td>-</td>
<td>3.50</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>360</td>
<td>-</td>
<td>CH₂</td>
<td>4-tolyl-NH-CO-CH₂</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>361</td>
<td>-</td>
<td>CH₂</td>
<td>4-OCH₃-Ph-NH-CO-CH₂</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>362</td>
<td>-</td>
<td>propyl</td>
<td>Ph</td>
<td>cyclopentyl</td>
<td>7.1</td>
<td>1,200</td>
</tr>
<tr>
<td>363</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-OH</td>
<td>CH₂-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>364</td>
<td>-</td>
<td>CH₂</td>
<td>CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>365</td>
<td>-</td>
<td>H</td>
<td>Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>366</td>
<td>-</td>
<td>H</td>
<td>Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>367</td>
<td>-</td>
<td>4-OCH₃-Ph</td>
<td>-</td>
<td>-</td>
<td>3,000</td>
<td>&gt;5,000</td>
</tr>
<tr>
<td>368</td>
<td>-</td>
<td>3-CH₃-Ph</td>
<td>-</td>
<td>-</td>
<td>1,400</td>
<td>&gt;5,000</td>
</tr>
<tr>
<td>369</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>2-Cl-Ph</td>
<td>630</td>
<td>2,500</td>
</tr>
<tr>
<td>370</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>4-CN-Ph</td>
<td>570</td>
<td>&gt;5,000</td>
</tr>
<tr>
<td>371</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>2-CF₃-Ph</td>
<td>1,900</td>
<td>&gt;5,000</td>
</tr>
<tr>
<td>372</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>CH₂-CH₂-CH₃</td>
<td>3-Cl-Ph</td>
<td>4,600</td>
<td>&gt;5,000</td>
</tr>
</tbody>
</table>
### Molecule Id, n, R, R1, R2, \( K_i \), Exp. \( K_i \), \( \text{Exp. (nM)} \), hA, hB, Exp. hA, \% displ. (nM), \% displ. hA, hB

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R1</th>
<th>R2</th>
<th>Exp. ( K_i ) (nM)</th>
<th>hA</th>
<th>hB</th>
<th>Exp. hA</th>
<th>% displ. (nM)</th>
<th>% displ. hA, hB</th>
</tr>
</thead>
<tbody>
<tr>
<td>373</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>13</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>374</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>28</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>375</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>F</td>
<td>&gt;5,000</td>
<td>28</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>376</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>14</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>377</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>28</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>378</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>28</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>379</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>14</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>380</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>28</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>381</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>14</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>382</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>28</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>383</td>
<td></td>
<td></td>
<td>CH₂-CH₂-CH₃</td>
<td>H</td>
<td>&gt;5,000</td>
<td>14</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
</tbody>
</table>

**Note:** The data represents experimental values for various molecules with different substitutions at positions \( R_1 \) and \( R_2 \), along with their respective \( K_i \) values and displacements at \( hA \) and \( hB \).
E.3  MODELS 1, 2, 3, internal test set.
<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. $K_i$ (nM) hA₁R or (% displ. 10 µM)</th>
<th>Exp. $K_i$ (nM) hA₂₄R or (% displ. 10 µM)</th>
<th>Exp. $K_i$ (nM) hA₂₅R or (% displ. 10 µM)</th>
<th>Exp. $K_i$ (nM) hA₃R or (% displ. 10 µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>384</td>
<td>3</td>
<td>CH₂OH</td>
<td>CH₂OH</td>
<td>H</td>
<td>432</td>
<td>0.19</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>385</td>
<td>3</td>
<td>CN</td>
<td>H</td>
<td>-</td>
<td>6,496</td>
<td>86</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>386</td>
<td>3</td>
<td>COOEt</td>
<td>H</td>
<td>-</td>
<td>4,941</td>
<td>4.0</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>387</td>
<td>3</td>
<td>C(NOH)NH₂</td>
<td>H</td>
<td>-</td>
<td>60</td>
<td>6.0</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>388</td>
<td>3</td>
<td>(NH)NH₂</td>
<td>H</td>
<td>-</td>
<td>4,965</td>
<td>4.4</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>389</td>
<td>2</td>
<td>H</td>
<td>H</td>
<td>-</td>
<td>549</td>
<td>1.1</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>390</td>
<td></td>
<td>OCH₃</td>
<td>OCH₃</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>38</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>391</td>
<td></td>
<td>H</td>
<td>OCH₂-O-CF₃-Ph</td>
<td>H</td>
<td>56</td>
<td>&gt;1,000</td>
<td>13</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>392</td>
<td></td>
<td>F</td>
<td>F</td>
<td>H</td>
<td>1,808</td>
<td>&gt;1,000</td>
<td>78</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>393</td>
<td></td>
<td>OCH₃</td>
<td>OH</td>
<td>H</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>342</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>394</td>
<td></td>
<td>Cl</td>
<td>H</td>
<td>-</td>
<td>448</td>
<td>&gt;1,000</td>
<td>39</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>395</td>
<td></td>
<td>H</td>
<td>OCH₃</td>
<td>-</td>
<td>1,440</td>
<td>&gt;1,000</td>
<td>81</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>396</td>
<td></td>
<td>OCH₃</td>
<td>H</td>
<td>-</td>
<td>1,175</td>
<td>&gt;1,000</td>
<td>58</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>397</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>398</td>
<td></td>
<td>n-C₃H₇</td>
<td>4-(ethoxycarbonyl)phenyl</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>32</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>399</td>
<td></td>
<td>n-C₃H₇</td>
<td>3,4-C₂₂-Ph</td>
<td>-</td>
<td>300</td>
<td>&gt;1,000</td>
<td>16</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>400</td>
<td></td>
<td>n-C₃H₇</td>
<td>3,4-OCH₃-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>12</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>401</td>
<td></td>
<td>n-C₃H₇</td>
<td>4-Br-Ph</td>
<td>-</td>
<td>168</td>
<td>&gt;1,000</td>
<td>93</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>402</td>
<td></td>
<td>n-C₃H₇</td>
<td>Ph</td>
<td>-</td>
<td>250</td>
<td>&gt;1,000</td>
<td>15</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>403</td>
<td></td>
<td>n-C₃H₇</td>
<td>4-F-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>55</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>404</td>
<td></td>
<td>CH₂CH=CH₂</td>
<td>Ph</td>
<td>-</td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td>24</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>Mol ld</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. Kᵢ (nM) hₐ₁R or (% displ. 10 µM)</td>
<td>Exp. Kᵢ (nM) hₐ₂₄R or (% displ. 10 µM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>------------</td>
<td>------------</td>
<td>----</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>-</td>
<td>OCH₂O</td>
<td>-</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>406</td>
<td>-</td>
<td>OCH₃</td>
<td>-</td>
<td></td>
<td>&gt;1,000</td>
<td>&gt;1,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>407</td>
<td>-</td>
<td>-</td>
<td>Ph</td>
<td></td>
<td>370</td>
<td>1,100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>408</td>
<td>-</td>
<td>-</td>
<td>4-Cl-Ph</td>
<td></td>
<td>740</td>
<td>1,800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>409</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>n-C₃H₇</td>
<td>3-OCH₃-Ph</td>
<td>810</td>
<td>&gt;5,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>n-C₃H₇</td>
<td>2-Cl-Ph</td>
<td>&gt;6,000</td>
<td>2,700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>411</td>
<td>-</td>
<td>n-C₃H₇</td>
<td>n-C₃H₇</td>
<td>2-CF₃-Ph</td>
<td>3,700</td>
<td>5,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>412</td>
<td>-</td>
<td>CH₂-CH₂-CH₃-CH₃</td>
<td>Ph₂-CH₂-CO</td>
<td></td>
<td>129</td>
<td>114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>413</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-CH₃</td>
<td>Ph-Ph₂-CH₂-CO</td>
<td></td>
<td>381</td>
<td>30,6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>414</td>
<td>-</td>
<td>CH₃</td>
<td>3-Cl-Ph₂-CH₂-CO</td>
<td></td>
<td>410</td>
<td>30,5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>415</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>Ph-Ph₂-CH₂-CO</td>
<td></td>
<td>622</td>
<td>52,5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>-</td>
<td>CH₃</td>
<td>4-Cl-Ph₂-CH₂-CO</td>
<td></td>
<td>1,850</td>
<td>156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>417</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-CH₃</td>
<td>α-naphthyl-CH₂-CO</td>
<td></td>
<td>170</td>
<td>11,7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>418</td>
<td>-</td>
<td>CH₃</td>
<td>4-CF₃-Ph₂-CH₂-CO</td>
<td></td>
<td>768</td>
<td>75,9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>419</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-CH₃</td>
<td>2-thienyl-CH₂-CO</td>
<td></td>
<td>70,1</td>
<td>4,15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>-</td>
<td>CH₂-CH₂-CH(CH₃)₂</td>
<td>3-Cl-Ph₂-CH₂-CO</td>
<td></td>
<td>13,7</td>
<td>1,86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>421</td>
<td>-</td>
<td>Ph-CH₂-CH₂-CH₂</td>
<td>α-naphthyl-CH₂-CO</td>
<td></td>
<td>241</td>
<td>73,6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>422</td>
<td>-</td>
<td>Ph-CH₂</td>
<td>Ph-CH₂-CH₂-CO</td>
<td></td>
<td>128</td>
<td>43,9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>423</td>
<td>-</td>
<td>CH₃</td>
<td>3,4-Cl₂-Ph-NH-CO</td>
<td></td>
<td>392</td>
<td>143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>424</td>
<td>-</td>
<td>CH₃</td>
<td>-</td>
<td></td>
<td>392</td>
<td>143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix E

**Models 1, 2, 3, internal test set: continued**

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>R</th>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>R&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM) hA&lt;sub&gt;2&lt;/sub&gt;R or (% displ. 10 μM)</th>
<th>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM) hA&lt;sub&gt;2&lt;/sub&gt;R or (% displ. 10 μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>425</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-F-Ph-NH-CO</td>
<td>-</td>
<td>700</td>
<td>120</td>
</tr>
<tr>
<td>426</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>2-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>300</td>
<td>133</td>
</tr>
<tr>
<td>427</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>3-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>140</td>
<td>215</td>
</tr>
<tr>
<td>428</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-CH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>400</td>
<td>12</td>
</tr>
<tr>
<td>429</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-Br-Ph-NH-CO</td>
<td>-</td>
<td>350</td>
<td>56</td>
</tr>
<tr>
<td>430</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-F-Ph-NH-CO</td>
<td>-</td>
<td>376</td>
<td>42</td>
</tr>
<tr>
<td>431</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>250</td>
<td>100</td>
</tr>
<tr>
<td>432</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>3-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>275</td>
<td>113</td>
</tr>
<tr>
<td>433</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>4-Cl-Ph-NH-CO</td>
<td>-</td>
<td>300</td>
<td>111</td>
</tr>
<tr>
<td>434</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>594</td>
<td>381</td>
</tr>
<tr>
<td>435</td>
<td>C&lt;sub&gt;2&lt;/sub&gt;H&lt;sub&gt;5&lt;/sub&gt;</td>
<td>4-SO&lt;sub&gt;3&lt;/sub&gt;H-Ph-NH-CO</td>
<td>-</td>
<td>&gt;10,000</td>
<td>249</td>
</tr>
<tr>
<td>436</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>3-Cl-Ph-NH-CO</td>
<td>-</td>
<td>30</td>
<td>1,220</td>
</tr>
<tr>
<td>437</td>
<td>n-C&lt;sub&gt;3&lt;/sub&gt;H&lt;sub&gt;7&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>10</td>
<td>2.51</td>
</tr>
<tr>
<td>438</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>201</td>
<td>120</td>
</tr>
<tr>
<td>439</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>251</td>
<td>1,010</td>
</tr>
<tr>
<td>440</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH=CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>1,531</td>
<td>176</td>
</tr>
<tr>
<td>441</td>
<td>H</td>
<td>4-OCH&lt;sub&gt;3&lt;/sub&gt;-Ph-NH-CO</td>
<td>-</td>
<td>796</td>
<td>545</td>
</tr>
<tr>
<td>442</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>10</td>
<td>2.8</td>
</tr>
<tr>
<td>443</td>
<td>(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;C=CH-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>H</td>
<td>-</td>
<td>5.42</td>
<td>0.8</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. Kᵢ (nM) hA₁R or (% displ. 10 µM)</td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>444</td>
<td>-</td>
<td>4-OCH₂CONHPh-4'-I-Ph</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>445</td>
<td>-</td>
<td>CH₃</td>
<td>H</td>
<td>SCH₃</td>
<td>8.4 1.2 10.3 35</td>
</tr>
<tr>
<td>446</td>
<td>-</td>
<td>CH₃</td>
<td>4-OCH₃-Ph-NH-CO</td>
<td>NH-CH₂CH₃</td>
<td>21 15 10.3 36</td>
</tr>
<tr>
<td>447</td>
<td>-</td>
<td>4-OH-Ph</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>448</td>
<td>-</td>
<td>4-ClPh</td>
<td>-</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>449</td>
<td>-</td>
<td>CH₂COCH₃</td>
<td>CH₂-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>450</td>
<td>-</td>
<td>CH₂CO₂Et</td>
<td>CH₂-Ph</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>451</td>
<td>-</td>
<td>CH₂CH₃</td>
<td>CH₂CH₃</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>452</td>
<td>-</td>
<td>H</td>
<td>C(CH₃)₃</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>453</td>
<td>-</td>
<td>H</td>
<td>cyclopropyl</td>
<td>-</td>
<td>350 &gt;1,000</td>
</tr>
<tr>
<td>454</td>
<td>-</td>
<td>cyclopropyl methyl</td>
<td>H</td>
<td>CF₃</td>
<td>&gt;6,000</td>
</tr>
<tr>
<td>455</td>
<td>-</td>
<td>propyl</td>
<td>H</td>
<td>Ph</td>
<td>350 &gt;5,000</td>
</tr>
<tr>
<td>456</td>
<td>-</td>
<td>n-C₄H₉</td>
<td>H</td>
<td>4-Cl-Ph</td>
<td>&gt;6,000</td>
</tr>
<tr>
<td>457</td>
<td>-</td>
<td>i-C₄H₉</td>
<td>H</td>
<td>4-CF₃-Ph</td>
<td>&gt;6,000</td>
</tr>
<tr>
<td>458</td>
<td>-</td>
<td>cyclopropyl methyl</td>
<td>H</td>
<td>4-CF₃-Ph</td>
<td>&gt;6,000</td>
</tr>
<tr>
<td>459</td>
<td>-</td>
<td>propyl</td>
<td>H</td>
<td>Ph</td>
<td>&gt;6,000</td>
</tr>
<tr>
<td>460</td>
<td>-</td>
<td>c-C₇H₁₁-CH₂CH₂</td>
<td>H</td>
<td>-</td>
<td>0.70 20.2 55.8 39</td>
</tr>
<tr>
<td>461</td>
<td>-</td>
<td>(CH₃)₂CH=CH₂-CH₂</td>
<td>CO-NH-C(CH₃)₃</td>
<td>-</td>
<td>523 &gt;10,000  &gt;10,000</td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R&lt;sub&gt;1&lt;/sub&gt;</td>
<td>R&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Exp. K&lt;sub&gt;i&lt;/sub&gt; (nM) hA&lt;sub&gt;A&lt;/sub&gt;R or (% displ. 10 µM)</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>----------------</td>
<td>----------------------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>462</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>CO-NH-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-</td>
<td>357</td>
</tr>
<tr>
<td>463</td>
<td>-</td>
<td>propyl</td>
<td>3-noradamantyl</td>
<td>-</td>
<td>500.6</td>
</tr>
<tr>
<td>464</td>
<td>-</td>
<td>propyl</td>
<td>(R)-1-phenylethyl</td>
<td>biphenyl</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>465</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-N(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>466</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>675</td>
</tr>
<tr>
<td>467</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3,4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>468</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3,4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>469</td>
<td>-</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>&gt;1,000</td>
</tr>
<tr>
<td>470</td>
<td>-</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-NH-CO-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
<td>140</td>
</tr>
<tr>
<td>471</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>2-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>2,910</td>
</tr>
<tr>
<td>472</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>3-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>6,870</td>
</tr>
<tr>
<td>473</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>6,760</td>
</tr>
<tr>
<td>474</td>
<td>-</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>2,4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>4,350</td>
</tr>
<tr>
<td>475</td>
<td>-</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>501</td>
</tr>
<tr>
<td>476</td>
<td>-</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2,4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>814</td>
</tr>
<tr>
<td>477</td>
<td>-</td>
<td>CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;(CH&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2,4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>6882</td>
</tr>
<tr>
<td>478</td>
<td>-</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>261</td>
</tr>
<tr>
<td>479</td>
<td>-</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>3-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>240</td>
</tr>
<tr>
<td>480</td>
<td>-</td>
<td>Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>4-Ch&lt;sub&gt;2&lt;/sub&gt;-Ph-CH&lt;sub&gt;2&lt;/sub&gt;O-Ph-CH&lt;sub&gt;2&lt;/sub&gt;-CO</td>
<td>-</td>
<td>360</td>
</tr>
</tbody>
</table>
### MODEL s 1, 2, 3, internal test set: continued

<table>
<thead>
<tr>
<th>Mol Id</th>
<th>n</th>
<th>R</th>
<th>R₁</th>
<th>R₂</th>
<th>Exp. $K_i$ (nM)</th>
<th>hₐ₂₅R or (% displ. 10 µM)</th>
<th>Exp. $K_i$ (nM)</th>
<th>hₐ₃₅R or (% displ. 10 µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>481</td>
<td>-</td>
<td>Ph-CH₂-CH₂</td>
<td>2₅₄-CH₃-Ph-CH₂-O-Ph-CH₂-CO</td>
<td>-</td>
<td>278</td>
<td>155</td>
<td>&gt;10,000</td>
<td>98.0</td>
</tr>
<tr>
<td>482</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>1,754</td>
<td>532</td>
<td>(100,000)</td>
<td>(29.8)</td>
</tr>
<tr>
<td>483</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>895</td>
<td>213</td>
<td>(7,140)</td>
<td>(75.5)</td>
</tr>
<tr>
<td>484</td>
<td>-</td>
<td>H</td>
<td>NH₂(CH₂)₄CH₂-NHBOc</td>
<td>-</td>
<td>3,559</td>
<td>447</td>
<td>(&gt;10,000)</td>
<td>(68.6)</td>
</tr>
<tr>
<td>485</td>
<td>-</td>
<td>H</td>
<td>NH₂(CH₂CH₂O)₂CH₂-CH₂-NHBOc</td>
<td>-</td>
<td>696</td>
<td>184</td>
<td>(30,000)</td>
<td>(74.0)</td>
</tr>
<tr>
<td>486</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>53</td>
<td>66</td>
<td>(10,000)</td>
<td>(49.4)</td>
</tr>
<tr>
<td>487</td>
<td>-</td>
<td>H</td>
<td>NH₂(CH₂CH₂O)₂CH₂-CH₂-NH₃⁺</td>
<td>-</td>
<td>&gt;10,000</td>
<td>64</td>
<td>(&gt;30,000)</td>
<td>(56.2)</td>
</tr>
<tr>
<td>488</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>87</td>
<td>173</td>
<td>(&gt;30,000)</td>
<td>1,240</td>
</tr>
<tr>
<td>489</td>
<td>-</td>
<td>H</td>
<td>-</td>
<td>-</td>
<td>212</td>
<td>71</td>
<td>(100,000)</td>
<td>(56.2)</td>
</tr>
<tr>
<td>490</td>
<td>-</td>
<td>H</td>
<td>NH₂(CH₂)₂-NH₃⁺</td>
<td>-</td>
<td>2,497</td>
<td>2,000</td>
<td>(&gt;100,000)</td>
<td>(64.5)</td>
</tr>
<tr>
<td>491</td>
<td>-</td>
<td>NH₂(CH₂)₂-NH₃⁺</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>(&gt;100,000)</td>
<td>(17.7)</td>
</tr>
<tr>
<td>Mol Id</td>
<td>R</td>
<td>R1</td>
<td>R2</td>
<td>(K_i) (nM)</td>
<td>(h_A) or (%) displ. 10 µM</td>
<td>(K_i) (nM)</td>
<td>(h_A) or (%) displ. 10 µM</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
<td>----------------------------</td>
<td>--------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>492</td>
<td>- H NH-(CH(_2))(_2)-NH(_3^+)</td>
<td>- NH(_2)-NH-(CH(_2))(_2)-NH(_3^+)</td>
<td>H</td>
<td>&gt;10,000</td>
<td>2,200</td>
<td>&gt;10,000</td>
<td>2,200</td>
<td></td>
</tr>
<tr>
<td>493</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>58</td>
<td>&gt;10,000 (2,200)</td>
<td>58</td>
<td>&gt;10,000 (2,200)</td>
<td></td>
</tr>
<tr>
<td>494</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>2,061</td>
<td>983</td>
<td>2,061</td>
<td>983</td>
<td></td>
</tr>
<tr>
<td>495</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>32,000</td>
<td>10,700</td>
<td>32,000</td>
<td>10,700</td>
<td></td>
</tr>
<tr>
<td>496</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>497</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>498</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>499</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>502</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>503</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>505</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>506</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>507</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>508</td>
<td>- H</td>
<td>- H</td>
<td>H</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td>&gt;10,000</td>
<td></td>
</tr>
<tr>
<td>Mol Id</td>
<td>n</td>
<td>R</td>
<td>R₁</td>
<td>R₂</td>
<td>Exp. $K_i$ (nM) $hA_1 R$ or (% displ. 10 µM)</td>
<td>Exp. $K_i$ (nM) $hA_2 B R$ or (% displ. 10 µM)</td>
<td>Exp. $K_i$ (nM) $hA_1 R$ or (% displ. 10 µM)</td>
<td>Exp. $K_i$ (nM) $hA_2 B R$ or (% displ. 10 µM)</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>509</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>CH₃-CH₂-NH-CO</td>
<td>-</td>
<td>77.2</td>
<td>223</td>
<td>(2,810)</td>
<td>0.60</td>
</tr>
<tr>
<td>510</td>
<td>-</td>
<td>CH₂-CH₂-CH₃</td>
<td>cC₆H₁₁-CH₂-NH-CO</td>
<td>-</td>
<td>13.8</td>
<td>258</td>
<td>(3,440)</td>
<td>3.53</td>
</tr>
<tr>
<td>511</td>
<td>-</td>
<td>cC₆H₁₁-CH₂</td>
<td>CH₃-O-CO-CH₂-cC₆H₁₁-NH-CO</td>
<td>-</td>
<td>52.5</td>
<td>5.74</td>
<td>(8,940)</td>
<td>164</td>
</tr>
<tr>
<td>512</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-CH₂-CH₃</td>
<td>CH₃-CH₂-O-CO-CH₂-Ph-NH-CO</td>
<td>-</td>
<td>54.7</td>
<td>4.77</td>
<td>(9,740)</td>
<td>18</td>
</tr>
<tr>
<td>513</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-CH₂-CH₃</td>
<td>CH₃CH₂CH(CH₃)-NH-CO</td>
<td>-</td>
<td>576</td>
<td>300</td>
<td>(&gt;100,000)</td>
<td>5.33</td>
</tr>
<tr>
<td>514</td>
<td>-</td>
<td>CH₂-CH₂-CH₂-CH₂-CH₃</td>
<td>4-CH₃-Ph-NH-CO</td>
<td>-</td>
<td>34.7</td>
<td>3.72</td>
<td>(2,910)</td>
<td>44.4</td>
</tr>
</tbody>
</table>
E.4 References


