Classification of Organic Compounds into Modes of Toxic Action

Den Naturwissenschaftlichen Fakultäten der

Friedrich-Alexander-Universität Erlangen-Nürnberg

zur Erlangung des Doktorgrades

vorgelegt von

Simon L. Spycher

aus Grabs/Schweiz

Als Dissertation genehmigt von
den Naturwissenschaftlichen Fakultäten der Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	18. Oktober 2005
Vorsitzender der Promotionskommission:	Prof. Dr. D.-P. Häder
Erstberichterstatter:	Prof. Dr. J. Gasteiger
Zweitberichterstatter:	Prof. Dr. T. Clark

This work would not have been possible without the know-how and encouragement of specialists from the most different fields of science. There is a large number of people I would like to thank. First of all my supervisor

Prof. Dr. J. Gasteiger

who gave me the opportunity to work on this fascinating subject in the first place and then for the support and inspiring discussions during the years in his group.

Then, I would like to thank my coauthors of the publications originating from this work for the excellent collaboration: Dr. Monika Nendza of the Analytisches Laboratorium Luhnstedt, Dr. Eric Pellegrini of our group, and PD Dr. Beate Escher of the Swiss Federal Institute of Aquatic Science and Technology (EAWAG).

I would also like to thank my colleagues of the team working on the multimedia teaching project Vernetztes Studium–Chemie (VS-C): Dr. Thomas Engel, Angelika Hofmann and especially Dr. Axel Schunk for the good collaboration in the Chemie für Mediziner multimedia module.

Additionally I would like to express thanks to my present and former colleagues in this group for the great working atmosphere and their help with numerous programming questions and especially those who adminstrated our large network of Unix- and Windows-machines: Prof. Dr. Joao Aires de Sousa, Prof. Dr. Fernando Batista Da Costa, Ulrike Burkard, Stephan Grell, Dr. Yongquan Han, Markus Hemmer, Dr. Achim Herwig, Alexander von Homeyer, Dimitar Hristozov, Dr. Qian-Nan Hu, Dr. Wolf-Dietrich Ihlenfeldt, Thomas Kleinöder, Dr. Thomas Kostka, Rastislav Krajcik, Dr. Sung Kwang Lee, Dr. Giorgi Lekishvili, Gisela Martinek, Jörg Marusczyk, Dr. Frank Oellien, Udo Ottmann, Dr. Matthias Pförtner, Martin Reitz, Dr. Oliver Sacher, Dr. Christian Scholten, Dr. Christof Schwab, Dr. Thomas Seidel, Dr. Markus Sitzmann, Dr. James Smith, Vladimir Sykora, Dr. Alexei Tarkhov, Dr. Andreas Teckentrup, Dr. Lothar Terfloth, Dr. Jaroslaw Tomczak, Thomas Tröger, Dr. Dietrich Trümbach, Dr. Dusica Vidovic, Dr. Jörg Wegner, Dr. Ai-Xia Yan, Dr. Jinhua Zhang.

Special thanks also to the following poeple from outside this group: Dr. Bjoern Reineking of the ETH-Zürich for his extensive support with R-calculations, Dr. Johannes Ranke of the University of Bremen, Dr. Said Hilal of the US-EPA, Dr. Aptula Aynur and Prof. Dr. Marc Cronin of the University of Liverpool, and especially Dr. Jean-Pierre Kocher of Molecular Networks GmbH for his help with the second publication and for his encouragements to scientific curiosity and debate.

I am also very indebted to our Secretaries Angela Döbler, Ulrike Scholz and Karin Holtzke who always made sure our group runs smoothly.

A big thank you to my girlfriend Cornelia for not despairing with her scientific boyfriend and finally to my parents Alfred and Hella and my brother Boris whose example and support guided me through my studies and always gave me new energy.

Contents

Chapter 1

Introduction

1.1 Motivation

Over the last century, chemistry has been one of the strongest driving forces for innovation in science, technology and economy. The enormous progress especially in the field of synthesis allowed the development of tens of thousands of new compounds and products with previously unknown efficiency and often with completely new properties. The innovations lead to new drugs, textiles, dyes, pesticides just to name a few applications. However, in the sixties and seventies it became increasingly clear that some products also had unwanted side effects not only on a local but also on a global scale [1]. In some cases these side effects came as complete surprise to the inventors, as an example the chlorofluorocarbons (CFC) were introduced as cooling gases especially because of their low toxicity and non-flammability. Decades later they revealed the unwanted property to catalytically reduce the concentration of stratospheric ozone and their use had to be severely restricted and partially banned.

The increasing public pressure and scientific concern led most countries to introduce laws for the protection of the environment and specifically to require manufacturers of chemicals to test compounds for persistence and toxic effects before they are brought onto the market. As an example, bringing a new chemical onto the market in the European Union requires environmental studies covering physico-chemical properties, biological degradation, environmental and toxicity tests for species belonging to different levels of biological organization etc. The costs of these studies might already reach several hundred thousand € for the base set [2], but might be much higher for products which are directly distributed in the environment, e.g., pesticides. On the other hand, there is big gap in the amount of infor-

mation required to manufacture so-called existing chemicals, i.e., compounds which were placed on the market before the registration requirements were imposed in 1981. The result is that for the largest portion of compounds actually manufactured in the EU, and to a similar extent in the US, there is only incomplete data about their environmental effects. As an example, for the 2747 High Production Volume Chemicals (yearly production volume > 1000 tons) the European Chemicals Bureau (ECB) has complete information only for 15% of the compounds, for 65 % of the compounds somewhat complete and for 21 % no data are available [3].

This situation does not only lead to worse decisions from an environmental viewpoint, but it also hampers innovation in chemical industry as "old" but possibly harmful compounds cause a lower financial burden to the manufacturer then the production of new and better designed compounds. With the White Paper for a Future Chemicals Policy the European Union has set forth a strategy to tackle this situation [4]. Under the REACH-program (**R**egistration, **E**valuation and **A**uthorisation of **Ch**emicals) about 30'000 existing chemicals shall be tested using a standardized set of tests in the period from 2005 to 2012 [5]. However, concerns against REACH have been raised from several sides: from the industry because of the additional financial burden and from animal protection groups because of the large number of test animals required.

Quantitative structure-activity relationships (QSAR) mathematically relate descriptors of a chemical structure to a biological activity and therefore can also be used to relate descriptors to toxic effects. QSAR methods are one area of the large field of chemoinformatics [6]. The development of QSARs in the field of toxicity prediction experienced a big surge through the nineties. The US-EPA and also the ECB pushed large experimental and QSAR modeling programs with the goal to develop methods that allow the rapid screening of large databases and eventually the reduction and replacement of animal testing [7, 8]. However, these high expectations soon had to be reduced given the enormous complexity of the task. Ten years after the start of these programs, QSAR was not considered by some experts any more as a tool for toxicity prediction within the regulatory framework of REACH [9]. Today, there is again a raising optimism that QSARs can at least complement experimental methods in the task of assessing a chemical's environmental effects. Pedersen *et al.* estimated a cost saving potential of 700–940 Mio € for the REACH program [10].

There is also a strong interest in toxicity QSAR in the field of drug design [11, 12].

Adverse effects of drugs are often detected in a very advanced phase of drug design which leads to costly drug candidate failures. The types of compounds studied in drug-design toxicity assessment and in environmental assessment differ. The environmental chemicals involve more often nonspecific interactions and sometimes highly reactive chemicals which would not be administered to patients. Thus, receptor based modeling is less frequent in environmental QSAR with the notable exception of estrogenic compounds. As a general tendency in the field of drug-design even subtler differences between compounds need to be modeled. After all, a drug is usually quite close to a poison. Again, one must note that the enormous complexity of the human body makes it highly unlikely that QSAR models can completely replace *in vitro* and *in vivo* tests. Nevertheless, already a rough screening method which increases the probability that a drug candidate is not toxic can be of great value.

1.2 Scientific Background

1.2.1 Chemoinformatics

Chemoinformatics experienced an enormous growth over the last decades. Starting with the core tasks of the electronic storage and retrieval of chemical structures it rapidly integrated and contributed to methods from other fields like informatics, mathematics, physical and organic chemistry, statistics and pattern recognition, pharmacology and biochemistry. The amazing breadth of this new field is reflected in the recent Handbook of Chemoinformatics [13].

Within chemoinformatics, all methods which relate descriptors to molecular properties like octanol-water partitioning constants or to a biological activity can be covered by the generic term structure-property relationships (SPR). Hence, QSAR methods are one possible type of SPRs as they relate descriptors to a biological activity. Some authors limit the definition of QSAR to linear free-energy relationships for a set of structural congeners like in the seminal works of Hansch and Fujita [14] and Hansch [15]. Today, there seems no commonly agreed definition, e.g. in the REACH documentation QSAR is used as a synonym for all types of SPRs. In the course of this work, any method which relates descriptors to a *quantitative* measure of a property will be referred to as QSAR regardless whether the relationship is linear or not. On the other hand, methods which relate descriptors to a *qualitative* mea-

sure of biological activity will be referred to as classification methods or structure-activity relationships (SAR).

Toxicity-SAR and QSAR are based on several fields of chemoinformatics as illustrated by Figure 1.1. Improvements in toxicity SAR and QSAR are triggered by advances in any one of the three depicted axes: chemistry, statistics or toxicology.

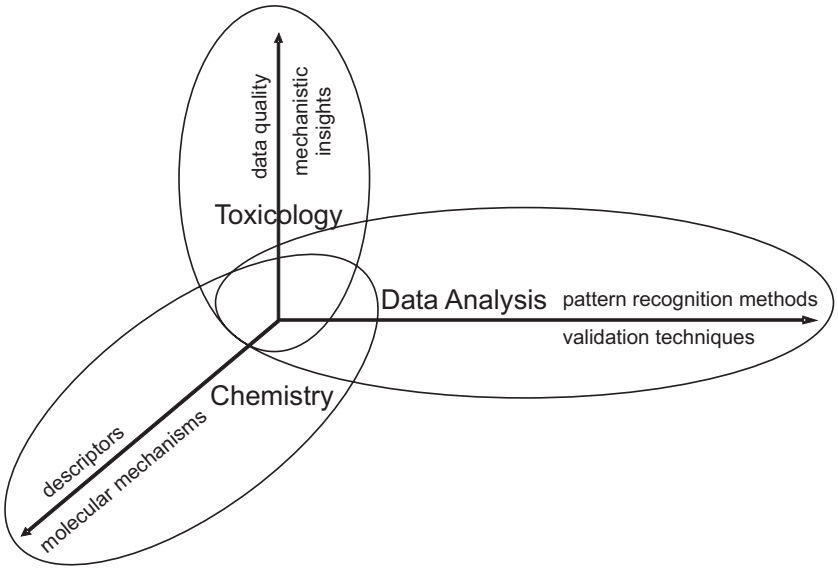


Fig. 1.1 Disciplines contributing to field of toxicity SAR and QSAR.

Chemistry: Progress in chemistry greatly increased the type and number of descriptors which can be used today. A descriptor is a numerical representation of a molecule and can be a scalar, a vector or even a matrix. It describes a feature of a molecule which can be related to the property of interest. Descriptors can be broken down into four types [16]: topological descriptors derived from the molecular graph, 3D-descriptors derived from molecular conformation(s), descriptors which require the calculation of the molecular wavefunction and as a fourth type descriptors which describe modifications of molecular structure (e.g. the acid dissociation constant, pK_{a}) or surrogate experimental measurements (e.g. partitioning coefficients). The Handbook of Molecular Descriptors lists over 3100 descriptors [17] and the MOSES-library developed in our group allows the calculation of over 200 atomic and vectorial descriptors [18] whose possible combinations increase in combinatorial fashion. The possible combinations of these descriptors leads to a multitude of possible descriptor combinations. An introduction into the calculation of molecular descriptors is given by Terfloth [19] and an advanced discussion by Gasteiger [6]. More details about the descriptors

used in this work will be given in the Materials and Methods sections of the three studies which comprise this thesis.

The second contribution of chemistry mentioned in Figure 1.1 is the knowledge about chemical reaction mechanims. As an example, electrophiles can bind to cellular nucleophiles through reactions like nucleophilic substitution, Schiff-base formation, Michael-type addition. Chemical reactivity models can be applied to describe such effects.

Data Analysis: In most cases, SAR and QSAR models are multivariate models, i.e., based on more than one descriptor. As many processes in toxicology follow nonlinear relationships, nonlinear modeling might be required. In cases where the functional relationship is completely unknown, the power of neural network approaches can come into play [20, 21]. Thus, the first contribution of data analysis depicted in Figure 1.1 can be defined as finding a pattern recognition method appropriate to the relationship inherently contained in the data. To take an example, it is not necessary or beneficial to model a linear relationship with neural networks. If, on the other hand, a large increase in predictive power is observed when neural networks are applied, this is a strong indicator that nonlinearities have to be considered for that kind of modeling problem. Additionally, statistical methods should be used to define what can be considered an "improvement", how to test it and also to put the improvement into a context, i.e., to judge how relevant the improvement is, which is much less trivial than it might sound.

The second contribution of data analysis methods shown in Figure 1.1 are validation techniques. They are of crucial importance for all inductive learning methods, because it can be shown that estimates of the determination coefficient, R^2, and also of the error rate of classification methods have an optimistic bias when calculated with data to fit the model (the training set) [21–23]. A number of methods can be used to estimate the predictive power of a model realistically, e.g. resampling methods like cross-validation or splitting the data set into training and test set [21, 24].

Another very important aspect is the applicability domain of a QSAR model. Almost all models are based on a limited subset of "chemical space". The consequence is that such a model should only be used for predictions within the domain of the model. If a model is based on only one descriptor the applicability domain would be the range between the minimum and the maximum of the descriptor. However, in the multivariate case this is not

a trivial problem. A comprehensive report on the currently used methods to determine a model's applicability domain has been published by Netzeva *et al.* [25].

Toxicology: Toxicity data are difficult to measure. This has been shown in several ring experiments where, depending on the species, differences of up to two orders of magnitude were observed [26]. This affects QSAR models in a way that data determined with different test protocols cannot be compared and data from different laboratories can be used for modeling only to a very limited extent [27]. Therefore, one indispensible contribution of toxicology to toxicity QSARs is the development of clearly defined endpoints and the generation of high quality data [28]. The definition of meaningful endpoints is a difficult task as the effects of toxicants are species-specific, organ-specific and time-dependent [29]. The second vital aspect is that toxicology can generate new insights into molecular mechanisms which in turn allow the focused search of descriptors and the interpretation of modeling results.

Currently several differing approaches coexist in the field of toxicity QSAR which will be described in more detail in the next section. A more detailed account of toxicity QSAR is given in a recent review [30] and an additional two covering specifically drug toxicity [12, 31]. A textbook introduction to the use of chemoinformatics methods for toxicity prediction is given by Kleinöder *et al.* [32].

1.2.2 The Mode of Toxic Action Approach

Three of the currently applied approaches to make quantitative predictions will be discussed in this section: the so-called Hansch-Analysis based on sets of clearly defined congeners [14, 15], global models using a tiered approach for database screening [33], and models based on common mode of toxic action (MOA).

Hansch-Analysis: The classic QSAR approach of Hansch was proposed for a series of indole derivates acting as plant growth regulators [15]. Their activity could be described by the following equation:

$$\log C^{-1} = k_1 \Pi + k_2 \sigma + k_3 \tag{1.1}$$

in which the inverse of C, the effect concentration of the toxicant is related to a hydrophobicity term, Π, and an electronic term (the Hammett substituent constant), σ. Steric terms can be added to this equation (typically Taft's steric parameter, E_s). This separation of descriptors

into hydrophobic, electronic and steric factor has proven enormously successful. In 2001 a database of QSARs following the Hansch approach contained 7100 QSAR equations for all kinds of chemical classes and many toxicity endpoints [34, 35]. However, the common restriction to structural congeners has the consequence that most of these QSARs are of quite local nature from a chemical diversity point of view (i.e., based on a series of compounds from the same chemical class). Additionally, the abundance of QSARs makes it also difficult to decide which equation to apply for someone who is not an expert in the Hansch approach.

Global Models: In an attempt to overcome such shortcomings another approach was proposed by Wang *et al.* [33]. They used a data set of 46'896 compounds with LD_{50}-values extracted from the Registry of Toxic Effects of Chemical Substances (RTECS) data base. In a tiered approach, they first used a coarse filter to group compounds into classes with similar structures and then for each of these classes developed quantitative models using CoMFA [36]. Similarity was defined using framework indicators and Tanimoto coefficients. However, some shortcomings of this approach might be that compounds binding to a given receptor do not necessarily have the same structural pattern nor have a high Tanimoto coefficient.

Modes of Toxic Action (MOA) approaches: MOA based approaches also try to overcome the shortcomings of approaches based on a chemical class [28], however with the notable difference that the class assignment is based on experimental data.

The term "mode of toxic action" refers to a more general effect or physiological response at a higher level of biological organization than the term "mechanism of toxic action" which is defined as the action of a toxicant at the molecular level

Figure 1.2 should illustrate how far MOA classes can be from what one generally percieves as structurally similar. While 2,6-dichlorophenol is a baseline toxicant (or polar narcotic) the structurally similar 2,3,4,5-tetrachlorophenol (2345TCP) has a completely different MOA, namely uncoupling of oxidative phosphorylation [37]. Thus, when following the Hansch approach one should not apply the same QSAR equation to these two compounds. Even more surprisingly the very dissimilar compound carbonyl cyanide *para*-trifluoromethoxyphenylhydrazone (FCCP) is also an uncoupler [38]. So 2345TCP and FCCP would pose a problem for global models based on classes of similar compounds like the one proposed by Wang *et al.* [33].

As 2345TCP and FCCP act via the same MOA it should be possible to model them in

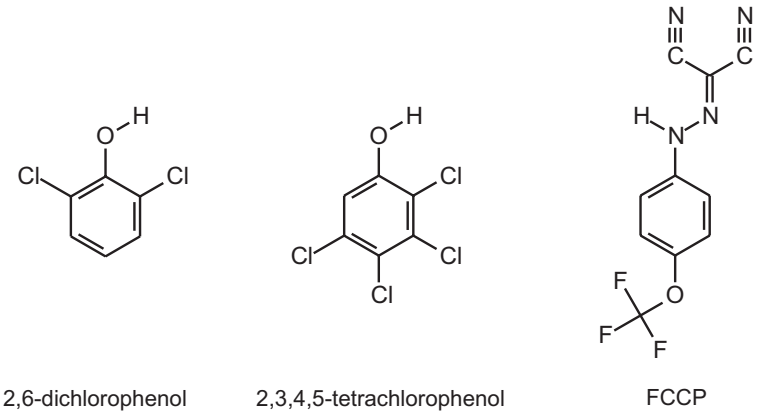


Fig. 1.2 Three structures with two different MOA.

the same QSAR equation. This ambitious plan of establishing MOA based QSARs was strongly pushed by the US-EPA for the last 10 years with the establishment of the fathead minnow acute toxicity data base [8, 39]. It contains 617 compounds with measured acute median lethal concentrations, LC_{50}, and about 280 high confidence MOA assignments with several experimental methods like joint toxicity evaluations, fish acute toxicity syndromes (FATS). However, the prerequisite for the prediction of unknown compounds with a MOA based approach is the successful classification into MOAs. The most relevant MOA based toxicity prediction system, ASTER (ASsessment Tool for Evaluating Risk) of the US-EPA selects the appropriate QSAR for a new compound on the basis of the occurrence of chemical fragments [8, 28]. Nevertheless, a fragment-based rule system is limited because it reduces a chemical structure to a specified substructure and ignores the other topological and potential electronic features of the entire compound which may influence its propensity to act under a given mode of toxic action [29]. Therefore, it is highly desirable to establish classification methods which could overcome this limitation.

A second important motivation to push for MOA classification is the assessment of mixture effects. In “real-life” risk assessment often complex mixtures of organic chemicals have to be compared, e.g., the effluent of industrial waste-water. The effect of compounds sharing the same MOA are concentration additive, thus although the single concentrations of waste-water are below the no-effect level they can add up to a significant toxic effect. MOA classification systems could help to asses the likeness of such an event.

Table 1.1 lists the four main studies, with four different MOA data sets, published so far.

Table. 1.1 Previous studies on the classification of compounds into MOA. Abbreviations of the statistical methods: PLS (Partial Least Squares), KNN (K-nearest neighbour), LVQ (Learning Vector Quantization), LDA (Linear Discriminant Analysis). OC: **O**verall **c**orrect classification rate, GC: **C**orrect classification rate within **G**roup (sensitivity, producers accuracy.)

Ref.	MOA	Size	Approach	Stat. Methods	Results for test set
Nouwen *et al.* (1997) [40]	Nonpolar narcosis Polar narcosis Reactives Specifically acting	512 (training) 205 (test) Diversity: High	**A priory class membership:** Structural fragments and expert judgement **Descriptors:** Presence or absence of 1429 fragments	PLS	OC: 67% (test set with 73 cmpds. after removal of salts, organic acids, esthers)
Basak *et al.* (1998) [41]	Narcotics I or II, Uncouplers, Neurotoxicants, Respiratory blockers and neurodepressants, AChE-Inhibition, Electrophile/Proelectrophile	220 (training) 63 (test) Diversity: High	**A priory class membership:** Expert decisions based on joint toxic action studies, physiological and behavioral response data and literature **Descriptors:** 151 topological indic. / 60 in final model	KNN LVQ LDA	OC: 72–95% GC: 0–100%
Nendza and Müller (2000) [42]	Nonpolar, polar nonspecific, Uncouplers, AChE-Inhibition, Photosynthesis-Inhibition, Resp. Blockers, SH-Alkylating, Reactives, Estrogenic	115 (training) Diversity: Very high	**A priory class membership:** Measured activity in *in vitro* tests (except reactives taken from Literature) **Descriptors:** 24 physico-chemical descriptors / 10 in final model	LDA PLS	OC: 89.6% (for training set) No test set
Aptula *et al.* (2002) [43]	Polar narcosis Uncouplers Proelectrophiles Soft electrophiles	111 (training) 110 (test) Diversity: Intermediate (only phenols)	**A priory class membership:** Structural rules derived from earlier studies [44] **Descriptors:** Physico-chemical: log K_{ow}, log D_{ow}, pK_{a}, LUMO, HOMO, H-Bond donor and acceptor counts	LDA	OC: 86–89% GC: 66–97%

The data set of Aptula *et al.* [43] was additionally studied by other authors [45–47] and further refined by the initial authors [48].

A look at Table 1.1 quickly makes clear that there are enormous differences in the current approaches. The most striking difference is in the toxicological methods used to determine the MOA (the *a priori* class membership). From this viewpoint, the study of Nendza and Müller [42] is the most interesting because the MOA was determined using a battery of MOA-specific *in vitro* tests. The benefit of such an approach is that it potentially shifts the definition of MOA from its physiological side towards a molecular mechanism side.

There is wide-spread agreement that quantitative data from different toxicity test protocols should not be merged or this should be done only with great caution [27]. Presently, qualitative toxicity data cannot be merged without caution either in the case of MOAs as is illustrated in Figure 1.3. The overlap of three data sets with different experimental MOA assignment was determined and an overlap of 13, 13 and 19 compounds, respectively, was found.