Appendix E  


Table E.1

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>hA(_1)R</td>
<td>11</td>
<td>12</td>
<td>32</td>
<td>10</td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td>hA(_2,4)R</td>
<td>26</td>
<td>7</td>
<td>27</td>
<td>5</td>
<td>0.84</td>
<td>0.79</td>
</tr>
<tr>
<td>hA(_2,9)R</td>
<td>28</td>
<td>6</td>
<td>27</td>
<td>4</td>
<td>0.87</td>
<td>0.82</td>
</tr>
<tr>
<td>hA(_3)R</td>
<td>29</td>
<td>7</td>
<td>29</td>
<td>2</td>
<td>0.93</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Table E.2

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>hA_{1}R</td>
<td>2</td>
<td>5</td>
<td>48</td>
<td>10</td>
<td>0.17</td>
<td>0.29</td>
</tr>
<tr>
<td>hA_{2,4}R</td>
<td>22</td>
<td>6</td>
<td>29</td>
<td>8</td>
<td>0.73</td>
<td>0.79</td>
</tr>
<tr>
<td>hA_{2,8}R</td>
<td>23</td>
<td>6</td>
<td>29</td>
<td>7</td>
<td>0.77</td>
<td>0.79</td>
</tr>
<tr>
<td>hA_{3}R</td>
<td>22</td>
<td>5</td>
<td>34</td>
<td>4</td>
<td>0.85</td>
<td>0.81</td>
</tr>
</tbody>
</table>
### Table E.3

<table>
<thead>
<tr>
<th>Classes</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>hA₁R</td>
<td>2</td>
<td>1</td>
<td>58</td>
<td>4</td>
<td>0.33</td>
<td>0.67</td>
</tr>
<tr>
<td>hA₂₄R</td>
<td>14</td>
<td>5</td>
<td>39</td>
<td>7</td>
<td>0.67</td>
<td>0.74</td>
</tr>
<tr>
<td>hA₂₈R</td>
<td>12</td>
<td>4</td>
<td>42</td>
<td>7</td>
<td>0.63</td>
<td>0.75</td>
</tr>
<tr>
<td>hA₃R</td>
<td>17</td>
<td>5</td>
<td>40</td>
<td>3</td>
<td>0.85</td>
<td>0.77</td>
</tr>
</tbody>
</table>
### Table E.4

<table>
<thead>
<tr>
<th>Classes</th>
<th>MODEL 1</th>
<th></th>
<th></th>
<th>MODEL 2</th>
<th></th>
<th></th>
<th>MODEL 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td></td>
</tr>
<tr>
<td>hA1R</td>
<td>0.54</td>
<td>1.00</td>
<td>0.18</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>hA24R</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
<td>0.92</td>
<td>0.82</td>
<td>1.00</td>
<td>0.82</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>hA29R</td>
<td>0.50</td>
<td>0.25</td>
<td>1.00</td>
<td>0.50</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>hA3R</td>
<td>1.00</td>
<td>0.67</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.28</td>
<td>1.00</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Accuracy(_{ML}(\alpha=1))</td>
<td>0.77</td>
<td>0.69</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F

Paper VI
Supplementary Information

F.1 TOXclass model, training set (no. refers to the reference database. Experimental and predicted classes: high $AT$, high acute toxicity; low $AT$, low acute toxicity).

F.2 MOA class model, training set (no. refers to the reference database. Experimental and predicted classes: A, baseline narcosis; B, polar narcosis; C, arylate and ester narcosis; D, electrophile or proelectrophile phosphorylation; E, neurodepressant; F, uncoupler of oxidative phosphorylation; G, central nervous system seizure mechanisms; H, AchE inhibition; I, respiratory blocker or inhibition).

F.3 Test set (no. refers to the reference database. Experimental and predicted classes by TOXclass model: high $AT$, high acute toxicity; low $AT$, low acute toxicity; experimental and predicted classes by MOA class model: A, baseline narcosis; B, polar narcosis; C, arylate and ester narcosis; D, electrophile or proelectrophile phosphorylation; E, neurodepressant; F, uncoupler of oxidative phosphorylation; G, central nervous system seizure mechanisms; H, AchE inhibition; I, respiratory blocker or inhibition).