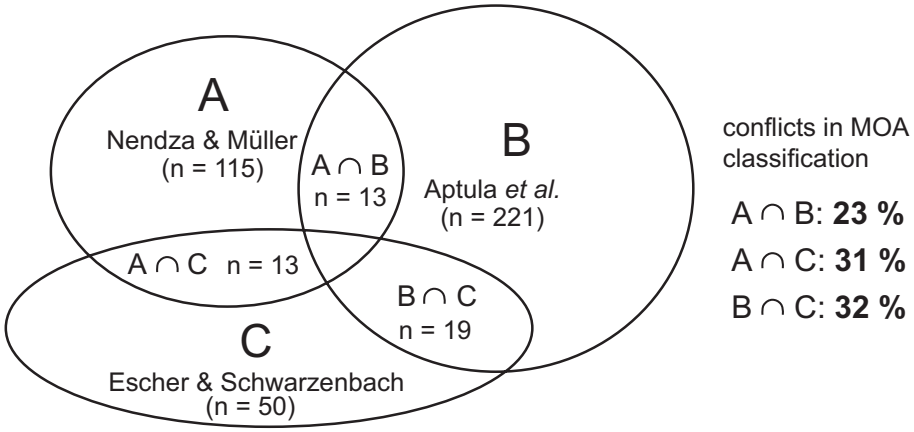


Fig. 1.3 Overlap and disagreement of experimental a priori MOA assignment, e.g., data set B and C had 19 compounds in common and six compounds or 32% of these 19 were not assigned to the same MOA. References: A [42], B [43] and C [37]. Data set C is not mentioned in Table 1.1, because it has not been used for classification studies so far.

The disagreement in experimental assignments ranges from 23% to 32%. As the number of overlapping compounds is rather small these numbers have a high uncertainty, but they make clear that, unfortunately, it is not possible to merge the data sets mentioned in Table 1.1.

Data set C [37] has a high percentage of disagreement with both data set A and B. It was mainly caused by compounds determined as uncouplers of phosphorylation. It is therefore

revealing to compare its test system with the test systems of the other two data sets. In data set A, uncouplers were classified by their activity in isolated mitochondria, in data set B, they were classified by rules derived from earlier studies [44] and in data set C, uncoupling was determined with a spectroscopic method [49, 50] which can differentiate between uncoupling and inhibition of oxidative phosphorylation. One can assume that the method used in data set C is the most sensitive and most specific one for the MOA of uncoupling. Thus, the notion that a compound has only one specific Mode of Action needs to be challenged. In fact, the transitions between MOAs are rather continuous than sharp. Compounds can even exhibit multiple modes of action, a fact that has not been considered in classification studies so far although the ASTER-system gives the user several suggestions for a compound's MOA [28]. However, modeling this continuous transition between MOA requires high quality data. Data set C also has quantitative data of high quality which have not been used in QSAR studies so far. Therefore, this data set might allow one to develop MOA based models in the full sense, i.e., first build a classification method and subsequently build quantitative models for each class.

On the level of the descriptors there seems to be a trend from high-dimensional structure representations in the two older studies with 1429 and 151 descriptors, respectively, towards lower dimensional representations based on quantum-mechanical and other whole-molecule descriptors with 24 and 6 descriptors, respectively in the two more recent studies.

Concerning the pattern recognition methods there is a tendency to favor linear methods like linear discriminant analysis and partial least squares. Only one data set has been additionally studied with K-nearest neighbour and learning vector quantization. Hence, there seems to be room for additional improvement by testing other nonlinear methods like counter-propagation neural networks [20].

1.3 Objectives and Structure of this Work

The central objective of this work as with all MOA classification studies is to overcome the limitation of fragment-based approaches. This work consists of three parts which can be grouped according to the three disciplines illustrated by Figure 1.1.

1.3.1 Data Analysis

Chapter 2 focuses mainly on the data analysis domain of toxicity SAR and QSAR consisting first of using appropriate classification methods and, second, of applying validation methods which do not lead to misleading conclusions. The data set of Nendza and Müller [42] was used in this study because it had not been validated before and because of its interesting experimental approach. The methodology developed in Chapter 2 shall be used in Chapters 3 and 4.

Questions regarding the classification methods: The available toxicity data sets are mostly of small size. Additionally, nonlinear effects can be expected. Therefore, the following questions will be investigated: Which methods help to establish MOA classification models with higher predictive power? Is it advantageous to use flexible learning methods like neural networks for such data sets?

Questions regarding validation: How the predictive power should be estimated is still a controversial issue. The small size of most toxicity data sets aggravates the problems of misleading validation studies. There is a majority of experts in the field of QSAR who claim that validation of QSARs always require splitting data sets into training and test set. On the other side, there is a general agreement amongst statisticians that this leads to less than optimal and often arbitrary estimates especially in the case of small data sets. This raises the following questions: Under which circumstances is it appropriate to use cross-validation to estimate a classifiers predictive power and when does it lead to misleading conclusions?

1.3.2 Chemistry

Chapter 3 explores different approaches to calculate descriptors. The data set of Aptula *et al.* [43] with 221 phenols was used for this purpose. The published model is based on quantum-mechanic and whole-molecule descriptors like log P. Models based on these descriptors will be compared to vectorial descriptors weighted with different atomic properties also called structure descriptors. If a model based on structure descriptors can reach predictive power comparable to previously published models this would mean a simplification concerning ease and speed of descriptor calculation.

Once such a model is established it can be used for exploratory studies of large data bases. In this work it was applied to screening the 3142 phenols of the open NCI database

with 248'259 compounds.

An additional aspect of validation methods covered in Chapter 3 is the determination of the applicability domain. As the validation methods are always based on limited data sets, they cannot foresee the case of test set which requires heavy extrapolation. Therefore, it is advisable to define a model's applicability domain. The basic principle of all measures for the applicability domain is to define a suitable distance measure between the training set data and the compounds to be predicted. A simple measure based on PCA-projections will be used on a test set and on the exploratory data set of the phenols from the open NCI database.

1.3.3 Toxicology

Chapters 4 and 5 focus on the advantages of an MOA-specific *in vitro* tests. A data set measured with a recently developed *in vitro* test system was compiled [37]. The advantages of this data set is that it included measured values for several descriptors which allow the exact quantification of the uptake into the membrane and thus, to calculate the concentration of toxicant at the site of action. Additionally, insights gained in the field of biophysics will be used to focus the search for appropriate descriptors.

The compounds belong to the two MOAs uncouplers of oxidative phosphorylation and baseline toxicants. For both MOA measured activity data of high quality are available. Therefore, the goal of Chapters 4 and 5 is to establish a classification model to distinguish these two MOAs and subsequently to build a quantitative model within each class.

Each chapter is followed by a comment putting the results in the general context of chemoinformatics. Additionally, suggestions for further research are made. In the conclusion of the work the results of the three chapters is put into the bigger perspective of the goal to establish an MOA based approach for screening large data bases.

The prediction of metabolism and the prediction of receptor mediated toxicity are also central for the prediction of toxicity, but these two fields were out of scope of this work, because they are vast fields of their own with their own methodology. There is some confidence that the enormous efforts made in receptor based drug-design can be used to effectively predict such types of toxicity [28, 51, 52].

References

[1] R. Carson, *Silent Spring*, Houghton Mifflin, Boston, **1984**.

[2] M. Fleischer, S. Kelm, D. Palm, "Regulation and Innovation in the Chemical Industry", Report EUR 19735 EN, Joint Research Center, European Commission, Institute for Prospective Technological Studies, Seville, **2000**.

[3] R. Allanou, B. G. Hansen, Y. van der Bilt, "Public Availability of Data on EU High Production Volume Chemicals", Report EUR 18996 EN, **1999**.
`http://ecb.jrc.it/DOCUMENTS/Existing-Chemicals/`

[4] "White Paper Strategy for a future Chemicals Policy", Commission of the European Communities, COM(2001) 88 final, **2001**.

[5] "Proposal Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)", COM(2003)644 final, Brussels, Belgium: European Commission, **2003**.
`http://europa.eu.int/eur-lex/en/com/pdf/2003/com2003_0644en.html`

[6] J. Gasteiger, "A Hierarchy of Structure Representations", in "Handbook of Chemoinformatics", J. Gasteiger (Ed.), Wiley-VCH, **2003**, Volume 3, 1034–1061.

[7] J. Hermens, S. Balaz, J. Damborsky, W. Karcher, M. Muller, W. Peijnenburg, A. Sabljic, M. Sjostrom, "Assessment of QSARS for predicting fate and effects of chemicals in the environment: an international European project", *SAR QSAR Environ. Res.* **1995,** *3,* 223–236.

[8] C. L. Russom, S. P. Bradbury, S. J. Broderius, D. E. Hammermeister, R. A. Drummond, "Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas)", *Environ. Toxicol. Chem.* **1997,** *16,* 948–967.

[9] R. Bias, "REACH und QSAR – gemeinsame Aufgabe von Industrie und Wissenschaft)", *Mitteil. Fachgr. Umweltch. u. Ökotox.* **2004,** 3.

[10] F. Pedersen, J. de Bruijn, S. Munn, K. van Leeuwen, "Assessment of additional testing needs under REACH. Effects of (Q)SARS, risk based testing and voluntary industry initiatives", Institute of Health and Consumer Protection (IHCP), Ispra, Italy, **2003**.

[11] P. Krajcsi, F. Darvas, "High-Throughput In Vitro Toxicology", in "High-Throughput ADMETox Estimation", F. Darvas, G. Dormán (Eds.), Eaton Publishing, **2002**, chapter 6, 75–81.

[12] M. T. D. Cronin, "Computational methods for the prediction of drug toxicity", *Curr. Opin. Drug Disc. Dev.* **2000,** *3,* 292–297.

[13] *Handbook of Chemoinformatics*, J. Gasteiger (ed.), 4 Volumes, Wiley-VCH, Weinheim, **2003**.

[14] C. Hansch, T. Fujita, "ρ-σ-π-Analysis. A method for the correlation of biological activity and chemical structure", *J. Am. Chem. Soc.* **1964,** *86,* 1616–1626.

[15] C. Hansch, "Quantitative approach to biochemical structure-activity relationships", *Acc. Chem. Res.* **1969,** *2,* 232–239.

[16] A. Bender, R. C. Glen, "Molecular similarity: a key technique in molecular informatics", *Org. Biomol. Chem.* **2004,** *2,* 3204–3218.

[17] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, **2000**.

[18] A. Herwig, T. Kleinöder, J. Marusczyck, L. Terfloth, J. Gasteiger, *Programmer's Guide - Molecular Structure Encoding System*, Computer-Chemie-Centrum und Institut für Organische Chemie, **2005**, version 1.0.

[19] L. Terfloth, "Calculation of Structure Descriptors", in "Chemoinformatics", J. Gasteiger, T. Engel (Eds.), Wiley-VCH, **2003**, chapter 8, 401–437.

[20] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, 2^{nd} Edition, Wiley-VCH, Weinheim, **1999**.

[21] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*, Springer series in statistics, Springer, New York, **2001**.

[22] B. Efron, "How Biased is the Apparent Error Rate of a Prediction Rule?", *JASA* **1986,** *81,* 461–470.

[23] B. Efron, R. Tibshirani, "Improvements on Cross-Validation: The .632+ Bootstrap Method", *JASA* **1997,** *92,* 548–560.

[24] F. E. Harrell, Jr., *Regression Modeling Strategies - With applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer-Verlag, New York, **2001**.

[25] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. van de Sandt, W. Tong, G. Veith, C. Yang, "Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure Activity Relationships – The Report and Recommendations of ECVAM Workshop 52", *ATLA* **2005,** *33,* 1–19.

[26] J. Blok, F. Balk, "Environmental Regulation in the European Community", in "Fundamentals of Aquatic Toxicology: Effects, Environmental Fate and Risk Assessment", G. M. Rand (Ed.), Taylor & Francis, UK, **1995**, book chapter 27, 775–802, second edition Edition.

[27] M. T. D. Cronin, T. W. Schultz, "Pitfalls in QSAR", *THEOCHEM* **2003,** *622,* 39–51.

[28] S. P. Bradbury, C. L. Russom, G. T. Ankley, T. W. Schultz, J. D. Walker, "Overview of data and conceptual approaches for derivation of quantitative structure-activity relationships for ecotoxicological effects of organic chemicals", *Environ. Toxicol. Chem.* **2003,** *22,* 1789–1798.

[29] S. P. Bradbury, "Predicting modes of toxic action from chemical structure: An overview", *SAR QSAR Environ. Res.* **1994,** *2,* 89–104.

[30] T. W. Schultz, M. T. D. Cronin, T. I. Netzeva, "The present status of QSAR in toxicology", *THEOCHEM* **2003,** *622,* 23–38.

[31] J. C. Dearden, "In silico prediction of drug toxicity", *J. Comput. Aided Mol. Des.* **2003,** *17,* 119–127.

[32] T. Kleinöder, A. Yan, S. Spycher, "Prediction of Properties of Compounds", in "Chemoinformatics", J. Gasteiger, T. Engel (Eds.), Wiley-VCH, Weinheim, **2003**, 487–514.

[33] J. Wang, L. Lai, Y. Tang, "Data mining of toxic chemicals: structure patterns and QSAR", *J. Mol. Model.* **1999,** *5,* 252–262.

[34] "C-QSAR", BioByte Corporation, Claremont, CA 91711, **2001**.
`http://www.biobyte.com`

[35] R. Garg, A. Kurup, C. Hansch, "Comparative QSAR: on the toxicology of the phenolic OH moiety", *Crit. Rev. Toxicol.* **2001,** *31,* 223–245.

[36] R. D. Cramer III, J. D. Bunce, D. E. Patterson, "Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies", *Quant. Struct.-Act. Relat.* **1988,** *7,* 18–25.

[37] B. I. Escher, R. P. Schwarzenbach, "Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms", *Aquat. Sci.* **2002,** *64,* 20–35.

[38] R. Benz, S. McLaughlin, "The molecular mechanism of action of the poroton inophore FCCP (carbonyl cyanide para-trifluoromethoxyphenylhydrazone)", *Biophys. J.* **1983,** *41,* 381–398.

[39] "EPA Fathead Minnow Aquatic Toxicity Database", U.S. Environmental Protection Agency, **2005**.
`http://www.epa.gov/nheerl/dsstox/Databases.html`

[40] J. Nouwen, F. Lindgren, B. Hansen, W. Karcher, "Classification of environmentally occurring chemicals using structural fragments and PLS discriminant analysis", *Environ. Sci. Technol.* **1997,** *31,* 2313–2318.

[41] S. C. Basak, G. D. Grunwald, G. E. Host, G. J. Niemi, S. P. Bradbury, "A comparative study of molecular similarity, statistical, and neural methods for predicting toxic modes of action", *Environ. Toxicol. Chem.* **1998,** *17,* 1056–1064.

[42] M. Nendza, M. Müller, "Discriminating toxicant classes by mode of action: 2. Physicochemical descriptors", *Quant. Struct.-Act. Relat.* **2000,** *19,* 581–598.

[43] A. O. Aptula, T. I. Netzeva, I. V. Valkova, M. T. D. Cronin, T. W. Schultz, R. Kühne, G. Schüürmann, "Multivariate discrimination between modes of toxic action of phenols", *Quant. Struct.-Act. Relat.* **2002,** *21,* 12–22.

[44] T. W. Schultz, G. D. Sinks, M. T. D. Cronin, "Identification of mechanisms of toxic action of phenols to Tetrahymena pyriformis from molecular descriptors", in "Quantitative Structure-Activity Relationships in Environmental Sciences-VII", F. Chen, G. Schüürmann (Eds.), SETAC Press: Pensacola, FL, **1997**, Proceedings of QSAR 96, Elsinore, DK, June 24-28, 1996, 329–337.

[45] S. Ren, "Use of molecular descriptors in separating phenols by three mechanisms of toxic action", *Quant. Struct.-Act. Relat.* **2002,** *21,* 486–492.

[46] S. Ren, P. D. Frymier, T. W. Schultz, "An exploratory study of the use of multivariate techniques to determine mechanisms of toxic action", *Ecotoxicol. Environ. Safety* **2003,** *55,* 86–97.

[47] S. Ren, "Ecotoxicity prediction using mechanism- and non-mechanism-based QSARs: a preliminary study", *Chemosphere* **2003,** *53,* 1053–1065.

[48] G. Schüürmann, A. O. Aptula, R. Kuehne, R. U. Ebert, "Stepwise Discrimination between Four Modes of Toxic Action of Phenols in the Tetrahymena pyriformis Assay", *Chem. Res. Toxicol.* **2003,** *16,* 974–987.

[49] B. I. Escher, M. Snozzi, K. Häberli, R. P. Schwarzenbach, "A new method for simultaneous quantification of uncoupling and inhibitory activity of organic pollutants in energy-transducing membranes", *Environ. Toxicol. Chem.* **1997,** *16,* 405–414.

[50] B. I. Escher, M. Snozzi, R. P. Schwarzenbach, "Uptake, Speciation, and Uncoupling Activity of Substituted Phenols in Energy Transducing Membranes", *Environ. Sci. Technol.* **1996,** *30,* 3071–3079.

[51] W. Tong, H. Fang, H. Hong, Q. Xie, R. Perkins, D. M. Sheehan, "Receptor-mediated toxicity: QSARs for estrogen receptor binding and priority setting of potential estrogenic endocrine disruptors", in "Predicting Chemical Toxicity and

Fate", Food and Drug Administration's National Center for Toxicological Research,Jefferson,AR,USA, Food and Drug Administration's National Center for Toxicological Research,Jefferson,AR,USA, **2004**.

[52] M. A. Lill, M. Dobler, A. Vedani, "In silico prediction of receptor-mediated environmental toxic phenomena - Application to endocrine disruption", *SAR QSAR Environ. Res.* **2005,** *16,* 149–169.

Chapter 2

Comparison of Different Classification Methods Applied to a Mode of Toxic Action Data Set

Simon Spycher[†], Monika Nendza[‡] and Johann Gasteiger[†]*

[†]Computer-Chemie-Centrum und Institut für Organische Chemie,
Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany

[‡]Analytisches Laboratorium, Bahnhofstraße 1, D-24816 Luhnstedt, Germany

Spycher, S., Nendza, M., Gasteiger, J. (2004), *QSAR & Comb. Sci.*, **23**, 779–791.[a]

Abstract

Successful discrimination of compounds by mode of toxic action (MOA) is a prerequisite for process-based quantitative structure-activity relationship (QSAR) approaches. A data set of 115 compounds comprising nine MOA classes and 24 descriptors has been studied with several classification methods: multinomial logistic regression (multinom), linear discriminant analysis (LDA), partial least squares (PLS), and counter-propagation neural networks (CPG NN). Variables were selected with stepwise methods and with a genetic algorithm (GA) for the CPG NN. Five-fold cross-validation was used for validating the models and the advantages and disadvantages of this validation method are critically discussed. Without variable selection the predictive power of the models ranges between 51% and 53% cross-

[a]This Chapter corresponds exactly to the original publication other than the last section after the reference list and the numbering.

validated overall correct classification. With appropriate parameter selection the predictive power slightly increased to 52–59%. The experimental data showed that a number of compounds were active in more than one MOA. Multinom and CPG NN models for multiple MOAs were derived for both single and multiple MOA data. The consideration of multiple MOAs resulted in a slight increase in predictive power even when all MOAs were modeled at the same time.

Abbreviations: CPG NN: counter-propagation neural networks; GA: genetic algorithm; λ: penalty factor; LDA: linear discriminant analysis; MOA: mode of toxic action; multinom: multinomial logistic regression; PLS: partial least squares; Penalized ML: penalized maximum likelihood; PCA: principal component analysis; PS II: photosystem II; r^2: Pearson correlation coefficient; R^2: coefficient of determination; s: standard deviation; Q^2: leave-one-out cross-validated coefficient of determination; QSAR: quantitative structure-activity relationship

2.1 Introduction

Difficulties in toxicity modeling, for high-throughput screening in the drug discovery process [1] as well as for environmental hazard and risk assessment of industrial chemicals, may be caused by the variability in experimental endpoints and the limited availability of consistent data. Because toxic effect concentrations are species-specific, organ-specific and time-dependent [2], they may vary substantially depending on species and type of test. As a consequence, a great diversity of models has been derived. Most are local in chemical domain (i.e., based on a series of compounds from the same chemical class), but global in toxic effects (e.g., lethality caused by different mechanisms). One approach is the development of QSAR models based on common mode of action (MOA) instead of chemical class. A MOA can be defined as a set of physiological and behavioral signs characterizing a type of adverse biological response, while a toxic mechanism is defined as the biochemical process underlying a given MOA [3]. Reliable classification methods by MOA are required to select appropriate QSARs for predictions and also may help to assess the effects of mixtures because compounds sharing the same MOA are generally concentration additive. Several data sets on MOA classification have been published [4–6]. The present study uses the data from

Ref. [5] because in this work MOAs were not assigned from structural features, but based on testing with a battery of MOA-specific *in vitro* assays. Though the data set is rather small (number of compounds, n = 115), the compounds are highly diverse in terms of chemical structure. A MOAclassification model using 10 physicochemical descriptors after variable selection by principal component analysis (PCA) and stepwise linear discriminant analysis (LDA) exhibited excellent fit [5]. It has however not been validated to date. Because initial tests indicated a rather low predictive power, this study investigates statistical tools to establish models with increased predictive performance. The objective is to compare the efficacy of different pattern recognition methods (multinomial logistic regression (multinom), linear discriminant analysis (LDA), partial least squares (PLS) and counter-propagation neural networks (CPG NN)) to address three major questions:

1. Are there pattern recognition methods that lead to more predictive classification models than others? Because many relationships in toxicology are intrinsically nonlinear, one might expect an improved performance if methods are used that model nonlinearities. However, it is very hard to say *a priori* that a certain method is suited best for modeling. Figure 2.1 illustrates a possible reason: the predictive power of any model depends, among other reasons, on the size of the data set. Thus, it is quite likely that the small data sets common in toxicity modeling result in models on the left end of the x-axis. In such a situation more “rigid” models with fewer degrees of freedom can outperform flexible nonlinear models, which can turn out to be too flexible in such a situation [7].

2. What is a suitable validation method for small data sets as often encountered with toxicity QSARs? Is it necessary and possible to integrate variable selection in the validation process? Both questions are currently controversially discussed with some strongly recommending to split data sets into training and test sets [8] and others favoring cross-validation [9]. A general work flow suitable for small data sets is presented, and the effects of integrating variable selection in the validation process are examined.

3. Do models benefit from considering multiple MOAs? Some compounds were measured to be active in more than one MOA, a fact that is discussed in detail in section 2.2.1. Simultaneous handling of multiple activities, so-called activity spectra, by

multinom and CPG NN is presented and the resulting models are evaluated relative to single MOA models.

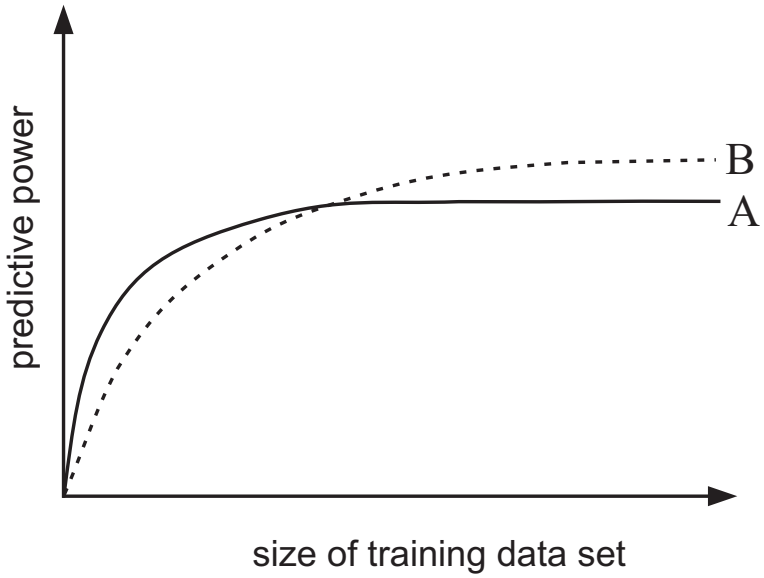


Fig. 2.1 Predictive power as a function of the size of a data set for two methods A and B. Method A uses less degrees of freedom for fitting the model than method B which is more flexible (e.g., a neural network method). For small training sets the predictive power of less flexible methods can be higher.