F.4 References.
## F.1 TOX class model: Training set

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>LC$_{50}$ (mg/l)</th>
<th>LC$_{50}$ (mmol/l)</th>
<th>Exp. Toxic class</th>
<th>Pred. Toxic class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4-(Hexyloxy)-m-anisaldehyde</td>
<td>61096-84-2</td>
<td>2.67</td>
<td>0.0113</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>2</td>
<td>5-Bromo-2-nitroanillin</td>
<td>98433-34-5</td>
<td>73.30</td>
<td>0.2660</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>4</td>
<td>p-Chlorophenyl-o-nitrophenyl ether</td>
<td>39145-47-6</td>
<td>1.92</td>
<td>0.0077</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>5</td>
<td>3'-Chloro-o-formotoluidide</td>
<td>71802-02-7</td>
<td>46.60</td>
<td>0.2750</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>6</td>
<td>Di-n-butylisophthalate</td>
<td>3126-90-7</td>
<td>0.90</td>
<td>0.0032</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>7</td>
<td>1,1-Diphenyl-2-propyn-1-ol</td>
<td>3923-52-2</td>
<td>11.10</td>
<td>0.0533</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>8</td>
<td>4,7-Dithiacedane</td>
<td>22017-97-4</td>
<td>7.52</td>
<td>0.0422</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>11</td>
<td>4,9-Dithiadecane</td>
<td>56348-39-1</td>
<td>2.90</td>
<td>0.0145</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>14</td>
<td>2-chloroethyl-N-cyclohexyl carbamate</td>
<td>31302-57-5</td>
<td>35</td>
<td>0.1700</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>15</td>
<td>Phenobarbital</td>
<td>50-06-6</td>
<td>4.84</td>
<td>0.0800</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>16</td>
<td>2,4-Dinitrophenol</td>
<td>51-28-5</td>
<td>10.90</td>
<td>0.0992</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>17</td>
<td>Urethane</td>
<td>51-79-6</td>
<td>52.40</td>
<td>58.8000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>18</td>
<td>Salicylic acid Na+</td>
<td>54-21-7</td>
<td>1.720</td>
<td>12.5000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>19</td>
<td>Benzamidine</td>
<td>55-21-4</td>
<td>661</td>
<td>5.4000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>21</td>
<td>1,1-Dimethyl hydrazine</td>
<td>57-14-7</td>
<td>7.85</td>
<td>0.1310</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>22</td>
<td>Pentobarbital sodium</td>
<td>57-33-0</td>
<td>49.50</td>
<td>0.1990</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>23</td>
<td>Amobarbital</td>
<td>57-43-2</td>
<td>85.40</td>
<td>0.3770</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>24</td>
<td>Caffeine</td>
<td>56-08-2</td>
<td>151</td>
<td>0.7780</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>25</td>
<td>2-Methyl-1,4-naphthoquinone</td>
<td>56-27-5</td>
<td>0.11</td>
<td>0.0006</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>26</td>
<td>2,3,4,6-Tetrachlorophenol</td>
<td>56-00-2</td>
<td>1.03</td>
<td>0.0044</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>27</td>
<td>4-Chloro-3-methyl phenol</td>
<td>56-50-7</td>
<td>5.47</td>
<td>0.0384</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC&lt;sub&gt;50&lt;/sub&gt; (mg/l)</td>
<td>LC&lt;sub&gt;50&lt;/sub&gt; (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------</td>
<td>-------</td>
<td>------------------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>28</td>
<td>Tolazoline hydrochloride</td>
<td>59-97-2</td>
<td>354</td>
<td>1.8000</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>29</td>
<td>Amphetamine sulfate</td>
<td>60-13-9</td>
<td>28.80</td>
<td>0.0782</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>30</td>
<td>Diethyl ether</td>
<td>60-23-7</td>
<td>2560</td>
<td>34.5000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>31</td>
<td>Strychnine hemisulphate salt</td>
<td>60-41-3</td>
<td>1.11</td>
<td>0.0011</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>32</td>
<td>Aniline</td>
<td>62-53-3</td>
<td>105</td>
<td>1.1300</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>33</td>
<td>Carbaryl (sevin)</td>
<td>63-25-2</td>
<td>8.75</td>
<td>0.0435</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>34</td>
<td>Ethanol</td>
<td>64-17-5</td>
<td>14700</td>
<td>319.0000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>35</td>
<td>Nicotine sulfate</td>
<td>65-30-5</td>
<td>13.80</td>
<td>0.0530</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>36</td>
<td>2-Hydroxybenzamide</td>
<td>65-45-2</td>
<td>101</td>
<td>0.7900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>38</td>
<td>Hexanal</td>
<td>66-25-1</td>
<td>17.50</td>
<td>0.1750</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>39</td>
<td>Dicumarol</td>
<td>66-76-2</td>
<td>5.11</td>
<td>0.0152</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>40</td>
<td>p-Phenoxynbenaldehyde</td>
<td>67-94-7</td>
<td>4.60</td>
<td>0.0232</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>41</td>
<td>Methanol</td>
<td>67-96-1</td>
<td>29400</td>
<td>917.0000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>42</td>
<td>2-Propanol</td>
<td>67-63-0</td>
<td>8680</td>
<td>144.0000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>43</td>
<td>Acetone</td>
<td>67-64-1</td>
<td>7160</td>
<td>123.0000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>44</td>
<td>Chloroform</td>
<td>67-66-3</td>
<td>70.70</td>
<td>0.5920</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>45</td>
<td>Methyl sulfoxide</td>
<td>67-68-5</td>
<td>34000</td>
<td>435.0000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>46</td>
<td>Hexachloroethane</td>
<td>67-72-1</td>
<td>1.42</td>
<td>0.0060</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>47</td>
<td>2,2'-Methylene bis(3,4,6-trichlorophenol)</td>
<td>70-30-1</td>
<td>0.02</td>
<td>0.0001</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>48</td>
<td>4'-Aminopropiophenone</td>
<td>70-69-0</td>
<td>14.6</td>
<td>0.5790</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>49</td>
<td>1-Propanol</td>
<td>71-23-8</td>
<td>4550</td>
<td>78.7000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>50</td>
<td>n-Butanol</td>
<td>71-36-3</td>
<td>1730</td>
<td>233.4000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC₅₀ (mg/l)</td>
<td>LC₅₀ (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>51</td>
<td>1-Pentanol</td>
<td>71-41-0</td>
<td>472</td>
<td>5.3550</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>52</td>
<td>Benzene</td>
<td>71-43-2</td>
<td>17.60</td>
<td>0.2250</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>53</td>
<td>1,1,1-Trichloroethane</td>
<td>71-55-6</td>
<td>47.30</td>
<td>0.3550</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>54</td>
<td>Thiophenol, sodium salt</td>
<td>71-73-8</td>
<td>26.20</td>
<td>0.0991</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>55</td>
<td>Acetonitrile</td>
<td>75-05-8</td>
<td>1644</td>
<td>40.0500</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>56</td>
<td>Ethanol</td>
<td>75-07-0</td>
<td>33.80</td>
<td>0.7670</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>57</td>
<td>Dichloromethane</td>
<td>75-09-2</td>
<td>330</td>
<td>3.8850</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>58</td>
<td>Iodoform</td>
<td>75-17-8</td>
<td>2.92</td>
<td>0.0074</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>60</td>
<td>2-Methyl-2-propanol</td>
<td>75-65-0</td>
<td>64.10</td>
<td>86.4800</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>61</td>
<td>2,2,2-Trifluoroethanol</td>
<td>75-80-8</td>
<td>119</td>
<td>1.1900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>62</td>
<td>3,3-Dimethyl-2-butanone</td>
<td>75-97-8</td>
<td>87</td>
<td>0.8690</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>63</td>
<td>Pentachloroethane</td>
<td>76-01-7</td>
<td>7.53</td>
<td>0.0372</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>64</td>
<td>5,5-Dimethylhydantoin</td>
<td>77-71-4</td>
<td>16400</td>
<td>128.5000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>65</td>
<td>3-Methyl-3-pentanol</td>
<td>77-74-7</td>
<td>672</td>
<td>65.770</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>66</td>
<td>3-Methyl-1-pentyn-3-ol</td>
<td>77-75-8</td>
<td>1220</td>
<td>12.4300</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>67</td>
<td>1-Ethynyl-cyclohexanol</td>
<td>78-27-3</td>
<td>256</td>
<td>2.0610</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>68</td>
<td>Tris(2-butoxyethyl) phosphate</td>
<td>78-51-3</td>
<td>11.20</td>
<td>0.0281</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>69</td>
<td>2-Methyl-1-propanol</td>
<td>78-88-1</td>
<td>1430</td>
<td>19.2500</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>70</td>
<td>1,2-Dichloropropane</td>
<td>78-87-5</td>
<td>127</td>
<td>1.1240</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>71</td>
<td>1,2-Diamino propane</td>
<td>78-94-4</td>
<td>1010</td>
<td>13.6300</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>72</td>
<td>2-Butanol</td>
<td>78-92-2</td>
<td>3670</td>
<td>49.5100</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>73</td>
<td>2-Butanone</td>
<td>78-93-3</td>
<td>3220</td>
<td>44.6600</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------------------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>74</td>
<td>1-Amino-2-propanol</td>
<td>78-96-6</td>
<td>25.20</td>
<td>33.5500</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>75</td>
<td>1,1,2-Trichloroethane</td>
<td>79-00-5</td>
<td>81.60</td>
<td>0.6120</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>76</td>
<td>Trichloroethylene</td>
<td>79-01-6</td>
<td>44.10</td>
<td>0.3360</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>77</td>
<td>Methyl acetate</td>
<td>79-20-9</td>
<td>357</td>
<td>4.8190</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>78</td>
<td>1,1,2,2-Tetrachloroethane</td>
<td>79-34-5</td>
<td>20.30</td>
<td>0.1210</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>79</td>
<td>beta-limonene</td>
<td>79-77-6</td>
<td>5.00</td>
<td>0.0065</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>80</td>
<td>4,4'-Isopropylidenediisocyanic acid (2,6-dichlorophenol)</td>
<td>79-95-8</td>
<td>1.33</td>
<td>0.0096</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>81</td>
<td>m-tert-Butylphenol</td>
<td>80-46-6</td>
<td>2.50</td>
<td>0.0158</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>82</td>
<td>1,8-Diamino-p-menthane</td>
<td>80-52-1</td>
<td>65.30</td>
<td>0.3830</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>83</td>
<td>alpha,alpha-2,6-Tetrachlorotoluene</td>
<td>81-19-6</td>
<td>0.97</td>
<td>0.0642</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>84</td>
<td>Aacenaphthene</td>
<td>83-32-9</td>
<td>1.73</td>
<td>0.0112</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>85</td>
<td>3-Methylindole</td>
<td>83-34-1</td>
<td>8.84</td>
<td>0.0074</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>86</td>
<td>Rotenone</td>
<td>84-70-4</td>
<td>0.0052</td>
<td>0.0000</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>88</td>
<td>Diphenyl phthalate</td>
<td>84-62-8</td>
<td>0.08</td>
<td>0.0003</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>90</td>
<td>Diethyl phthalate</td>
<td>84-66-2</td>
<td>3.180</td>
<td>0.1430</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>91</td>
<td>Dil-n-butylthiophthalate</td>
<td>84-74-2</td>
<td>1.00</td>
<td>0.0036</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>94</td>
<td>Salicylanilide</td>
<td>87-17-2</td>
<td>3.95</td>
<td>0.0185</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>95</td>
<td>Hexachloro-1,3-butadiene</td>
<td>87-68-3</td>
<td>0.09</td>
<td>0.0003</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>96</td>
<td>Pentachlorophenol</td>
<td>87-86-5</td>
<td>0.2450</td>
<td>0.0009</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>97</td>
<td>2,4,6-Trichlorophenol</td>
<td>88-06-2</td>
<td>4.80</td>
<td>0.0248</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>98</td>
<td>3-Trifluoromethyl-4-nitrophenol</td>
<td>88-30-2</td>
<td>9.14</td>
<td>0.0411</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>99</td>
<td>Anthranilamide</td>
<td>88-68-6</td>
<td>395</td>
<td>2.9010</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>100</td>
<td>2-Nitrophenol</td>
<td>88-75-5</td>
<td>160</td>
<td>1.1500</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>101</td>
<td>2-sec-Butyl-4,6-dinitrophenol (dinoseb)</td>
<td>88-85-7</td>
<td>0.5350</td>
<td>0.0022</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>102</td>
<td>Salicylaldehyde</td>
<td>90-02-8</td>
<td>2.30</td>
<td>0.0188</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>103</td>
<td>1-Naphthol</td>
<td>90-15-3</td>
<td>4.63</td>
<td>0.0321</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>104</td>
<td>2-Phenylphenol</td>
<td>90-43-7</td>
<td>6.15</td>
<td>0.0361</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>106</td>
<td>3,5-Dibromosalicylaldehyde</td>
<td>90-59-5</td>
<td>0.85</td>
<td>0.0030</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>107</td>
<td>Naphthalene</td>
<td>91-20-4</td>
<td>6.14</td>
<td>0.0170</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>108</td>
<td>Quinoline</td>
<td>91-22-5</td>
<td>77.80</td>
<td>0.5020</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>109</td>
<td>N,N-Diethyldicyclohexylamine</td>
<td>91-65-6</td>
<td>21.40</td>
<td>0.1380</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>110</td>
<td>N,N-Diethylaminiline</td>
<td>91-96-7</td>
<td>16.40</td>
<td>0.100</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>111</td>
<td>2-(N-Ethyl-m-toluidino)ethanol</td>
<td>91-88-3</td>
<td>52.90</td>
<td>0.2950</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>113</td>
<td>1-Benzoylacetone</td>
<td>93-91-1</td>
<td>1.10</td>
<td>0.0008</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>114</td>
<td>Ethyl-$p$-aminobenzoate</td>
<td>94-00-4</td>
<td>35.70</td>
<td>0.2160</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>115</td>
<td>Piperine (aliphatic)</td>
<td>94-62-2</td>
<td>7.84</td>
<td>0.0275</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>116</td>
<td>2,4-Dihydroxybenzaldehyde</td>
<td>95-01-2</td>
<td>13.10</td>
<td>0.0550</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>117</td>
<td>o-Xylene</td>
<td>95-47-6</td>
<td>16.40</td>
<td>0.1540</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>118</td>
<td>o-Cresol</td>
<td>95-48-7</td>
<td>14.00</td>
<td>0.1250</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>119</td>
<td>1,2-Dichlorobenzene</td>
<td>95-50-1</td>
<td>9.47</td>
<td>0.0640</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>120</td>
<td>2-Chloroaniline</td>
<td>95-51-2</td>
<td>5.74</td>
<td>0.0450</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>121</td>
<td>2-Fluorotoluene</td>
<td>95-52-3</td>
<td>19.40</td>
<td>0.1760</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>122</td>
<td>2-Chlorophenol</td>
<td>95-57-8</td>
<td>11.40</td>
<td>0.0887</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>123</td>
<td>1,2,4-Trimethylbenzene</td>
<td>95-63-6</td>
<td>7.72</td>
<td>0.0642</td>
<td>highAT</td>
<td>highAT</td>
</tr>
</tbody>
</table>
### TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>LC$_{50}$ (mg/l)</th>
<th>LC$_{50}$ (mmol/l)</th>
<th>Exp. class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>3,4-Dichlorotoluene</td>
<td>95-75-0</td>
<td>2.91</td>
<td>0.0181</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>125</td>
<td>3,4-Dichloroaniline</td>
<td>95-76-1</td>
<td>7.57</td>
<td>0.0467</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>126</td>
<td>Allyl methacrylate</td>
<td>96-05-9</td>
<td>0.99</td>
<td>0.0079</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>128</td>
<td>2-Methylbutylaldehyde</td>
<td>96-17-3</td>
<td>9.97</td>
<td>0.1160</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>129</td>
<td>1,2,3-Trichloropropane</td>
<td>96-18-4</td>
<td>57.70</td>
<td>0.3910</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>130</td>
<td>3-Pentanone</td>
<td>96-22-0</td>
<td>1540</td>
<td>17.8800</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>132</td>
<td>2-Butanone oxime</td>
<td>96-29-7</td>
<td>843</td>
<td>9.6760</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>133</td>
<td>2-(Diisopropylamino)ethanol</td>
<td>96-80-0</td>
<td>201</td>
<td>1.3840</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>134</td>
<td>2,4-Dinitroaniline</td>
<td>97-02-9</td>
<td>14.80</td>
<td>0.0808</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>136</td>
<td>p-tert-Butylphenol</td>
<td>98-54-7</td>
<td>843</td>
<td>9.6760</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>137</td>
<td>Isopropylbenzene</td>
<td>98-82-8</td>
<td>6.32</td>
<td>0.0526</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>138</td>
<td>Nitrobenzene</td>
<td>98-86-2</td>
<td>1162</td>
<td>1.3480</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>139</td>
<td>Aacetophenone</td>
<td>98-95-3</td>
<td>119</td>
<td>0.0670</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>140</td>
<td>m-Aminoacetophenone</td>
<td>99-03-6</td>
<td>382</td>
<td>22.6900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>141</td>
<td>m-Nitrotoluene</td>
<td>99-04-1</td>
<td>75.00</td>
<td>0.1870</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>142</td>
<td>N,N-Dimethyl-p-toluidine</td>
<td>99-97-8</td>
<td>48.90</td>
<td>0.3630</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>143</td>
<td>m-Nitroaniline</td>
<td>100-01-6</td>
<td>125</td>
<td>0.0505</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>144</td>
<td>p-Nitrophenol</td>
<td>100-02-7</td>
<td>4.80</td>
<td>0.3220</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>145</td>
<td>p-Dimethylaminobenzaldehyde</td>
<td>100-10-7</td>
<td>45.70</td>
<td>0.3060</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>146</td>
<td>1,4-Dinitrobenzene</td>
<td>100-25-4</td>
<td>0.71</td>
<td>0.0042</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>147</td>
<td>N,N-Diethylethanolamine</td>
<td>100-37-8</td>
<td>1780</td>
<td>15.1900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
</tbody>
</table>
### TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>$LC_{50}$ (mg/l)</th>
<th>$LC_{50}$ (mmol/l)</th>
<th>Exp. class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>148</td>
<td>Ethylbenzene</td>
<td>100-41-4</td>
<td>10.50</td>
<td>0.0989</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>149</td>
<td>Benzylamine</td>
<td>100-46-9</td>
<td>102</td>
<td>0.03520</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>150</td>
<td>Benzaldehyde</td>
<td>100-52-7</td>
<td>9.87</td>
<td>0.09930</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>151</td>
<td>N-Methylaniline</td>
<td>100-61-8</td>
<td>100</td>
<td>0.3330</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>152</td>
<td>Cyclohexanone oxime</td>
<td>100-64-1</td>
<td>208</td>
<td>1.8830</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>153</td>
<td>2-Cyanopyridine</td>
<td>100-70-9</td>
<td>726</td>
<td>6.9730</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>154</td>
<td>2-Ethylpyridine</td>
<td>100-71-0</td>
<td>414</td>
<td>3.8640</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>155</td>
<td>Solketal</td>
<td>100-78-8</td>
<td>16700</td>
<td>126.4000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>156</td>
<td>Hexamethylene tetramine (aliphatic)</td>
<td>100-97-0</td>
<td>46800</td>
<td>355.2000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>157</td>
<td>Phenyl ether</td>
<td>101-84-8</td>
<td>4.000</td>
<td>0.0235</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>158</td>
<td>N-Ethyl-m-toluidine</td>
<td>102-77-2</td>
<td>49.50</td>
<td>0.3660</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>159</td>
<td>Tripropylamine</td>
<td>102-60-2</td>
<td>50.90</td>
<td>0.3550</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>160</td>
<td>Triethanolamine</td>
<td>102-71-6</td>
<td>11800</td>
<td>79.0900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>161</td>
<td>Benzyl-tert-butanol</td>
<td>103-05-9</td>
<td>66.40</td>
<td>0.3400</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>162</td>
<td>1-(2-Hydroxyethyl)piperazine</td>
<td>103-76-4</td>
<td>461.00</td>
<td>46.2400</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>163</td>
<td>N,N-Dimethylaniline</td>
<td>103-88-3</td>
<td>37.80</td>
<td>0.2800</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>164</td>
<td>4-Acetamidophenol</td>
<td>103-00-2</td>
<td>81.4</td>
<td>5.3950</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>165</td>
<td>4-Butylaniline</td>
<td>104-13-2</td>
<td>10.20</td>
<td>0.0690</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>166</td>
<td>Nonylphenol (mixed)</td>
<td>25154-52-3</td>
<td>0.14</td>
<td>0.0006</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>167</td>
<td>2-Ethyl-1-hexanol</td>
<td>104-76-7</td>
<td>28.20</td>
<td>0.2170</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>168</td>
<td>4-Chlorobenzaldehyde</td>
<td>104-88-1</td>
<td>2.20</td>
<td>0.0156</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>169</td>
<td>5-Ethyl-2-methylpyridine</td>
<td>104-90-5</td>
<td>81.10</td>
<td>0.6690</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. Toxic class</td>
<td>Pred. Toxic class</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>172</td>
<td>5-Diethylamino-2-pentanone</td>
<td>105-14-6</td>
<td>3.26</td>
<td>2.1370</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>173</td>
<td>Diethyl malonate</td>
<td>105-53-3</td>
<td>14.70</td>
<td>0.0918</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>174</td>
<td>2,4-Dimethylphenol</td>
<td>105-67-9</td>
<td>16.60</td>
<td>0.1360</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>175</td>
<td>Dibutyl fumarate</td>
<td>105-75-0</td>
<td>0.63</td>
<td>0.0028</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>176</td>
<td>Dibutyl adipate</td>
<td>105-99-7</td>
<td>3.64</td>
<td>0.0141</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>177</td>
<td>p-Bromonitroaniline</td>
<td>106-40-1</td>
<td>47.50</td>
<td>0.2760</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>178</td>
<td>p-Xylene</td>
<td>106-42-3</td>
<td>8.87</td>
<td>0.0635</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>179</td>
<td>4-Methylphenol (p-cresol)</td>
<td>106-44-5</td>
<td>16.50</td>
<td>0.1530</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>180</td>
<td>4-Chloroaniline</td>
<td>106-47-8</td>
<td>31.40</td>
<td>0.2460</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>181</td>
<td>4-Chlorophenol</td>
<td>106-48-9</td>
<td>6.11</td>
<td>0.0475</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>182</td>
<td>4-Toluamide</td>
<td>106-49-0</td>
<td>160</td>
<td>1.4930</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>183</td>
<td>Isobutyl acrylate</td>
<td>106-53-8</td>
<td>2.10</td>
<td>0.0164</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>184</td>
<td>1-Bromopropane</td>
<td>106-61-5</td>
<td>67.30</td>
<td>0.5470</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>185</td>
<td>Acrolein</td>
<td>107-02-8</td>
<td>0.017</td>
<td>0.0003</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>186</td>
<td>1,2-Dichloroethane</td>
<td>107-06-2</td>
<td>136</td>
<td>1.3740</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>187</td>
<td>2-Chloroethanol</td>
<td>107-07-3</td>
<td>53.70</td>
<td>0.6670</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>188</td>
<td>Propylamine</td>
<td>107-10-8</td>
<td>308</td>
<td>5.2110</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>189</td>
<td>Propionitrile</td>
<td>107-12-0</td>
<td>1520</td>
<td>27.6000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>190</td>
<td>Chloroacetanilic</td>
<td>107-14-2</td>
<td>1.35</td>
<td>0.0178</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>191</td>
<td>Ethylenediamine</td>
<td>107-15-3</td>
<td>220</td>
<td>3.0610</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>192</td>
<td>Allyl alcohol</td>
<td>107-18-6</td>
<td>0.32</td>
<td>0.0055</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>193</td>
<td>2-Propyn-1-ol</td>
<td>107-19-7</td>
<td>1.48</td>
<td>0.0284</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC(_{50}) (mg/l)</td>
<td>LC(_{50}) (mmol/l)</td>
<td>Exp. T.ox class</td>
<td>Pred. T.ox class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>-------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>194</td>
<td>Acetaldoxime</td>
<td>107-29-9</td>
<td>76</td>
<td>1.2870</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>195</td>
<td>2-Methyl-2,4-pentanediol</td>
<td>107-41-5</td>
<td>107.00</td>
<td>90.5400</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>196</td>
<td>tert-Octylamine</td>
<td>107-45-9</td>
<td>24.60</td>
<td>0.1900</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>197</td>
<td>tert-Butyl sulfide</td>
<td>107-47-1</td>
<td>29.10</td>
<td>0.1900</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>198</td>
<td>2-Pentanone</td>
<td>107-87-9</td>
<td>1240</td>
<td>14.4000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>199</td>
<td>4-Methyl-2-pentanone</td>
<td>108-10-4</td>
<td>522</td>
<td>5.2120</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>200</td>
<td>Isopropyl ether</td>
<td>108-20-3</td>
<td>786</td>
<td>7.6000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>201</td>
<td>Toluene</td>
<td>108-88-3</td>
<td>33.90</td>
<td>0.3680</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>202</td>
<td>4-Picoline</td>
<td>108-28-4</td>
<td>403</td>
<td>4.3270</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>203</td>
<td>Chlorobenzene</td>
<td>108-90-7</td>
<td>16.90</td>
<td>0.1500</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>204</td>
<td>Cyclohexanol</td>
<td>108-03-0</td>
<td>704</td>
<td>7.0290</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>205</td>
<td>Cyclohexanone</td>
<td>108-04-1</td>
<td>621</td>
<td>6.3270</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>206</td>
<td>Phenol</td>
<td>108-88-5</td>
<td>32.70</td>
<td>0.3470</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>207</td>
<td>3-Picoline</td>
<td>108-29-4</td>
<td>144</td>
<td>1.5460</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>208</td>
<td>1-Methylpiperazine</td>
<td>109-01-3</td>
<td>2300</td>
<td>22.9000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>209</td>
<td>2-Picoline</td>
<td>109-06-8</td>
<td>897</td>
<td>9.6320</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>210</td>
<td>2-Methylpiperazine</td>
<td>109-07-0</td>
<td>2240</td>
<td>22.3000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>211</td>
<td>Propyl acetate</td>
<td>109-66-4</td>
<td>60</td>
<td>0.5870</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>212</td>
<td>1,3-Dibromopropene</td>
<td>109-64-8</td>
<td>2.09</td>
<td>0.0104</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>213</td>
<td>1-Bromobutane</td>
<td>109-65-9</td>
<td>36.70</td>
<td>0.3690</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>214</td>
<td>Butylamine</td>
<td>109-75-9</td>
<td>268</td>
<td>3.6640</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>215</td>
<td>Allyl cyanide</td>
<td>109-75-1</td>
<td>182</td>
<td>2.7130</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC&lt;sub&gt;50&lt;/sub&gt; (mg/l)</td>
<td>LC&lt;sub&gt;50&lt;/sub&gt; (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>216</td>
<td>1,3-Diaminopropane</td>
<td>109-76-2</td>
<td>11.90</td>
<td>16.0500</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>217</td>
<td>Malononitrile</td>
<td>109-77-3</td>
<td>0.56</td>
<td>0.0085</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>218</td>
<td>2-Methoxyethylamine</td>
<td>109-85-3</td>
<td>524</td>
<td>6.9760</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>219</td>
<td>Diethylamine</td>
<td>109-89-7</td>
<td>855</td>
<td>11.6900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>220</td>
<td>Pyrrole</td>
<td>109-97-7</td>
<td>210</td>
<td>3.1300</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>221</td>
<td>Tetrahydrofuran</td>
<td>109-99-0</td>
<td>2160</td>
<td>29.9600</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>222</td>
<td>Fur'an</td>
<td>110-00-0</td>
<td>61</td>
<td>0.8900</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>223</td>
<td>4-Butyl disulfide</td>
<td>110-06-5</td>
<td>1.37</td>
<td>0.0077</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>224</td>
<td>5-Methyl-2-hexanone</td>
<td>110-12-3</td>
<td>159</td>
<td>1.2920</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>225</td>
<td>Diethyl sebacate</td>
<td>110-40-7</td>
<td>2.72</td>
<td>0.0105</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>226</td>
<td>2-Heptanone</td>
<td>110-43-0</td>
<td>131</td>
<td>1.1470</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>227</td>
<td>Hexane</td>
<td>110-54-3</td>
<td>2.50</td>
<td>0.0280</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>228</td>
<td>1,4-Dichlorobutane</td>
<td>110-55-6</td>
<td>51.60</td>
<td>0.1000</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>229</td>
<td>Amylamine</td>
<td>110-58-7</td>
<td>1.77</td>
<td>0.0310</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>230</td>
<td>Valeraldehydye</td>
<td>110-62-3</td>
<td>12.90</td>
<td>0.1500</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>231</td>
<td>2-Butyne-1,4-diol</td>
<td>110-65-4</td>
<td>53.60</td>
<td>0.6230</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>232</td>
<td>2-(Ethylamino)ethanol</td>
<td>110-73-6</td>
<td>1480</td>
<td>16.6000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>233</td>
<td>Cyclohexane</td>
<td>110-82-7</td>
<td>4.53</td>
<td>0.0538</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>234</td>
<td>Pyridine</td>
<td>110-86-1</td>
<td>90.80</td>
<td>1.3020</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>235</td>
<td>t-Tricane</td>
<td>110-88-3</td>
<td>5590</td>
<td>66.0500</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>236</td>
<td>6-Methyl-5-hepten-2-one</td>
<td>110-93-4</td>
<td>85.70</td>
<td>0.6790</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>237</td>
<td>2-Octanone</td>
<td>111-13-7</td>
<td>36</td>
<td>0.2810</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>238</td>
<td>2-Ethoxyethyl acetate</td>
<td>111-15-9</td>
<td>42.10</td>
<td>0.3190</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>239</td>
<td>1-Bromohexane</td>
<td>111-25-1</td>
<td>3.45</td>
<td>0.0209</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>240</td>
<td>Hexylamine</td>
<td>111-26-2</td>
<td>56.60</td>
<td>0.5590</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>241</td>
<td>Diethanolamine</td>
<td>111-42-2</td>
<td>47.10</td>
<td>44.8000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>242</td>
<td>2-Hydroxyethyl ether</td>
<td>111-46-6</td>
<td>75.200</td>
<td>708.6000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>243</td>
<td>n-Propyl sulfide</td>
<td>111-47-7</td>
<td>21.70</td>
<td>0.1840</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>244</td>
<td>n-Octylamine</td>
<td>111-68-2</td>
<td>21.80</td>
<td>0.1890</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>245</td>
<td>1,4-Dicyanobutane</td>
<td>111-69-3</td>
<td>19.30</td>
<td>17.8500</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>246</td>
<td>1-Heptanol</td>
<td>111-70-6</td>
<td>34.50</td>
<td>0.2970</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>247</td>
<td>1-Hexanol</td>
<td>111-83-1</td>
<td>8.3800</td>
<td>0.0813</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>248</td>
<td>1-Octanol</td>
<td>111-86-1</td>
<td>5.19</td>
<td>0.0402</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>249</td>
<td>1-Undecanone</td>
<td>111-87-5</td>
<td>13.50</td>
<td>0.1040</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>250</td>
<td>1-Decanol</td>
<td>111-90-0</td>
<td>205.00</td>
<td>197.5000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>251</td>
<td>2-2-Hydroxyethoxy)ethanol</td>
<td>111-90-0</td>
<td>104</td>
<td>0.0570</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>252</td>
<td>Nonanoic acid</td>
<td>112-05-0</td>
<td>1.50</td>
<td>0.0088</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>253</td>
<td>2-Undecanone</td>
<td>112-12-9</td>
<td>2.16</td>
<td>0.0150</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>254</td>
<td>Nonylamine</td>
<td>112-20-9</td>
<td>2.16</td>
<td>0.0150</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>255</td>
<td>Triethylene glycol</td>
<td>112-27-5</td>
<td>68.9000</td>
<td>458.8000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>256</td>
<td>1-Decanol</td>
<td>112-30-1</td>
<td>2.40</td>
<td>0.0152</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>257</td>
<td>Propoxur (Baygon)</td>
<td>114-26-1</td>
<td>8.80</td>
<td>0.0421</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>258</td>
<td>2-Methyl-3-butyne-2-ol</td>
<td>115-19-5</td>
<td>32.90</td>
<td>39.1100</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>259</td>
<td>2,2,2-Trichloroethanol</td>
<td>115-20-8</td>
<td>29.9</td>
<td>2.0010</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
</tbody>
</table>
### TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>LC&lt;sub&gt;50&lt;/sub&gt; (mg/l)</th>
<th>LC&lt;sub&gt;50&lt;/sub&gt; (mmol/l)</th>
<th>Exp. class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>261</td>
<td>Dicofol (Keithane)</td>
<td>115-32-2</td>
<td>0.6030</td>
<td>0.0017</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>262</td>
<td>Triprenyl phosphate</td>
<td>115-86-6</td>
<td>0.87</td>
<td>0.0027</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>264</td>
<td>Aldicarb</td>
<td>116-06-3</td>
<td>0.8610</td>
<td>0.0045</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>267</td>
<td>Phenyl salicylate</td>
<td>118-55-8</td>
<td>1.18</td>
<td>0.0055</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>268</td>
<td>Ethyl salicylate</td>
<td>118-61-6</td>
<td>20.20</td>
<td>0.1220</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>269</td>
<td>2,4,6-Trichlorophenol</td>
<td>118-79-6</td>
<td>6.54</td>
<td>0.0198</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>270</td>
<td>4-Amino-2-nitrophenol</td>
<td>119-34-6</td>
<td>36.20</td>
<td>0.2350</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>271</td>
<td>Benzophenone</td>
<td>119-61-9</td>
<td>14.70</td>
<td>0.0807</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>272</td>
<td>N-Phenylmethanolamine</td>
<td>120-07-0</td>
<td>735</td>
<td>4.0560</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>273</td>
<td>4-Diethylamino)benzaldehyde</td>
<td>120-21-8</td>
<td>23.90</td>
<td>0.1350</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>274</td>
<td>Catechol</td>
<td>120-80-9</td>
<td>9.22</td>
<td>0.0837</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>275</td>
<td>1,2,4-Trichlorobenzene</td>
<td>120-82-1</td>
<td>2.99</td>
<td>0.0165</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>276</td>
<td>2,4-Dichlorophenol</td>
<td>120-83-2</td>
<td>7.75</td>
<td>0.0475</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>277</td>
<td>2,4-Dinitrotoluene</td>
<td>121-14-2</td>
<td>24.30</td>
<td>0.1330</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>278</td>
<td>3-Ethoxy-4-hydroxybenzaldehyde</td>
<td>121-32-4</td>
<td>87.60</td>
<td>0.5270</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>279</td>
<td>Vanillin</td>
<td>121-23-5</td>
<td>83.80</td>
<td>0.5510</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>280</td>
<td>N,N-Dimethylaniline</td>
<td>121-69-7</td>
<td>64.10</td>
<td>0.2990</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>281</td>
<td>1-Chloro-3-nitrobenzene</td>
<td>121-73-3</td>
<td>18.80</td>
<td>0.1190</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>283</td>
<td>2-Chloro-4-nitroaniline</td>
<td>121-87-9</td>
<td>20.10</td>
<td>0.1160</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>284</td>
<td>p-Isopropyl benzaldehyde</td>
<td>122-03-2</td>
<td>6.62</td>
<td>0.0447</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>285</td>
<td>Diphenylamine</td>
<td>122-93-4</td>
<td>3.79</td>
<td>0.0224</td>
<td>high AT</td>
<td>high AT</td>
</tr>
</tbody>
</table>
### TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance name</th>
<th>Substance CASRN</th>
<th>LC50 (mg/l)</th>
<th>LC50 (mmol/l)</th>
<th>Exp. class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>286</td>
<td>2-Phenoxyethanol</td>
<td>122-99-6</td>
<td>344</td>
<td>2.4000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>287</td>
<td>4-Ethylphenol</td>
<td>123-07-9</td>
<td>10.40</td>
<td>0.00851</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>288</td>
<td>2-Methylth甲醛</td>
<td>123-15-9</td>
<td>18.80</td>
<td>0.1880</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>289</td>
<td>2,4-Pentanedione</td>
<td>123-54-6</td>
<td>135</td>
<td>1.3480</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>290</td>
<td>Ethyl hexanoate</td>
<td>123-66-0</td>
<td>8.90</td>
<td>0.0617</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>291</td>
<td>Butanal</td>
<td>123-72-8</td>
<td>14.70</td>
<td>0.2040</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>292</td>
<td>Butyraldehyde</td>
<td>123-86-4</td>
<td>18.00</td>
<td>0.1550</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>293</td>
<td>1,4-Dioxane</td>
<td>123-91-1</td>
<td>10000</td>
<td>116.9000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>294</td>
<td>Dodecylamine</td>
<td>124-22-1</td>
<td>0.1030</td>
<td>0.0006</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>296</td>
<td>Tributyl phosphate</td>
<td>136-73-8</td>
<td>9.48</td>
<td>0.00356</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>297</td>
<td>5,5-Dimethyl-1,3-cyclohexanedione</td>
<td>128-81-8</td>
<td>11500</td>
<td>82.0400</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>298</td>
<td>1-Chloro-2-propanol</td>
<td>127-00-4</td>
<td>245</td>
<td>2.5910</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>299</td>
<td>Tetrachloroethylene</td>
<td>127-18-4</td>
<td>16.50</td>
<td>0.0995</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>300</td>
<td>2-Phenyl-3-buten-2-ol</td>
<td>127-66-2</td>
<td>11.3</td>
<td>0.7780</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>301</td>
<td>2,6-Di-tert-butyl-4-methylphenol</td>
<td>28-37-0</td>
<td>0.3630</td>
<td>0.0017</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>302</td>
<td>Saccharin sodium salt hydrate</td>
<td>82385-12-0</td>
<td>8300</td>
<td>82.0000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>303</td>
<td>Dibenzo[b,d]furran</td>
<td>123-64-9</td>
<td>1.50</td>
<td>0.0089</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>304</td>
<td>Phenyl 4-aminosalicylate</td>
<td>123-11-9</td>
<td>4.76</td>
<td>0.0208</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>305</td>
<td>N,N-Diethyl-m-toluamide</td>
<td>134-62-3</td>
<td>11.0</td>
<td>0.5750</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>306</td>
<td>Propionic acid, sodium salt</td>
<td>137-30-6</td>
<td>4790</td>
<td>49.6000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>307</td>
<td>1-(2-Aminoethyl) piperazine</td>
<td>140-31-8</td>
<td>2190</td>
<td>16.9500</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>308</td>
<td>Dibutyl succinate</td>
<td>141-06-7</td>
<td>4.46</td>
<td>0.0194</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC\textsubscript{50} (mg/l)</td>
<td>LC\textsubscript{50} (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>309</td>
<td>Diethyl adipate</td>
<td>141-28-6</td>
<td>18.00</td>
<td>0.0899</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>310</td>
<td>2-Aminoethanol</td>
<td>141-43-5</td>
<td>2070</td>
<td>33,8900</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>311</td>
<td>Ethyl acetate</td>
<td>141-78-6</td>
<td>230</td>
<td>2.6100</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>312</td>
<td>2,6-Dimethylmorpholine</td>
<td>141-91-3</td>
<td>387</td>
<td>3.2600</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>313</td>
<td>m-Diethylbenzene</td>
<td>141-03-5</td>
<td>4.15</td>
<td>0.0309</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>314</td>
<td>1,3-Dichloropropene</td>
<td>142-28-9</td>
<td>111</td>
<td>0.0820</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>315</td>
<td>Hexanoic acid</td>
<td>142-62-1</td>
<td>320</td>
<td>2.7550</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>316</td>
<td>Hexyl acetate</td>
<td>142-92-7</td>
<td>4.10</td>
<td>0.0305</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>317</td>
<td>Butyl ether</td>
<td>142-06-1</td>
<td>32.30</td>
<td>0.2480</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>318</td>
<td>1-Nonanol</td>
<td>143-08-8</td>
<td>5.70</td>
<td>0.0305</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>319</td>
<td>Di-n-hexylamine</td>
<td>143-16-8</td>
<td>0.78</td>
<td>0.0012</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>321</td>
<td>o-Vanillin</td>
<td>148-23-8</td>
<td>2.10</td>
<td>0.0158</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>322</td>
<td>3-Methoxyphenol</td>
<td>150-19-6</td>
<td>74</td>
<td>0.5900</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>323</td>
<td>4-Methoxyphenol</td>
<td>150-76-5</td>
<td>110</td>
<td>0.8800</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>324</td>
<td>p-Dimethoxybenzene</td>
<td>150-78-7</td>
<td>11.7</td>
<td>0.8470</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>325</td>
<td>2,3-Dimethylfuran</td>
<td>271-89-6</td>
<td>14.00</td>
<td>0.1190</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>326</td>
<td>1,4-Diazabicyclo[2.2.2]octane</td>
<td>280-57-9</td>
<td>17.50</td>
<td>15.4200</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>327</td>
<td>Adamanate</td>
<td>281-23-2</td>
<td>0.28</td>
<td>0.0021</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>329</td>
<td>Secobarbital, sodium salt</td>
<td>309-43-3</td>
<td>23.60</td>
<td>0.0907</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>330</td>
<td>Bromacil</td>
<td>314-40-9</td>
<td>186</td>
<td>0.7120</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>331</td>
<td>2,5-Dinitrophenol</td>
<td>329-71-5</td>
<td>3.36</td>
<td>0.0182</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------</td>
<td>---------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>332</td>
<td>Diuron</td>
<td>330-54-1</td>
<td>14.20</td>
<td>0.0609</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>333</td>
<td>p-Fluorophenyl ether</td>
<td>330-93-8</td>
<td>1.13</td>
<td>0.0055</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>335</td>
<td>1-Fluoro-4-nitrobenzene</td>
<td>350-16-9</td>
<td>28.40</td>
<td>0.2010</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>336</td>
<td>alpha, alpha, alpha-Trifluoro-m-tolunitrile</td>
<td>368-77-4</td>
<td>47.70</td>
<td>0.2790</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>337</td>
<td>4-Fluoroaniline</td>
<td>371-40-4</td>
<td>16.90</td>
<td>0.1520</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>338</td>
<td>2-Chloro-6-fluorobenzaldehyde</td>
<td>387-45-1</td>
<td>9.41</td>
<td>0.0593</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>339</td>
<td>alpha, alpha, alpha-Tetrafluoro-o-toluidine</td>
<td>393-39-5</td>
<td>29.00</td>
<td>0.1650</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>340</td>
<td>2-Fluorobenzaldehyde</td>
<td>416-73-6</td>
<td>1.35</td>
<td>0.0109</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>341</td>
<td>alpha, alpha, alpha-Trifluoro-o-tolunitrile</td>
<td>417-00-9</td>
<td>42.30</td>
<td>0.2470</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>342</td>
<td>alpha, alpha, alpha-Trifluoro-m-tolualdehyde</td>
<td>454-89-7</td>
<td>0.9240</td>
<td>0.0533</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>343</td>
<td>4-Fluoro-N-methylaniline</td>
<td>479-59-6</td>
<td>38.40</td>
<td>0.2070</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>344</td>
<td>[(1S)-endo]-(-)-Borneol</td>
<td>464-45-9</td>
<td>65.20</td>
<td>0.4100</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>345</td>
<td>(1S)-(-)-Camphor</td>
<td>464-18-2</td>
<td>17.00</td>
<td>0.1120</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>347</td>
<td>Cineole</td>
<td>470-82-6</td>
<td>102</td>
<td>0.0610</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>348</td>
<td>Neoabietic acid</td>
<td>471-77-2</td>
<td>1.49</td>
<td>0.0049</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>350</td>
<td>2,3-Dihydrobenzofuran</td>
<td>496-16-2</td>
<td>81.70</td>
<td>0.6800</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>351</td>
<td>exo-Norboreneol</td>
<td>497-37-0</td>
<td>228</td>
<td>2.0330</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>352</td>
<td>Norbornylene</td>
<td>498-66-8</td>
<td>10.00</td>
<td>0.1000</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>353</td>
<td>2,6-Pyrindinedicarboxylic acid</td>
<td>499-83-2</td>
<td>322</td>
<td>1.5270</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>354</td>
<td>3-Pyridinecarboxaldehyde</td>
<td>500-22-1</td>
<td>16.40</td>
<td>0.1530</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>355</td>
<td>5-Norbornene</td>
<td>502-56-7</td>
<td>31.00</td>
<td>0.2180</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC_{50} (mg/l)</td>
<td>LC_{50} (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>356</td>
<td>2,3-Dimethyl-1,3-butadiene</td>
<td>513-81-5</td>
<td>6.91</td>
<td>0.0841</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>357</td>
<td>Abietic acid</td>
<td>514-10-3</td>
<td>2.38</td>
<td>0.0079</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>358</td>
<td>Flavone</td>
<td>525-82-6</td>
<td>3.50</td>
<td>0.0157</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>359</td>
<td>2,4,6-Trimethylphenol</td>
<td>527-60-6</td>
<td>13.00</td>
<td>0.0954</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>360</td>
<td>3-Methylcatechol</td>
<td>528-19-1</td>
<td>44.70</td>
<td>0.3820</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>361</td>
<td>3-Methylphenol</td>
<td>529-30-4</td>
<td>52.90</td>
<td>0.4400</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>362</td>
<td>Benzoic acid, sodium salt</td>
<td>532-32-1</td>
<td>484</td>
<td>3.2300</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>363</td>
<td>1,4-Dinitro- o-cresol</td>
<td>534-52-1</td>
<td>1.73</td>
<td>0.0087</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>364</td>
<td>4-Aminobenzene</td>
<td>538-88-1</td>
<td>1.71</td>
<td>0.0115</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>365</td>
<td>tert-Butyl acetate</td>
<td>540-88-5</td>
<td>327</td>
<td>2.8150</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>366</td>
<td>1,2-Dichlorobenzene</td>
<td>541-73-1</td>
<td>8.03</td>
<td>0.0546</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>367</td>
<td>1,2-Dichloroacetic acid</td>
<td>542-75-6</td>
<td>0.2390</td>
<td>0.0022</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>368</td>
<td>n-Butyl sulfide</td>
<td>544-10-1</td>
<td>3.58</td>
<td>0.0245</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>369</td>
<td>2'-Hydroxy-4'-methoxyacetophenone</td>
<td>552-14-0</td>
<td>69.50</td>
<td>0.4180</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>370</td>
<td>4-Nitrobenzaldehyde</td>
<td>552-89-6</td>
<td>14.40</td>
<td>0.0953</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>371</td>
<td>3-Nitrobenzaldehyde</td>
<td>555-16-8</td>
<td>10.10</td>
<td>0.0668</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>372</td>
<td>3-Methyl-2-butanone</td>
<td>563-80-4</td>
<td>864</td>
<td>10.0300</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>373</td>
<td>2,6-Dinitrophenol</td>
<td>573-56-8</td>
<td>39.70</td>
<td>0.2160</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>374</td>
<td>1,2-Dibromoethane</td>
<td>583-55-9</td>
<td>4.05</td>
<td>0.0172</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>375</td>
<td>N-allylaniline</td>
<td>588-06-2</td>
<td>35.00</td>
<td>0.2700</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>376</td>
<td>4-Ethylaniline</td>
<td>589-14-2</td>
<td>73</td>
<td>0.6020</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>377</td>
<td>Isovaleraldehyde</td>
<td>590-86-3</td>
<td>3.25</td>
<td>0.0377</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC(_{50}) (mg/l)</td>
<td>LC(_{50}) (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>---------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>378</td>
<td>2-Hexanone</td>
<td>591-78-6</td>
<td>428</td>
<td>4.2730</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>379</td>
<td>2,4-Hexadiene</td>
<td>502-46-1</td>
<td>20.00</td>
<td>0.2430</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>380</td>
<td>2-Tridecanone</td>
<td>503-08-8</td>
<td>0.36</td>
<td>0.0018</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>382</td>
<td>Maceol</td>
<td>506-85-0</td>
<td>0.12</td>
<td>0.0004</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>384</td>
<td>1,2-Dimethylpropylamine</td>
<td>508-74-3</td>
<td>284</td>
<td>3.2580</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>385</td>
<td>2,4-Dimethyl-3-pentanol</td>
<td>600-36-2</td>
<td>163</td>
<td>1.4030</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>386</td>
<td>N,N-Diphenylformamide</td>
<td>607-00-1</td>
<td>30.40</td>
<td>0.1540</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>388</td>
<td>Diethyl benzylmalonate</td>
<td>607-81-8</td>
<td>5.43</td>
<td>0.0217</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>389</td>
<td>Pentabromophenol</td>
<td>608-71-9</td>
<td>0.0930</td>
<td>0.0002</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>390</td>
<td>2,4,6-Trilodophenol</td>
<td>609-23-4</td>
<td>1.21</td>
<td>0.0026</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>391</td>
<td>2,4-Dimethoxybenzaldehyde</td>
<td>613-45-6</td>
<td>20.10</td>
<td>0.1210</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>392</td>
<td>2-Acetimidophenol</td>
<td>614-80-2</td>
<td>27.00</td>
<td>0.1790</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>393</td>
<td>2-Chloro-4-methylaniline</td>
<td>615-65-6</td>
<td>35.90</td>
<td>0.2540</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>394</td>
<td>4-Ethoxy-2-nitroaniline</td>
<td>616-86-4</td>
<td>28.00</td>
<td>0.1430</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>395</td>
<td>Methyl p-nitrobenzoate</td>
<td>619-50-1</td>
<td>23.80</td>
<td>0.1310</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>396</td>
<td>4-Nitrobenzamide</td>
<td>619-80-7</td>
<td>133</td>
<td>0.8010</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>397</td>
<td>4-Nitrophenyl phenyl ether</td>
<td>620-88-2</td>
<td>2.65</td>
<td>0.0123</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>398</td>
<td>Benzyl sulfoxide</td>
<td>621-08-9</td>
<td>80.10</td>
<td>0.3890</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>399</td>
<td>3-Acetimidophenol</td>
<td>621-42-1</td>
<td>11.30</td>
<td>7.4750</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>400</td>
<td>4-(2-Hydroxyethyl)morpholine</td>
<td>622-80-2</td>
<td>27.10</td>
<td>20.6000</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>401</td>
<td>alpha,alpha',alpha'-Dichloro-p-xylene</td>
<td>623-25-6</td>
<td>0.0300</td>
<td>0.0002</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>403</td>
<td>2,5-Dimethylfuran</td>
<td>625-86-5</td>
<td>71.10</td>
<td>0.7400</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------</td>
<td>--------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>404</td>
<td>1,5-Dichloropentane</td>
<td>628-76-2</td>
<td>25.30</td>
<td>0.1790</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>405</td>
<td>1-Bromoheptane</td>
<td>625-04-9</td>
<td>1.47</td>
<td>0.0082</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>406</td>
<td>Propyl disulfide</td>
<td>628-19-6</td>
<td>2.62</td>
<td>0.0170</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>407</td>
<td>1,6-Dicyanohexane</td>
<td>629-40-3</td>
<td>528</td>
<td>3.8770</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>408</td>
<td>2,3,4-Trichloroaniline</td>
<td>634-67-3</td>
<td>3.64</td>
<td>0.0185</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>409</td>
<td>5-Chloroacetylene</td>
<td>635-03-8</td>
<td>0.77</td>
<td>0.0049</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>410</td>
<td>4-Propylphenol</td>
<td>645-56-7</td>
<td>1.10</td>
<td>0.0008</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>411</td>
<td>Pentfluorobenzaldehyde</td>
<td>653-37-2</td>
<td>1.10</td>
<td>0.0056</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>412</td>
<td>2,2-Dichloroacetamide</td>
<td>684-72-7</td>
<td>241</td>
<td>1.8830</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>413</td>
<td>1-Methyl heptylamine</td>
<td>694-16-3</td>
<td>5.19</td>
<td>0.0402</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>414</td>
<td>2-Decanone</td>
<td>694-54-9</td>
<td>4.83</td>
<td>0.0309</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>415</td>
<td>Pentyl ether</td>
<td>694-55-2</td>
<td>3.14</td>
<td>0.0198</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>416</td>
<td>4-Methylisoxazole</td>
<td>693-93-6</td>
<td>1390</td>
<td>16.7300</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>417</td>
<td>2-Methylimidazole</td>
<td>693-98-1</td>
<td>286</td>
<td>3.1830</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>418</td>
<td>2-Adamantanone</td>
<td>700-58-3</td>
<td>60.80</td>
<td>0.4050</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>419</td>
<td>gamma-Decanolactone</td>
<td>706-14-9</td>
<td>18.00</td>
<td>0.1600</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>420</td>
<td>4,6-Dimethoxy-2-hydroxybenzaldehyde</td>
<td>708-76-9</td>
<td>2.68</td>
<td>0.0147</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>421</td>
<td>Propanil</td>
<td>709-08-8</td>
<td>8.50</td>
<td>0.0394</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>422</td>
<td>2,4,6-Tri-tert-butylphenol</td>
<td>732-26-3</td>
<td>0.0009</td>
<td>0.0002</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>423</td>
<td>3,4-Dichloro-1-butenol</td>
<td>760-23-6</td>
<td>8.18</td>
<td>0.0654</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>424</td>
<td>N,N-Dibutylformamide</td>
<td>761-65-0</td>
<td>89.30</td>
<td>0.5680</td>
<td>lowAT</td>
<td>highAT</td>
</tr>
<tr>
<td>425</td>
<td>2-Butyn-1-ol</td>
<td>764-01-2</td>
<td>10.10</td>
<td>0.1440</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
</tbody>
</table>
**TOXclass model, training set: continued**