2.2 Materials and Methods

2.2.1 Data

The original data set contained 115 compounds assigned to nine MOAs, kindly provided by Andrea Wenzel of the Fraunhofer Institute for Molecular Biology and Applied Ecology, Schmallenberg, Germany. The compounds were tested with six different *in vitro* assays specified in Table 2.1 [10]. The analysis of the *in vitro* effect concentrations showed substantial differences regarding the specificity of the assays. MOA assignment was unequivocal if the assay measured activity on a specific target, such as acetylcholinesterase or estrogenic receptors. It was not as clear when general phenomena were measured like inhibition of photosynthesis by decrease of O_2-production in algae. Reexamination of the data resulted in removing 11 substances, leaving a reduced data set of 104 compounds in seven MOAs (Table 2.1).

Table. 2.1 Type of MOA, specific assay and number of compounds for each MOA after reassessment of the laboratory data.

MOA	Name	*In vitro* Assay	Number of cpds.
1.	Nonspecific nonreactive toxicants	Neutralred assay	32
2.	Uncouplers of oxidative phosphorylation	Mitochondria O_2-consumption	25
3.	Inhibitors of photosystem II	Algae O_2-production	9
4.	Inhibitors of acetylcholinesterase	AChE enzyme assay	14
5.	Thiol-alkylating agents	GSH assay	9
6.	Reactives	Neutral red assay	7
7.	Estrogenic compounds	Yeast receptor assay	8
			Total 104

It has to be emphasized that compounds were only eliminated to obtain a consistent data set and not because they had an unfavorable fit to a model:

1. The MOA “Respiratory Blockers” was eliminated for statistical reasons, because it contains only three compounds. This does not allow one to validate predictions for this MOA.

2. Five inhibitors of photosynthesis were eliminated because they interact with specific plant cell targets, but not with photosystem II (PS II): dinoterb (uncoupler of oxidative phosphorylation [11]), oxyfluorfen (inhibitor of protoporphyrinogen oxidase [11], nitrofen (inhibition of energy transfer in the ADP synthase reaction [12] or inhibitor of protoporphyrinogen oxidase [11]) and the two compounds 1,3-dinitrobenzene and p-chloromercuribenzoic acid for which no information was found to assign them to a specific target within the plant cell. The remaining compounds of this MOA are known inhibitors of PS II [11, 13, 14]. The narrowing of the MOA “Inhibition of Photosynthesis” to “Inhibition of PS II” seems justified if one considers how different the possible targets within a plant cell can be, and that specialists in the field distinguish up to 20 different modes of inhibiting photosynthesis [11].

3. The MOAs “Nonpolar Nonspecific Toxicity” and “Polar Nonspecific Toxicity” were merged to one MOA “Nonspecific Nonreactive Toxicants” following the approach of Basak et al. [15] who discriminated these two classes in a later step.

4. The comparison of the Molfiles with the CAS registry numbers of the substances used in the *in vitro* tests showed three cases of disagreement: thiazafluron, isobutylacrylate

and benzylbutylphtalate were eliminated because the calculated descriptors were not for those structures for which the activity had been measured.

Many compounds assigned to a certain MOA showed considerable activity also in other *in vitro* assays particularly uncoupling of oxidative phosphorylation and inhibition of photosynthesis. Figure 2.2 illustrates this phenomenon with the activity data of the uncouplers.

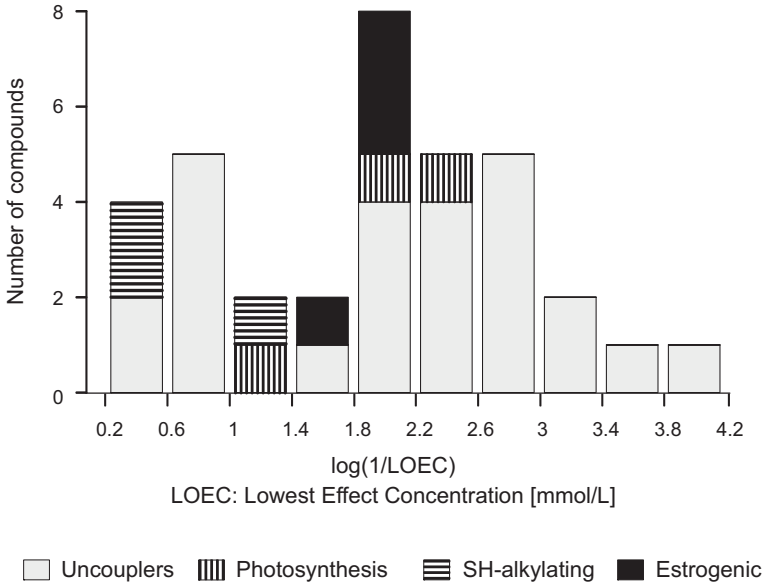


Fig. 2.2 Distribution of measured *in vitro* activities for uncouplers of oxidative phosphorylation. 25 compounds were determined to be uncouplers in the original study. Ten compounds, some of them highly active in other assays, showed also considerable activity in the uncoupling assay..

The light grey shaded bars show the activity of the 25 compounds determined as uncouplers in the original study. The other bars represent compounds determined to belong to other MOAs, e.g., four estrogenic compounds were highly active in the yeast receptor assay, but they nevertheless showed considerable activity in the uncoupler test, too. There are two possible reasons for the phenomenon shown in Figure 2.2: the compounds assigned to other MOAs are either active in more than one MOA and thus do have multiple MOAs, or the *in vitro* assays are not specific enough. A comparison with other published experimental data did not answer this question [6, 16], but revealed a high variability between different experimental MOA assignments. However, if these compounds really have multiple MOAs, and are not only artefacts of the experimental methods, classification models should benefit from considering this fact.

24 physicochemical descriptors calculated in [5] were used for the present study. A PCA

allowed to group them into electronic charge parameters, steric bulk, lipophilicity, and orbital energies. In the original study the final model consisted of 10 descriptors: log K_{ow}, HOMO, V^{+}, Q_{AV}, $\mathrm{H}^{+}_{\mathrm{MAX}}$, MR, MW, $\mathrm{D}_{\mathrm{EFF}}$, SASA, SAVOL.

2.2.2 Preprocessing

Scaling: All descriptors were mean centered and scaled to unity, because the standard deviation of the descriptors varied by almost four orders of magnitude and some methods like PLS are sensitive to the choice of the scaling technique. Correlation analysis: To detect highly correlated pairs of descriptors (X-variables) and to obtain a general impression of the correlation of the descriptors, the following steps were taken:

1. identify variables j and k with the highest squared Pearson correlation coefficient, r^2,
2. calculate the sum, S, of the correlation coefficients with all other variables for both j and k,
3. eliminate the variable with the higher S,
4. repeat steps 1–3 until the highest remaining r^2 falls below a specified level, r^2_{limit}.

A value of r^2_{limit} = 0.9 was chosen to remove highly correlated limit variables. As long as r^2 limit is high, it is of secondary importance which variable is removed, as both variables contain almost the same information. To eliminate less correlated variables, other variable selection methods should be used. Removing highly collinear variables is particularly important with LDA and logistic regression, whereas projection methods like PLS are assumed not to be affected by collinearity.

2.2.3 Variable Selection and Pattern Recognition Methods

2.2.3.1 Linear Discriminant Analysis (LDA)

LDA is the most common and well established method [17, 18] used in previous MOA classification studies. It is often used with stepwise variable selection. If a variable meets certain criteria (P-values) it is included or excluded from the model in an automated procedure. LDA-models were calculated using SPSS [19] using the following options: equal

prior probabilities, pooled covariance matrices and variable selection with forward stepwise (maximised F-ratio with P-inclusion 0.05 and P-exclusion 0.1).

2.2.3.2 Multiple logistic regression (multinom)

This method [7, 20, 21] can be placed somewhere between LDA and back-propagation neural networks. It is less flexible but more transparent than back-propagation and less prone to overfitting. On the other hand, it is based on fewer assumptions than LDA [7]. In order to prevent overfitting, maximum likelihood estimates were penalized: By subtracting a penalization term from the log likelihood (Eq. 1), too large parameter estimates are avoided [18, 21, 22].

$$\log L^{\lambda} = \log L - 1/2 \cdot \lambda \sum_{i=1}^{p} (s_i \beta_i)^2 \tag{2.1}$$

L denotes the usual likelihood function, λ a penalty factor, p the number of variables, s_i scale factors to make $s_i\beta_i$ unitless and β_i the regression coefficients. The higher the value of λ, the smaller are the estimates of β_i. The purpose of penalization is not to make a variable selection, but to avoid that the regression coefficients get too large and the model adapts to smoothly to the data. Logistic regression models were calculated using the VR-library in R [23]. By default, λ was set to 0.01 and stepwise backward (function stepAIC with k = 2) was used for variable selection.

2.2.3.3 Partial Least Squares (PLS)

PLS algorithms have been previously used in MOA classification studies [5, 24]. From orthogonal linear combinations of X, predictions of Y are obtained. PLS can handle highly correlated, noisy, and numerous X-variables and simultaneously predict several response variables, Y [25–27]. For PLS discriminant analysis (PLS-DA), the Y-variables are coded as a matrix of dummy vectors with 1 for actives and 0 for nonactives, e.g., a compound i that is active only in one assay (e.g., test number 2) is written as $Y_i = (0, 1, 0, 0, 0, 0, 0)$. In PLS model building, no variable selection was used. The number of components was increased as long as Q^2 (cross-validated R^2) increased. PLS models were calculated using Simca [28].

2.2.3.4 Counter Propagation Neural Networks (CPG NN)

CPG NN are based on Kohonen neural networks, also called Self-Organizing Maps, but were extended for modeling tasks, i.e., relating objects with their properties [29, 30]. Classifications are based on the projection of objects from multivariate space onto 2D clusters. The class prediction for new objects then depends on the 2D cluster the object is projected onto. This approach was successfully applied for the classification of different olive oils by fatty acid contents [31]. In this study we used a CPG NN architecture with multiple output layers as described earlier [32, 33] that to date has not been subjected to a model comparison study. Instead of projecting the compounds onto one single map, each MOA was modeled in a separate output layer. The Y-variables were represented by dummy vectors as described for PLS-DA. This multiple output layer approach improves prediction quality and interpretability in case of conflicts, i.e., when two compounds with different MOAs are assigned to the same neuron. Network dimensions of 7×5, with rectangular topology and 4300 cycles of training were chosen. CPG NN were calculated using SONNIA [34]. A genetic algorithm was used for variable selection with the CPG NN [35]. The common GA procedure of splitting the data set into three subsets, a training set for fitting the models, an evaluation set for determining the fitness of the individuals, and a test set for determining the predictive power, was not applicable because of the small size of the data set. Therefore, the same compounds were used for training and for evaluation of the fitness. Thus, one of the two necessary splits of the data set was avoided. To prevent the network just memorizing the data without conflicts, small network dimensions, i.e., a ratio of one neuron to four compounds, was used. This means that the GA had to force at least 4 compounds into each neuron. Large errors result if compounds from different classes are placed into the same neuron, and only variables that help to avoid such misclassifications have high fitness in the GA population. The additional parameters were set as follows: smallest RMS as fitness function, number of variables 7–12, population 1200, generations 80, crossover-probability 0.7 and permutation probability 0.05.

2.2.4 Multiple MOA

A comparative assessment of single vs. multiple MOA models was made using multinom and CPG NN based on activities in the uncoupling assay for 106 compounds: 25 definite

uncouplers, 10 compounds assigned to other MOAs, but also active in the uncoupling assay (cf. Figure 2.2), and 71 inactive chemicals. The data set comprised two previously excluded substances: dinoterb and p-chloromercuribenzoic acid which inhibit photosynthesis (but not PSII) and showed distinct activity as uncouplers. For single MOA models, compounds were assigned according to their predominant activity either as "uncouplers" or "others", for multiple MOA models all compounds with effects in the uncoupling assay were grouped as "uncouplers" and all inactives in this test as "others". In addition to these 2-class models, potential multiple activities were accounted for by a 7-class CPG NN model with the previously mentioned ten compounds being active as both uncouplers and their originally assigned MOA, e.g., the estrogenic compound 4,4'-dihydroxybiphenyl was coded by a response variable Y = (0,1,0,0,0,0,1) with uncoupling activity in column 2 and estrogenic activity in column 7.

2.2.5 Performance criteria

The criteria for assessing the quality of a quantitative model are usually the coefficient of determination, R^2, and the standard deviation, s. For classification models with multiple classes the most common method is to calculate a confusion matrix of the actual class versus predicted class. The diagonal elements of this matrix contain the number of correct classifications in each class and the sum is used to calculate the overall correct classification rate (or its counterpart the error rate, also called misclassification rate). However, this measure is not independent of prevalence, i.e., data sets with one class much larger than the others are easier to predict. If, e.g., a two class model with 95% of one and 5% of the other class assigns all compounds to the first class, it is completely useless despite an overall correct classification rate of 95%. Therefore, two additional criteria were used: sensitivity and Cohen's kappa. Sensitivity of class A is obtained by dividing the number of correctly predicted cases of A by the total number of class A. However, this measure can also be misleading if a model assigns too many cases to class A, the predictive power for other classes suffers. Cohen's kappa is independent of prevalence [36, 37]. It can be used with two class models as a measure of the proportion of all possible cases of actives and inactives that are predicted correctly after accounting for chance effects. Values of 0.0–0.4 are considered to indicate slight to fair model performance, values of 0.4–0.6 moderate, 0.6–0.8 substantial and 0.8–1.0 almost perfect.

2.2.6 Model validation

If performance criteria are calculated on the training data only, they will be optimistically biased. Thorough validation covers the major components of the model building process, i.e., selecting variables, choosing model parameters and model fitting. Usually the data set is split into a training set (model development) and a test set (model validation). However, this means a loss of expensive experimental data which could be used for developing better-founded models [20]. A possible solution is to use cross-validation [18, 38] or bootstrap methods [39–41]. Bootstrapping was shown to provide estimates of predictive power closest to the true values [42], but it requires that all important steps of model building are automated in order to repeat the model building process numerous times.

Initial leave-one-out cross-validation, after scaling, dimensionality reduction with PCA and variable selection with stepwise LDA [5], was done by training the model n-times on all data except for one compound and a prediction was made for that compound. However, because the variable selection used the entire data set, the requirement to make predictions only for compounds not used for model building is violated. In particular, stepwise variable selection is known to introduce a strong bias, if it is applied to the whole data set and performance criteria are calculated afterwards with some validation method [20]. Thus, leave-one-out cross-validation is not per se overoptimistic [42], but its estimate can have a strong optimistic bias if the methods are "tuned" using the response variables and thereafter validated. As a consequence, these estimates should only be used to see first trends and not for model comparison.

A possible approach to overcome this deficiency is a workflow based on k-fold cross-validation. The data set is split into k parts, then the model building process is carried out using k-1 parts of the data set and a prediction is made for the remainder. The process is repeated k times until each compound has been predicted once. Recommended values of k vary between 5 and 10 [18], but in some studies smaller values, as low as two, have been used as well. The crucial point for avoiding optimistic bias in cross-validation is to mimic the whole process leading to the final model, or as much of it as possible [43]. Figure 2.3 illustrates how this was implemented in this study: variable selection and model fitting are made in each of the k possible cross-validation training sets and then the prediction of the corresponding test set is made. Thus, the central source of bias is eliminated [20], assuming that preprocessing has little influence on the bias (it might have some influence on

the variance of estimated error rate, though). Of course repeating variable selection k times requires additional calculation time especially in the case of a GA, but with k = 5 the effort is manageable.

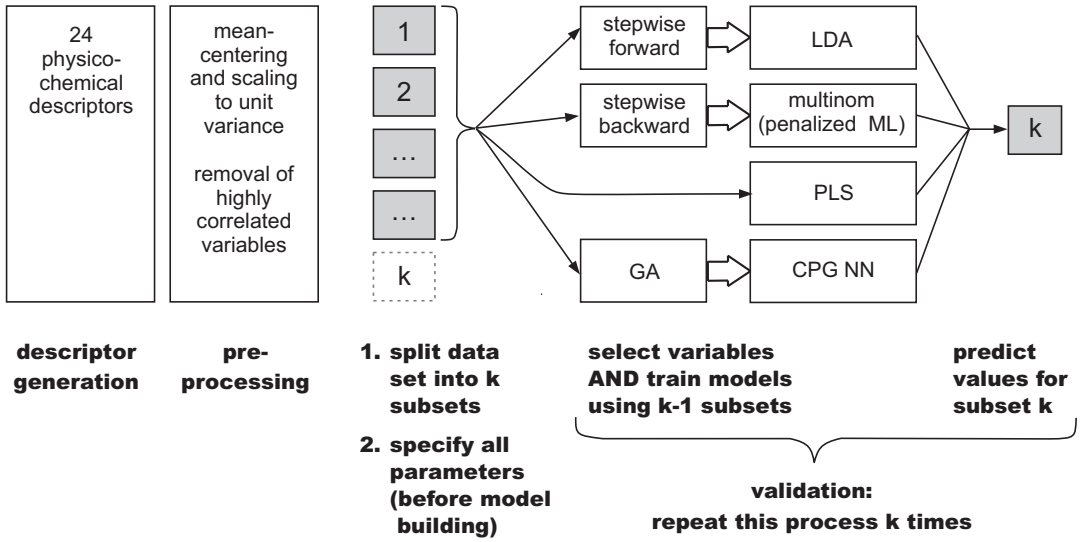


Fig. 2.3 Workflow for k-fold cross-validation. The variable selection step is integrated into the validation and thus a central source of bias in the estimates is eliminated. Abbreviatons: Penalized Maximum Likelihood (Penalized ML), Genetic Algorithm (GA).

Depending on the subset used, different variables might be selected. However, this is no real disadvantage of the strategy, as the main goal is to simulate how a model will perform with new data. The issue of which variables to use for the "final model" has been discussed in detail by Sauerbrei, recommending repeatedly splitting the data set and then determining which variables are selected most frequently during that process [44]. For the sake of simplicity the variables selected with the whole data were taken in this study. Note that this was only made in order to study which variables are selected and not for estimating model performance.

A decision to be made is how to split the data set. Statistics textbooks usually recommend repeated random splitting [18, 20]: the data set is split into k subsets, the performance criteria are calculated k times leaving out each subset once, and then this process is repeated d times with d ranging anywhere from 5 to 100. In this study the data set was not split randomly, but in a balanced way: within each MOA compounds were sorted according to their *in vitro*

toxicity and then alternately distributed to the k subsets, with the same number of compounds of the same MOA in each subset.

2.3 Results

2.3.1 Preliminary studies

As a first and easy to implement test, the predictive power of the four methods was assessed with leave-one-out crossvalidation. The full data set of 115 compounds with nine MOAs and the ten descriptors selected in the original study were used for building the models. The results are given in Table 2.2, with apparent percentage denoting the overall correct classification percentage calculated when the whole data set is used for training the model.

Table. 2.2 Apparent and leave-one-out cross-validated overall correct classification rates.

	LDA separate	LDA pooled	multinom	PLS 3 (6) components	CPG NN
Apparent %	89.6	60.9	67.0	41.7 (44.3)	–
Leave-one-out %	28.7	44.3	47.8	31.3 (34.8)	33.0

With LDA, using separate group-covariance matrices and prior probabilities depending on group size, a leave-one-out estimate of 29% overall correct classifications was calculated. A closer look revealed numerical problems in the covariance matrix calculations (singularities). With pooled covariance matrices the estimate of the predictive power increased to 44%. Equal prior probabilities were used for the calculation of this estimate, however there was only little difference to group size dependent prior probabilities. Multinom, PLS and CPG NN did not yield substantially higher estimates of predictive power. Both LDA and multinom suffer from serious overfitting as indicated by the large differences between the apparent and the crossvalidated percentages. No apparent percentage was calculated for CPG NN because this type of neural network would have a high apparent overall correct classification percentage after training by its very nature and not because it is overfitted. The PLS-model was less overfitted but had low predictive power. This is in agreement with the very low values for both cumulated R^2Y and cumulated Q^2Y of 0.14 and 0.11 respectively, where the coefficient of determination, R^2Y measures the fraction of the sum of squares of all the Ys explained by all extracted components and Q^2Y is the equivalent for cross-validated predictions.

2.3.2 Reassessment

Using the reduced data set (n = 104, seven MOAs, see section 2.1), correlation analysis with r^2_{limit} = 0.9 identified four highly correlated pairs of variables (r^2 > 0.98). From each of these pairs, one variable was removed, namely HARD, SAVOL, V^- and V_{TOT}, leaving 20 variables in the analysis. The correlation analysis procedure described in section 2.2 was also used to assess the degree of intercorrelation among the remaining X-variables in the data set. All variables were virtually eliminated until only one pair was left, which were EN and H^+_{MAX}. This allows one to visualize a correlation matrix in a condensed form as shown in Figure 2.4.

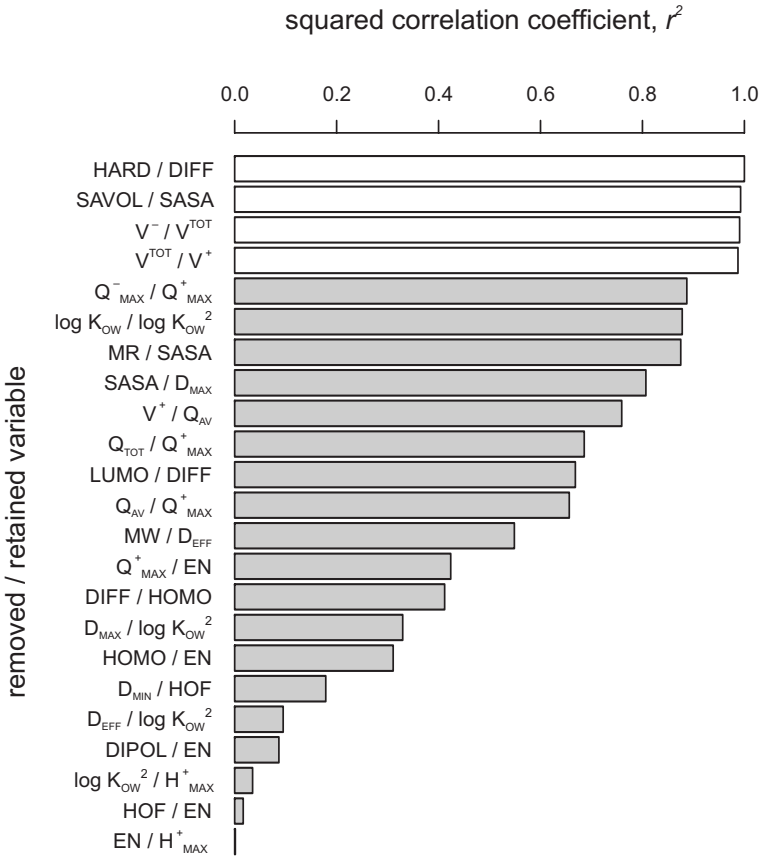


Fig. 2.4 Barplot of the stepwise removal of highly correlated variables. The white bars indicate the four pairs of highly correlated variables. In each step the pair with the highest correlation is determined and one variable is removed (the one on the left side). HARD, SAVOL, V^- and V_{TOT} were removed completely from the analysis.

The advantage of this graphical representation is that it immediately gives an impression of the degree of intercorrelation, an information which is very hard to see as soon as a

correlation matrix has more than just a few variables. For the present data set one can see that nine variables lie in the interval between an r^2_{limit} of 0.5 and 0.9. Thus, about half of the variables contain quite similar information. The likely consequence is that variable selection will vary a lot, depending on the cross-validation subset used. Methods that are affected by correlated variables will have to be used with caution. Following the procedure of Steyerberg et al. [43], models based on all 20 variables as well as models obtained with variable selection were compared (Table 2.3).

Table. 2.3 Apparent and 5-fold cross-validated (cv) overall correct classification rates without and with variable selection.