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>LC(_{50}) (mg/l)</th>
<th>LC(_{50}) (mmol/l)</th>
<th>Exp. class</th>
<th>Toxic class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>426</td>
<td>2,5-Dimethyl-2,4-hexadiene</td>
<td>764-13-6</td>
<td>3.78</td>
<td>0.0343</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>427</td>
<td>1-Adamantanamine</td>
<td>768-94-5</td>
<td>25.00</td>
<td>0.1650</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>428</td>
<td>3-Cyanoo-4,6-dimethyl-2-hydroxypyridine</td>
<td>769-28-8</td>
<td>157.00</td>
<td>1.0600</td>
<td>lowAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>429</td>
<td>2,3,4,5,6-Pentahexafluorobenzene</td>
<td>429</td>
<td>37.10</td>
<td>0.2030</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>430</td>
<td>Triphenylphosphine oxide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>431</td>
<td>2-Hydroxyethyl acrylate</td>
<td>818-61-1</td>
<td>4.80</td>
<td>0.0141</td>
<td>highAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>432</td>
<td>1-Octyn-3-ol</td>
<td>818-72-4</td>
<td>0.1130</td>
<td>0.0503</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>433</td>
<td>2-Naphthole</td>
<td>821-55-6</td>
<td>15.20</td>
<td>0.1070</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>434</td>
<td>trans-1,2-Dichlorocyclohexane</td>
<td>822-86-6</td>
<td>18.40</td>
<td>0.1200</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>p-Phenylbenzaldehyde</td>
<td>831-82-3</td>
<td>4.05</td>
<td>0.0286</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>436</td>
<td>2-Hydroxyethyl methacrylate</td>
<td>868-77-9</td>
<td>227</td>
<td>1.7440</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>437</td>
<td>3-Bromothiophene</td>
<td>872-31-1</td>
<td>6.19</td>
<td>0.0380</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>438</td>
<td>2,4-Dichlorobenzaldehyde</td>
<td>874-82-0</td>
<td>1.80</td>
<td>0.0103</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>439</td>
<td>Ethyl 3-amino-1-methoxymethanesulfonic acid</td>
<td>886-96-2</td>
<td>79.00</td>
<td>0.3020</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>440</td>
<td>1,1,3,3,3-Hexafluoro-2-propanol</td>
<td>920-66-1</td>
<td>244</td>
<td>1.4520</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>1,5-Hexadien-4-ol</td>
<td>924-41-4</td>
<td>38.10</td>
<td>0.3880</td>
<td>highAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>442</td>
<td>cis-3-Hexen-1-ol</td>
<td>928-96-1</td>
<td>381</td>
<td>3.8040</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>443</td>
<td>trans-3-Hexen-1-ol</td>
<td>928-97-2</td>
<td>271</td>
<td>2.7060</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>444</td>
<td>2-Acrylo-1,3-butadiene</td>
<td>932-16-1</td>
<td>157</td>
<td>1.2750</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>445</td>
<td>4-Phenylpyridine</td>
<td>939-23-1</td>
<td>16.10</td>
<td>0.1040</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>Phenyl sulfoxide</td>
<td>945-51-7</td>
<td>87.30</td>
<td>0.3320</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>447</td>
<td>2-Hydroxypropyl acrylate</td>
<td>999-61-1</td>
<td>3.34</td>
<td>0.0260</td>
<td>highAT</td>
<td>lowAT</td>
<td></td>
</tr>
</tbody>
</table>
## TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>LC$_{50}$ (mg/l)</th>
<th>LC$_{50}$ (mmol/l)</th>
<th>Exp. class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>454</td>
<td>2-Amino-5-bromopyridine</td>
<td>1072-97-5</td>
<td>177</td>
<td>1.0230</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>455</td>
<td>Diethyl benzylphosphonate</td>
<td>1080-32-6</td>
<td>336</td>
<td>1.4720</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>457</td>
<td>4-Acetyl pyridine</td>
<td>1121-54-9</td>
<td>168</td>
<td>1.3870</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>458</td>
<td>Methyl p-chlorobenzoate</td>
<td>1126-46-1</td>
<td>11.00</td>
<td>0.0640</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>459</td>
<td>Butyl phenyl ether</td>
<td>1126-79-0</td>
<td>5.77</td>
<td>0.0380</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>460</td>
<td>Methyl 4-cyanobenzoate</td>
<td>1128-35-7</td>
<td>46.80</td>
<td>0.2900</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>461</td>
<td>Tetrachlorocatecho</td>
<td>1196-55-6</td>
<td>1.27</td>
<td>0.0051</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>462</td>
<td>alpha-Bromo-2,5-dimethoxycetophenone</td>
<td>1201-21-3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>464</td>
<td>3,4-Dimethyl-1-pentyn-3-ol</td>
<td>1482-15-1</td>
<td>205</td>
<td>1.8280</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>465</td>
<td>N-Vinylcarbazole</td>
<td>1484-13-5</td>
<td>0.0032</td>
<td>0.0000</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>466</td>
<td>3-Benzoxylaniline</td>
<td>1484-26-0</td>
<td>9.14</td>
<td>0.0459</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>467</td>
<td>Carbofuran</td>
<td>1763-66-2</td>
<td>0.8440</td>
<td>0.0038</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>468</td>
<td>tert-Butyl methyl ether</td>
<td>1634-04-4</td>
<td>672</td>
<td>7.6230</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>469</td>
<td>1,2-Decadiene</td>
<td>1641-16-1</td>
<td>0.29</td>
<td>0.0021</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>470</td>
<td>p-Phenylnaphthol</td>
<td>1689-82-3</td>
<td>1.17</td>
<td>0.0000</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>471</td>
<td>3,5-Diido-4-hydroxybenzonitrile</td>
<td>1689-83-4</td>
<td>6.80</td>
<td>0.0183</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>472</td>
<td>3,5-Dibromo-4-hydroxybenzonitrile</td>
<td>1689-84-5</td>
<td>12.60</td>
<td>0.0455</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>473</td>
<td>2-Hydroxyacetonic acid</td>
<td>1740-19-8</td>
<td>2.10</td>
<td>0.0070</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>474</td>
<td>tert-Butyl alcohol</td>
<td>1745-81-9</td>
<td>15.00</td>
<td>0.1120</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>475</td>
<td>4-Bromohexaldehyde</td>
<td>1761-61-1</td>
<td>1.30</td>
<td>0.0005</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>476</td>
<td>3-Chloro-2,4-dichloromethyl-1-propene</td>
<td>1871-57-4</td>
<td>0.19</td>
<td>0.0015</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>477</td>
<td>3,5-Dichloro-4-hydroxybenzonitrile</td>
<td>1891-95-8</td>
<td>24.30</td>
<td>0.1290</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>478</td>
<td>Di-n-butylterephthalate</td>
<td>1062-75-0</td>
<td>0.59</td>
<td>0.0021</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>479</td>
<td>4,4'-Dihydroxydiphenylether</td>
<td>1965-09-9</td>
<td>5.78</td>
<td>0.0286</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>480</td>
<td>2,6-Dichlorobenzamide</td>
<td>2008-58-4</td>
<td>469</td>
<td>2.1680</td>
<td>low AT</td>
<td>high AT</td>
</tr>
<tr>
<td>481</td>
<td>n-Decylamine</td>
<td>2016-57-1</td>
<td>1.03</td>
<td>0.0006</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>482</td>
<td>Aminocarb</td>
<td>2032-56-9</td>
<td>1.95</td>
<td>0.0094</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>483</td>
<td>2,4,5-Trichloroimidazole</td>
<td>2034-23-2</td>
<td>6.12</td>
<td>0.0201</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>485</td>
<td>(+)-L-Pentyn-2-ol</td>
<td>2117-11-5</td>
<td>35.10</td>
<td>0.1170</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>486</td>
<td>4-Chlorocatechol</td>
<td>2138-22-9</td>
<td>1.58</td>
<td>0.0109</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>487</td>
<td>Methyl 2,4-dihydroxybenzoate</td>
<td>2150-47-2</td>
<td>45.80</td>
<td>0.2720</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>489</td>
<td>Pentachloropyridine</td>
<td>2176-62-7</td>
<td>0.47</td>
<td>0.0019</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>490</td>
<td>(1R,2S,5R)-(-)-menthol</td>
<td>2216-51-5</td>
<td>18.90</td>
<td>0.1210</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>491</td>
<td>1-([3]-Toluenesulfonyl)imidazole</td>
<td>2232-08-8</td>
<td>41.80</td>
<td>0.1880</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>492</td>
<td>2',4'-Dichloroacetophenone</td>
<td>2234-16-1</td>
<td>11.70</td>
<td>0.0689</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>493</td>
<td>n-Octyl cyanide</td>
<td>2243-26-8</td>
<td>5.25</td>
<td>0.0377</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>494</td>
<td>alpha, alpha, alpha-4-Tetrafluoro-n-toluidine</td>
<td>2357-47-3</td>
<td>30.10</td>
<td>0.1600</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>495</td>
<td>trans-2-Phenyl-1-cyclohexanol</td>
<td>2362-61-0</td>
<td>44.40</td>
<td>0.2520</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>496</td>
<td>2-Ethoxyethyl methacrylate</td>
<td>2370-63-0</td>
<td>27.70</td>
<td>0.1750</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>497</td>
<td>2,4,6-Trimethylphenol</td>
<td>2416-91-6</td>
<td>8.20</td>
<td>0.0602</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>498</td>
<td>n-Undecyl cyanide</td>
<td>2437-25-4</td>
<td>0.43</td>
<td>0.0024</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>499</td>
<td>o-Methoxybenzamide</td>
<td>2438-77-2</td>
<td>120</td>
<td>0.7940</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>501</td>
<td>Tetrahydrofurfurylmethacrylate</td>
<td>2450-24-5</td>
<td>34.70</td>
<td>0.2010</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>502</td>
<td>4,5-Dichloroguaiacol</td>
<td>2460-49-3</td>
<td>4.47</td>
<td>0.0202</td>
<td>high AT</td>
<td>high AT</td>
</tr>
</tbody>
</table>
### TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>LC$_{50}$ (mg/l)</th>
<th>LC$_{50}$ (mmol/l)</th>
<th>Exp. tox class</th>
<th>Pred. tox class</th>
</tr>
</thead>
<tbody>
<tr>
<td>503</td>
<td>Benzyl methacrylate</td>
<td>2495-37-6</td>
<td>4.67</td>
<td>0.0265</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>504</td>
<td>Hexyl acrylate</td>
<td>2499-05-8</td>
<td>1.11</td>
<td>0.0071</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>505</td>
<td>o-(tert-butyl)-Phenyl-N-methylcarbamate</td>
<td>2606-83-7</td>
<td>10.00</td>
<td>0.0482</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>507</td>
<td>1-Benzylpiperazine</td>
<td>2750-28-6</td>
<td>47.40</td>
<td>0.2690</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>508</td>
<td>3-(5-Pyridyl)-1-propanol</td>
<td>2859-67-8</td>
<td>150</td>
<td>1.0930</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>509</td>
<td>Tridecylamine</td>
<td>2869-34-3</td>
<td>0.0654</td>
<td>0.0003</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>510</td>
<td>2-Amino-4-chlorobenzophenone</td>
<td>2894-51-1</td>
<td>2.12</td>
<td>0.0092</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>511</td>
<td>Methyl 2,5-dichlorobenzoate</td>
<td>2895-00-3</td>
<td>14.00</td>
<td>0.0583</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>513</td>
<td>5-Bromoanillin</td>
<td>2973-76-4</td>
<td>59.70</td>
<td>0.2580</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>514</td>
<td>Cyclohexyl acrylate</td>
<td>3066-71-5</td>
<td>1.48</td>
<td>0.0006</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>516</td>
<td>Triethyl nitritocarboxylate</td>
<td>3206-31-3</td>
<td>13.30</td>
<td>0.0570</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>517</td>
<td>4,5-Dichlorocatechol</td>
<td>3428-24-8</td>
<td>0.89</td>
<td>0.0050</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>518</td>
<td>2,3,5,6-Tetrachloroaniline</td>
<td>3481-20-7</td>
<td>0.27</td>
<td>0.0012</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>519</td>
<td>2,6-Diphenylpyridine</td>
<td>3558-69-8</td>
<td>0.21</td>
<td>0.0009</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>520</td>
<td>1,3-Dichloro-4,6-dinitrobenzene</td>
<td>3698-83-7</td>
<td>0.0456</td>
<td>0.0002</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>523</td>
<td>2,3-Dimethylfuranaldehyde</td>
<td>3914-76-1</td>
<td>16.00</td>
<td>0.1400</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>524</td>
<td>2-Decyn-1-ol</td>
<td>4117-14-0</td>
<td>1.07</td>
<td>0.0009</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>525</td>
<td>5-Chloro-2-pyridinol</td>
<td>4214-70-3</td>
<td>11.40</td>
<td>8.8000</td>
<td>low AT</td>
<td>low AT</td>
</tr>
<tr>
<td>526</td>
<td>Isopropyl disulfide</td>
<td>4253-88-8</td>
<td>8.31</td>
<td>0.0533</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>527</td>
<td>2,4,5-Trimethoxybenzaldehyde</td>
<td>4450-80-0</td>
<td>49.50</td>
<td>0.2320</td>
<td>high AT</td>
<td>high AT</td>
</tr>
<tr>
<td>528</td>
<td>Isopropyl methacrylate</td>
<td>4655-34-9</td>
<td>38.00</td>
<td>0.2090</td>
<td>high AT</td>
<td>low AT</td>
</tr>
<tr>
<td>529</td>
<td>1-Hexen-3-ol</td>
<td>4798-44-1</td>
<td>30.40</td>
<td>0.3040</td>
<td>high AT</td>
<td>low AT</td>
</tr>
</tbody>
</table>
### TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>LC&lt;sub&gt;50&lt;/sub&gt; (mg/l)</th>
<th>LC&lt;sub&gt;50&lt;/sub&gt; (mmol/l)</th>
<th>Exp. class</th>
<th>Tox class</th>
<th>Pred. Tox class</th>
</tr>
</thead>
<tbody>
<tr>
<td>530</td>
<td>2,3,4,5-Tetrachlorophenol</td>
<td>4901-51-3</td>
<td>0.41</td>
<td>0.0018</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>531</td>
<td>1,2-bis(4-Pyridyl)ethane</td>
<td>4916-57-8</td>
<td>151</td>
<td>0.8200</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>532</td>
<td>1,3-Diethyl-2-thiobarbituric acid</td>
<td>5217-47-0</td>
<td>4510</td>
<td>22.5200</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>533</td>
<td>Dimethyl nitroterephthalate</td>
<td>5292-45-5</td>
<td>6.52</td>
<td>0.0273</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>534</td>
<td>5-Chloro-2-mercaptobenzothiazole</td>
<td>5331-91-9</td>
<td>3.21</td>
<td>0.0159</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>535</td>
<td>Dimethyl aminoterephthalate</td>
<td>5372-81-6</td>
<td>8.94</td>
<td>0.0427</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>536</td>
<td>3,6-Dithiaoctane</td>
<td>5205-75-5</td>
<td>60.20</td>
<td>0.4100</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>537</td>
<td>Dimethylaminopropyl chloride, hydrochloride</td>
<td>5407-04-5</td>
<td>133</td>
<td>0.8410</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>538</td>
<td>4'-Chloro-3'-nitroacetophenone</td>
<td>5465-65-6</td>
<td>5.50</td>
<td>0.0276</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>539</td>
<td>2-Amino-4-Chloro-6-methylpyrimidine</td>
<td>5683-33-0</td>
<td>127</td>
<td>1.0400</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>540</td>
<td>2,4-Dimethylpyridine</td>
<td>5673-07-4</td>
<td>20.20</td>
<td>0.1330</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>541</td>
<td>2-Dimethylaminopyridine</td>
<td>5683-33-0</td>
<td>127</td>
<td>1.0400</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>542</td>
<td>2,2-Dimethyl-1-propylamine</td>
<td>5813-64-9</td>
<td>475</td>
<td>5.4400</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>543</td>
<td>Isopimaric acid</td>
<td>5835-28-7</td>
<td>0.87</td>
<td>0.0029</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>544</td>
<td>2-Amino-5-chlorobenzonitrile</td>
<td>5922-60-1</td>
<td>28.60</td>
<td>0.1870</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>545</td>
<td>1,1,1-Trichloro-2-methyl-2-propanol(hydrate) (2:1)</td>
<td>6001-64-5</td>
<td>135</td>
<td>0.3620</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>546</td>
<td>2-Dodecanone</td>
<td>6175-49-1</td>
<td>1.18</td>
<td>0.0064</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>547</td>
<td>4-Dimethylaminocinnamaldehyde</td>
<td>6203-18-5</td>
<td>5.90</td>
<td>0.0397</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>548</td>
<td>1,3,5-Trichloro-2,4-dinitrobenzene</td>
<td>6284-83-9</td>
<td>0.2220</td>
<td>0.0008</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>550</td>
<td>2-Chloro-5-nitrobenzaldehyde</td>
<td>6361-21-3</td>
<td>3.87</td>
<td>0.0209</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>551</td>
<td>2-Chloro-6-methylbenzonitrile</td>
<td>6575-09-3</td>
<td>15.10</td>
<td>0.0996</td>
<td>highAT</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>552</td>
<td>2-Bromo-3-pyridinol</td>
<td>6602-32-0</td>
<td>469</td>
<td>2.6970</td>
<td>lowAT</td>
<td>lowAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. class</td>
<td>Pred. class</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------</td>
<td>-------</td>
<td>------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>2-Chloro-3-pyridinol</td>
<td>6636-78-8</td>
<td>622</td>
<td>4.8010</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>Tripropargylamine</td>
<td>6921-28-5</td>
<td>296</td>
<td>2.2560</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>N,N-bis[2,2-Diethoxyethyl]methylamine</td>
<td>6948-86-3</td>
<td>635</td>
<td>2.4110</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>556</td>
<td>1,4-bis[3-Aminopropyl]piperazine</td>
<td>7200-38-3</td>
<td>3100</td>
<td>15.4700</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>557</td>
<td>3-Hydroxy-3,7,11-trimethyl-1,6,10-dodecatriène</td>
<td>7212-44-4</td>
<td>1.43</td>
<td>0.0064</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>558</td>
<td>1-(2-Chloroethyl)pyrrolidine-HCl</td>
<td>7250-67-1</td>
<td>153</td>
<td>0.0000</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>559</td>
<td>n-Undecylamine</td>
<td>7307-55-3</td>
<td>0.21</td>
<td>0.0012</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>560</td>
<td>1-Heptyl-3-ol</td>
<td>7383-19-9</td>
<td>1.76</td>
<td>0.0157</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>561</td>
<td>p-Ethoxybenzaldehyde</td>
<td>10031-82-0</td>
<td>28.10</td>
<td>0.1870</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>562</td>
<td>[(R)-endo]-(+)-3-Bromocamphor</td>
<td>10383-06-8</td>
<td>68.30</td>
<td>0.2960</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>563</td>
<td>Resmethrin</td>
<td>10453-86-8</td>
<td>0.0002</td>
<td>0.0000</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>565</td>
<td>alpha, alpha', alpha', alpha'-Tetramethylsiline</td>
<td>13280-15-9</td>
<td>0.4370</td>
<td>0.0001</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>566</td>
<td>2',3',4'-Trichlorocyclohexone</td>
<td>13608-87-2</td>
<td>2.00</td>
<td>0.0090</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>567</td>
<td>2',3',4'-Trichloroacetophenone</td>
<td>13908-73-4</td>
<td>199</td>
<td>0.9470</td>
<td>lowAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>568</td>
<td>Diethyl chloroacetate</td>
<td>14064-10-9</td>
<td>0.95</td>
<td>0.0049</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>569</td>
<td>N-Ethylbenzaldehyde</td>
<td>14321-27-8</td>
<td>57.10</td>
<td>0.4220</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>571</td>
<td>4-Bromophenyl 3-pyridyl ketone</td>
<td>14548-55-9</td>
<td>20.40</td>
<td>0.0778</td>
<td>lowAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>572</td>
<td>4-Benzoylpyridine</td>
<td>14548-86-0</td>
<td>103</td>
<td>0.5620</td>
<td>lowAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>573</td>
<td>2,2,5,5'-Tetrame thy1,1,1,-dihydroxybutane</td>
<td>15015-13-9</td>
<td>168</td>
<td>1.3100</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>574</td>
<td>3-Hydroxy-2-nitropyridine</td>
<td>15128-92-2</td>
<td>167</td>
<td>1.1920</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>575</td>
<td>Alachlor</td>
<td>15972-60-8</td>
<td>5.00</td>
<td>0.0185</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>576</td>
<td>4-Octylaniline</td>
<td>16245-79-7</td>
<td>0.12</td>
<td>0.0006</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC₅₀ (mg/l)</td>
<td>LC₅₀ (mmol/l)</td>
<td>Exp. class</td>
<td>Toxic class</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>577</td>
<td>Methomyl</td>
<td>16752-77-5</td>
<td>2.11</td>
<td>0.0130</td>
<td>highAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>578</td>
<td>6-Chloro-2-pyridinol</td>
<td>16879-02-0</td>
<td>21.4</td>
<td>1.6520</td>
<td>lowAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>579</td>
<td>3-Amino-5,6-dimethyl-1,2,4-triazine</td>
<td>17584-12-2</td>
<td>962</td>
<td>7.6680</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>581</td>
<td>4-(Diethylamino)salicylaldehyde</td>
<td>17754-90-4</td>
<td>5.36</td>
<td>0.0277</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>582</td>
<td>6-Chloro-2-picoline</td>
<td>18308-63-3</td>
<td>232</td>
<td>1.8190</td>
<td>lowAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>583</td>
<td>3,6-Dimethyl-1-heptyl-3-ol</td>
<td>19540-98-5</td>
<td>49.00</td>
<td>0.3490</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>584</td>
<td>2,4,5-Trimethylazéole</td>
<td>20662-84-4</td>
<td>419</td>
<td>4.0400</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>585</td>
<td>4-Dimethylamino-3-methyl-2-butanone</td>
<td>22104-62-7</td>
<td>8.50</td>
<td>0.0658</td>
<td>highAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>586</td>
<td>Oxamyl</td>
<td>23135-22-0</td>
<td>6.78</td>
<td>0.0309</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>589</td>
<td>2,6-Diisopropylaniline</td>
<td>24541-01-5</td>
<td>15.30</td>
<td>0.0863</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>593</td>
<td>2-Methyl-3,3,4,4-tetrafluoro-2-butanone</td>
<td>25553-26-2</td>
<td>582</td>
<td>3.6350</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>594</td>
<td>(+)-α-s-cis-Butylanine</td>
<td>13962-284-6</td>
<td>275</td>
<td>3.7600</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>595</td>
<td>N-(3-Methoxypropyl)-3,4,5-trimethoxybenzylamine</td>
<td>31274-54-9</td>
<td>136</td>
<td>0.5050</td>
<td>lowAT</td>
<td>lowAT</td>
<td></td>
</tr>
<tr>
<td>596</td>
<td>2-[Bromomethyl]tetrahydro-2H-pyran</td>
<td>31473-82-5</td>
<td>205</td>
<td>1.1450</td>
<td>lowAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>597</td>
<td>4-Decylaniline</td>
<td>37529-90-9</td>
<td>0.0620</td>
<td>0.0003</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>598</td>
<td>4-Hexylcyclohexane</td>
<td>39005-57-2</td>
<td>3.01</td>
<td>0.0156</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>599</td>
<td>Methyl 4-chloro-2-nitrobenzoate</td>
<td>42087-98-9</td>
<td>27.70</td>
<td>0.1280</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>5-Hydroxy-2-nitrobenzaldehyde</td>
<td>42454-68-8</td>
<td>41.90</td>
<td>0.2510</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>601</td>
<td>Fenvalerate</td>
<td>51630-58-1</td>
<td>0.0051</td>
<td>0.0000</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>602</td>
<td>Permethrin</td>
<td>50815-53-1</td>
<td>0.0160</td>
<td>0.0000</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
<tr>
<td>603</td>
<td>3,8-Dithiadiene</td>
<td>54576-32-8</td>
<td>6.06</td>
<td>0.0340</td>
<td>highAT</td>
<td>highAT</td>
<td></td>
</tr>
</tbody>
</table>
## TOXclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>LC50 (mg/l)</th>
<th>LC50 (mmol/l)</th>
<th>Exp. class</th>
<th>Pred. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>604</td>
<td>2’-(Octyloxy)acetanilide</td>
<td>55762-61-5</td>
<td>0.45</td>
<td>0.0017</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>605</td>
<td>p-(tert-Butyl)benzamide</td>
<td>56108-12-4</td>
<td>31.90</td>
<td>0.1800</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>606</td>
<td>2,5-Dithiadecane</td>
<td>56348-40-4</td>
<td>10.10</td>
<td>0.0566</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>607</td>
<td>dl-3-Butyn-2-ol</td>
<td>67337-13-5</td>
<td>11.70</td>
<td>0.1670</td>
<td>highAT</td>
<td>lowAT</td>
</tr>
<tr>
<td>610</td>
<td>3-(4-tert-Butylphenoxy)benzaldehyde</td>
<td>69770-23-6</td>
<td>0.37</td>
<td>0.0015</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>611</td>
<td>Flucythrinate</td>
<td>70124-77-5</td>
<td>0.0002</td>
<td>0.0000</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>613</td>
<td>3-(3,4-Dichlorophenoxy)benzaldehyde</td>
<td>79124-76-8</td>
<td>0.30</td>
<td>0.0011</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>614</td>
<td>2,4-Dinitro-1-naphthol sodium salt dihydrate (Martius yellow)</td>
<td>101836-92-4</td>
<td>4.24</td>
<td>0.0145</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>615</td>
<td>t-Butylstyrene</td>
<td>1746-23-2</td>
<td>0.49</td>
<td>0.0051</td>
<td>highAT</td>
<td>highAT</td>
</tr>
<tr>
<td>617</td>
<td>Chloromethyl styrene</td>
<td>30030-25-2</td>
<td>0.31</td>
<td>0.0020</td>
<td>highAT</td>
<td>highAT</td>
</tr>
</tbody>
</table>
### F.2 MOA class model: Training set

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>Exp. class</th>
<th>MOA class</th>
<th>Pred. class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>p-Chlorophenyl-o-nitrophenyl ether</td>
<td>39145-47-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Di-n-butylinophosphate</td>
<td>31263-90-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1,1-Diphenyl-2-propyn-1-ol</td>
<td>39233-52-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4,7-Dithiadecane</td>
<td>22037-97-4</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Phenobarbital</td>
<td>50-06-6</td>
<td>E</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2,4-Dinitrophenol</td>
<td>51-28-5</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Urethane</td>
<td>51-76-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1,1-Dimethyl hydrazine</td>
<td>57-14-7</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Pentobarbital sodium</td>
<td>57-33-0</td>
<td>E</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Amobarbital</td>
<td>57-48-2</td>
<td>E</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Caffeine</td>
<td>58-08-2</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>2-Methyl-1,4-naphthoquinone</td>
<td>58-27-5</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>4-Chloro-3-methyl phenol</td>
<td>59-50-7</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Tolazoline hydrochloride</td>
<td>59-97-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Amphetamine sulfate</td>
<td>60-13-9</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Diethyl ether</td>
<td>60-23-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Strychnine hemisulphate salt</td>
<td>60-41-3</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Aniline</td>
<td>62-53-3</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Carbaryl (sevin)</td>
<td>63-25-2</td>
<td>H</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Ethanol</td>
<td>64-17-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Nicotine sulfate</td>
<td>65-30-5</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### MOAclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. class</th>
<th>MOA</th>
<th>Pred. class</th>
<th>MOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>Hexanal</td>
<td>66-25-1</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Methanol</td>
<td>67-56-1</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>2-Propanol</td>
<td>67-63-0</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Hexachloroethane</td>
<td>67-72-1</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>4'-Aminopropiophenone</td>
<td>70-09-9</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>1-Propanol</td>
<td>71-23-8</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>n-Butanol</td>
<td>71-26-3</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>1-Pentanol</td>
<td>71-41-0</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>1,1,1-Trichloroethane</td>
<td>71-55-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Thiopental, sodium salt</td>
<td>71-73-8</td>
<td>E</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>Dichloromethane</td>
<td>75-09-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>2-Methyl-2-propanol</td>
<td>75-65-0</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>Pentachloroethane</td>
<td>76-01-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>3-Methyl-3-pentanol</td>
<td>77-74-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>3-Methyl-1-pentyn-3-ol</td>
<td>77-75-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>1-Ethynyl-cyclohexanol</td>
<td>78-27-3</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>2-Methyl-1-propanol</td>
<td>78-83-1</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>1,2-Dichloropropane</td>
<td>78-87-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>1,2-Diaminopropane</td>
<td>78-00-0</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>2-Butanol</td>
<td>78-92-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>1,1,2-Trichloroethane</td>
<td>79-00-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>Trichloroethylene</td>
<td>79-01-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MOA class model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. MO class</th>
<th>Pred. MO class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>Methyl acetate</td>
<td>79-20-9</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>1,1,2,2-Tetrachloroethane</td>
<td>79-34-5</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>\beta-ionone</td>
<td>79-77-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>3-Methylindole</td>
<td>83-34-1</td>
<td>D</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>Rotenone</td>
<td>83-73-4</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Hexachloro-1,3-butadiene</td>
<td>87-68-3</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Pentachlorophenol</td>
<td>87-86-5</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>2,4,6-Trichlorophenol</td>
<td>88-06-2</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>3-Trifluoromethyl-1-nitrophenol</td>
<td>88-30-2</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>1-Naphthol</td>
<td>90-15-3</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>Naphthalene</td>
<td>91-20-3</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>N,N-Diethylaniline</td>
<td>91-66-7</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>2-(N-Ethyl-\text{-}m-toluidino)ethanol</td>
<td>91-88-3</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>Ethyl (p)-aminobenzoate</td>
<td>94-09-7</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>(o)-Cresol</td>
<td>95-48-7</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>2-Chloroaniline</td>
<td>95-51-2</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>2-Chlorophenol</td>
<td>95-57-8</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>3,4-Dichloroaniline</td>
<td>95-76-1</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>Allyl methacrylate</td>
<td>96-05-9</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>1,2,3-Trichloropropane</td>
<td>96-18-4</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>2-(Dipropylamino)ethanol</td>
<td>96-80-0</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>133</td>
<td>2,4-Dinitroaniline</td>
<td>97-02-9</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>
### MOAclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance name</th>
<th>CASRN</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>138</td>
<td>Acetophenone</td>
<td>98-86-2</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>142</td>
<td>N,N-Dimethyl-p-toluidine</td>
<td>99-97-8</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>144</td>
<td>p-Nitrophenol</td>
<td>100-42-7</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>147</td>
<td>N,N-Diethylethanolamine</td>
<td>100-57-8</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>150</td>
<td>Benzenediahyde</td>
<td>100-52-7</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>153</td>
<td>2-Cyanopyridine</td>
<td>100-70-9</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>154</td>
<td>2-Ethylpyridine</td>
<td>100-71-0</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>161</td>
<td>Benzyl-tert-butanol</td>
<td>103-05-9</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>169</td>
<td>2-Ethyl-1-hexanol</td>
<td>104-76-7</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>171</td>
<td>5-Ethyl-2-methylpyridine</td>
<td>104-90-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>174</td>
<td>2,4-Dimethylphenol</td>
<td>106-47-9</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>177</td>
<td>p-Bromoaniline</td>
<td>106-10-1</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>180</td>
<td>4-Chloroaniline</td>
<td>106-17-8</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>182</td>
<td>4-Toluidine</td>
<td>106-19-0</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>183</td>
<td>Isobutyl acrylate</td>
<td>106-63-8</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>184</td>
<td>1-Bromopropane</td>
<td>106-94-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>185</td>
<td>Acrolein</td>
<td>107-42-8</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>186</td>
<td>1,2-Dichloroethane</td>
<td>107-06-2</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>187</td>
<td>2-Chloroethanol</td>
<td>107-47-3</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>189</td>
<td>Propionitrile</td>
<td>107-12-0</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>192</td>
<td>Allyl alcohol</td>
<td>107-18-6</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>193</td>
<td>2-Propyn-1-ol</td>
<td>107-19-7</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>
MOAclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. class</th>
<th>MOA class</th>
<th>Pred. class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>195</td>
<td>2-Methyl-2,4-pentanediol</td>
<td>107-41-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>4-Methyl-2-pentanone</td>
<td>108-10-1</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>Isopropyl ether</td>
<td>108-20-3</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>202</td>
<td>4-Picoline</td>
<td>108-80-1</td>
<td>B</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>204</td>
<td>Cyclohexanol</td>
<td>108-93-0</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>206</td>
<td>Phenol</td>
<td>108-95-2</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>209</td>
<td>2-Picoline</td>
<td>109-06-8</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>212</td>
<td>1,3-Dibromopropane</td>
<td>109-64-8</td>
<td>D</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>1-Bromobutane</td>
<td>109-65-9</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>1,3-Diaminopropane</td>
<td>109-76-2</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>226</td>
<td>2-Heptanone</td>
<td>110-43-0</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>1,4-Dichlorobutane</td>
<td>110-26-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>Valeraldehyde</td>
<td>110-62-3</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>231</td>
<td>2-Butyne-1,4-diol</td>
<td>110-65-6</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>234</td>
<td>Pyridine</td>
<td>110-96-1</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>2-Octanone</td>
<td>111-13-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>2-Ethoxymethyl acetate</td>
<td>111-15-9</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>1-Bromohexane</td>
<td>111-25-1</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>1-Hexanol</td>
<td>111-27-3</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>243</td>
<td>2-Hydroxyethyl ether</td>
<td>111-46-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>247</td>
<td>1-Heptanol</td>
<td>111-70-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>248</td>
<td>1-Bromooctane</td>
<td>111-83-1</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### MOA class model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1-Octanol</td>
<td>111-87-5</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>251</td>
<td>2-(2-Ethoxyethoxy)ethanol</td>
<td>111-90-0</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>253</td>
<td>2-Undecanone</td>
<td>112-12-9</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>255</td>
<td>Triethylene glycol</td>
<td>112-25-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>1-Undecanol</td>
<td>112-30-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>258</td>
<td>Propoxur (Baygon)</td>
<td>114-26-1</td>
<td>H</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>259</td>
<td>2-Methyl-3-buten-2-ol</td>
<td>115-19-5</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>2,2,2-Trichloroethanol</td>
<td>115-20-8</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>261</td>
<td>Dicofol (Kelthane)</td>
<td>115-32-2</td>
<td>G</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>264</td>
<td>Aldicarb</td>
<td>116-06-3</td>
<td>H</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>268</td>
<td>Ethyl salicylate</td>
<td>118-61-6</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>269</td>
<td>2,4,6-Trichlorophenol</td>
<td>118-73-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>270</td>
<td>4-Amino2-nitrophenol</td>
<td>119-34-6</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>272</td>
<td>N-Phenyldiethanolamine</td>
<td>120-07-0</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>273</td>
<td>4-(Dethylamino)benzaldehyde</td>
<td>120-21-8</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>274</td>
<td>Catechol</td>
<td>120-80-9</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>1,2,4-Trichlorobenzene</td>
<td>120-82-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>276</td>
<td>2,4-Dichlorophenol</td>
<td>120-83-2</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>280</td>
<td>N,N-Dimethylaniline</td>
<td>121-60-7</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>283</td>
<td>2-Chloro-4-nitroaniline</td>
<td>121-87-9</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>284</td>
<td>p-Isopropyl benzaldehyde</td>
<td>122-03-2</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>286</td>
<td>2-Phenoxyethanol</td>
<td>122-56-4</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>
## MOA class model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>288</td>
<td>2-Methylvaleraldehyde</td>
<td>123-15-9</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>296</td>
<td>Tributyl phosphate</td>
<td>126-73-8</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>298</td>
<td>1-Chloro-2-propanol</td>
<td>127-00-4</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>299</td>
<td>Tetrachloroethylene</td>
<td>127-18-4</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>300</td>
<td>2-Phenyl-3-butyn-2-ol</td>
<td>127-66-2</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>304</td>
<td>Phenyl 4-aminosalicylate</td>
<td>133-11-9</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>314</td>
<td>1,3-Dichloropropane</td>
<td>142-28-9</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>317</td>
<td>Butyl ether</td>
<td>142-96-1</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>318</td>
<td>1-Nonanol</td>
<td>143-08-8</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>322</td>
<td>3-Methoxyphenol</td>
<td>150-15-6</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>323</td>
<td>4-Methoxyphenol</td>
<td>150-76-5</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>324</td>
<td>p-Dimethoxybenzene</td>
<td>150-78-7</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>325</td>
<td>2,3-Benzofuran</td>
<td>271-89-6</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>329</td>
<td>Secobarbital, sodium salt</td>
<td>309-43-3</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>331</td>
<td>2,5-Dinitrophenol</td>
<td>329-71-5</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>339</td>
<td>$\alpha,\alpha,\alpha$-Tetrafluoro-$\alpha$-toluidine</td>
<td>393-90-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>341</td>
<td>$\alpha,\alpha,\alpha$-Trifluoro-$\alpha$-tolunitrile</td>
<td>447-90-9</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>347</td>
<td>Cineole</td>
<td>470-82-6</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>350</td>
<td>2,3-Dihydrobenzofuran</td>
<td>496-16-3</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>356</td>
<td>2,3-Dimethyl-1,3-butadiene</td>
<td>513-81-5</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>358</td>
<td>Flavone</td>
<td>525-82-6</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>359</td>
<td>2,4,6-Trimethylphenol</td>
<td>527-69-6</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>
### MOA class model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance</th>
<th>Chemical name</th>
<th>CASRN</th>
<th>Exp. MOA</th>
<th>Pred. MOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>363</td>
<td>4,6-Dinitro-o-cresol</td>
<td>534-52-1</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>367</td>
<td>1,3-Dichloropropene</td>
<td>542-75-6</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>371</td>
<td>4-Nitrobenzaldehyde</td>
<td>555-16-8</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>373</td>
<td>2,6-Dinitrophenol</td>
<td>573-56-8</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>377</td>
<td>Isovaleraldehyde</td>
<td>590-86-3</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>378</td>
<td>2-Hexanone</td>
<td>561-78-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>379</td>
<td>2,4-Hexadiene</td>
<td>592-46-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>2,4-Dimethyl-3-pentanol</td>
<td>600-36-2</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>389</td>
<td>Pentabromophenol</td>
<td>608-71-9</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>393</td>
<td>2-Chloro-4-methylaniline</td>
<td>615-65-6</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>394</td>
<td>4-Ethoxy-2-nitroaniline</td>
<td>616-86-4</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>396</td>
<td>4-Nitrobenzamide</td>
<td>619-80-7</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>398</td>
<td>Benzyl sulfoxide</td>
<td>621-08-9</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>4-(2-Hydroxyethyl)morpholine</td>
<td>622-40-2</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>403</td>
<td>2,5-Dimethylfuran</td>
<td>625-86-5</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>404</td>
<td>1,5-Dichloropentane</td>
<td>628-76-2</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>1-Bromohexane</td>
<td>629-34-9</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>406</td>
<td>Propyl disulfide</td>
<td>629-19-6</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>415</td>
<td>Pentyl ether</td>
<td>693-65-2</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>4-Methylloxazide</td>
<td>693-93-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>417</td>
<td>2-Methylimidazole</td>
<td>693-98-1</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>419</td>
<td>gamma-Decanolactone</td>
<td>706-14-9</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>
### MOA class model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>423</td>
<td>3,4-Dichloro-1-butene</td>
<td>760-23-6</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>425</td>
<td>2-Butyn-1-ol</td>
<td>764-01-2</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>426</td>
<td>2,5-Dimethyl-2,4-hexadiene</td>
<td>764-13-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>428</td>
<td>3-Cyano-4,6-dimethyl-2-hydroxypyridine</td>
<td>769-28-8</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>429</td>
<td>2,3,4,5,6-Pentafluoromellitine</td>
<td>771-60-8</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>433</td>
<td>2-Hydroxyethyl acrylate</td>
<td>818-61-1</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>434</td>
<td>1-Octyn-3-ol</td>
<td>818-72-4</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>2-Nonanone</td>
<td>821-35-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>436</td>
<td>trans-1,2-Dichlorocyclohexane</td>
<td>822-86-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>439</td>
<td>3-Bromo-thiophene</td>
<td>872-51-1</td>
<td>D</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>443</td>
<td>Ethyl 3-aminobenzene methanesulfonic acid salt</td>
<td>886-86-2</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>444</td>
<td>1,1,1,3,3,3-Hexafluoro-2-propanol</td>
<td>920-66-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>445</td>
<td>1,5-Hexadien-3-ol</td>
<td>924-41-4</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>3-Butyn-1-ol</td>
<td>927-74-2</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>447</td>
<td>cis-3-Hexen-1-ol</td>
<td>928-66-1</td>
<td>A</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>448</td>
<td>trans-3-Hexen-1-ol</td>
<td>928-97-2</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>4-Phenyllpyridine</td>
<td>903-23-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>451</td>
<td>Phenyl sulfide</td>
<td>915-51-7</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>453</td>
<td>2-Hydroxypropyl acrylate</td>
<td>999-61-1</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>454</td>
<td>2-Amino-5-bromopyridine</td>
<td>1072-97-5</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>455</td>
<td>Diethyl benzylphosphonate</td>
<td>1080-32-6</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>457</td>
<td>4-Acetyl pyridine</td>
<td>1122-54-9</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>Exp. class</td>
<td>MOA class</td>
<td>Pred. class</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------</td>
<td>-----------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>458</td>
<td>Methyl p-chlorobenzoate</td>
<td>1126-46-1</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>459</td>
<td>Butyl phenyl ether</td>
<td>1126-73-0</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>461</td>
<td>Tetrachlorocatehol</td>
<td>1198-55-6</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>464</td>
<td>3,4-Dimethyl-1-pentyn-3-ol</td>
<td>1482-15-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>465</td>
<td>N-Vinylcarbazole</td>
<td>1484-13-5</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>467</td>
<td>Carbofuran</td>
<td>1563-06-2</td>
<td>H</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>469</td>
<td>1,9-Decadiene</td>
<td>1647-16-1</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>470</td>
<td>p-Phenylaoxonol</td>
<td>1688-83-3</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>476</td>
<td>3-Chloro-2-chloromethyl-1-propene</td>
<td>1871-57-4</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>480</td>
<td>2,6-Dichlorobenzamide</td>
<td>2008-56-4</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>482</td>
<td>Aminocarb</td>
<td>2032-29-9</td>
<td>H</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>485</td>
<td>(+-) 4-Pentyn-2-ol</td>
<td>2117-11-5</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>486</td>
<td>4-Chlorocatehol</td>
<td>2138-22-9</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>490</td>
<td>(1R,2S,5R)-(+) -menthol</td>
<td>2216-51-5</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>492</td>
<td>2',4'-Dichlorocetophene</td>
<td>2234-16-4</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>493</td>
<td>n-Octyl cyanide</td>
<td>2243-27-8A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>494</td>
<td>alpha, alpha, alpha-L-Tetrafluoro-m-toluidine</td>
<td>2357-47-3</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>495</td>
<td>trans-2-Phenyl-1-cyclohexanol</td>
<td>2362-26-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>496</td>
<td>2-Ethoxethyl methacrylate</td>
<td>2370-63-0</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>Tetrahydrofururyl methacrylate</td>
<td>2455-24-5</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>503</td>
<td>Benzyl methacrylate</td>
<td>2495-37-6</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>510</td>
<td>2-Amino-4'-chlorobenzophenone</td>
<td>2894-51-1</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>
### MOAclass model, training set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>CASRN</th>
<th>Exp. class</th>
<th>Pred. class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>514</td>
<td>Cyclohexyl acrylate</td>
<td>3066-71-5</td>
<td>D</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>518</td>
<td>2,3,5,6-Tetrachloroaniline</td>
<td>3481-20-7</td>
<td>F</td>
<td>F</td>
<td>A</td>
</tr>
<tr>
<td>526</td>
<td>Isopropyl disulfide</td>
<td>4253-88-8</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>528</td>
<td>Isopropyl methacrylate</td>
<td>4655-34-9</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>529</td>
<td>1-Hexen-3-ol</td>
<td>4798-44-1</td>
<td>D</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>532</td>
<td>1,3-Dimethyl-2-thiobarbituric acid</td>
<td>5217-47-0</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>536</td>
<td>3,6-Dithiaoctane</td>
<td>5395-75-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>539</td>
<td>2-Amino-4-chloro-6-methylpyrimidine</td>
<td>5600-21-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>540</td>
<td>2,6-Dimethoxytoulene</td>
<td>5673-07-4</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>541</td>
<td>2-Dimethylaminopyridine</td>
<td>5683-33-0</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>545</td>
<td>1,1,1-Trichloro-2-methyl-2-propanol(hydrate)</td>
<td>6001-64-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>546</td>
<td>2-Dodecanone</td>
<td>61 75-49-1</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>547</td>
<td>4-Dimethylaminocinnamaldehyde</td>
<td>6200-18-5</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>554</td>
<td>Tripropargylamine</td>
<td>6621-28-5</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>555</td>
<td>N,N-bis(2,2-Diethoxyethyl)methylanine</td>
<td>6948-86-3</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>557</td>
<td>3-Hydroxy-3,7,11-trimethyl-1,6,10-dodecatriene</td>
<td>7212-44-4</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>558</td>
<td>1-(2-Chloroethyl)pyrrolidine-HCl</td>
<td>7250-67-1</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>560</td>
<td>1-Heptyn-3-ol</td>
<td>7388-19-9</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>563</td>
<td>Resmethrin</td>
<td>10453-86-8</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>567</td>
<td>2',3',4'-Trimethoxyacetophenone</td>
<td>13909-73-4</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>571</td>
<td>4-Bromophenyl 3-pyridyl ketone</td>
<td>14544-45-9</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>573</td>
<td>2,2,5,5-Tetramethyltetrahydrofuran</td>
<td>15045-44-9</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>
**MOA class, training set: continued**