		LDA pooled	multinom	PLS	CPG NN
Full variable set (20 variables)	apparent %	76.9	94.2	54.8	–
	5-fold cv %	52.9	51.0	52.9	51.9
With variable selection	apparent %	69.2	82.7	–	–
	5-fold cv %	50.0	53.8	–	55.8
	Number of variables selected	9	9	–	11

Cross-validated overall correct classification rates for full models vary between 51.0% and 52.9%. With a small data set of 104 compounds overfitting is a problem for multinom and LDA when all 20 variables are used, as the apparent overall correct classification rate of 94.2% and 76.9%, respectively reveal. As statistical models should be based on at least 10 events per variable [43], which would limit the number of descriptors to 10 or 11, reducing the number of variables is indispensable for these two methods. If variable selection was applied in the way shown in Figure 2.3, i.e., if was integrated into the cross-validation process, the estimated predictive power decreased slightly for LDA, but increased for CPG NN with genetic algorithms and multinom with stepwise backward selection. Table 2.4 shows the variables selected with the different methods.

In all cases the number of variables could approximately be halved. Four variables were selected with each of the three selection methods: log K_{ow}, HOMO, $\mathrm{Q_{AV}}$ and $\mathrm{H^{+}_{MAX}}$. Each of them belongs to a different PCA cluster representing lipophilicity, reactivity, electronic charge and other parameters. Apart from these four variables there was no consistent pattern regarding the selection of the different methods.

Table. 2.4 Variables (**X**) selected by the different methods.

Cluster	Variable	Ref. [5]	LDA	multinom	CPG NN
Electronic charge	V^+	**X**	**X**	**X**	
	Q_{AV}	**X**	**X**	**X**	**X**
	Q^+_{MAX}		**X**	**X**	
	Q_{TOT}				
	Q^-_{MAX}		**X**		
	DIPOL				
Steric bulk	SASA	**X**			**X**
	D_{MAX}				
	MR	**X**		**X**	**X**
	MW	**X**	**X**		**X**
	D_{EFF}	**X**			
	D_{MIN}				
Reactivity	HOMO	**X**	**X**	**X**	**X**
	DIFF				**X**
Lipophilicity	log ${K_{ow}}^2$		**X**		
	log K_{ow}	**X**	**X**	**X**	**X**
Others	EN				**X**
	LUMO				**X**
	HOF			**X**	**X**
	H^+_{MAX}	**X**	**X**	**X**	**X**

2.3.3 Optimization

To reduce the overfitting of models due to too many X-variables and to minimize the gap between apparent and cross-validated percentage, with pooled LDA having a gap of 19.2% (69.2% apparent to 50% cross-validated) and multinom one of 28.9% (82.7% and 53.8%), the use of more stringent inclusion/exclusion levels (P-values) in the case of LDA and the increase of the penalty factor λ in the case of multinom were explored. Table 2.5 shows the results obtained when more stringent, i.e., lower, P-values were used for variable selection with forward stepwise LDA.

Decreasing the default P-values of 0.05/0.1, reduced the number of variables, but resulted in a distinct loss of predictive power and the difference between apparent and cross-validated percentage even increased further. If variable selection is not integrated into the validation step, the estimates of predictive power become optimistically biased. Comparing the correct classification rates of LDA with different P-values if variable selection is either

Table. 2.5 Apparent and 5-fold cross-validated overall correct classification rates for LDA models with and without variable selection at varied P-values for inclusion/exclusion of variables.

	0.05/0.1	0.025/0.03	0.01/0.025
Apparent %	69.2	68.3	68.3
Variable selection included 5-fold cv %	50.0	45.2	39.4
Variable selection not included 5-fold cv %	57.7	55.8	61.5
Selected variables	9	8	7

integrated or not integrated in the validation workflow, reveals substantial optimistic bias (8% up to 30%) if only model fitting is subjected to cross-validation. This underlines the necessity for covering the entire process of model building in the validation workflow, in order to obtain realistic estimates of predictive power.

For multinom the gap between apparent and crossvalidated percentages was particularly high (Table 2.3). In order to reduce this gap, the penalty factor λ in equation 2 was varied from 0.001 to 0.5, with the latter indicating a very strong penalization (Figure 2.5).

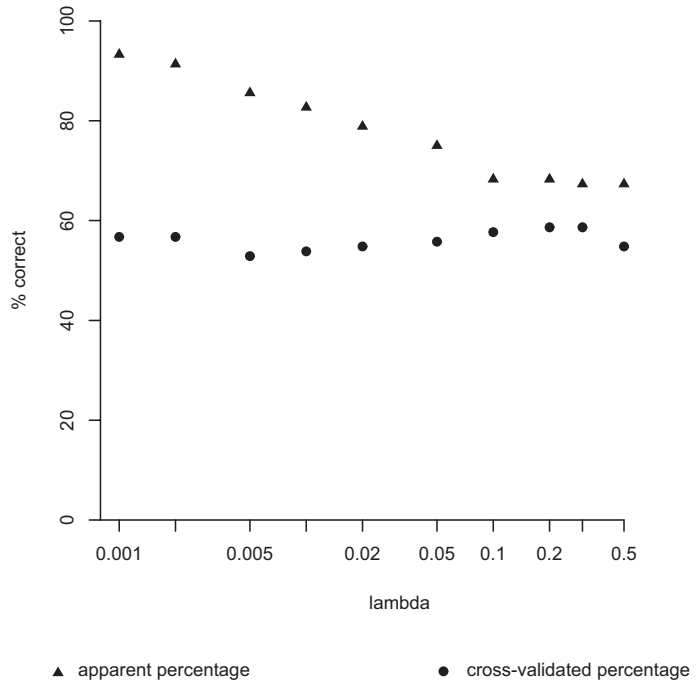


Fig. 2.5 Apparent and cross-validated overall correct classification rates for multinomial logistic regression with increasing penalty factors λ.

Strong penalization ($\lambda \geq 0.1$) resulted in reduced model complexity because the number of variables selected with stepwise backward selection could be reduced to only five:

log K_{ow}, MW, HOMO, HOF and H^{+}_{MAX}. The gap between apparent and cross-validated percentages could be reduced to < 10% while the predictive power increased to 58.6%.

Therefore, a high value of 0.1 for the penalty factor λ was used for subsequent multinom models. Confusion matrices of predicted versus actual class memberships for the best models (highest cross-validated overall correct classification rate) derived by LDA, multinom, PLS and GA CPG NN were comparatively analyzed to obtain a detailed picture of their performance (Tables 2.6a–d). For the multinom model the following parameters were used: stepAIC and penalised maximum likelihood using a λ of 0.1. For the LDA model P-values of 0.05/0.1 were used. Numbers in columns indicate the experimental assignments, numbers in rows the models assignment, e.g., 15 nonspecific nonreactives were predicted to be nonspecific nonreactive, four to be uncouplers, zero to be inhibitors of PS II, etc.

All models exhibited a good sensitivity for nonspecific nonreactive toxicants, i.e., a high fraction of nonspecific nonreactive toxicants is actually predicted as such. Sensitivity for uncouplers ranged between 48% and 68%, for inhibitors of AChE between 57% and 86% and varied strongly for other MOAs with fewer compounds. It was intermediate for estrogenic compounds, low for SH-alkylat-ing agents and very low for reactives. Sensitivity for inhibitors of PS II is very low as well, thus, the modeling did not benefit from narrowing the MOA-definition from inhibitors of photosynthesis to inhibitors of PS II. Because known inhibitors are of considerable structural variability, the herbicide binding niche of PS II can be assumed to be rather large [14]. As a consequence, reliable classification models for PS II may be difficult to establish without more data, regardless of the quality of descriptors. A remarkable observation is that SH-alkylating chemicals and reactives are hardly ever confounded with each other but mostly with nonspecific nonreactive toxicants.

Differences between the methods are considerable. LDA has the most even distribution between classes, i.e., each class has quite similar sensitivity, PLS only predicts the three large classes and no compounds at all were assigned to the smaller classes. CPG NN also shows a tendency to favor the larger classes which was further increased by the GA. The problems of the PLS model were due to a shift in the distributions of the smaller classes. While the larger classes had predicted values ranging from 0 to 1, the classes with few compounds had narrower distributions, shifted towards negative values, e.g., predicted values for estrogenic compounds ranged from . 0.30 to 0.46. Range scaling the predictions between 0 and 1 did inverse the problem such that most compounds were now predicted as belonging to the 2

Table. 2.6 Confusion matrices of the best models in terms of overall correct classifications. Abbreviations: Nonspecific nonreactive (Nonsp), Uncouplers of oxidative phosporylation (Unc), Inhibitors of photosystem II (PS II), Inhibitors of acetylcholinesterase (AChE), SH-alkylating agents (SH), Reactives (Reac), Estrogenic compounds (Est).

a) LDA pooled		experimental MOA						
		Nonsp	Unc	PS II	AChE	SH	Reac	Est
predicted MOA	Nonsp	15	0	0	0	0	3	0
	Unc	4	12	0	1	2	0	2
	PS II	0	2	4	5	1	0	0
	AChE	2	2	4	8	0	1	0
	SH	3	1	0	0	5	1	0
	Reac	4	3	0	0	1	2	0
	Est	4	5	1	0	0	0	6

b) Multinom		experimental MOA						
		Nonsp	Unc	PS II	AChE	SH	Reac	Est
predicted MOA	Nonsp	23	2	0	0	4	5	0
	Unc	3	13	5	0	2	0	1
	PS II	0	2	1	1	0	0	0
	AChE	1	2	3	12	0	0	0
	SH	3	1	0	1	3	1	0
	Reac	1	2	0	0	0	1	0
	Est	1	3	0	0	0	0	7

c) PLS		experimental MOA						
		Nonsp	Unc	PS II	AChE	SH	Reac	Est
predicted MOA	Nonsp	26	5	0	0	6	6	4
	Unc	4	17	2	2	2	0	4
	PS II	0	0	0	0	0	0	0
	AChE	2	3	7	12	1	1	0
	SH	0	0	0	0	0	0	0
	Reac	0	0	0	0	0	0	0
	Est	0	0	0	0	0	0	0

d) GA CPG NN		experimental MOA						
		Nonsp	Unc	PS II	AChE	SH	Reac	Est
predicted MOA	Nonsp	26	5	0	0	5	5	1
	Unc	2	15	3	2	2	1	2
	PS II	0	1	0	1	0	0	0
	AChE	0	3	5	11	0	1	1
	SH	2	0	0	0	2	0	0
	Reac	5	1	0	1	0	0	0
	Est	1	2	0	1	0	0	4

MOAs with few compounds.

2.3.4 Multiple MOA

Table 2.7 shows the results of the single versus multiple MOA comparisons using the cross-validated predictions. For multinom three variables for single MOA (Q_{AV}, Q^{-}_{MAX}, D_{EFF}) and four variables for multiple MOA (Q_{AV}, Q^{+}_{MAX}, D_{EFF}, H^{+}_{MAX}) were selected with stepwise backward selection. Considering that Q^{+}_{MAX} and Q^{-}_{MAX} are highly correlated, the same parameters of charge and steric bulk properties are used.

Table. 2.7 Results of multiple MOA comparison.

	Single multinom	Multiple multinom	Single CPG NN	Multiple CPG NN	Multiple CPG NN with 7 classes
Percent overall correct classification	78.3	73.6	78.3	74.5	74.5
Sensitivity	36.0	45.7	48.0	54.3	51.40
Kappa	0.31	0.35	0.37	0.40	0.39
Selected variables	3	4	No variable selection performed		

The main difference between the multinom 2-class models (“uncouplers” vs. “others”) is the inclusion of the additional variable H^{+}_{MAX} for multiple activities. A neuron to pattern ratio of 1.0 was used for the multiple MOA models with CPG NN as it resulted in better predictions than the 0.4 ratio used for the previous models. For both multinom and CPG NN, the overall correct classifications dropped for multiple actives, however, the numbers are not directly comparable because the sizes of the classes differ. At the same time, sensitivity and Cohen’s kappa increased if multiple MOAs were accounted for. The 7-class CPG NN model predicts multiple MOA activity spectra simultaneously without loss of predictive power. Uncoupling activity vs. other predominant MOAs alone or in combination is correctly classified for 74.5% of 106 chemicals. 4-n-Nonylphenol, e.g., was actually predicted having estrogenic as well as uncoupling activities as was observed in the *in vitro* experiments. Table 2.8 details the cross-validated predicted probabilities by single and multiple MOA models for the 10 compounds exhibiting multiple MOAs. Apparently the estrogenic compounds fit into the general pattern of uncouplers, while SH-alkylating chemicals do not.

There is no recognizable trend for the three inhibitors of photosynthesis except that dinoterb is confirmed to be an uncoupler which is in agreement with literature data [11]. It is even predicted to be an uncoupler in single MOA models and thus “misclassified” in these models. For both multinom as well as CPG NN models the 10 additional active com-

Table. 2.8 Single MOA versus multiple MOA models: comparison of cross-validated predictions for the 10 compounds with uncoupling activity as additional MOA (cf. Figure 2.2). Values > 0.5 active and < 0.5 nonacative as uncoupler.

	Single multinom	Multiple multinom	Single CPG NN	Multiple CPG NN	Compound
Inhibitors of	0.79	0.85	0.88	0.89	Dinoterb
photosynthesis	0.29	0.63	0.14	0.20	p-Chloromercuribenzoic acid
	0.05	0.09	0.03	0.03	Fenuron
SH-alkylating	0.04	0.22	0.00	0.02	2-Hydroxyethylacrylate
	0.00	0.03	0.00	0.00	Formaldehyde
	0.13	0.30	0.05	0.70	Diethylfumarate
Estrogenic	0.24	0.48	0.53	0.59	Bisphenol A
	0.18	0.42	0.25	0.26	4,4'-Dihydroxybiphenyl
	0.07	0.52	0.00	0.61	4-n-Nonylphenol
	0.05	0.31	0.01	0.87	4-n-Octylphenol

pounds were predicted better with single MOA models, while the predictions for the 25 compounds with predominant uncoupling activity improved through the consideration of the 10 additional actives.

2.4 Discussion

A data set characterized by three major attributes – many classes for the size of the data set, many descriptors for the size of the data set and many descriptors with a rather high intercorrelation – was used to compare the performance of different pattern recognition methods (LDA, multinom, PLS and CPG NN) for classifying chemicals by MOA. Considering only the estimates of predictive power (Table 3) the four models do not differ much and all classify about 55% of the compounds correctly. However, the crossvalidated percentage of correct classification is not the only criterion to judge the quality of a model. Another desirable feature is to use as few descriptors as possible and as a consequence, particularly for multinom and LDA, there should be only a small gap between apparent and crossvalidated overall correct classification percentage [45]. If, e.g., one linear model has 80% apparent and 60% crossvalidated overall correct classification while another has 65% and 58%, respectively, the second model should rather be used for predictions because the gap is much smaller and

the model is probably less overfitted. Thus, appropriate variable selection makes sense also under this aspect. Table 2.3 shows, that even after variable selection the gap is still very high for LDA and multinom, when the default parameters for variable selection are used.

Further reduction of the number of independent variables, i.e., consuming less degrees of freedom, should improve robustness and performance of models based on small data sets 2.1). However, more stringent inclusion/exclusion levels (P-values) for variable selection with LDA did reduce the complexity of the model, but also resulted in a loss of predictive power, which is in accordance with other studies [43]. Apparently, stepwise descriptor selection procedures with LDA can cause problems when data sets have a high fraction of highly correlated variables. On the other hand, increasing the multinom penalty factor λ allowed to build the model with the highest predictive power and the smallest number of variables. The fact that a regression method with only five variables had the best predictive performance is a strong indicator that the possibility to punish a model for being too complex makes multinom well suited for this type of data set.

Major variability was observed concerning the variables selected with the different methods. Apart from log K_{ow}, HOMO, Q_{AV} and $\mathrm{H}^{+}_{\mathrm{MAX}}$, each method uses different variables particularly to represent electronic charge and steric bulk. The main reason is probably the high correlation between many descriptors (cf. Figure 2.4). The question remains why no model could exceed the level of 60% correct classifications. The first reason could be that, although covering a wide range of chemical properties, the current set of 20 variables appears to still miss some important features related to the MOA of the compounds. This seems especially true for reactives and SH-alkylating compounds which could hardly be discriminated from nonspecific nonreactive toxicants. Another reason for insufficient predictive power could be that the experimental MOA assignment is still too crude, and that successful classification models need to distinguish further subgroups. For example the uncouplers might have to be divided into weak organic acids and different other types of uncouplers, because their effects are based on different mechanisms [46]. This problem, however, cannot be solved without additional experimental work.

The second objective of this study was to apply a validation method suited for such a small data set. Fivefold cross-validation might be a little bit too pessimistic [47], but it has the advantage of being straightforward to implement. The validation workflow must cover all relevant steps of model building, e.g., the effect of including variable selection proved

to be substantial (Table 2.5). If variable selection is based on the entire data set instead of the k-fold cross-validation subset, an overoptimistic bias of 8% up to 30% is introduced into the estimates of predictive power, depending on the inclusion/exclusion levels (P-values) for variable selection. The model obtained with the lowest Pvalues is particularly treacherous because with 61.5% it seemingly had the highest cross-validated overall correct classification of all models and it also had a small gap between apparent and cross-validated estimates of only 6.8%. With variable selection integrated into cross-validation the percentage drastically dropped to 39.4%. Such effects might be one of the sources of optimistic bias currently criticized in cross-validation [8] and leading to model failures when "real" external predictions have to be made. However, such mistakes can be avoided if all model building parameters are specified before fitting the models and then the whole process of model building is integrated into the validation. If, on the other hand, cross-validation is used in an iterative manner in order to improve the crossvalidated predictions, inevitably optimistic bias is introduced into the estimates of predictive power and serious model misinterpretation can result. Such failures are possible with split data sets as well if variables are selected using the whole data set. However, if it is not possible to specify all model building parameters, e.g. because graphical methods are used in an interactive manner which cannot be integrated into the cross-validation process, then splitting the data set into training and test set is indispensable and the validation approach presented here is not applicable.

The small size of data sets is a general problem for QSARs in toxicology. Future work should address this issue with suited validation methods. The necessity of comprehensive validation methods is underlined by the following approximation: the correct classification rate, and thus also the probability of misclassification, pmc, of a dataset of n compounds has a binomial (n, pmc) distribution. Thus, pmc has a variance of

$$\mathrm{pmc} \cdot (1 - \mathrm{pmc})/n \leq 1/4n \tag{2.2}$$

With the data set of this study with n being approximately 100 and a pmc of 0.4, the standard deviation, s is estimated as

$$s = \sqrt{0.4 \times 0.6/100} \approx 0.05 \tag{2.3}$$

for the simplest case of a two class model. Applying two standard deviations as an approximate 95% confidence interval, an overall correct classification percentage of $60 \pm 10\%$

results. Therefore, the application of optimized validation techniques to reduce the high variance of estimates with small data sets, having a low bias at the same time, are of major interest in QSAR studies of toxicity. More elaborated methods to determine confidence intervals would also be useful, because they help to recognize the limits of models [42]. Then it will be possible to regard improvements in a context, i.e., a rather large increase in estimated predictive power with a model based on 100 compounds can be purely random while a small increase can be highly significant if the model is based on 1000 compounds. The multiple MOA approach is promising and the inclusion of the ten compounds with multiple activities seems to slightly increase the predictive power for both logistic regression and CPG NN. The question of interest was whether these compounds really are multiple MOAs or whether the *in vitro* assays are not specific enough. The results shown in Table 2.8 show that probably both cases occur. While some compounds like dinoterb fit the pattern of uncouplers other compounds rather seem to interfere with the experimental system through other interactions. This explains also why the increase of predictive power was not more pronounced. The variables selected with the stepwise methods are in agreement with physiological models of this MOA [46]. The presented neural network approach would allow one to predict a whole spectrum of activities of a compound. The limiting part, however, is again the availability of sufficient high quality data.

2.5 Conclusions

Throughout this study the distinction of methods and models has been emphasized. Models always consist of a data set with specific characteristics like size and type of distribution and of a method used to build the model with the data. In most cases, it is not possible to say *a priori* which method works best for a given data set, and, thus, it is useful to apply several pattern recognition methods and then to compare the models:

1. The best models by any of the four different statistical methods based on the current set of descriptors have similar overall predictive power and do not exceed 60% correct classifications. Thus, inappropriate selection of statistical methods is not the reason for lack of success in MOA discriminations. Next steps towards reliable classifications have to test improved descriptors and/or extended toxicological data sets.

2. Overfitting is a problem particularly with small data sets with highly intercorrelated

variables, not only with neural networks but also with conservative less flexible methods such as LDA. Multinom with strong penalization ($\lambda \geq 0.1$) resulted in a stable model with the smallest number of variables (log K_{ow}, MW, HOMO, HOF, $\mathrm{H}^{+}_{\mathrm{MAX}}$) and the highest predictive power (58.6%).

3. Variable selection is the crucial step in model building and must be integrated into the validation workflow to avoid too optimistic estimates of predictive power.

4. Cross-validation is a potent means to assess the predictive power of models. Problems encountered with leave-one-out cross-validation due to insufficient consideration of variable selection can be easily overcome by using k-fold cross-validation over the entire model-building process (variable selection and model fitting).

5. The classification models benefit from considering multiple MOAs.

In MOA classifications, the simultaneous use of two or three methods – ideally a linear and a nonlinear method – seems advisable. The combination of thorough validation, as presented in this paper, and better methods for assessing the uncertainty of predictions with improved descriptors and extended toxicological data sets will be used in further studies to approach the main goal of MOA classification as a decision support system for predictive QSAR applications, as needed under the REACH process.

Acknowledgement

The authors thank Andrea Wenzel and Martin Müller of the Fraunhofer Institute for Molecular Biology and Applied Ecology, Schmallenberg, Germany, for providing their original data and for valuable discussions and comments.

References

[1] P. Krajcsi, F. Darvas, “High-Throughput In Vitro Toxicology”, in “High-Throughput ADMETox Estimation”, F. Darvas, G. Dormán (Eds.), Eaton Publishing, **2002**, chapter 6, 75–81.

[2] S. P. Bradbury, “Predicting modes of toxic action from chemical structure: An overview”, *SAR QSAR Environ. Res.* **1994,** *2,* 89–104.

[3] G. M. Rand, P. G. Wells, L. S. McCarty, "Introduction to Aquatic Toxicology", in "Fundamentals of Aquatic Toxicology: Effects, Environmental Fate and Risk Assessment", G. M. Rand (Ed.), Taylor & Francis, UK, **1995**, 3–67, 2^{nd} Edition.

[4] C. L. Russom, S. P. Bradbury, S. J. Broderius, D. E. Hammermeister, R. A. Drummond, "Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas)", *Environ. Toxicol. Chem.* **1997,** *16,* 948–967.

[5] M. Nendza, M. Müller, "Discriminating toxicant classes by mode of action: 2. Physicochemical descriptors", *Quant. Struct.-Act. Relat.* **2000,** *19,* 581–598.

[6] A. O. Aptula, T. I. Netzeva, I. V. Valkova, M. T. D. Cronin, T. W. Schultz, R. Kühne, G. Schüürmann, "Multivariate discrimination between modes of toxic action of phenols", *Quant. Struct.-Act. Relat.* **2002,** *21,* 12–22.

[7] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*, Springer series in statistics, Springer, New York, **2001**.

[8] A. Golbraikh, A. Tropsha, "Beware of q2!", *J. Mol. Graph.* **2002,** *20,* 269–276.

[9] D. M. Hawkins, S. C. Basak, D. Mills, "Assessing model fit by cross-validation", *J. Chem. Inf. Comput. Sci.* **2003,** *43,* 579–586.

[10] A. Wenzel, M. Nendza, P. Hartmann, R. Kanne, "Test battery for the assessment of aquatic toxicity", *Chemosphere* **1997,** *35,* 307–322.

[11] R. R. Schmidt, "HRAC classification of herbicides according to mode of action", in "The 1997 Brighton Crop Protection Conference: Weeds proceedings of an international conference organised by the British Crop Protection Council", Farnham, Surrey: BCPC, **1997**, Volume 3, 1133–1140.

[12] W. Hayes, Jr, E. Laws, Jr. (eds.), *Handbook of Pesticide Toxicology*, Volume 3, Academic Press Inc., San Diego, CA, **1991**.

[13] J. R. Corbett, K. Wright, A. Baillie, *The Biochemical Mode of Action of Pesticides*, Academic Press, London, **1984**.