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>Exp. class</th>
<th>MOA class</th>
<th>Pred. class</th>
<th>MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>574</td>
<td>3-Hydroxy-2-nitropyridine</td>
<td>15128-82-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>577</td>
<td>Methomyl</td>
<td>16752-77-5</td>
<td>H</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>578</td>
<td>6-Chloro-2-pyridinol</td>
<td>16879-02-0</td>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>581</td>
<td>4-(Diethylamino)salicylaldehyde</td>
<td>17754-90-4</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>583</td>
<td>3,6-Dimethyl-1-heptyn-3-ol</td>
<td>19540-99-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>584</td>
<td>2,4,5-Trimethylpyrazole</td>
<td>20062-81-4</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>587</td>
<td>m-Bromobenzamide</td>
<td>22726-00-7</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>588</td>
<td>Oxyanil</td>
<td>23135-22-0</td>
<td>H</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>589</td>
<td>2,6-Diisopropylazine</td>
<td>24144-04-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>593</td>
<td>2-Methyl-3,3,4,4-tetrafluoro-2-butanol</td>
<td>29653-28-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>596</td>
<td>2-(Bromomethyl)tetrahydro-2H-pyran</td>
<td>34724-82-5</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>598</td>
<td>4-Hexyloxanil</td>
<td>39906-57-2</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>601</td>
<td>Fenvalerate</td>
<td>51630-58-1</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>602</td>
<td>Permethrin</td>
<td>52645-53-1</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>605</td>
<td>p-(tert-Butyl)benzamide</td>
<td>56100-12-4</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>607</td>
<td>dl-3-Butyn-2-carboxylic acid</td>
<td>65337-1-3-5</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>610</td>
<td>3-(4-tert-Butylphenox)benzaldehyde</td>
<td>69770-26-6</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>611</td>
<td>Flucythrinate</td>
<td>70124-77-5</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>613</td>
<td>3-(3,4-Dichlorophenox)benzaldehyde</td>
<td>79124-76-8</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>614</td>
<td>2,4-Dinitro-1-naphthol sodium salt dihydrate (Martius yellow)</td>
<td>101896-92-1</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>615</td>
<td>1-Butylstyrene</td>
<td>17462-23-2</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>617</td>
<td>Chloromethyl styrene</td>
<td>30090-25-2</td>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### F.3 Test set