[14] M. Devine, S. O. Duke, C. Fedtke, *Physiology of Herbicide Action*, Prentice Hall, Englewood Cliffs, NJ, **1993**.

[15] S. C. Basak, G. D. Grunwald, G. E. Host, G. J. Niemi, S. P. Bradbury, "A comparative study of molecular similarity, statistical, and neural methods for predicting toxic modes of action", *Environ. Toxicol. Chem.* **1998,** *17,* 1056–1064.

[16] B. I. Escher, R. P. Schwarzenbach, "Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms", *Aquat. Sci.* **2002,** *64,* 20–35.

[17] W. R. Dillon, M. Goldstein, *Multivariate Analysis - Methods and Applications,* John Wiley & Sons, Inc., New York, **1984**.

[18] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, **1996**.

[19] "SPSS for Windows", **2002**, version 11.5.1, Chicago, SPSS Inc.
`http://www.spss.com`

[20] F. E. Harrell, Jr., *Regression Modeling Strategies - With applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer-Verlag, New York, **2001**.

[21] D. W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, Inc., New York, **2000**, 2^{nd} Edition.

[22] S. le Cessie, J. C. van Houwelingen, "Ridge estimators in logistic regression", *Appl. Statist.* **1992,** *41,* 191–201.

[23] "R: A language and environment for statistical computing", Version 1.7.1, R Foundation for Statistical Computing, Vienna, Austria, **2003**.
`http://www.r-project.org`

[24] J. Nouwen, F. Lindgren, B. Hansen, W. Karcher, "Classification of environmentally occurring chemicals using structural fragments and PLS discriminant analysis", *Environ. Sci. Technol.* **1997,** *31,* 2313–2318.

[25] L. Eriksson, E. Johansson, "Multivariate design and modeling in QSAR", *Chemom. Intell. Lab. Sys.* **1996,** *34,* 1–19.

[26] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, *An Introduction to Multi- and Megavariate Data Analysis*, Umetrics AB, Umea, **2000**.

[27] L. Eriksson, H. Antti, E. Holmes, E. Johansson, T. Lundstedt, J. Shockcor, S. Wold, "Partial Least Squares (PLS) in Chemoinformatics", in "Handbook of Chemoinformatics", J. Gasteiger (Ed.), Wiley-VCH, Weinheim, **2003**, Volume 3, 1134–1166.

[28] "Simca-P", **2001**, version 9.0, Umea, Umetrics AB.
`http://www.umetrics.com`

[29] J. Zupan, M. Novic, J. Gasteiger, "Neural networks with counter-propagation learning strategy used for modeling", *Chemom. Intell. Lab. Sys.* **1995,** *27,* 175–187.

[30] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, 2^{nd} Edition, Wiley-VCH, Weinheim, **1999**.

[31] J. Zupan, M. Novic, X. Li, J. Gasteiger, "Classification of multicomponent analytical data of olive oils using different neural networks", *Anal. Chim. Acta* **1994,** *292,* 219–234.

[32] J. Gasteiger, A. Teckentrup, L. Terfloth, S. Spycher, "Neural networks as data mining tools in drug design", *J. Phys. Org. Chem.* **2003,** *16,* 232–245.

[33] T. Kleinöder, A. Yan, S. Spycher, "Prediction of Properties of Compounds", in "Chemoinformatics", J. Gasteiger, T. Engel (Eds.), Wiley-VCH, Weinheim, **2003**, 487–514.

[34] "SONNIA - Self Organizing Neural Network for Information Analysis", **2002**, version 4.1, Erlangen, MolNet GmbH.
`http://www.mol-net.com`

[35] A. Teckentrup, "Einsatzmöglichkeiten selbstorganisierender neuronaler Netze in der Wirkstofforschung", PhD-Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, **2000**.

[36] A. D. Forbes, "Classification-algorithm evaluation: five performance measures based on confusion matrices", *J. Clin. Monit.* **1995,** *11,* 189–206.

[37] S. Manel, H. C. Williams, S. J. Ormerod, "Evaluating presence-absence models in ecology: the need to account for prevalence", *J. Appl. Ecol.* **2001,** *38,* 921–931.

[38] A. P. Worth, M. D. Barratt, B. J. Houston, "The Validation of Computational Prediction Techniques", *Alternatives To Laboratory Animals (ATLA)* **1998,** *26,* 241–247.

[39] B. Efron, "Bootstrap methods: another look at the jackknife", *Ann. Stat.* **1979,** *7,* 1–26.

[40] R. Wehrens, H. Putter, L. M. C. Buydens, "The bootstrap: a tutorial", *Chemom. Intell. Lab. Sys.* **2000,** *54,* 35–52.

[41] J. Carpenter, J. Bithell, "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians", *Stat. Med.* **2000,** *19,* 1141–1164.

[42] B. Efron, R. Tibshirani, "Improvements on Cross-Validation: The .632+ Bootstrap Method", *JASA* **1997,** *92,* 548–560.

[43] E. W. Steyerberg, M. J. C. Eijkemans, F. E. J. Harrell, H. J. F., "Prognostic Modeling with Logistic Regression Analysis: In Search of a Sensible Strategy in Small Data Sets", *Med. Decis. Making* **2001,** *21,* 45–56.

[44] W. Sauerbrei, "The use of resampling methods to simplify regression models in medical statistics", *Appl. Statist.* **1999,** *48,* 313–329.

[45] M. T. D. Cronin, T. W. Schultz, "Pitfalls in QSAR", *THEOCHEM* **2003,** *622,* 39–51.

[46] H. Terada, "Uncouplers of oxidative phosphorylation", *Environ. Health Perspect.* **1990,** *87,* 213–218.

[47] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", in "International Joint Conference on Artificial Intelligence (IJCAI), Montréal, Québec, Canada, August 20-25, 1995", Morgan Kaufmann Publishers, **1995**, 1137–1145.

2.6 Further Discussion of Data Analysis Methods

Some statements of this study are in disagreement with the view of many members of the scientific community. Therefore, some clarifications and specific comments in addition to the published work is given in this section.

The most important clarification concerns cross-validation. Many authors stress that splitting the data set into training and test is the only way to establish a reliable QSAR model for predictive purposes [8,48,49]. This is the complete opposite from the general opinion in statistics, and machine learning where authors agree about the more representative results obtained with resampling methods [7,20,50]. Thus, the question is, where this discrepancy comes from. Is it possible that the type of data that are dealt with in chemoinformatics and QSAR are different from the data sets used for machine learning benchmark studies? Is there a difference in the nature of SAR and QSAR data sets which does not allow to use cross-validation methods while in other disciplines they can be used. Or to adapt a figure of speech used in optimization theory [51]: is there no free lunch for chemists while there is for other scientists?

The following considerations might help to resolve this discrepancy. In the majority of other disciplines, variables (or descriptors) carry a price, i.e., in medical statistics blood-pressure has to be measured or social factors have to be collected by asking the patients. On the other hand, in chemoinformatics the calculation of descriptors is not financially constricted and in many studies an initial set of hundreds or even thousands of descriptors is generated. If in such a case, descriptor selection is made using the whole data set, there will be a strong bias in the estimates of predictive power. Table 2.5 of the previous publication showed an overestimation of the predictive power by already 20%. This overestimation is not limited to the specific method of linear discriminant analysis. When the experiment of Table 2.5 was repeated for multinomial logistic regression similar numbers were obtained with an overestimation of the predictive power ranging between 6 and 14%. If there is already such a strong overestimation with only 20 variables then the situation is probably much worse in cases with hundreds of descriptors. However, this bias is not an argument against cross-validation as this will also affect estimates gained by splitting the data set (although to a lesser extent) unless the data set is split *before* the variable selection is performed.

Indeed the studies that should prove the failure of cross-validation are exactly of such a kind. As an example, in one of these studies an extensive automated random search with

kNN was used to maximize the cross-validated R^2 (or Q^2) [8]. Then, the obtained Q^2 was compared with the R^2-values from external test sets. The bad correlation between Q^2 and R^2 of the external test sets lead to the conclusion that a high Q^2 is not a reliable indicator for high predictive power and cross-validation cannot estimate the predictive power of models. However, in view of the results presented in Table 2.5 this conclusion must be seen primarily as a failure of accounting for the effect of variable selection. This claim got some support at about the same time the work from Chapter 2 was published by an article by Hawkins [52] in which the effect of variable selection was investigated in a more systematic manner. In this article, [52], cross-validated R^2 without accounting for variable selection is called "naive" R^2 and it was concluded that properly conducted cross-validation is a good tool to assess a model's predictive power. This is especially true for small data set where splitting often leads to very small test sets. For small tests sets, the results determined by splitting the data set obtain a very arbitrary nature, because a different split of the data set can lead to totally different results.

Thus, the question returns if cross-validation can readily be used to estimate the predictive-power of SAR and QSAR models. Or to use the figure of speech again: is there is free lunch for chemists after all? Unfortunately, there is a major problem with including variable selection into cross-validation. It requires that all model building steps are included. This is not just a programming problem, but mainly a problem of how modeling works in practice. Often one makes a model with some initial guess, but then by analyzing the residuals develops new hypotheses and ideas for new descriptors or new types of parameterizations to calculate a descriptor. This iterative workflow cannot be accounted for, because one gradually works towards a better model using more and more "knowledge" contained in the data set. In such a case it is unavoidable to split the data set into a training and a test set and the model must be derived completely by using the training set. Another typical case is that some method is highly dependent on the choice of parameters. Often these parameters are tuned over and over until the model fits well. However, the way that led to the final parameter choice is often forgotten. In such a case, again, testing with an external set is absolutely required.

If, on the other hand, it is possible to define the whole descriptors set, *before* model building starts, like in the previous study, then appropriately conducted cross-validation is a good method to determine the predictive power. The studies of the remaining chapters were such cases, because the descriptors where almost completely defined in advance of the model

building process. However, to come back to the figure of speech one must say that in view of the special nature of SAR and QSAR modeling there is, unfortunately, very rarely free lunch for chemists. Of course, the results presented in Table 2.5 cannot be called a systematic investigation, but they hint at a problem which deserves thorough investigation. A suggestion for a future study would be to take a large data set and systematically vary the size and the number of descriptors. Then, the influence of variable selection with both approaches, cross-validation and splitting, could be investigated. If, indeed, variable selection turns out to have such strong effects as suggested by Hawkins [52] such a paper would deserve great attention in the QSAR community which suffers badly from models without predictive power.

There is another aspect about predictive power that neither cross-validation nor splitting into training and test set can cover: the applicability domain. Especially with small data sets one cannot cover the whole distribution of descriptors, i.e., new candidates of compounds can be extrapolations and these will most probably fail. In order to warn the modeler or the future user of a model of such extrapolations, methods about the applicability domain are required in addition to proper model validation. One possible method for the determination of an applicability domain will be presented in the next chapter.

Additional References:

[48] M. Bohac, B. Loeprecht, J. Damborsky, G. Schüürmann, "Impact of orthogonal signal correction (OSC) on the predictive ability of CoMFA models for the ciliate toxicity of nitrobenzenes", *Quant. Struct.-Act. Relat.* **2002,** *21,* 3–11.

[49] P. Gramatica, P. Pilutti, E. Papa, "Validated QSAR Prediction of OH Tropospheric Degradation of VOCs: Splitting into Training-Test Sets and Consensus Modeling", *J. Chem. Inf. Comput. Sci.* **2004,** *44,* 1794–1802.

[50] J. J. Faraway, "Data Splitting Strategies for Reducing the Effect of Model Selection on Inference", Technical Report #259 of the Department of Statistics, University of Michigan, **1995**.
`http://www.stat.lsa.umich.edu/~faraway/techrep.html`

[51] D. H. Wolpert, W. G. Macready, "No Free Lunch Theorems for Optimization", Technical Report SFI-TR-95-02-010, Santa Fe Institute, Sante Fe, NM, USA, **1995**.
`http://www.no-free-lunch.org/WoMa95.pdf`

[52] D. M. Hawkins, "The problem of overfitting", *J. Chem. Inf. Comput. Sci.* **2004,** *44,* 1–12.

Chapter 3

Use of Structure Descriptors to Discriminate between Modes of Toxic Action of Phenols

Simon Spycher, Eric Pellegrini and Johann Gasteiger*
Computer-Chemie-Centrum und Institut für Organische Chemie, Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany
Spycher, S., Pellegrini, E., Gasteiger, J. (2005), *J. Chem. Inf. Model.*, **45**, 200–208.[a]

Abstract

Two classification models were developed based on a data set of 220 phenols with four associated Modes of Toxic Action (MOA). Counter-propagation neural networks (CPG NN) and multinomial logistic regression (multinom) were used as classification methods. The combination of topological autocorrelation of empirical π-charge and σ-electronegativity and of surface autocorrelation of hydrogen-bonding potential resulted in a 21 dimensional model that successfully discriminated between the four MOAs. Its overall predictive power was estimated to 92% using 5-fold cross-validation. Subsequently, a simple score for the distance to the training data was used to determine the prediction space of the model and used in an exploratory study on the phenols contained in the open NCI database. The use of a prediction space metric proved indispensable for the screening of such a diverse database.

[a]This Chapter corresponds exactly to the original publication other than the last section after the reference list and the numbering.

The prediction space covered by the proposed model is still of rather local nature which is either caused by the limited diversity and size of the training set or by the high dimensionality of the descriptors.

3.1 Introduction

Toxicity may be one of the most difficult properties to model, especially for high-throughput screening in the drug discovery process [1], but also for environmental risk assessment of industrial chemicals. The difficulties for developing effective *in silico* models are caused by two factors. First, the abundance of possible endpoints that will have to be modeled (ranging from *in vivo* data to specific enzyme tests or microarrays). Second, the scarcity of consistent public data that prevents one from building data sets that span a wide range of the chemical space.

Several approaches are currently applied in order to make quantitative predictions: local models based on sets of clearly defined congeners [2], global models using flexible machine learning techniques [3], and models based on common mode of toxic action (MOA). A MOA can be defined as a common set of physiological and behavioral signs characterizing a type of adverse biological response, while a toxic mechanism is defined as the chemical process underlying a given MOA [4].

If one wants to make predictions for a new compound, a central question for each approach is which quantitative model to apply. For models based on congeners the answer seems rather simple, but for MOA based models a successful classification into the different MOAs is the prerequisite for successful quantitative prediction. The well known MOA based ASTER system (ASsessment Tool for Evaluating Risk) of the US EPA [5] selects the appropriate QSAR for a new compound on the basis of the occurrence of chemical fragments. The limitation of this fragment-based method is to reduce a chemical structure to a specified substructure and ignoring the other topological and potential electronic features of the entire compound which may influence its MOA [6]. With the aim to overcome this limitation, several MOA classification studies were published in the last years [7–10].

The quantitative prediction and also the classification of reactive MOAs requires descriptors reflecting their reactive nature [11], which in most cases requires quantum chemical methods. The most frequently used descriptors comprise atomic charges, molecular orbital

energies and superdelocalizabilities [12], which are usually calculated with semiempirical methods. The calculation of such descriptors is still slow which hampers the screening of large databases. We have developed a set of empirical descriptors that model reactivity [13, 14] and therefore should also be applicable to model toxicity involving reactive chemicals.

The MOA data set of Schüürmann *et al.* was used [10] which is based on the original publication of Aptula *et al.* [9], but has some small changes and additional descriptors. It contains 220 phenols classified into four MOAs. Several classification models based on various statistical methods had been derived for this data set using quantum mechanical based descriptors and additional whole molecule descriptors [9, 10, 15, 16].

In the present study, classification models were derived based on empirical atomic physicochemical descriptors encoded by topological autocorrelation, A and surface autocorrelation vectors to structure-based vectorial descriptors. These descriptors from now on will be referred to as structure descriptors. Counter-propagation neural networks (CPG NN) and penalized multinomial logistic regression (multinom) were used to build the classification models. This study has the following objectives:

- Development of MOA classifiers based on empirical descriptors.
- Comparison of the predictive power of structure descriptors with descriptors used in earlier studies.
- Application of the models to a potential screening data set.

3.2 Materials and Methods

3.2.1 The Data Sets

3.2.1.1 Training Set

The MOAs of the 220 phenols [10] were assigned following structural rules developed earlier using growth inhibition assays with *Tetrahymena pyriformis* [17]. First, a short characterization of each MOA and the corresponding structural rules proposed by Schultz *et al.* [17] will be given.

Uncouplers of oxidative phosphorylation inhibit the coupling between the electron transport and the phosphorylation reactions which take place in mitochondrial membranes. Thus, they inhibit ATP synthesis in energy producing membranes [18].Uncouplers are hydrophobic weak acids with the ability to transport protons through proton-impermeable membranes (some uncouplers which are not weak acids are reported, but their mechanism is not clear [18]). Schüürmann *et al.* observed uncoupling to be limited to compounds with pK_{a}-values between 3.8 and 8.5 [19]. The structural rule suggested for uncouplers was: more than one nitro group, or more than one cyano group, or more than three halogen groups, or a single nitro group and more than one halogen group.

Precursors to soft electrophiles (proelectrophiles) are toxicants whose action is based upon their metabolic transformation to an electrophile [20, 21]. The transformation takes place by oxidation in a one- or two-electron process. The products have several possible targets [21]. The following rule was suggested for proelectrophiles: two or more hydroxy groups in ortho or para position and at least one unsubstituted aromatic carbon atom, or an amino group in ortho or para position to the hydroxy group and at least one unsubstituted aromatic carbon group.

Soft electrophiles alkylate essential protein thiol or amino groups or produce oxidative stress by free radical formation [9]. Suggested mechanisms for the reactivity are either direct reaction with biological nucleophiles or previous reduction of nitro groups to highly reactive nitroso groups [22, 23]. The suggested rule was a single nitro group, but not more than one halogen atom.

By default, phenols not assigned to any of these three MOAs were assigned to the MOA polar narcotic. The underlying mechanism and the target of polar narcosis is still a matter of debate. It is either caused by the presence of the toxicants in lipid components of the membrane or involves membrane-bound protein target-sites [24].

To summarize, one should note that there are considerable differences in the current molecular understanding of the four MOAs. While there is a clear understanding of both the mechanism and the target of uncouplers [25], there is much less understanding of proelectrophiles and soft electrophiles. For these two MOAs the abundance of possible targets causes problems for both classification and quantitative prediction (e.g., correlation of toxicity with the reaction rates with specific nucleophiles like GSH [26]).

In this study, two compounds of the training set were assigned to a different MOA than

in [10]. 3-aminophenol and 5-amino-2-methoxyphenol were reassigned as polar narcotics (both previously proelectrophiles, but the amino group is in meta not in para or ortho position). Additionally, 4-acetamidophenol (proelectrophile) and 4-nitrosophenol (soft electrophile) are not covered by the structural rules for their MOAs, but were left in their original MOA because of their high excess toxicity.

Thus, the training data set used in this study contains 220 compounds (155 polar narcotics, 19 uncouplers, 24 proelectrophiles and 22 soft electrophiles).

3.2.1.2 Test Sets

Once the final model based on the 220 phenols was established, two additional data sets were used as external tests sets. A first MOA test set was extracted from a publication using the same MOA assignment rules based on *Tetrahymena pyriformis* assays [27]. After removal of the duplicates and compounds with other MOA not covered by the training set a data set containing 25 compounds was obtained. This data set consisted of 17 polar narcotics, one uncoupler and 7 soft electrophiles. A second test set was used for an exploratory study. It consisted of all 3142 monocyclic phenols found among the 248'259 compounds in the open NCI database (release 1, 1999). The subset was cleaned of 261 compounds that were either salts or contained elements other than C,H,O,N,S,P,X (X = Halogens), of 9 compounds which could not be processed by the PETRA package [14], of 13 compounds which match the structural rules of pro-redox cyclers (as defined in [17]), 263 internal duplicates and 182 compounds contained already in the training set. The final set then comprised 2414 phenols for which the MOA was assigned according to the rules described above. This resulted in a data set of 1770 narcotics, 73 uncouplers, 429 proelectrophiles and 142 soft electrophiles. The assumption behind this exploratory study is that the rules are valid even for this very diverse data set (with limitations given in the Discussion).

The prediction of several thousands of compounds requires methods which are fast and easy to parameterize and have a high predictive power. A simple scheme of the workflow used to achieve this goal is presented in Figure 3.1.

The methods are explained in detail in the next three sections.

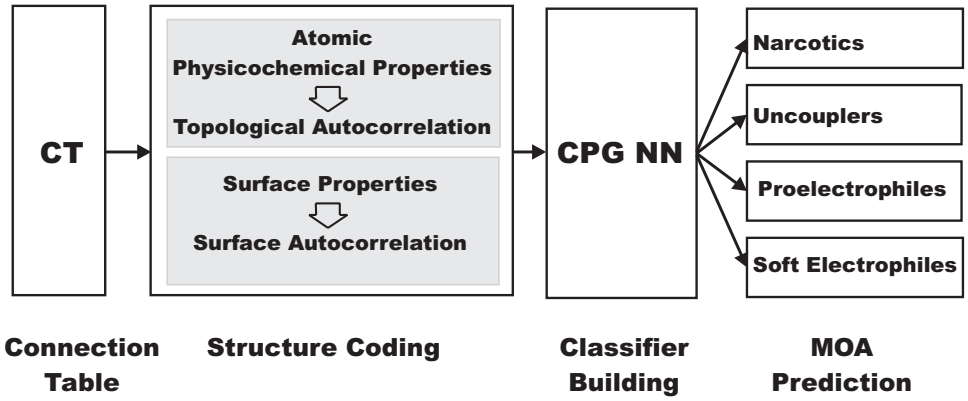


Fig. 3.1 Workflow used for the development of MOA classifiers suitable to screen large structure databases.

3.2.2 Calculation of Atomic Physicochemical Properties

In the present work, atomic physicochemical properties were calculated using PETRA [14]. This program package comprises various empirical methods for the calculation of physicochemical properties in organic molecules. A wide panel of effects is covered ranging from atomic properties (e.g., total, σ- and π-charge distribution, atomic polarizability, lone-pair, σ- and π-electronegativities) to molecular properties (e.g., heat of formation, molecular polarizability).

3.2.3 Encoding of Atomic Physicochemical Properties

The molecules atomic physicochemical properties were encoded using topological and surface autocorrelation vectors, A. They are both vectorial descriptors that integrate a physicochemical property inside an atomic distance-based function that is sampled on a fixed number of components. Therefore, the number of vector elements is independent of the size of the molecule.

An A vector can be expressed in various ways such as topological, 3D, or surface autocorrelation vectors [30–34] depending on the kind of distance that is used for their computation. On a general point of view, they are expressed through the following formula:

$$A(r) = \sum_{i=1}^{N} \sum_{j=i}^{N} p_i p_j f\left(r - r_{ij}\right) \tag{3.1}$$

Depending on the kind of A vector, the meaning of the different variables of Equation 3.1 changes a little bit. Indeed, in the case of a topological or 3D distance A vector, r_{ij} is the

topological or 3D distance between atoms i and j, and p_i and p_j are an atomic physicochemical property of atoms i and j. In the case of a $surface\ A$ vector, r_{ij} is the distance between two surface points i and j, and p_i and p_j are physicochemical properties of points i and j calculated on the molecular surface. Several kinds of properties can be calculated on the surface of a molecule such as electrostatic potential or hydrophobic and H-bonding potentials. In the present study, the hydrogen-bonding potential (HBP) was used according to a force field formulation of this potential [35]. The nature of the functional form f also changes regarding the A vector under consideration. In the case of a topological A vector, it is a Dirac function that, for each sampled topological distance r, considers only the topological distances r_{ij} strictly equal to r. For 3D distance A or $surface\ A$ vectors, it is a combination of Heavyside functions that, for each sampled real distance r, considers a range of distance over r. In the present work, topological A vectors with various properties p and $surface\ A$ vector were used. Hydrogen atoms were excluded for topological autocorrelation vectors. Autocorrelation vectors were always sampled over five topological distances for A vector, and nine 3D distance intervals from 1 to 7.75 Å for $surface\ A$ vector, respectively.

Single 3D conformations of the 220 phenols were generated using CORINA [36]. The A vectors were calculated with AUTOCORR [37] and the $surface\ A$ vectors with SURFACE [38]. Depending on the atomic properties calculated the standard deviations differed by several orders of magnitude. Therefore each vector element was autoscaled, i.e. mean centered and scaled to unit variance.

3.2.4 Classification Methods

3.2.4.1 Counter-Propagation Neural Networks

Counter-propagation neural networks (CPG NN) are an extension of Kohonen Self-Organizing Maps (SOM) where one or more output layers are added to the Kohonen input layer [39]. These types of neural networks have been applied for both classification studies [40] and for the prediction of continuous values (e.g. IR spectra, H-NMR spectra [41, 42]). The architecture of CPG NN is illustrated in Figure 3.2.

The input layers represent the descriptors (X-variables), while the output layers represent the properties to predict (Y-variables). The Y-variables are coded as a matrix of dummy vectors with 1 for actives and 0 for nonactives, i.e. a compound *i* active as uncoupler is coded

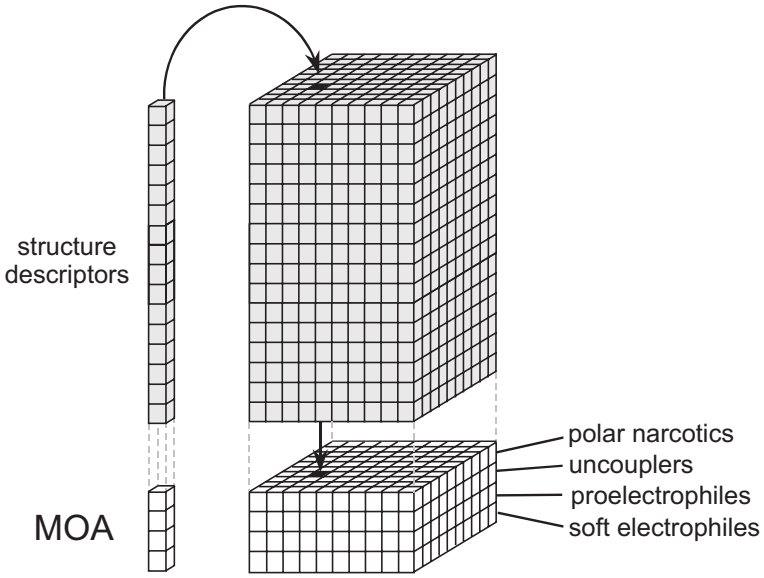


Fig. 3.2 Architecture of the CPG NN for the classification of toxicants with four output layers, one for each MOA.

as $Y_i = (0, 1, 0, 0)$. During training, the weights of both input and output layers are adapted, but only the input layers are used to determine the winning neuron. Thus, the unsupervised character of the SOM is preserved, i.e. the X-variables are organized independently of the Y-variable and therefore can not be fitted to nonexistent connections between X-variables (descriptors) and Y-variables just by adding additional descriptors. At the end of training, the weights obtained for the output layers are used as the predicted values for the properties under consideration. At this point the architecture shown in Figure 3.2 becomes clear: each layer of the output block corresponds to one MOA. This allows to study Kohonen maps of active versus nonactive compounds for each MOA. Predictions for new compounds are made by determining the winning neuron, defined as the neuron with the smallest Euclidian distance between its weight vector and the X-variables of the compound. A rectangular network topology was used, with a size of 11*11 neurons and a training time of 50 epochs. CPG NNs were calculated using SONNIA [43].

3.2.4.2 Logistic Regression

Logistic regression is a type of regression analysis where the dependent variable Y is a dummy variable (coded 0, 1). A logistic regression model solves the equation:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_p x_p \qquad (3.2)$$

where π is the probability that the event $Y = 1$ occurs, the β are the coefficients of the regression model with p descriptors, x. Equation 2 is the so-called logit function. A compound was assigned to the MOA with the highest predicted probability. The Maximum Likelihood Estimation method (MLE) is used in order to derive the regression coefficients, β_i. In order to reduce overfitting, a penalization term can be added to the Maximum Likelihood Estimation [44, 45]. The penalization term has the following formula:

$$\log\left(L^{\lambda}\right) = \log\left(L\right) - \frac{1}{2}\lambda\sum_{i=1}^{p}(s_i\beta_i)^2 \tag{3.3}$$

where L is the usual likelihood function, λ a predefined penalty, s_i a scale factor chosen to make $s_i\beta_i$ unitless. In practice, appropriate values for λ can vary in the range of 0.001 to 0.1 [46]. In the present study, a value of $\lambda = 0.1$ was chosen as it proved suitable in an earlier study with a data set that was prone to overfitting [47]. Logistic regression can be extended from the basic two classes to a model with *k* classes by using *k*-1 logit functions. The likelihood function used to determine the coefficients of the multinomial regression function β_{ik} is then determined from the summation including all logit functions. The Y-variables are coded as dummy matrix as described in the subsection above. A detailed description of the method is given in [48].

Logistic regression models were calculated using R [49], with the VR-add-on package containing the extension to multinomial logistic regression models.

3.2.4.3 Model Validation

k-fold cross-validation was used for model validation [50]. The data set was randomly divided into k subsets with equal percentages of each MOA present in each subset [51]. Then a model was fitted taking $k - 1$ of these subsets as a training set and the remaining one as a test set. The procedure is repeated k times until each subset has been used once as a test set. Usually, k takes values between 5 and 10 [46]. Due to the random splitting the estimates may vary, therefore the whole procedure was repeated B times (recommended values vary between 5 and 100). In this study, k was set to five, and B to 19 leading to 95 model fitting runs. The confusion matrix of the average cross-validated predictions was computed at the end of the procedure. From the confusion matrix the overall correct classification rate and sensitivities were derived. Sensitivity is obtained by dividing the number of correctly predicted compounds of a given MOA with the total number of compounds of this MOA.

3.2.4.4 Determination of Prediction Space

In order to determine the space covered by our models, we used a procedure based on the Hotelling's T^2 statistic. First, a PCA was performed on the training set. Then the loadings matrix of the *training* set was used to calculate the scores of the *external* set. Using these scores and a given confidence level α, a compound i of the external set was decided to belong to the prediction space if its Hotelling's score, T_i, satisfied the following criterion:

$$T_i^2 < \frac{A(N^2-1)}{N(N-A)} F(p=\alpha) \tag{3.4}$$

where $F(p = \alpha)$ is a tabulated value for a F distribution using a confidence level α, A is the number of principal components used to build the Hotelling's test, N the number of compounds of the training set and

$$T_i^2 = \sum_{a=1}^{A} \frac{t_{ia}^2}{s_a^2} \tag{3.5}$$

where s_a^2 is the variance explained by principal component a and t_{ia} is the score of compound i for principal component a [52]. The goal of using this statistical test, is to get a value for a compound's distance to the model. One can expect predictions to improve for compounds with small values of T^2 and to deteriorate for compounds with large values. If the T^2 of a given compound falls outside the limit - for a given confidence level - the compound is outside the confidence region and no predictions for this compound should be made or only with caution.

As PCA is sensitive to outliers, strong outliers in the training data have a strong influence on position and extent of the prediction space. In order to reduce such effects the following measure was taken. In a first step, the 99% confidence region of the training data was determined. Compounds falling outside this region were then excluded and the PCA loadings of the remaining training data were recalculated and used to determine the Hotelling's score as described above. This measure leads to a narrower definition of model space and hence to more compounds rejected from prediction, but to more robust values for the score.

3.3 Results

3.3.1 CPG NN-Models Based on Structure Descriptors

3.3.1.1 Models Based on one Physicochemical Property

As an initial test, 6-dimensional topological A vectors were calculated with seven different atomic physicochemical properties. Seven CPG NN models were trained each with one type of property represented with A vectors. Table 3.1 shows the results of the 5-fold cross-validation for the eight models, with overall correct classification rates ranging from 70.9 to 84.1%.

Table. 3.1 Predictive Power of seven CPG NN models each using a different type of atomic physicochemical descriptor represented with A vectors. Abbreviations: Polar narcotics (N), uncouplers of oxidative phosphorylation (U), proelectrophiles (P), soft electrophiles (S).

descriptors	correct prediction (%)				
	overall	N	U	P	S
identity function, A1	70.9	93.6	57.9	0	0
partial charge, q_{tot}	74.6	94.1	31.6	33.3	18.2
π charge [53], q_π	84.1	98.1	42.1	25.0	86.4
σ charge [54], q_σ	76.4	94.8	63.2	37.5	0
σ electronegativity [55], χ_σ	82.3	96.8	78.9	0	72.3
π electronegativity [56], χ_π	84.1	100.0	68.4	0	77.3
polarizability [55], α	70.9	97.4	26.3	0	0

When the overall correct classification rates are judged the strong prevalence of the 155 polar narcotics needs to be reminded, i.e. a (worthless) classifier predicting all the 220 phenols to be narcotics would result in 70.5% correct predictions. The model based on polarizability, α seems pretty close to such a case. The sensitivities for the specific MOAs U, P and S vary strongly from one descriptor to another. In the case of proelectrophiles, no descriptor is able to increase the sensitivity.

3.3.1.2 Models Based on an Additional Physicochemical Property

In a second step, an additional A vector was added to q_π-A, the best performing descriptor in terms of reaching a high sensitivity for one of the three specific MOAs (i.e. uncouplers,

proelectrophiles and soft electrophiles). q_π-A could classify 86.4% of the soft electrophiles correctly, but additional descriptors were necessary for uncouplers and proelectrophiles. The descriptor with the second-highest sensitivity was therefore added, namely the χ_σ-A with a sensitivity for uncouplers of 78.9%. Figure 3.3 shows the four output layers of a network trained with q_π-A and χ_σ-A. It reveals a spread of the polar narcotics over the structural space, while the other MOAs tend to cluster to various extents. The strongest tendency to cluster could be observed for soft electrophiles and the least for proelectrophiles. The largest fraction of the conflicts is caused by proelectrophiles conflicting with polar narcotics. A typical conflict was between catechol and 2-methoxyphenol. Apparently, for an oxygen or a nitrogen atom, the chosen descriptor set could not retrieve if the atom was a hydrogen bond donor or not.

The estimate of overall predictive power increased only slightly through the addition of χ_σ-A to 87.3%. This was due to an increase in sensitivity for uncouplers, which increased from 42.1% (q_π alone) to 68.4% (q_π combined with χ_σ), while sensitivities for proelectrophiles remained low with 29.2%. Thus, the next descriptors should increase sensitivity for this MOA. Analysis of the misclassification of the cross-validated predictions confirmed the pattern observed in the maps: proelectrophiles were predominantly misclassified as polar narcotics (all of the 17 misclassified proelectrophiles) and polar narcotics as proelectrophiles (2 of 3 misclassifications).

3.3.1.3 Improvement of Proelectrophile Classification by Adding Surface Autocorrelation Descriptors

The problem in discriminating proelectrophiles from polar narcotics had been encountered in previous studies [9, 15] as well and it was the reason that the number of H-bond-donors, N_{Hdon}, was introduced as a descriptor. In order to keep the concept of hydrogen bond, but to introduce structural information, 9-dimensional $surface\ A$ vectors of hydrogen bonding potential ($surface\ A$) were calculated. CPG NN models were calculated with a combination of q_π-A, χ_σ-A and $surface\ A$. Figure 3.4 shows the resulting maps, one for each output layer of the CPG NN, respectively for each MOA.

The most striking difference to Figure 3.3 is the decrease in conflicts. Uncouplers (row 2) and soft electrophiles (row 4) now form almost perfect clusters with no conflicting neurons. Interestingly, even with a very low number of conflicts the proelectrophiles did not form a

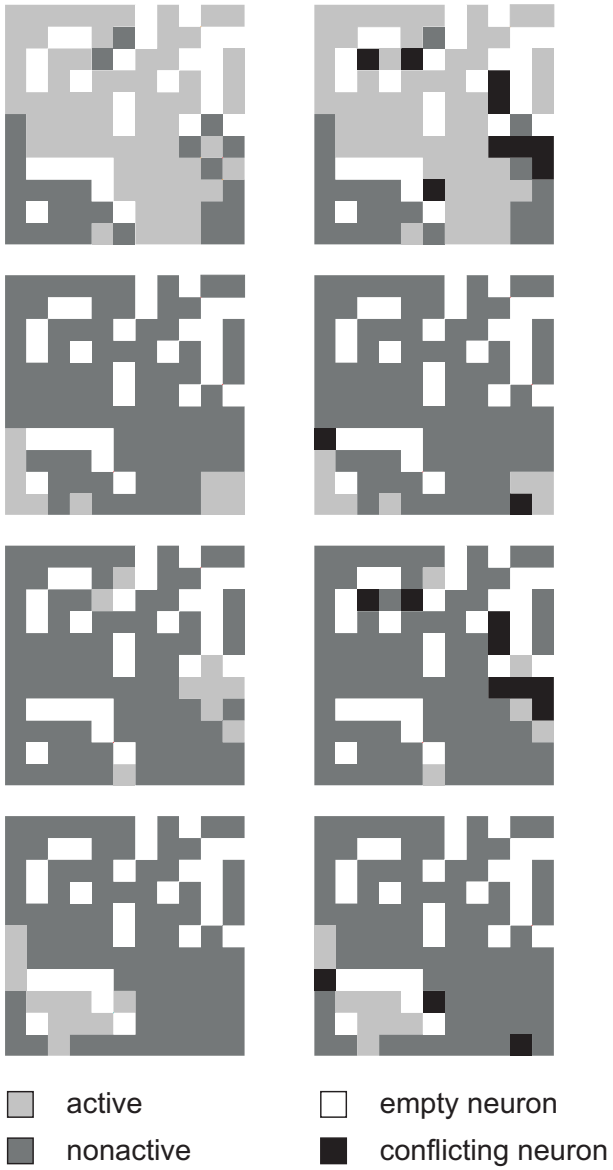


Fig. 3.3 Projections of compounds represented with q_π- and χ_σ-A vectors into four maps, corresponding to the four MOA polar narcotics, uncouplers, proelectrophiles and soft electrophiles (top down). Active molecules in light grey, nonactives in dark grey (colored according to the most frequent output). The maps on the right hand side additionally show conflicting neurons(black), i.e. neurons containing compounds from different MOAs. White squares indicate empty neurons.

distinct cluster. Thus, the CPG NN seemed to model different subgroups of this MOA. A confusion matrix with cross-validated predictions of the resulting models is shown in Table 3.2.

The addition of $surface\ A$ dramatically increased the sensitivity for proelectrophiles and also improved the sensitivity for polar narcotics. The overall correct classification rate

Fig. 3.4 Projections of compounds represented with q_π-A, χ_σ-A and $surface\ A$ vectors into four maps, corresponding to the MOAs 1-4, respectively from top to down, polar narcotics, uncouplers, pro-electrophiles and soft electrophiles. Active molecules in light grey, nonactives in dark grey (colored according to the most frequent output). The right hand side maps are the same than the left hand side ones but with the conflicting neurons displayed (black). White squares indicate empty neurons.

increased to 92.3%. This combination of two six-dimensional A vectors and one nine-dimensional $surface\ A$ vector represents the final 21 dimensional model which was subsequently used in this study.

Table. 3.2 Confusion matrix of 5-fold cross-validated predictions of a CPG NN with q_π-A and χ_σ-A vectors and $surface\ A$.

	Act N	Act U	Act P	Act S
Pred N	152	2	5	2
Pred U	1	13	0	1
Pred P	2	0	19	0
Pred S	0	4	0	19
Sensitivity	98.1	68.4	79.2	86.4
Overall Prediction				92.3

3.3.2 Multinomial Logistic Regression-Models Based on Structure Descriptors

The descriptors chosen during the CPG NN model building were also tested with multinom as an alternative classification method. Model parameters used are described in the methods section. When the whole data set was used for fitting the model a correct classification rate of 98.1% was calculated (apparent correct classification), compared to 92.7% estimated with cross-validation. With 21 descriptors for 220 compounds (respectively 176 during 5-fold cross-validation) the number of observations per descriptor is at the lower end and thus, overfitting can become a problem. However, the difference of 5.4 % between apparent and cross-validated correct classification rate is not much higher than the differences obtained using the six descriptors of Schüürmann *et al.* (cf. subsection 3.3.3). The confusion matrix with cross-validated predictions of the resulting multinom models is shown in Table 3.3.

Table. 3.3 Confusion matrix of 5-fold cross-validated predictions of a multinomial logistic regression model with q_π-A and χ_σ-A vectors and $surface\ A$.

	Act N	Act U	Act P	Act S
Pred N	150	0	5	1
Pred U	1	18	0	4
Pred P	4	0	19	0
Pred S	0	1	0	17
Sensitivity	96.8	94.7	79.2	77.3
Overall Prediction				92.7

Like in the CPG NN models almost all misclassifications occur either between polar narcotics and proelectrophiles or uncouplers and soft electrophiles. The comparison with Table 3.2 shows, that the difference to the CPG NN model is within this pattern, i.e. more uncouplers are predicted correctly, however, more soft electrophiles are predicted as uncouplers.

3.3.3 Models Based on Previously Published Descriptors

Models based on structure descriptors were compared to two sets of previously published descriptors. First, models based on descriptors published by Aptula *et al.* were evaluated [9]. The following five descriptors were used: octanol/water partition coefficient, log P; energy of the highest occupied molecular orbital, E_{HOMO}; energy of the lowest unoccupied molecular orbital, E_{LUMO}; negative logarithm of the ionization constant, pK_{a}; and the number of hydrogen-bond donors, N_{Hdon}. Cross-validation of the CPG NN model resulted in 86.8% correct classifications. The multinom model resulted in 87.7% correct classifications and 91.4% when the whole data set was used for fitting the model (apparent correct classification). Secondly, the following six descriptors calculated by Schüürmann *et al.* were evaluated [10]: log P, E_{HOMO}, E_{LUMO}, N_{Hdon}, highest electrophilic delocalizability of a carbon atom, D^{E}_{C-max}, and average carbon atom nucleophilic delocalizability, D^{N}_{C-av}). Cross-validation of the CPG NN model resulted in 91.4% correct classifications. The multinom resulted in 91.8% correct classifications with an apparent correct classification of 95.9%. Sensitivities of this multinom model amounted to N = 96.7%, U = 78.9%, P = 87.5%, S = 72.7%.

3.3.4 Predictions of External Data Sets

The multinom model using the full set of 21 descriptors was tested with the two external data sets described in the methods section. The two external data sets were compiled after the final choice of variables and model parameters was made. As a first step the predictions space spanned by the training set was determined as described in section 2.6. Figure 3.5 shows the training set scores for the first two principal components (PC) with the ellipse defined by equation 4 calculated for 2 PCs at a 95% confidence level.

The compounds show a good spread over space. However, some compounds with specific MOA are placed outside the prediction space at the chosen confidence level. Uncouplers

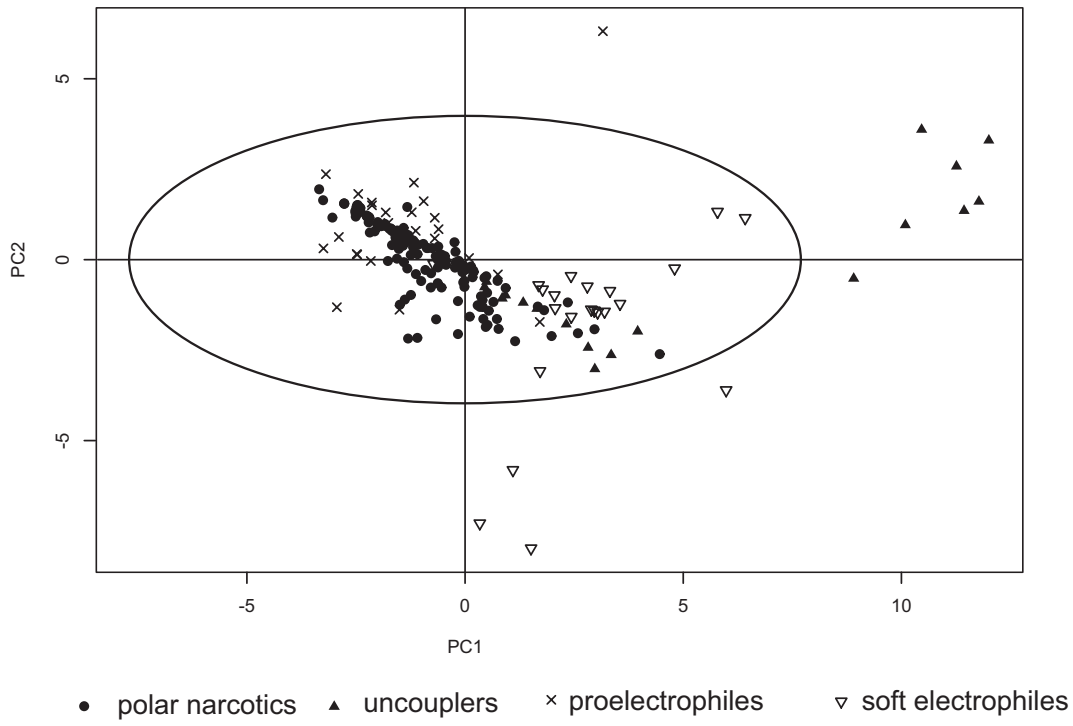


Fig. 3.5 PCA-score plot of the training set with 21 descriptors. The ellipse defines the 95% confidence region.

are specially affected with all 7 dinitrophenols excluded. The figure also shows that the prediction space is defined in a global and rather conservative way. For the analysis of the external data sets a prediction space was defined using the first 5 PCs of the training set to calculate the Hotelling's score (equation 5).

The first external data set consisted of 25 compounds extracted from the study of Cronin *et al.* [27]. Intuitively these compounds can be described as being structurally close to the training data. Two compounds fell outside the 95% confidence region defined by the Hotelling's score, which confirms that the majority of the compounds are close to the training set. The multinom model classified 84% of the 25 compounds correctly and 91% correctly if the two rejected compounds (both misclassified) are not counted.

The second external set consisted of the 2414 phenols extracted from the open NCI-database. This selection can be described as being structurally extremely diverse and thus, one can expect much more conflicts between predictions and the previously assigned MOA. This allowed us to study how the quality of the predictions changed with increasing distance to the training set compounds. This effect was studied by gradually increasing the confidence region from a very low value of 40% up to 99.9%. The larger the region the more compounds are assigned to be inside the models prediction space. Table 3.4 shows the overall correct

classifications and the sensitivities at different choices of the prediction space including the full external data set.