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>LC$_{50}$ (mg/l)</th>
<th>LC$_{50}$ (mmol/l)</th>
<th>Exp. Tox class</th>
<th>Pred. Tox class</th>
<th>MOA confidence</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4-(Hexyloxy)-m-anisaldehyde</td>
<td>61096-84-2</td>
<td>2.67</td>
<td>0.0113</td>
<td>high AT</td>
<td>high AT</td>
<td>Low</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>5-Bromo-2-nitrovanillin</td>
<td>98-131-3</td>
<td>73.3</td>
<td>0.266</td>
<td>high AT</td>
<td>high AT</td>
<td>Low</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>5</td>
<td>3'-Chloro-o-formotoluidide</td>
<td>71802-02-7</td>
<td>46.6</td>
<td>0.275</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>11</td>
<td>4,9-Dithiadiocane</td>
<td>56348-39-1</td>
<td>2.99</td>
<td>0.0145</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>14</td>
<td>2-chloroethyl-N-cyclohexyl carbamate</td>
<td>31502-57-5</td>
<td>35</td>
<td>0.17</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>Salicylic acid Na+</td>
<td>54-21-7</td>
<td>1.20</td>
<td>0.0285</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>19</td>
<td>Benzamide</td>
<td>55-21-0</td>
<td>661</td>
<td>5.46</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>26</td>
<td>2,4,4'-Tetrachlorophenol</td>
<td>58-90-2</td>
<td>1.03</td>
<td>0.0014</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>F</td>
</tr>
<tr>
<td>36</td>
<td>2-Hydroxybenzamide</td>
<td>65-45-2</td>
<td>101</td>
<td>0.736</td>
<td>low AT</td>
<td>low AT</td>
<td>Low</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>39</td>
<td>Dicumyl</td>
<td>66-76-2</td>
<td>5.11</td>
<td>0.0152</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>40</td>
<td>p-Phenoxylbenzaldehyde</td>
<td>67-36-7</td>
<td>4.6</td>
<td>0.0282</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>43</td>
<td>Acetone</td>
<td>67-64-1</td>
<td>7100</td>
<td>123</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>44</td>
<td>Chloroform</td>
<td>67-66-3</td>
<td>70.7</td>
<td>0.592</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>45</td>
<td>Methyl sulfide</td>
<td>67-68-5</td>
<td>34000</td>
<td>435</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>47</td>
<td>2,2'-Methylene bis(3,4,6-trichlorophenol)</td>
<td>76-30-4</td>
<td>0.021</td>
<td>5.16E-05</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>G</td>
</tr>
<tr>
<td>52</td>
<td>Benene</td>
<td>71-43-2</td>
<td>17.6</td>
<td>0.225</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>55</td>
<td>Acetonitrile</td>
<td>75-05-8</td>
<td>1644</td>
<td>40.05</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>56</td>
<td>Ethanol</td>
<td>75-07-0</td>
<td>33.8</td>
<td>0.767</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>58</td>
<td>Iodoform</td>
<td>75-47-8</td>
<td>2.92</td>
<td>0.0074</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>61</td>
<td>2,2,2-Trifluoroethanol</td>
<td>75-86-8</td>
<td>119</td>
<td>1.19</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Sub stance CASRN</td>
<td>$LC_{50}$ (mg/l)</td>
<td>$LC_{50}$ (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>62</td>
<td>3,3-Dimethyl-2-butanone</td>
<td>75-97-8</td>
<td>87</td>
<td>0.869</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>64</td>
<td>5,5-Dimethylhydantoin</td>
<td>77-71-4</td>
<td>1600</td>
<td>128.5</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>68</td>
<td>Tris(2-butoxyethyl) phosphate</td>
<td>78-51-3</td>
<td>11.2</td>
<td>0.0281</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>73</td>
<td>2-Butanone</td>
<td>78-83-3</td>
<td>3220</td>
<td>44.66</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>74</td>
<td>1-Amino-2-propylol</td>
<td>78-96-6</td>
<td>2520</td>
<td>33.55</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>80</td>
<td>4,4'-Isoproplylidenebis(2,6-dichlorophenol)</td>
<td>70-95-8</td>
<td>1.33</td>
<td>0.0036</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>81</td>
<td>$p$-tert-Propylphenol</td>
<td>80-46-6</td>
<td>2.50</td>
<td>0.0158</td>
<td>high AT</td>
<td>high AT</td>
<td>High</td>
<td>A; B</td>
<td>A</td>
</tr>
<tr>
<td>82</td>
<td>1,4-Diamino-$p$-menthane</td>
<td>80-52-4</td>
<td>65.30</td>
<td>0.383</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>83</td>
<td>alpha,alpha-2,6-Tetrachlorotoluene</td>
<td>81-19-6</td>
<td>0.97</td>
<td>0.0042</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>84</td>
<td>Aacenaphthene</td>
<td>83-32-9</td>
<td>1.73</td>
<td>0.0112</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>88</td>
<td>Diphenyl phthalate</td>
<td>84-62-8</td>
<td>0.08</td>
<td>0.00002</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>G</td>
</tr>
<tr>
<td>90</td>
<td>Diethyl phthalate</td>
<td>84-66-2</td>
<td>31.8</td>
<td>0.143</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>91</td>
<td>Di-$n$-butyl orthophthalate</td>
<td>84-74-2</td>
<td>1</td>
<td>0.0036</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>94</td>
<td>Salicylamide</td>
<td>87-17-2</td>
<td>3.95</td>
<td>0.0185</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>99</td>
<td>Anthranilamide</td>
<td>88-68-6</td>
<td>305</td>
<td>2.501</td>
<td>low AT</td>
<td>low AT</td>
<td>Low</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>100</td>
<td>2-Nitrophenol</td>
<td>88-75-5</td>
<td>160</td>
<td>1.15</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>101</td>
<td>2-sec-Butyl-4-dinitrophenol (dinoseb)</td>
<td>88-85-7</td>
<td>0.535</td>
<td>0.0022</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>F</td>
<td>A</td>
</tr>
<tr>
<td>102</td>
<td>Salicylaldehyde</td>
<td>50-22-8</td>
<td>2.3</td>
<td>0.188</td>
<td>high AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>104</td>
<td>2-Phenylphenol</td>
<td>50-43-7</td>
<td>6.15</td>
<td>0.0361</td>
<td>high AT</td>
<td>high AT</td>
<td>High</td>
<td>A; B</td>
<td>D</td>
</tr>
<tr>
<td>106</td>
<td>3,5-Dibromosalicylaldehyde</td>
<td>50-50-5</td>
<td>0.85</td>
<td>0.0030</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>108</td>
<td>Quinoline</td>
<td>91-22-5</td>
<td>77.80</td>
<td>0.602</td>
<td>low AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC_{50} (mg/l)</td>
<td>LC_{50} (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>109</td>
<td>N,N-Diethylcyclohexylamine</td>
<td>91-65-6</td>
<td>21.4</td>
<td>0.138</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>113</td>
<td>1-Benzoylacetone</td>
<td>93-91-4</td>
<td>1.1</td>
<td>0.0068</td>
<td>highAT</td>
<td>highAT</td>
<td>Low</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>115</td>
<td>Piperine (aliphatic)</td>
<td>94-62-2</td>
<td>7.84</td>
<td>0.0275</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>116</td>
<td>2,4-Dihydroxybenzaldehyde</td>
<td>95-01-2</td>
<td>13.1</td>
<td>0.095</td>
<td>highAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>117</td>
<td>a-Xylene</td>
<td>95-47-6</td>
<td>16.4</td>
<td>0.154</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>119</td>
<td>1,2-Dichlorobenzene</td>
<td>95-50-1</td>
<td>9.47</td>
<td>0.064</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>121</td>
<td>2-Fluorotoluene</td>
<td>95-52-3</td>
<td>19.4</td>
<td>0.176</td>
<td>highAT</td>
<td>lowAT</td>
<td>Low</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>123</td>
<td>1,2,4-Trimethylbenzene</td>
<td>95-65-6</td>
<td>7.72</td>
<td>0.0642</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>124</td>
<td>3,4-Dichlorotoluene</td>
<td>95-75-0</td>
<td>2.91</td>
<td>0.0181</td>
<td>highAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>127</td>
<td>2,3-Dibromopropanol</td>
<td>96-13-9</td>
<td>7.1</td>
<td>0.326</td>
<td>highAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>128</td>
<td>2-Methylbutanaldehyde</td>
<td>96-17-3</td>
<td>9.97</td>
<td>0.116</td>
<td>highAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>130</td>
<td>3-Pentanone</td>
<td>96-22-0</td>
<td>1540</td>
<td>1.788</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>131</td>
<td>2-Butanone oxime</td>
<td>96-23-7</td>
<td>843</td>
<td>9.676</td>
<td>lowAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>134</td>
<td>2,2'-Methylenebis(4-chlorophenol)</td>
<td>97-23-4</td>
<td>0.31</td>
<td>0.0011</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>136</td>
<td>p-Isoamylacetophenol</td>
<td>98-51-4</td>
<td>5.15</td>
<td>0.0333</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>137</td>
<td>Isovaleraldehyde</td>
<td>98-82-8</td>
<td>6.32</td>
<td>0.0526</td>
<td>highAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>139</td>
<td>Nitrobenzene</td>
<td>98-95-3</td>
<td>119</td>
<td>0.967</td>
<td>lowAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>140</td>
<td>m-Aminooacetophenone</td>
<td>99-03-6</td>
<td>382</td>
<td>2.826</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>141</td>
<td>m-Nitrotoluene</td>
<td>99-08-1</td>
<td>25.6</td>
<td>0.187</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>143</td>
<td>p-Nitroaniline</td>
<td>100-01-6</td>
<td>125</td>
<td>0.905</td>
<td>lowAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>145</td>
<td>p-Dimethylaminobenzaldehyde</td>
<td>100-10-7</td>
<td>45.70</td>
<td>0.306</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>
null
<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>LC$_{50}$ (mg/l)</th>
<th>LC$_{50}$ (mmol/l)</th>
<th>Exp. tox class</th>
<th>Pred. tox class</th>
<th>MOA confidence</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>178</td>
<td>p-Xylene</td>
<td>106-42-3</td>
<td>8.87</td>
<td>0.0835</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>179</td>
<td>4-Methylphenol (p-cresol)</td>
<td>106-44-5</td>
<td>16.5</td>
<td>0.153</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>181</td>
<td>4-Chlorophenol</td>
<td>106-48-9</td>
<td>6.11</td>
<td>0.0475</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>188</td>
<td>Propyamine</td>
<td>107-10-8</td>
<td>308</td>
<td>5.211</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>190</td>
<td>Chloroacetanilide</td>
<td>107-14-2</td>
<td>1.35</td>
<td>0.0178</td>
<td>high AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>I</td>
<td>D</td>
</tr>
<tr>
<td>191</td>
<td>Ethylenediamine</td>
<td>107-15-3</td>
<td>220</td>
<td>3.661</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>194</td>
<td>Acetaldehyde</td>
<td>107-20-9</td>
<td>76</td>
<td>1.287</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>196</td>
<td>tert-Butylamine</td>
<td>107-45-9</td>
<td>24.6</td>
<td>0.19</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>197</td>
<td>tert-Butyl sulfide</td>
<td>107-47-1</td>
<td>29.1</td>
<td>0.199</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>198</td>
<td>2-Pentanone</td>
<td>107-87-9</td>
<td>1240</td>
<td>14.4</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>201</td>
<td>Toluene</td>
<td>108-88-3</td>
<td>33.90</td>
<td>0.368</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>203</td>
<td>Chlorobenzene</td>
<td>108-00-7</td>
<td>16.9</td>
<td>0.15</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>205</td>
<td>Cyclohexanone</td>
<td>108-94-1</td>
<td>621</td>
<td>6.227</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>207</td>
<td>3-Picolin</td>
<td>108-90-4</td>
<td>144</td>
<td>1.546</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>208</td>
<td>1-Methylpyrrolidene</td>
<td>108-01-3</td>
<td>2300</td>
<td>22.96</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>210</td>
<td>2-Methylpyrrolidene</td>
<td>108-07-9</td>
<td>2240</td>
<td>22.36</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>211</td>
<td>Propyl acetate</td>
<td>108-90-4</td>
<td>60</td>
<td>0.587</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>214</td>
<td>Butylamine</td>
<td>108-73-5</td>
<td>268</td>
<td>3.664</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>215</td>
<td>Allyl cyanide</td>
<td>108-75-1</td>
<td>182</td>
<td>2.713</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>217</td>
<td>Malononitrile</td>
<td>109-77-3</td>
<td>0.56</td>
<td>0.0085</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>218</td>
<td>2-Methoxyethylamine</td>
<td>109-85-3</td>
<td>524</td>
<td>6.976</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>( LC_{50} ) (mg/l)</td>
<td>( LC_{50} ) (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>219</td>
<td>Diethylamine</td>
<td>100-89-7</td>
<td>855</td>
<td>11.69</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>220</td>
<td>Pyridine</td>
<td>109-97-7</td>
<td>210</td>
<td>3.13</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>Tetrahydrofuran</td>
<td>108-09-9</td>
<td>2160</td>
<td>29.96</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>222</td>
<td>Furan</td>
<td>110-00-9</td>
<td>61</td>
<td>0.896</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate D</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>223</td>
<td>t-Butyl disulfide</td>
<td>110-00-5</td>
<td>1.37</td>
<td>0.0077</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>5-Methyl-2-hexanone</td>
<td>110-12-3</td>
<td>150</td>
<td>1.392</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>Diethyl sebacate</td>
<td>110-06-5</td>
<td>1480</td>
<td>16.6</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>227</td>
<td>Hexane</td>
<td>110-54-3</td>
<td>2.5</td>
<td>0.029</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>229</td>
<td>Amylamine</td>
<td>110-58-7</td>
<td>177</td>
<td>2.031</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>232</td>
<td>2-(Ethylamino)ethanol</td>
<td>110-73-6</td>
<td>1480</td>
<td>16.6</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>233</td>
<td>Cyclohexane</td>
<td>110-08-2</td>
<td>4.53</td>
<td>0.0538</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>235</td>
<td>r-Trisylate</td>
<td>110-88-3</td>
<td>5050</td>
<td>66.05</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>236</td>
<td>6-Methyl-2-hepten-2-one</td>
<td>110-89-0</td>
<td>86.7</td>
<td>0.679</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>240</td>
<td>Hexylamine</td>
<td>111-21-2</td>
<td>56.6</td>
<td>0.539</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>242</td>
<td>Diethanolamine</td>
<td>111-22-2</td>
<td>47.1</td>
<td>44.8</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>244</td>
<td>n-Propyl sulfide</td>
<td>111-17-7</td>
<td>21.7</td>
<td>0.184</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>245</td>
<td>n-Butylamine</td>
<td>111-48-8</td>
<td>21.8</td>
<td>0.189</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>246</td>
<td>1,4-Dicyanoobutane</td>
<td>111-49-3</td>
<td>1530</td>
<td>17.85</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>249</td>
<td>Octylamine</td>
<td>111-86-4</td>
<td>5.19</td>
<td>0.0402</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>252</td>
<td>Nonanoic acid</td>
<td>112-20-5</td>
<td>104</td>
<td>0.657</td>
<td>low AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>254</td>
<td>Nonylamine</td>
<td>112-20-9</td>
<td>2.16</td>
<td>0.015</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>262</td>
<td>Triphe nyl phosphate</td>
<td>115-86-6</td>
<td>0.87</td>
<td>0.0027</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>G</td>
</tr>
<tr>
<td>267</td>
<td>Phen yl salicylate</td>
<td>118-55-8</td>
<td>1.18</td>
<td>0.0055</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate C</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>271</td>
<td>Benzophenone</td>
<td>119-61-9</td>
<td>14.7</td>
<td>0.0807</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>277</td>
<td>2,4-Dinitroto nene</td>
<td>121-14-2</td>
<td>24.3</td>
<td>0.133</td>
<td>high AT</td>
<td>high AT</td>
<td>Low D</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td>3-Ethoxy-4-hydroxybenz aldehyde</td>
<td>121-32-4</td>
<td>87.6</td>
<td>0.527</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>279</td>
<td>Vanillin</td>
<td>121-33-5</td>
<td>83.8</td>
<td>0.551</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate D</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>281</td>
<td>1-Chloro-3-nitrobenzene</td>
<td>121-73-3</td>
<td>18.8</td>
<td>0.119</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>284</td>
<td>$p$-Isopropyl benz aldehyde</td>
<td>122-60-2</td>
<td>6.62</td>
<td>0.0447</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>287</td>
<td>4-Ethylphenol</td>
<td>123-07-5</td>
<td>10.4</td>
<td>0.0851</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>289</td>
<td>2,4-Pentanedione</td>
<td>123-54-6</td>
<td>135</td>
<td>1.348</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate D</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>Ethyl hexanone</td>
<td>123-66-0</td>
<td>8.9</td>
<td>0.0617</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate C</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>291</td>
<td>Butanal</td>
<td>123-72-8</td>
<td>14.7</td>
<td>0.204</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>292</td>
<td>Butyl acetate</td>
<td>123-86-4</td>
<td>18</td>
<td>0.155</td>
<td>high AT</td>
<td>low AT</td>
<td>Moderate C</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>293</td>
<td>1,4-Dio xane</td>
<td>123-91-1</td>
<td>10300</td>
<td>116.9</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>294</td>
<td>Dodecylamine</td>
<td>124-22-1</td>
<td>0.103</td>
<td>0.0006</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>297</td>
<td>5,5-Dimethyl-1,3-cyclohex anedione</td>
<td>130-81-8</td>
<td>11500</td>
<td>82.04</td>
<td>low AT</td>
<td>low AT</td>
<td>Low A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>2,6-Di-tert-butyl-L-methylphenol</td>
<td>128-37-0</td>
<td>0.363</td>
<td>0.0016</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>302</td>
<td>Saccharin sodium salt hydrate</td>
<td>82388-12-0</td>
<td>82300</td>
<td>82</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>303</td>
<td>Dibenzofuran</td>
<td>132-61-9</td>
<td>1.5</td>
<td>0.0089</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>F</td>
</tr>
<tr>
<td>305</td>
<td>N,N-Diethyl-m-toluamide</td>
<td>131-62-3</td>
<td>110</td>
<td>0.575</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>306</td>
<td>Propionic acid, sodium salt</td>
<td>137-40-6</td>
<td>4700</td>
<td>49.86</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC&lt;sub&gt;50&lt;/sub&gt; (mg/l)</td>
<td>LC&lt;sub&gt;50&lt;/sub&gt; (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>-------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>307</td>
<td>1-(2-Aminoethyl)piperazine</td>
<td>140-31-8</td>
<td>21.90</td>
<td>16.95</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>308</td>
<td>Dibutyl succinate</td>
<td>141-03-7</td>
<td>4.46</td>
<td>0.0194</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>309</td>
<td>Diethyl adipate</td>
<td>141-28-6</td>
<td>18</td>
<td>0.0889</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>310</td>
<td>2-Aminoethanol</td>
<td>141-43-5</td>
<td>2070</td>
<td>33.89</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>311</td>
<td>Ethyl acetate</td>
<td>141-78-6</td>
<td>230</td>
<td>2.61</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>312</td>
<td>2,6-Dimethyl morpholine</td>
<td>141-91-3</td>
<td>38.7</td>
<td>3.36</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>313</td>
<td>m-Diethylenzene</td>
<td>141-03-5</td>
<td>4.15</td>
<td>0.0309</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>315</td>
<td>Hexanoic acid</td>
<td>142-62-1</td>
<td>320</td>
<td>2.755</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>316</td>
<td>Hexyl acetate</td>
<td>142-62-7</td>
<td>4.4</td>
<td>0.0305</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>319</td>
<td>Di-n-hexylamine</td>
<td>143-14-8</td>
<td>0.78</td>
<td>0.0042</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>321</td>
<td>o-Vanillin</td>
<td>148-23-8</td>
<td>2.40</td>
<td>0.0158</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>326</td>
<td>1,4-Diazabicyclo[2.2.2]octane</td>
<td>280-57-9</td>
<td>17.30</td>
<td>15.42</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>327</td>
<td>Adamantine</td>
<td>281-23-2</td>
<td>0.28</td>
<td>0.0021</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>330</td>
<td>Bromacil</td>
<td>314-04-9</td>
<td>186</td>
<td>0.712</td>
<td>low AT</td>
<td>low AT</td>
<td>Low</td>
<td>A</td>
<td>H</td>
</tr>
<tr>
<td>332</td>
<td>Diuron</td>
<td>330-54-1</td>
<td>14.2</td>
<td>0.0869</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>333</td>
<td>p-Fluorophenyl ether</td>
<td>330-93-8</td>
<td>1.13</td>
<td>0.0055</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>335</td>
<td>1-Fluoro-4-nitrobenzene</td>
<td>350-45-9</td>
<td>28.4</td>
<td>0.201</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>356</td>
<td>alpha,alpha,alpha-Trifluoro-m-tolunitrile</td>
<td>368-77-7</td>
<td>47.70</td>
<td>0.279</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>357</td>
<td>4-Fluorobenzaldehyde</td>
<td>371-40-4</td>
<td>16.9</td>
<td>0.152</td>
<td>high AT</td>
<td>high AT</td>
<td>Low</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>338</td>
<td>2-Chloro-6-fluorobenzaldehyde</td>
<td>387-45-1</td>
<td>9.41</td>
<td>0.0553</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>340</td>
<td>2-Fluorobenzaldehyde</td>
<td>446-52-6</td>
<td>1.35</td>
<td>0.0109</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>( LC_{50} ) (mg/l)</td>
<td>( LC_{50} ) (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>342</td>
<td>alpha, alpha, alpha-Trifluoro-m-tolualdehyde</td>
<td>454-89-7</td>
<td>0.924</td>
<td>0.0053</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>343</td>
<td>4-Flouro-N-methylaniline</td>
<td>459-56-6</td>
<td>38.4</td>
<td>0.307</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A; B</td>
<td>A</td>
</tr>
<tr>
<td>345</td>
<td>[(1S)-endo]-(-)-Borneol</td>
<td>461-45-9</td>
<td>63.2</td>
<td>0.41</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>346</td>
<td>(1S)-(−)-Camphor</td>
<td>461-48-2</td>
<td>17</td>
<td>0.112</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>348</td>
<td>Neoabietic acid</td>
<td>471-77-2</td>
<td>1.49</td>
<td>0.0049</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>351</td>
<td>exo-Norborneol</td>
<td>497-37-0</td>
<td>228</td>
<td>2.033</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>352</td>
<td>Norbornylene</td>
<td>498-66-8</td>
<td>10</td>
<td>0.106</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>353</td>
<td>2,6-Pyridinedicarboxylic acid</td>
<td>499-84-2</td>
<td>322</td>
<td>1.927</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd D</td>
<td></td>
</tr>
<tr>
<td>354</td>
<td>3-Pyridinecarboxaldehyde</td>
<td>500-22-1</td>
<td>16.4</td>
<td>0.153</td>
<td>high AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>D D</td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>5-Norborneol</td>
<td>502-56-7</td>
<td>31</td>
<td>0.218</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A A</td>
<td></td>
</tr>
<tr>
<td>357</td>
<td>Abietic acid</td>
<td>514-10-3</td>
<td>2.38</td>
<td>0.0079</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>360</td>
<td>α-Tolualdehyde</td>
<td>520-19-1</td>
<td>44.7</td>
<td>0.382</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A A</td>
<td></td>
</tr>
<tr>
<td>361</td>
<td>α-Toluene</td>
<td>528-20-4</td>
<td>52.9</td>
<td>0.44</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A D</td>
<td></td>
</tr>
<tr>
<td>362</td>
<td>Bencotic acid, sodium salt</td>
<td>532-32-1</td>
<td>484</td>
<td>3.359</td>
<td>low AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>364</td>
<td>Amylbencene</td>
<td>538-98-1</td>
<td>1.71</td>
<td>0.0115</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A D</td>
<td></td>
</tr>
<tr>
<td>365</td>
<td>tert-Butyl acetate</td>
<td>540-49-3</td>
<td>327</td>
<td>2.915</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A A</td>
<td></td>
</tr>
<tr>
<td>366</td>
<td>1,2-Dichlorobenzene</td>
<td>541-73-1</td>
<td>8.03</td>
<td>0.0546</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>368</td>
<td>n-Butyl sulfide</td>
<td>544-40-1</td>
<td>3.58</td>
<td>0.0245</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>369</td>
<td>2'-Hydroxy-4'-methoxycetophenone</td>
<td>552-41-0</td>
<td>69.5</td>
<td>0.418</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A A</td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>α-Nitrobenzaldehyde</td>
<td>552-88-6</td>
<td>14.4</td>
<td>0.0952</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D B</td>
<td></td>
</tr>
<tr>
<td>372</td>
<td>3-Methyl-2-butene</td>
<td>563-80-1</td>
<td>864</td>
<td>10.03</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A A</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------------</td>
<td>-------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>374</td>
<td>1,2-Dibromobenzene</td>
<td>583-53-9</td>
<td>4.05</td>
<td>0.0172</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>375</td>
<td>N-Allylaniline</td>
<td>589-09-3</td>
<td>35.9</td>
<td>0.27</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>376</td>
<td>4-Ethylaniline</td>
<td>589-16-2</td>
<td>73</td>
<td>0.602</td>
<td>lowAT</td>
<td>highAT</td>
<td>High</td>
<td>A; B</td>
<td>B</td>
</tr>
<tr>
<td>380</td>
<td>2-Tridecanone</td>
<td>593-08-8</td>
<td>0.36</td>
<td>0.0018</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>382</td>
<td>Mancol</td>
<td>596-85-0</td>
<td>0.12</td>
<td>0.0004</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>384</td>
<td>1,2-Dimethylpropylamine</td>
<td>598-74-3</td>
<td>28.4</td>
<td>3.298</td>
<td>lowAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>387</td>
<td>N,N-Diphenylformamide</td>
<td>607-00-1</td>
<td>30.4</td>
<td>0.154</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>388</td>
<td>Diethyl benzylmalonate</td>
<td>607-81-8</td>
<td>5.43</td>
<td>0.0217</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>390</td>
<td>2,4,6-Trichlorophenol</td>
<td>600-27-4</td>
<td>1.21</td>
<td>0.0026</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A; F</td>
<td></td>
</tr>
<tr>
<td>391</td>
<td>2,4-Dimethoxybenzaldehyde</td>
<td>613-45-6</td>
<td>20.1</td>
<td>0.121</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>392</td>
<td>2-Acetamidophenol</td>
<td>614-80-2</td>
<td>27</td>
<td>0.179</td>
<td>highAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>395</td>
<td>Methyl p-nitrobenzaldehyde</td>
<td>619-50-1</td>
<td>23.8</td>
<td>0.131</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>397</td>
<td>4-Nitrophenyl phenyl ether</td>
<td>620-48-2</td>
<td>2.65</td>
<td>0.0123</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>399</td>
<td>3-Acetamidophenol</td>
<td>621-22-1</td>
<td>1.30</td>
<td>7.475</td>
<td>lowAT</td>
<td>lowAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>401</td>
<td>alpha, alpha-Dichloro-p-xylene</td>
<td>623-25-6</td>
<td>0.039</td>
<td>0.0002</td>
<td>highAT</td>
<td>highAT</td>
<td>Low</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>407</td>
<td>1,6-Dicyanohexane</td>
<td>628-79-3</td>
<td>5.28</td>
<td>3.877</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>408</td>
<td>2,4,6-Trichlorophenol</td>
<td>634-67-3</td>
<td>3.64</td>
<td>0.0185</td>
<td>highAT</td>
<td>highAT</td>
<td>High</td>
<td>A; B</td>
<td>B</td>
</tr>
<tr>
<td>409</td>
<td>5-Chloroacetophenol</td>
<td>635-64-8</td>
<td>0.77</td>
<td>0.0049</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>410</td>
<td>4-Propyphenol</td>
<td>645-56-7</td>
<td>11</td>
<td>0.0808</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>411</td>
<td>Pentfluorobenzaldehyde</td>
<td>653-57-2</td>
<td>1.1</td>
<td>0.0056</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>412</td>
<td>2,2-Dichloroacetamide</td>
<td>683-72-7</td>
<td>24.1</td>
<td>1.883</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Moderate</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>413</td>
<td>1-Methyl heptylamine</td>
<td>693-16-3</td>
<td>5.19</td>
<td>0.0162</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------</td>
<td>---------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>414</td>
<td>2-Decanone</td>
<td>693-54-9</td>
<td>4.83</td>
<td>0.0309</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>418</td>
<td>2-Adamantanone</td>
<td>700-58-3</td>
<td>60.8</td>
<td>0.405</td>
<td>high AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>420</td>
<td>4,6-Dimethoxy-2-hydroxybenzaldehyde</td>
<td>708-76-9</td>
<td>2.68</td>
<td>0.0147</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>421</td>
<td>Propanil</td>
<td>708-98-8</td>
<td>8.6</td>
<td>0.0394</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>422</td>
<td>2,4,6-Tri-tert-butylphenol</td>
<td>732-26-3</td>
<td>0.0009</td>
<td>0.0002</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>424</td>
<td>N,N-Dimethylformamide</td>
<td>761-65-9</td>
<td>89.3</td>
<td>0.568</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>427</td>
<td>1-Adamantanilimine</td>
<td>768-04-5</td>
<td>25</td>
<td>0.165</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>431</td>
<td>Triclosan</td>
<td>701-28-6</td>
<td>53.7</td>
<td>0.193</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>437</td>
<td>p-Phenoxycarboxil</td>
<td>831-82-3</td>
<td>4.95</td>
<td>0.0266</td>
<td>high AT</td>
<td>high AT</td>
<td>High</td>
<td>A; B</td>
<td>A</td>
</tr>
<tr>
<td>438</td>
<td>2-Hydroxyethyl methacrylate</td>
<td>868-77-9</td>
<td>227</td>
<td>1.744</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>440</td>
<td>2,4-Dichlorobenzaldehyde</td>
<td>874-42-0</td>
<td>1.8</td>
<td>0.0103</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>449</td>
<td>2-Acetyl-1-methylpyrrole</td>
<td>902-21-1</td>
<td>157</td>
<td>1.275</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>460</td>
<td>Methyl 4-cyanobenzoate</td>
<td>1129-35-7</td>
<td>46.8</td>
<td>0.29</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>462</td>
<td>alpha-Chloro-2,3,5-trimethoxyacetophenone</td>
<td>1204-21-3</td>
<td>0.0066</td>
<td>0.0025</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>466</td>
<td>3-Benzoxylalcohol</td>
<td>1484-26-0</td>
<td>9.14</td>
<td>0.0459</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>468</td>
<td>tert-Butyl methyl ether</td>
<td>1463-26-1</td>
<td>672</td>
<td>7.623</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>471</td>
<td>3,2-Diiodo-4-hydroxybenzonitrile</td>
<td>1689-53-4</td>
<td>6.8</td>
<td>0.0183</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>F</td>
</tr>
<tr>
<td>472</td>
<td>3,2-Dibromo-4-hydroxybenzonitrile</td>
<td>1689-84-5</td>
<td>12.6</td>
<td>0.0455</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>-</td>
<td>nd</td>
</tr>
<tr>
<td>473</td>
<td>Dehydroabietic acid</td>
<td>1740-19-8</td>
<td>2.1</td>
<td>0.0070</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>474</td>
<td>2-Allylphenol</td>
<td>1745-81-9</td>
<td>15</td>
<td>0.112</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>475</td>
<td>5-Bromosalicylaldehyde</td>
<td>1761-61-1</td>
<td>1.3</td>
<td>0.0065</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC$_{50}$ (mg/l)</td>
<td>LC$_{50}$ (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>477</td>
<td>3,5-Dichloro-4-hydroxybenzonitrile</td>
<td>1891-00-8</td>
<td>24.3</td>
<td>0.129</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>478</td>
<td>Di-n-butylterephthalate</td>
<td>1962-75-0</td>
<td>0.59</td>
<td>0.0021</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>479</td>
<td>4,4’-Dihydroxydiphenyl ether</td>
<td>1965-00-9</td>
<td>5.78</td>
<td>0.0286</td>
<td>high AT</td>
<td>high AT</td>
<td>High-Moderate</td>
<td>A; B</td>
<td>A</td>
</tr>
<tr>
<td>481</td>
<td>n-Decylamine</td>
<td>2016-57-1</td>
<td>1.03</td>
<td>0.0065</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>483</td>
<td>2,4,5-tribromomidazole</td>
<td>2034-22-2</td>
<td>6.12</td>
<td>0.0201</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>486</td>
<td>4-Chlorocatechol</td>
<td>2138-22-9</td>
<td>1.58</td>
<td>0.0109</td>
<td>high AT</td>
<td>low AT</td>
<td>Low</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>487</td>
<td>Methyl 2,4-dihydroxybenzoate</td>
<td>2150-47-2</td>
<td>45.8</td>
<td>0.272</td>
<td>high AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>491</td>
<td>1-(p-Toluenesulfonyl)imidazole</td>
<td>2232-08-8</td>
<td>41.8</td>
<td>0.188</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>G</td>
</tr>
<tr>
<td>497</td>
<td>2,4,6-Trimethylphenol</td>
<td>2416-94-6</td>
<td>8.2</td>
<td>0.0032</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>498</td>
<td>n-Undecyl cyanide</td>
<td>2437-25-4</td>
<td>0.43</td>
<td>0.0024</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>499</td>
<td>o-Methoxybenzamide</td>
<td>2438-77-2</td>
<td>130</td>
<td>0.794</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>500</td>
<td>2,4-Dichlorobenzamide</td>
<td>2447-79-2</td>
<td>95.6</td>
<td>0.503</td>
<td>low AT</td>
<td>low AT</td>
<td>Low</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>502</td>
<td>4,5-Dichloroguaiacol</td>
<td>2440-19-3</td>
<td>4.47</td>
<td>0.0232</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>504</td>
<td>Hexyl acrylate</td>
<td>2499-95-8</td>
<td>1.11</td>
<td>0.0071</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>505</td>
<td>p-((tert-buty1)-Phenyl-N-methylcarbamate</td>
<td>2504-82-7</td>
<td>10</td>
<td>0.0482</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>507</td>
<td>1-Benzylpiperazine</td>
<td>2759-28-6</td>
<td>47.4</td>
<td>0.269</td>
<td>high AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>508</td>
<td>3-(3-Pyridyl)-1-propanol</td>
<td>2609-67-8</td>
<td>150</td>
<td>1.093</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>509</td>
<td>Tridecylamine</td>
<td>2608-43-1</td>
<td>0.0054</td>
<td>0.0003</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>511</td>
<td>Methyl 2,4-dichlorobenzoate</td>
<td>2905-23-3</td>
<td>14</td>
<td>0.0083</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>513</td>
<td>5-Bromovanillin</td>
<td>2974-76-4</td>
<td>59.7</td>
<td>0.258</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>B</td>
</tr>
</tbody>
</table>
## Test set: continued

<table>
<thead>
<tr>
<th>No.</th>
<th>Substance chemical name</th>
<th>Substance CASRN</th>
<th>LC\textsubscript{50} (mg/l)</th>
<th>LC\textsubscript{50} (mmol/l)</th>
<th>Exp. Tox class</th>
<th>Pred. Tox class</th>
<th>MOA confidence</th>
<th>Exp. MOA class</th>
<th>Pred. MOA class</th>
</tr>
</thead>
<tbody>
<tr>
<td>516</td>
<td>Triethyl nitrilotriacetic acid</td>
<td>32006-31-3</td>
<td>13.3</td>
<td>0.057</td>
<td>highAT</td>
<td>highAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low C</td>
</tr>
<tr>
<td>517</td>
<td>4,5-Dihydroxyacetophenone</td>
<td>3428-24-8</td>
<td>0.89</td>
<td>0.0050</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate D</td>
<td>Moderate A</td>
</tr>
<tr>
<td>519</td>
<td>2,6-Diphenylpyridine</td>
<td>3558-69-8</td>
<td>0.21</td>
<td>0.0009</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate D</td>
<td>Moderate A</td>
</tr>
<tr>
<td>520</td>
<td>1,3-Dichloro-4,6-dinitrobenzene</td>
<td>3698-83-7</td>
<td>0.0456</td>
<td>0.0002</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate D</td>
<td>Moderate A</td>
</tr>
<tr>
<td>523</td>
<td>2,3-Dimethylvaleraldehyde</td>
<td>3944-76-1</td>
<td>16</td>
<td>0.14</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>524</td>
<td>2-Decyn-1-ol</td>
<td>4117-1-40</td>
<td>1.07</td>
<td>0.0070</td>
<td>highAT</td>
<td>highAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>525</td>
<td>5-Chloro-2-pyridinol</td>
<td>4214-79-3</td>
<td>1140</td>
<td>8.8</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>527</td>
<td>2,4-Dibromoacetophenoxide</td>
<td>4460-86-0</td>
<td>49.5</td>
<td>0.272</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>530</td>
<td>2,4,5-Trichlorophenol</td>
<td>4901-51-3</td>
<td>0.41</td>
<td>0.0018</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>531</td>
<td>1,2-Bis(4-Pyridyl)ethane</td>
<td>4916-57-8</td>
<td>151</td>
<td>0.82</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>533</td>
<td>Dimethyl nitroterephthalate</td>
<td>5299-24-5</td>
<td>6.5</td>
<td>0.0273</td>
<td>highAT</td>
<td>highAT</td>
<td>Low C</td>
<td>Low A</td>
<td>Low C</td>
</tr>
<tr>
<td>534</td>
<td>5-Chloro-2-mercaptobenzo thiole</td>
<td>5331-21-9</td>
<td>3.21</td>
<td>0.0159</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>535</td>
<td>Dimethyl aminoterephthalate</td>
<td>5372-81-6</td>
<td>8.94</td>
<td>0.0402</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>537</td>
<td>Dimethylaminopropyl chloride, hydrochloride</td>
<td>5407-91-5</td>
<td>133</td>
<td>0.841</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>538</td>
<td>4-Chloro-3-nitrocresol</td>
<td>5465-23-5</td>
<td>5.5</td>
<td>0.0276</td>
<td>highAT</td>
<td>highAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>542</td>
<td>2,2-Dimethyl-1-propylamine</td>
<td>5813-61-9</td>
<td>475</td>
<td>3.449</td>
<td>lowAT</td>
<td>lowAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>543</td>
<td>Isopropionic acid</td>
<td>5835-26-7</td>
<td>0.87</td>
<td>0.0059</td>
<td>highAT</td>
<td>highAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>544</td>
<td>2-Aminoisobutyrophenone</td>
<td>5922-20-1</td>
<td>20.6</td>
<td>0.187</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>549</td>
<td>1,3,5-Trichloro-2,4-dinitrobenzene</td>
<td>6284-83-9</td>
<td>0.222</td>
<td>0.0008</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>550</td>
<td>2-Fluoro-5-nitrobenzaldehyde</td>
<td>6361-21-3</td>
<td>3.87</td>
<td>0.0209</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate A</td>
<td>Moderate A</td>
<td>Moderate A</td>
</tr>
<tr>
<td>551</td>
<td>2-Chloro-6-methylbenzonitrile</td>
<td>6575-00-3</td>
<td>15.1</td>
<td>0.0096</td>
<td>highAT</td>
<td>highAT</td>
<td>Low D</td>
<td>Low B</td>
<td>Low B</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>Substance CASRN</td>
<td>LC50 (mg/l)</td>
<td>LC50 (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>Exp. MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>552</td>
<td>2-Bromo-3-pyridinol</td>
<td>6602-32-0</td>
<td>469</td>
<td>2.695</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>553</td>
<td>2-Chloro-3-pyridinol</td>
<td>6636-78-8</td>
<td>622</td>
<td>4.801</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>555</td>
<td>1,4-bis(3-Aminopropyl)piperazine</td>
<td>7209-38-3</td>
<td>31.00</td>
<td>15.47</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>559</td>
<td>n-Undecylamine</td>
<td>7307-55-3</td>
<td>0.21</td>
<td>0.0012</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>561</td>
<td>p-Ethoxynitrobenzaldehyde</td>
<td>10031-82-0</td>
<td>28.1</td>
<td>0.187</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>562</td>
<td>1[(R)-endo]-(+)-3-Bromo camphor</td>
<td>10293-06-8</td>
<td>68.5</td>
<td>0.206</td>
<td>-</td>
<td>-</td>
<td>nd</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>565</td>
<td>alpha,alpha,alpha&quot;,alpha&quot;-Tetra bromo-o-xylene</td>
<td>13209-15-9</td>
<td>0.437</td>
<td>0.0010</td>
<td>high AT</td>
<td>high AT</td>
<td>Low</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>566</td>
<td>2',3',4'-Trichloroacetophenone</td>
<td>13608-87-2</td>
<td>2</td>
<td>0.0089</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>568</td>
<td>Diethyl chloromalonate</td>
<td>14064-10-9</td>
<td>0.95</td>
<td>0.0019</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>C</td>
</tr>
<tr>
<td>569</td>
<td>N-Ethylbenzylamine</td>
<td>14321-27-8</td>
<td>5.71</td>
<td>0.422</td>
<td>high AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>572</td>
<td>4-Benzoylpyridine</td>
<td>14548-46-0</td>
<td>1.03</td>
<td>0.562</td>
<td>low AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>575</td>
<td>Alachlor</td>
<td>15072-60-8</td>
<td>5</td>
<td>0.0185</td>
<td>high AT</td>
<td>high AT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>576</td>
<td>4-Octylaniline</td>
<td>16215-73-7</td>
<td>0.12</td>
<td>0.0006</td>
<td>high AT</td>
<td>high AT</td>
<td>Low</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>579</td>
<td>3-Amino-5,6-dimethyl-1,2,4-triazine</td>
<td>17584-12-2</td>
<td>952</td>
<td>7.668</td>
<td>low AT</td>
<td>low AT</td>
<td>Moderate</td>
<td>B</td>
<td>H</td>
</tr>
<tr>
<td>582</td>
<td>6-Chloro-2-piperine</td>
<td>18360-63-3</td>
<td>232</td>
<td>1.819</td>
<td>low AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>585</td>
<td>4-Dimethylamino-3-methyl-2-butanone</td>
<td>2210-62-7</td>
<td>8.5</td>
<td>0.0658</td>
<td>low AT</td>
<td>high AT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
<tr>
<td>594</td>
<td>(+-)sec-Butylamine</td>
<td>13852-81-6</td>
<td>275</td>
<td>3.76</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>D</td>
</tr>
<tr>
<td>595</td>
<td>N-(3-Methoxypropyl)-3,4,5-trimethoxybenzylamine</td>
<td>34274-01-9</td>
<td>136</td>
<td>0.505</td>
<td>low AT</td>
<td>low AT</td>
<td>-</td>
<td>nd</td>
<td>H</td>
</tr>
<tr>
<td>597</td>
<td>4-Decylaniline</td>
<td>37529-30-9</td>
<td>0.062</td>
<td>0.0003</td>
<td>high AT</td>
<td>high AT</td>
<td>Low</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>No.</td>
<td>Substance chemical name</td>
<td>CASRN</td>
<td>LCso (mg/l)</td>
<td>LC50 (mmol/l)</td>
<td>Exp. Tox class</td>
<td>Pred. Tox class</td>
<td>MOA confidence</td>
<td>MOA class</td>
<td>Pred. MOA class</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td>599</td>
<td>Methyl 4-chloro-2-nitrobenzoate</td>
<td>42087-80-9</td>
<td>27.7</td>
<td>0.128</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>600</td>
<td>5-Hydroxy-2-nitrobenzaldehyde</td>
<td>42454-06-8</td>
<td>41.9</td>
<td>0.251</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>B</td>
</tr>
<tr>
<td>603</td>
<td>3,8-Dithiadecane</td>
<td>54576-32-8</td>
<td>6.06</td>
<td>0.034</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>604</td>
<td>2'-(Octyloxy)acetanilide</td>
<td>55792-61-5</td>
<td>0.45</td>
<td>0.0017</td>
<td>highAT</td>
<td>highAT</td>
<td>Moderate</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>606</td>
<td>2,9-Dithiadecane</td>
<td>56348-40-4</td>
<td>10.1</td>
<td>0.0566</td>
<td>highAT</td>
<td>highAT</td>
<td>-</td>
<td>nd</td>
<td>A</td>
</tr>
</tbody>
</table>
F.4 References