Table. 3.4 Overall correct classification rate and sensitivities with increasingly large prediction spaces with a threshold of 100% indicating the full test set. The number of included compounds is in parenthesis.

threshold	overall	N	U	P	S
40	83.5 (508)	85.3 (442)	100.0 (1)	70.3 (64)	100.0 (1)
60	80.3 (640)	82.9 (549)	66.7 (3)	63.9 (83)	80.0 (5)
80	76.7 (803)	80.1 (663)	42.9 (7)	59.1 (110)	73.9 (23)
90	74.2 (938)	77.7 (773)	45.5 (11)	55.5 (119)	68.6 (35)
95	72.6 (1024)	76.3 (843)	41.7 (12)	52.7 (131)	68.4 (38)
99	69.0 (1208)	73.5 (989)	42.9 (14)	45.0 (160)	64.4 (45)
99.9	65.5 (1419)	71.1 (1130)	46.7 (15)	39.4 (216)	58.6 (58)
100	56.2 (2414)	65.0 (1770)	37.0 (73)	30.1 (429)	34.5 (142)

Two issues are covered by Table 3.4: first, the behavior of the Hotelling's T^2 statistic used to determine the prediction space and second, the quality of the predictions themselves. The total number of compounds included into the models prediction space ranged from 508 to 1419 of the totally 2414 compounds, i.e. a large fraction of this external test set cannot be described by the model, e.g., the majority of the compounds containing sulfur and all compounds containing phosphorus (with both elements not present in the training set). The original proportions of the test set, 73.3% polar narcotics, 3.0% uncouplers, 17.8% proelectrophiles and 5.9% soft electrophiles were more or less maintained even when a small prediction space was chosen. The only exception was the uncouplers which on most levels were over-proportionally excluded from the model space with a proportion of only 1% or less. At the 99% level 14 uncouplers were selected and 59 excluded. All 55 dinitro- or trinitrophenols fell outside the prediction space which caused the low number of uncouplers included. Concerning the quality of the predictions, it is obvious that both overall correct classification and sensitivities follow the expected trend to decrease with increasing distance to the model. For the narrowest definition of prediction space the sensitivities correspond to the estimates obtained with cross-validation. However, for larger confidence regions the sensitivities of the specific MOAs rapidly decreases, especially for uncouplers. The reason for the low sensitivity for the 14 uncouplers included in the 99% confidence region was the large number of compounds combining structural features from several MOA. There were, e.g., six uncou-

plers with two hydroxy-groups and 4 halogen groups which were predominantly predicted as proelectrophiles. If these cases of overlapping MOA-features are left out the sensitivity for uncouplers increases to 75%, however with only 8 compounds remaining in the test set. The reason for the decreasing sensitivity of proelectrophile classification is the size dependence of the surface autocorrelation coefficient. The training set proelectrophiles consisted mainly of phenols with few small substituents (while the training set polar narcotics also contained phenols with long alkyl chains). As phenols with large substituents were increasingly included with larger confidence regions, the misclassification of such compounds as polar narcotics was the main reason for the decrease in sensitivity for proelectrophiles. The sensitivity of soft electrophiles remains fairly constant over the first few confidence regions (up to the 95% confidence region) and is quite close to cross-validation estimate of 77.3%. The most predominant structural pattern of misclassified compounds was the presence of a nitro group plus an electron withdrawing group at the ring which lead to misclassifications as uncouplers.

3.4 Discussion

This study should achieve three objectives: 1. development of a classifier suitable for database screening, 2. comparison of this classifier with existing approaches, 3. application of the classifier in exploratory study on a large data set.

The approach of building MOA classifiers presented in this study and illustrated by Figure 3.1 proved to be straightforward and fast. The processing of a structure through the workflow takes less than a second of calculation time on a desktop machine (PIII, 2.2 GHz). The estimates of the predictive power determined with cross-validation are high. For the full 21 descriptor model with q_π-A, χ_σ-A and hydrogen bonding potential ($surface\ A$) they amounted to 92.3 and 92.7% overall correct classification with CPG NN, respectively with multinomial logistic regression. These high estimates were additionally supported by the good predictions for the small external set compiled from literature. Thus, the first stated goal to use empirical descriptors which can be rapidly calculated to build a model with high predictive power is met. The remaining misclassifications between uncouplers and soft-electrophiles do not necessarily indicate a model weakness, but could point towards an overlap of these two MOAs. This would be in agreement with specific *in vitro* tests for uncouplers

where considerable activity could be observed for compounds like 4-nitrophenol which is a soft electrophile in the present data set [57]. Overlapping MOAs were also suggested by Katrizky *et al.* [23].

The estimates for the models based on the six descriptors of Schüürmann *et al.* amounted to 91.4% for CPG NN and 91.8% for multinom, respectively. The differences to models based on empirical descriptors are very small and do not allow to judge which models have higher predictive power. The same is valid for the sensitivities of the three specific MOAs where differences between all models amount usually to differences of one or two compounds (with the exception of uncouplers in the multinom model of Table 3.3). The important finding is, however, that for this data set it was possible to replace quantum-chemical descriptors with empirical descriptors coded with autocorrelation.

The similarity of the predictions of structure descriptors and previously published descriptors suggests that these contain similar information. Schüürmann *et al.* developed in their study a hierarchical scheme which descriptors contribute to separate which MOAs [10]. On the top level of their hierarchy they used E_{LUMO} to separate two MOA groups from each other, i.e. one group with low E_{LUMO} values (uncouplers and soft electrophiles) from one with higher values (polar narcotics and proelectrophiles). A similar pattern could be observed for the model based on q_π-A and χ_σ-A. In these models, misclassifications occurred either between polar narcotics and proelectrophiles or between uncouplers and soft electrophiles while the two groups of MOA were well separated. To investigate the relation between these descriptors a PLS model with the Y-variable E_{LUMO} and the 12 X-variables q_π-A and χ_σ-A was fitted. With 4 components a coefficient of determination, R^2 of 0.84 and a cross-validated R^2 of 0.81 were be obtained. Thus, the descriptors do not only lead to similar predictions in classifications, but correlate well also as continuous variables. Both types of descriptors are a global measure of electrophilicity decrease in the following order uncouplers $>$ soft electrophiles $>>$ proelectrophiles $>$ polar narcotics. Also the subsequent separation within these two blocks partly follows the line of argumentation of Schüürmann *et al.*. They used E_{HOMO}, N_{Hdon} and D^E_{C-max} for separating polar narcotics from proelectrophiles while in this study the same end was accomplished using $surface\ A$.

$Surface\ A$ lead to major improvements in the classification of pro-electrophiles because it describes hydrogen-bond acceptor or donor patterns that occur in the molecule. In that sense, it contains similar information as N_{Hdon} used in previous studies [9, 10]. The ques-

tion for both descriptors is how to relate them to the underlying mechanisms. The case for N_{Hdon} as a descriptor is made on the basis that the distinct feature of polar narcotics is their combination of hydrophobicity and hydrogen bond donor capacity [10]. Thus, the $surface\ A$ vector also seems to contain this information. On the other hand, Aptula *et al.* noted that N_{Hdon} might simply reflect the structural rule for the *a priori* assignment of proelectrophiles, since it essentially counts the number of hydroxy and amino groups [9]. Although both N_{Hdon} and $surface\ A$ vectors highly correlate with the response variable it is hard to prove that they reflect the underlying mechanisms. The only way to prove it would be to find additional data especially less "typical" compounds for each MOA.

The investigation with the two external sets proved to be particularly interesting in several aspects. The first aspect is the prediction space defined by the methodology based on the Hotelling's T^2 statistic of PCA scores. The conclusions for this aspect are: First, the definition of a prediction space is essential for any model, if it is used to screen a large and diverse data base. It prevents that predictions are made for compounds not covered by the training set. Second, the number of compounds included in the prediction space was surprisingly small even for large confidence regions. In other words, the models developed are still rather local in nature. Two factors cause this effect: the first factor is that the training set spans a limited part of the chemical space taken by the 2414 phenols and the second factor is that the high dimensionality of structure descriptors automatically leads to a low sample density in the multidimensional space. The latter phenomenon is commonly referred to as the "curse of dimensionality" [50]. Third, the ellipsoid formed by the definition of predictions space is a rather global model which does not capture local clusters towards the outer limits. Such an effect occurred with uncouplers where all dinitrophenols were excluded from the model space because of their q_π-autocorrelation vectors exhibited very high values. This deficiency should by addressed in future studies, e.g. by using bootstrapped distance corrections [58] which can take into account such local effects and also take into account characteristics of a classifier.

The second aspect is the change of the prediction quality with increasing size of the confidence regions. Although sensitivities were high for the narrowest choice of prediction space, it dropped surprisingly fast, considering that as an initial guess one would allow all predictions up to a 95% confidence level. Two reasons can cause this effect. First, the violation of assumptions behind a statistical test, i.e. the Hotelling's T^2 test used here cannot be under-

stood directly as a probability of error, but just as a score and the appropriate limit has to be determined using a test set. Second, the rules used to assign the MOA have their limits and cannot be applied to all compounds of the extremely diverse external test set. For example phenols with carboxy-groups assigned by the rules as uncouplers are probably not enough lipophilic to enter the mitochondrial membrane [18, 25], or on the other hand some well known uncouplers (e.g., 2,6-di-t-butyl-4-(2,2-dicyanovinyl)phenol) [18] are not assigned as such and also many proelectrophiles [59] are not covered by the rules, respectively would be assigned as polar narcotics. Thus, the data set, although suited for explanatory studies, poses some dangers for the development of predictive models since there is a high probability that descriptors are selected that mainly reflect the rules used to build the data set. With increasing distance to the training set for which the rules were derived such conflicts will increase. Therefore, the study with the 2414 phenols keeps an exploratory nature.

Regarding the comparison between structure descriptors used in this study and descriptors used in previous studies their advantages and disadvantages can summarized as follows: structure descriptors are much simpler and much more rapid to calculate, and they can be used for modeling even when the underlying mechanisms are not known (which is partially the case in the present data set, e.g. for proelectrophiles). However, they cannot be directly linked to a chemical mechanism as easily as whole molecule descriptors e.g., a log P value. The exploratory study gave also indications that their rather high dimensionality, leads to models with rather local character. It would be interesting to repeat the experiment using the six descriptors proposed by Schüürmann *et al.* in order to determine weather these lowerdimensional models have confidence regions containing more compounds. If this is confirmed, the price for going from the sixdimensional space to the 21 dimensions would be that larger training sets are needed to cover the same space.

3.5 Conclusion

In this study MOA classifiers based on empirical structure descriptors were derived. The calculation of such descriptors is very fast which makes them ideally suited for screening of databases. The estimates of predictive power turned out to be equal to previously published quantum mechanical based descriptors. In some cases the information contained in structure descriptors could be directly related to the whole molecule descriptors, which is an additional

justification to use empirical structure descriptors. The determination of a models prediction space proved indespensable for screening studies with structurally diverse databases. In an exploratory study using a large external test set the models showed to have limited prediction spaces, but to have satisfactory quality within this space.

Acknowledgements

Eric Pellegrini would like to thank the Federation of European Biochemical Societies (FEBS) for granting this work. Simon Spycher would like to thank the Bundesministerium für Bildung und Forschung (BMBF) for funding (Systems Biology, project no. FZJ 0313080 C). We also thank Aynur Aptula for providing the data set and for valuable comments.

References

[1] P. Krajcsi, F. Darvas, "High-Throughput In Vitro Toxicology", in "High-Throughput ADMETox Estimation", F. Darvas, G. Dormán (Eds.), Eaton Publishing, **2002**, chapter 6, 75–81.

[2] C. Hansch, "Quantitative approach to biochemical structure-activity relationships", *Acc. Chem. Res.* **1969,** *2,* 232–239.

[3] J. Wang, L. Lai, Y. Tang, "Data mining of toxic chemicals: structure patterns and QSAR", *J. Mol. Model.* **1999,** *5,* 252–262.

[4] G. M. Rand, P. G. Wells, L. S. McCarty, "Introduction to Aquatic Toxicology", in "Fundamentals of Aquatic Toxicology: Effects, Environmental Fate and Risk Assessment", G. M. Rand (Ed.), Taylor & Francis, UK, **1995**, 3–67, 2^{nd} Edition.

[5] C. L. Russom, S. P. Bradbury, S. J. Broderius, D. E. Hammermeister, R. A. Drummond, "Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas)", *Environ. Toxicol. Chem.* **1997,** *16,* 948–967.

[6] S. P. Bradbury, "Predicting modes of toxic action from chemical structure: An overview", *SAR QSAR Environ. Res.* **1994,** *2,* 89–104.

[7] S. C. Basak, G. D. Grunwald, G. E. Host, G. J. Niemi, S. P. Bradbury, "A comparative study of molecular similarity, statistical, and neural methods for predicting toxic modes of action", *Environ. Toxicol. Chem.* **1998,** *17,* 1056–1064.

[8] M. Nendza, M. Müller, "Discriminating toxicant classes by mode of action: 2. Physicochemical descriptors", *Quant. Struct.-Act. Relat.* **2000,** *19,* 581–598.

[9] A. O. Aptula, T. I. Netzeva, I. V. Valkova, M. T. D. Cronin, T. W. Schultz, R. Kühne, G. Schüürmann, "Multivariate discrimination between modes of toxic action of phenols", *Quant. Struct.-Act. Relat.* **2002,** *21,* 12–22.

[10] G. Schüürmann, A. O. Aptula, R. Kuehne, R. U. Ebert, "Stepwise Discrimination between Four Modes of Toxic Action of Phenols in the Tetrahymena pyriformis Assay", *Chem. Res. Toxicol.* **2003,** *16,* 974–987.

[11] O. G. Mekenyan, G. D. Veith, "The electronic factor in QSAR: MO-parameters, competing interactions, reactivity and toxicity", *SAR QSAR Environ. Res.* **1994,** *2,* 129–143.

[12] M. Karelson, V. S. Lobanov, A. R. Katritzky, "Quantum-Chemical Descriptors in QSAR/QSPR Studies", *Chem. Rev.* **1996,** *96,* 1027–1043.

[13] J. Gasteiger, M. Marsili, M. G. Hutchings, H. Saller, P. Löw, P. Röse, K. Rafeiner, "Models for the Representation of Knowledge about Chemical Reactions", *J. Chem. Inf. Comput. Sci.* **1990,** 467–476.

[14] "PETRA - Parameter Estimation for the Treatment of Reactivity Applications", Version 3.1, MolNet GmbH, Erlangen, **2002**.
`http://www.mol-net.com`
`http://www2.chemie.uni-erlangen.de/services/petra/index.html`

[15] S. Ren, "Determining the mechanisms of toxic action of phenols to Tetrahymena pyriformis", *Environ. Toxicol.* **2002,** *17,* 119–127.

[16] S. Ren, H. Kim, "Comparative Assessment of Multiresponse Regression Methods for Predicting the Mechanisms of Toxic Action of Phenols", *J. Chem. Inf. Comput. Sci.* **2003,** *43,* 2106–2110.

[17] T. W. Schultz, G. D. Sinks, M. T. D. Cronin, "Identification of mechanisms of toxic action of phenols to Tetrahymena pyriformis from molecular descriptors", in "Quantitative Structure-Activity Relationships in Environmental Sciences-VII", F. Chen, G. Schüürmann (Eds.), SETAC Press: Pensacola, FL, **1997**, Proceedings of QSAR 96, Elsinore, DK, June 24-28, 1996, 329–337.

[18] H. Terada, "Uncouplers of oxidative phosphorylation", *Environ. Health Perspect.* **1990,** *87,* 213–218.

[19] G. Schüürmann, R. K. Somashekar, U. Kristen, "Structure-activity relationships for chloro- and nitrophenol toxicity in the pollen tube growth test", *Environ. Toxicol. Chem.* **1996,** *15,* 1702–1708.

[20] R. L. Lipnick, "Outliers: their origin and use in the classification of molecular mechanisms of toxicity", *Sci. Total Environ.* **1991,** *109-110,* 131–153.

[21] T. I. Netzeva, A. O. Aptula, S. H. Chaudary, J. C. Duffy, T. W. Schultz, G. Schueuermann, M. T. D. Cronin, "Structure-activity relationships for the toxicity of substituted poly-hydroxylated benzenes to Tetrahymena pyriformis: Influence of free radical formation", *QSAR Comb. Sci.* **2003,** *22,* 575–582.

[22] D. W. Roberts, "An analysis of published data on fish toxicity of nitrobenzene and aniline derivatives", in "QSAR in Environmental Toxicology", K. L. E. Kaiser (Ed.), D. Reidel Publishing Company, **1987**, Proc. 2^{nd} Int. Workshop, 295–308.

[23] A. R. Katritzky, P. Oliferenko, A. Oliferenko, A. Lomaka, M. Karelson, "Nitrobenzene toxicity: QSAR correlations and mechanistic interpretations", *J. Phys. Org. Chem.* **2003,** *16,* 811–817.

[24] A. P. van Wezel, A. Opperhuizen, "Narcosis due to environmental pollutants in aquatic organisms: residue-based toxicity, mechanisms, and membrane burdens", *Crit. Rev. Toxicol.* **1995,** *25,* 255–279.

[25] B. I. Escher, R. Hunziker, R. P. Schwarzenbach, J. C. Westall, "Kinetic Model To Describe the Intrinsic Uncoupling Activity of Substituted Phenols in Energy Transducing Membranes", *Environ. Sci. Technol.* **1999,** *33,* 560–570.

[26] A. Harder, B. I. Escher, R. P. Schwarzenbach, "Applicability and Limitation of QSARs for the Toxicity of Electrophilic Chemicals", *Environ. Sci. Technol.* **2003,** *37,* 4955–4961.

[27] M. T. D. Cronin, A. O. Aptula, J. C. Duffy, T. I. Netzeva, P. H. Rowe, I. V. Valkova, T. W. Schultz, "Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis", *Chemosphere* **2002,** *49,* 1201–1221.

[28] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Oxford Univ. Press, New York, N. Y, **1987**.

[29] J. Karle, "Applications of Mathematics to Structural Chemistry", *J. Chem. Inf. Comput. Sci.* **1994,** *34,* 381–390.

[30] G. Moreau, P. Broto, "The autocorrelation of a topological structure: a new molecular descriptor", *Nouveau Journal de Chimie* **1980,** *4,* 359–360.

[31] G. Moreau, P. Broto, "Autocorrelation of molecular structures. Application to SAR studies", *Nouveau Journal de Chimie* **1980,** *4,* 757–764.

[32] M. Wagener, J. Sadowski, J. Gasteiger, "Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks", *J. Am. Chem. Soc.* **1995,** *117,* 7769–7775.

[33] L. Terfloth, "Calculation of Structure Descriptors", in "Chemoinformatics", J. Gasteiger, T. Engel (Eds.), Wiley-VCH, **2003**, chapter 8, 401–437.

[34] J. Gasteiger, A. Teckentrup, L. Terfloth, S. Spycher, "Neural networks as data mining tools in drug design", *J. Phys. Org. Chem.* **2003,** *16,* 232–245.

[35] A. Vedani, D. W. Huhta, "A new force field for modeling metalloproteins", *J. Am. Chem. Soc.* **1990,** *112,* 4759–4767.

[36] "CORINA", Version 2.4, MolNet GmbH, Erlangen.
`http://www.mol-net.com`
`http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html`

[37] "AUTOCORR", Version 2.4, MolNet GmbH, Erlangen.
`http://www.mol-net.com`

[38] "SURFACE", Version 1.1, MolNet GmbH, Erlangen.
`http://www.mol-net.com`

[39] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, **1999**, 2^{nd} Edition.

[40] J. Zupan, M. Novic, X. Li, J. Gasteiger, "Classification of multicomponent analytical data of olive oils using different neural networks", *Anal. Chim. Acta* **1994,** *292,* 219–234.

[41] P. Selzer, J. Gasteiger, H. Thomas, R. Salzer, "Rapid access to infrared reference spectra of arbitrary organic compounds: scope and limitations of an approach to the simulation of infrared spectra by neural networks", *Chem. Eur. J.* **2000,** *6,* 920–927.

[42] J. A. de Sousa, M. C. Hemmer, J. Gasteiger, "Prediction of 1H NMR Chemical Shifts Using Neural Networks", *Anal. Chem.* **2002,** *74,* 80–90.

[43] "SONNIA - Self Organizing Neural Network for Information Analysis", **2002**, version 4.1, Erlangen, MolNet GmbH.
`http://www.mol-net.com`

[44] S. le Cessie, J. C. van Houwelingen, "Ridge estimators in logistic regression", *Appl. Statist.* **1992,** *41,* 191–201.

[45] F. E. Harrell, Jr., *Regression Modeling Strategies - With applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer-Verlag, New York, **2001**.

[46] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, **1996**.

[47] S. Spycher, M. Nendza, J. Gasteiger, "Comparison of different classification methods applied to a mode of toxic action data set", *QSAR Comb. Sci.* **2004,** *23,* 779–791.

[48] D. W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, Inc., New York, **2000**, 2^{nd} Edition.

[49] "R: A language and environment for statistical computing", Version 1.7.1, R Foundation for Statistical Computing, Vienna, Austria, **2003**.
`http://www.r-project.org`

[50] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*, Springer series in statistics, Springer, New York, **2001**.

[51] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", in "International Joint Conference on Artificial Intelligence (IJCAI), Montréal, Québec, Canada, August 20-25, 1995", Morgan Kaufmann Publishers, **1995**, 1137–1145.

[52] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, *An Introduction to Multi- and Megavariate Data Analysis*, Umetrics AB, Umea, **2000**.

[53] J. Gasteiger, M. Marsili, "Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges", *Tetrahedron* **1980,** *36,* 3219–3222.

[54] J. Gasteiger, H. Saller, "Calculation of the charge distribution in conjugated systems by quantification of the mesomerism concept", *Angew. Chem.* **1985,** *97,* 699–701.

[55] J. Gasteiger, M. G. Hutchings, "Quantification of effective polarizability. Applications to studies of x-ray photoelectron spectroscopy and alkylamine protonation", *Journal of the Chemical Society, Perkin Transactions 2: Physical Organic Chemistry (1972-1999)* **1984,** 559–564.

[56] M. G. Hutchings, J. Gasteiger, "Residual electronegativity - an empirical quantification of polar influences and its application to the proton affinity of amines", *Tetrahedron Letters* **1983,** *24,* 2541–2544.

[57] B. I. Escher, R. P. Schwarzenbach, "Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms", *Aquat. Sci.* **2002,** *64,* 20–35.

[58] B. Efron, R. Tibshirani, "Improvements on Cross-Validation: The .632+ Bootstrap Method", *JASA* **1997,** *92,* 548–560.

[59] R. Garg, A. Kurup, C. Hansch, "Comparative QSAR: on the toxicology of the phenolic OH moiety", *Crit. Rev. Toxicol.* **2001,** *31,* 223–245.

3.6 Further Discussion of Structure Representation

In this section some additional comments are given. They should put the published work into the perspective of the whole thesis.

The main advantage of structure descriptors over previously published descriptors is that they can easily and rapidly be calculated. This allowed the rapid exploratory screening of the 3142 monocyclic phenols contained in the NCI database without major efforts. However, the determination of the model's prediction space indicated that structure descriptors have a price, namely their high dimensionality.

A specific example for the price of using structure descriptors is the problem with proelectrophiles mentioned in Section 3.3.4. The training set did not contain proelectrophiles with larger alkyl-substituents, while there were a large number of polar narcotics with small to very large alkyl-substituents. Such a situation is a problem for the use of $surface\ A$ vectors as shown in Figure 3.6.

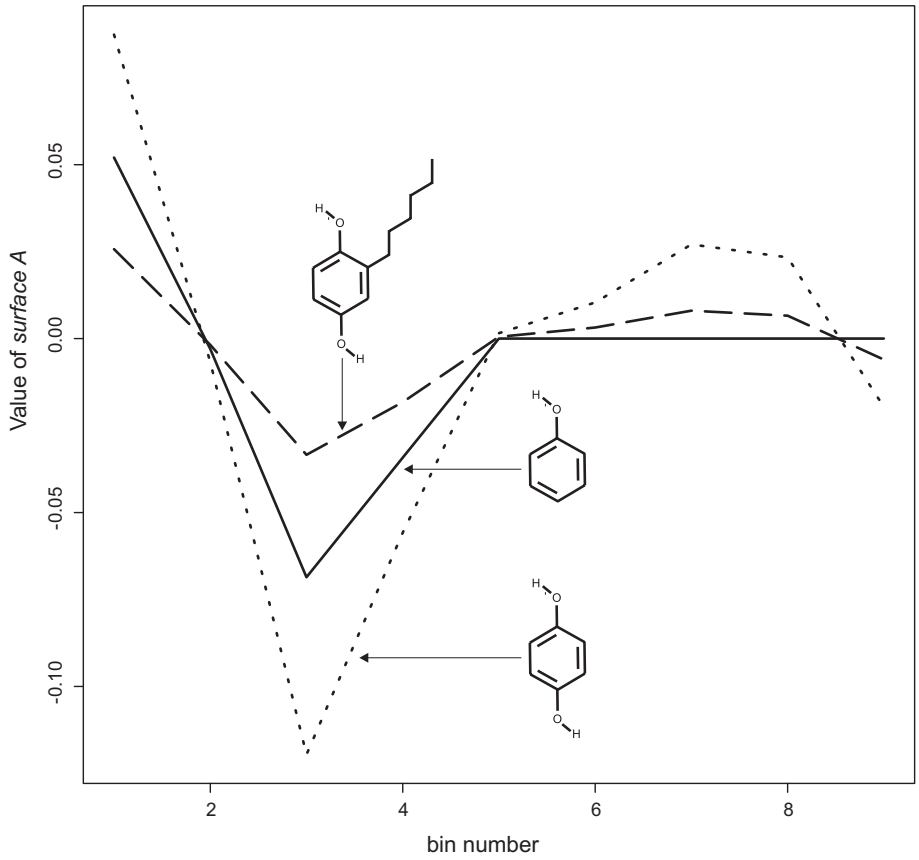


Fig. 3.6 Surface autocorrelation of hydrogen-bonding vectors for the three structures phenol, hydroquinone and 2-hexylhydroquinone. Autocorrelation coded from 1 Å to 7.75 Å in 9 bins.

The two compounds with two OH-groups, hydroquinone and 2-hexylhydroquinone, are proelectrophiles (at least according to the rules derived in the experimental studies of Schultz *et al.* [17]). On the other hand, phenol acts as a polar narcotic. However, the $surface\ A$ vectors of hydroquinone and 2-hexylhydroquinone are quite dissimilar while the polar narcotic phenol is somewhere in between these to proelectrophiles. This is a consequence of the surface autocorrelation algorithm which samples randomly over the whole surface of a molecule and then averages the resulting sums. A large alkyl-substituent increases the size of a molecule and therefore the main peak (at bin number three) of 2-hexylhydroquinone is much smaller than that of hydroquinone although the peaks are at the same position. Thus, this type of descriptor is strongly size dependent. One might say although structure descriptor have a physico-chemical basis they have a strong component of structural similarity descriptors. There are two ways to deal with this problem. The first one is to choose highly balanced training sets really covering all representative compounds of *each* class. The second way to resolve such problems of size-dependence and of dimensionality in general is to use available knowledge about the mechanisms. If, as an example, the reaction center of all compounds is known, vectorial descriptors can be centered on the reaction center, e.g., with atomic Radial Distribution Functions (ARDF):

$$A_i(r) = \sum_{j,j\neq i}^{N} p_i p_j e^{-B(r-r_{ij})^2} \tag{3.6}$$

where N indicates the number of atoms, p an atomic property, B a temperature factor and r_{ij} an interatomic distance. Such ARDF descriptors are not influenced by substituents which are outside its sampling interval, e.g. for hydroquinone and 2-hexylhydroquinone of Figure 3.6 the ARDF could be centered on the oxygen atom without ortho-substituents. If a short sampling interval is taken these two compounds have the same ARDF-vectors. However, in the data set used in this chapter it is not always clear on which atom the reaction takes place and for the first two MOAs there is not even a reaction center because of their unspecific toxic action.

Thus, given the enormous complexity of the processes of phenol toxicity [60], progress towards more mechanistic models can be made only in small steps. Each MOA in itself needs to be better understood to draw advantages of the mechanistic insights. As long as such knowledge is only partially available one can justify the molecular similarity oriented

approach of this chapter as long as it comes along with a good method determining the model's prediction space.

The prediction space method based on Hotelling's T^2 statistic used in Chapter 3 could still be improved. Other methods than Hotelling's as described in a recent report could be applied [61]. Most of all, it would be interesting to make an analogous study to the one presented in Table 3.4 with a large data set of real experimentally measured class assignments instead of the class assignments based on the presently used rules.

The issue of using available knowledge will be investigated more thoroughly in the next chapter, dealing with one MOA in particular, namely uncouplers of oxidative phosphorylation. For this MOA there is a vast amount of toxicological and bio-physical studies especially with artifical membranes. However, these studies went largely unnoticed in QSAR. This knowledge shall be used for a focused search of descriptors reflecting the underlying mechanisms.

Additional References:

[60] I. M. C. M. Rietjens, C. den Besten, R. P. Hanzlik, P. J. van Bladeren, "Cytochrome P450-Catalyzed Oxidation of Halobenzene Derivatives", *Chem. Res. Toxicol.* **1997,** *10,* 629–635.

[61] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. van de Sandt, W. Tong, G. Veith, C. Yang, "Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure Activity Relationships – The Report and Recommendations of ECVAM Workshop 52", *ATLA* **2005,** *33,* 1–19.

Chapter 4

A QSAR Model for the Intrinsic Activity of Uncouplers of Oxidative Phosphorylation

Simon Spycher[†], Beate I. Escher[‡] and Johann Gasteiger[†]*

[†]Computer-Chemie-Centrum und Institut für Organische Chemie,
Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany

[‡]Department of Environmental Toxicology UTOX, Swiss Federal Institute of Aquatic Science and Technology, EAWAG, CH-8600 Dübendorf, Switzerland

Spycher, S., Escher, B. I., Gasteiger, J. (2005), submitted to *Chem. Res. Toxicol.* [a]

Abstract

A quantitative structure-activity relationship (QSAR) has been derived for the prediction of the activity of phenols in uncoupling oxidative and photo-phosphorylation. 21 compounds with experimental data for uncoupling activity, as well as for the acid dissociation constant, pK_{a}, and for partitioning constants of the neutral and the charged species into model membranes were analyzed. From these measured data the effective concentration in the membrane was derived, which allowed the study of the intrinsic activity of uncouplers within the membrane. A linear regression model for the intrinsic activity could be established using the following three descriptors: solvation free energies of the anions, an estimate for het-

[a]This Chapter corresponds exactly to the submitted manuscript other than the last section after the reference list and the numbering.

erodimer formation describing transport processes and pK_{a} values describing the speciation of the phenols. In a next step, the aqueous effect concentrations were modelled by combining the model for the intrinsic uncoupling activity with descriptors accounting for the uptake into membranes. Results obtained with experimental membrane-water partitioning data were compared with the results obtained with experimental octanol-water partition coefficients, log K_{ow}, and with calculated log K_{ow} values. The properties of these different measures of lipophilicity were critically discussed.

4.1 Introduction

Energy-transducing membranes are proton-impermeable membranes which can establish a pH gradient necessary for the production of ATP from inorganic phosphate and ADP by ATP synthase (ATPase). Two types of energy-transducing membranes can be distinguished: mitochondria and chloroplasts in photosynthetic organisms. While these two types of energy-transducing membranes have different sources of energy, they are both able to couple two reactions: electron transport and phosphorylation [1].

Wallace and Starkow [2] defined uncoupling as any energy-dissipating process competing for energy with routine functions of energy-transducing membranes, thus indicating a metabolically futile wasting of energy. In this study we limit the notion of uncouplers to the most important group of uncouplers: weak acids which have the common Mode of Action (MOA) that they can act as proton-shuttles transporting protons across the membrane [3]. Thereby they dissipate the pH-gradient across the membrane by providing an alternate entry path for protons. The most potent uncouplers are so efficient that the ATP-production is completely uncoupled at toxicant concentrations which are up to 20 times below the number of ATPases [1]. This observation, amongst others, led to a widespread agreement that these toxicants do not specifically inhibit a target of energy-transducing membranes, but rather act as catalysts.

When it comes to describing such processes the choice of the appropriate structure representation is the key to the successful development of structure-activity relationships [4]. Chemoinformatic tools today allow the rapid calculation of thousands of descriptors. These descriptors can be related by inductive learning methods to the property or activity of interest. However, this abundance of available descriptors makes descriptor selection a difficult task

and also poses the danger of chance correlations especially for the small data sets common in toxicity QSAR. One of the most efficient ways for focusing the search for appropriate descriptors is to build on the available knowledge of the processes taking place at the molecular level. Fortunately, in the case of uncouplers there is a good understanding of the processes involved in toxicity as will be shown in detail in the section on toxic mechanisms.

A large number of QSARs for weak acids acting as uncouplers have been published. The models based on data from specific mitochondrial tests (in-vivo models are omitted for brevity) cover a wide range of chemical classes ranging from phenols [5–9] to compounds with NH-acidic groups like diarylamines [10], salicylanilides [11], and pyrroles [12]. All of these QSARs are based on a term for lipophilicity describing the uptake into the membrane and most cases on an additional term describing the acidity of the toxicants. Data sets including compounds with bulky ortho-substituents also required steric descriptors to describe the shielding of the negative charge on the oxygen or nitrogen atom, respectively, of the anion [6]. Lipophilicity has been described by the octanol-water partitioning coefficient, log P_{ow}, (or log K_{ow} in environmental sciences) or the membrane-water partitioning coefficient, log K_{mw}, measured with liposomes [7]. Interestingly, although the uptake of the anions into the membrane is of crucial importance to our knowledge no successful attempts to use pH-dependent partitioning coefficients, log D, have been reported for the mode of action of uncoupling. However, there is some agreement that toxicity of phenolic compounds cannot be attributed only to the neutral species alone [13, 14].

The structural requirements for high uncoupling activity can be summarized as follows: an acidic group for the delivery of a proton, bulky lipophilic groups for a better uptake of the charged species into the membrane and a strong electron withdrawing moiety for the delocalization of the charge of the anion [1].

In this study, a data set consisting of 21 phenols with experimental data for uncoupling activity will be examined. Additionally, experimental data for pK_{a} and for the pH-dependent membrane-water partitioning coefficient, log D_{mw} are available. Therefore it is possible to calculate the concentrations of the neutral and the charged species in the membranes and, thus, to clearly separate the modeling of the uptake into the membrane from the activity within the membrane - the so-called intrinsic activity. The goal of this study is to develop a regression model for the activity within the membrane, $\mathrm{EC_m}$, and then later to search for ways to relate the predictions for $\mathrm{EC_m}$ to aqueous effect concentration, $\mathrm{EC_w}$ as illustrated

by Figure 4.1.

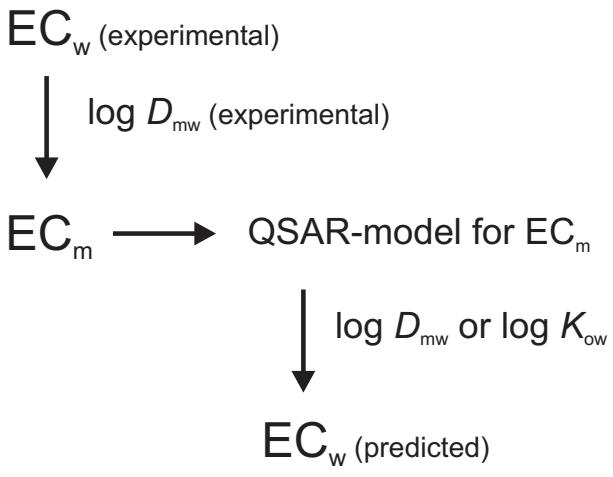


Fig. 4.1 Scheme for the approach to separately model intrinsic activity, $\mathrm{EC_m}$, and uptake using log D_{mw} or log K_{ow}.

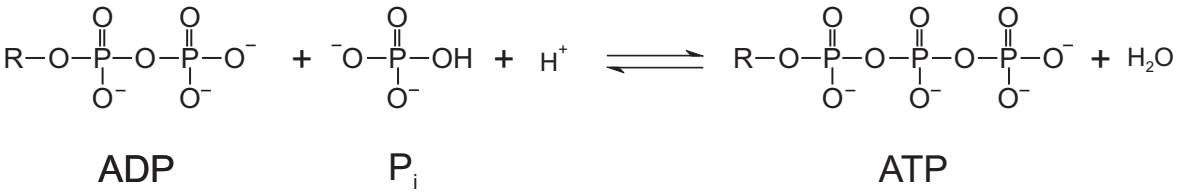
As experimental and calculated values for log K_{ow} are available as well, it will be additionally possible to study the error introduced by using octanol as an approximation for real membranes and the additional error introduced by calculation methods for log K_{ow}.

4.2 Materials and Methods

4.2.1 Toxic Mechanisms

Energy-transducing membranes such as mitochondria and photosynthetic membranes couple the oxidation of NADH with the phosphorylation of ATP, an endergonic reaction shown in Scheme 1.

Scheme 1. Synthesis of ATP.



The chemi-osmotic theory of Mitchell [15] could explain how these reactions are coupled by the generation of an electrochemical proton gradient. In addition, the theory consistently

explained the toxic effect of some lipophilic compounds which can dissipate the electrochemical proton gradient, so-called uncouplers. Figure 4.2 shows the current understanding of uncouplers acting by the proton shuttle mechanism as reviewed by McLaughlin *et al.* [3] and Terada [1]. Uncouplers can carry protons across energy-transducing membranes. In mitochondria the protons pass the inner mitochondrial membrane without passing across the F_0F_1ATPase, avoiding the production of ATP. Thus, in uncoupled mitochondria, oxidation continues, driving protons out of the mitochondria and generating heat, but no ATP is synthesized, no energy is stored, because the uncouplers short-circuits the system [12]. The observed symptom in isolated mitochondria is an increase in O_2-consumption.

The *in vitro* test system Kinspec, which was used in this study, is based on chromatophores extracted from the photosynthetic bacterium *Rhodobacter sphaeroides* [16]. In the Kinspec test a brief "single-turnover" flash of light causes a build-up of membrane potential and the subsequent decay is monitored using time-resolved spectroscopy [16, 17]. The presence of uncouplers acting by the proton shuttle mechanism accelerates the decay of the chromatophore membrane potential, because of their ability to transport protons across the membrane. This allows the determination of the activity of different uncouplers. The endpoint for the toxic effect of uncoupling in this system [18] is the concentration of toxicant needed to induce an observed pseudo-first order rate constant, k_{obs}, of 0.5 s^{-1}. The measurements in the Kinspec system are in good agreement with other *in vitro* tests based on oxygen evolution of isolated mitochondria or submitochondrial particles [7, 9, 16]. Miyoshi and Fujita also found excellent correlations between uncoupling data measured with different test systems like rat liver mitochondria, flight muscles of house flies, and spinach chloroplasts [6].

A closer look at the structure of bilayer membranes yields the key to the approach used in this QSAR study and to the choice of descriptors. The crucial property is the change in hydrophobicity across the membrane as illustrated in Figure 4.3 by the change of the relative dielectric constant, ε_{r}. It is a consequence of the arrangement of the phosphatidylcholine molecules in the bilayer with polar head groups forming the outer side of the membrane and the alkyl chains forming the hydrophobic core.

The relative dielectric constant, ε_{r}, drops from a value of 80 in water to values around 70 on the membrane surface [20] to 30 at the interface between polar head groups and hydrocarbon core [20] to values around 2 in the hydrophobic core [19, 21]. This arrangement leads to large differences in the Gibb's free energy for a molecule depending on the po-

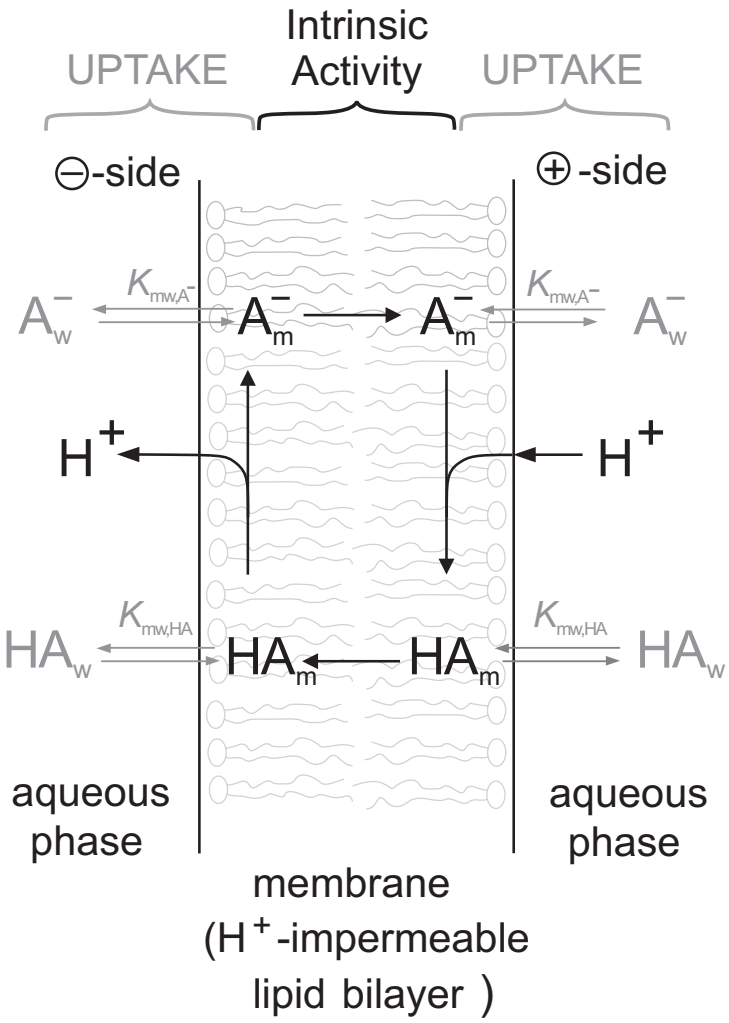


Fig. 4.2 Uncoupling process of weak acids acting as uncouplers in photosynthetic membranes (black letters): through diffusion weak acids can transport protons from the +-side to the −-side where they release a proton. Driven by the membrane potential the anions migrate back and can pick up a new proton. Note that in mitochondrial membranes the +-side is the cytosol and the −-side is the inner matrix, while in photosynthetic membranes as depicted here the opposite is the case. The grey letters indicate the partitioning constants of the uptake into the membrane and will be discussed in detail in the next section. The grey shaded circles indicate the polar head groups of the phosphatidylcholine molecules, while the ondulated lines indicate the fatty acid alkyl chains.

sition in the membrane. As illustrated in the free energy schemes of Figure 4.3, charged hydrophobic compounds and also polar hydrophobic compounds have the lowest energy in the region of the polar head groups, i.e., this is the region where such compounds are mainly accomodated. For a given molecule, the energy required to cross the hydrophobic core is substantially higher for the anionic species than for the neutral species.

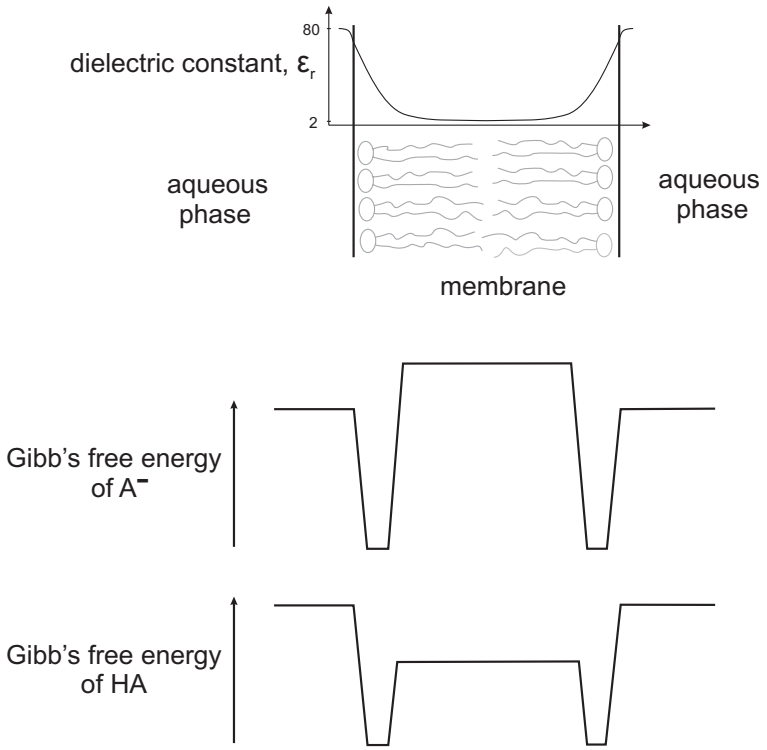


Fig. 4.3 Change in the relative dielectric constant, ε_{r}, and in the Gibb's free energy of the anion and the neutral species across the membrane bilayer for the compound carbonyl cyanide *m*-chlorophenylhydrazone, CCCP (energy schemes adapted from Refs. [19] and [3]).

The schemes also make clear that the uptake *into* the membrane and the permeation *through* the membrane are different processes. If uptake into the target site (in our case, the energy-transducing membrane) and biological activity at the target site (in our case, the transport of protons through the membrane) are treated separately, insights into the toxicants intrinsic activity can be gained. This separation of uptake and intrinsic activity is the basis for the QSAR model proposed in this work.

Equation 4.1 describes how the aqueous effect concentration, $\mathrm{EC_w}$, and the intrinsic effect concentrations within the membrane, $\mathrm{EC_m}$, are related to the pH-dependent partition coefficient, D(pH), which is usually referred to by its logarithmic form as log D. All log D values in this study were calculated for a pH-value of 7.

$$\mathrm{EC_m}/\mathrm{EC_w} = D \tag{4.1}$$

If toxicity is expressed as the negative logarithm of the effect concentration one obtains the following expression

$$\log 1/\mathrm{EC_m} = \log 1/\mathrm{EC_w} - logD \tag{4.2}$$

which makes clear that a compound with a high aqueous toxicity, log $1/\mathrm{EC_w}$, but a low tendency to partition into the membrane, i.e., a low log D, has a high intrinsic uncoupling activity. On the other hand a compound with comparable log $1/\mathrm{EC_w}$ and large log D values is intrinsically not very active.

The central questions to address are, first, whether only the neutral species, HA or also the charged species, $\mathrm{A^-}$, are of relevance for a given mode of action and, secondly, how to describe log D. Therefore, in the next section, some approaches to model $\mathrm{EC_m}$ are presented.

Description of the Uptake of Weak Acids: The most common approach in QSAR is to model the uptake by the octanol-water partitioning constant, log K_{ow}. However, for charged species the uptake becomes pH-dependent and cannot be modeled that simply. QSARs for uncoupling activity of weak acids are such a case. In some models for the uptake of weak acids it is assumed that only the neutral species, HA, is taken up into the membrane. The fraction of neutral species in water, f_{HA} can be expressed as

$$f_{\mathrm{HA}} = \frac{1}{1 + 10^{\,\mathrm{pH-p}K_{\mathrm{a}}}} \tag{4.3}$$

for a given pH-value. Then, the pH-dependent distribution coefficient between octanol and water, D_{ow}, can be expressed by

$$D_{\mathrm{ow}} = f_{\mathrm{HA}} \cdot K_{\mathrm{ow}} \tag{4.4}$$

However, the assumption in Equation 4.4 that only the neutral species partitions into octanol is valid only for a few pH-units away from the pK_{a} value [22].

As most uncouplers have a pK_{a} value which is only a few log units away from the pH-value, this error is small. However, the most serious flaw of using Equation 4.4 is that the bulk solvent octanol is not a good model for biological membranes in the case of polar and especially in the case of charged molecules. Experiments with liposomes, which are artificial bilayer vesicles consisting of phospholipids, showed that anions sorb into membranes at concentrations which are by orders of magnitude larger than those measured in the solvent octanol [21, 23–25]. Liposomes are a good approximation to real membranes, as their partition coefficients correspond well with measured partition coefficients to the lipid fraction in biological membranes [24].

If the liposome-water partition coefficients of the neutral species, $K_{\mathrm{mw,HA}}$ and also of the charged species, $K_{\mathrm{mw,A^-}}$, are known, it is possible to calculate the exact concentration of each species in each phase. The subscript m denotes the membrane phase approximated by liposomes, HA denotes the neutral species and $\mathrm{A^-}$ the anion. The formula for the pH-dependent distribution coefficient between membranes and water, D_{mw} can then be expressed by

$$D_{\mathrm{mw}} = K_{\mathrm{mw,HA}} \cdot f_{\mathrm{HA}} + K_{\mathrm{mw,A^-}} \cdot (1 - f_{\mathrm{HA}}) \tag{4.5}$$

The pH-dependence of log D_{mw} described by Equation 4.5 for the example of 2,4-dinitro-o-cresol (DNOC) is the upper line shown in Figure 4.4. On the other hand the lower line shows the experimental log D_{ow} of DNOC and the dotted line the estimated log D_{ow} using Equation 4.4.

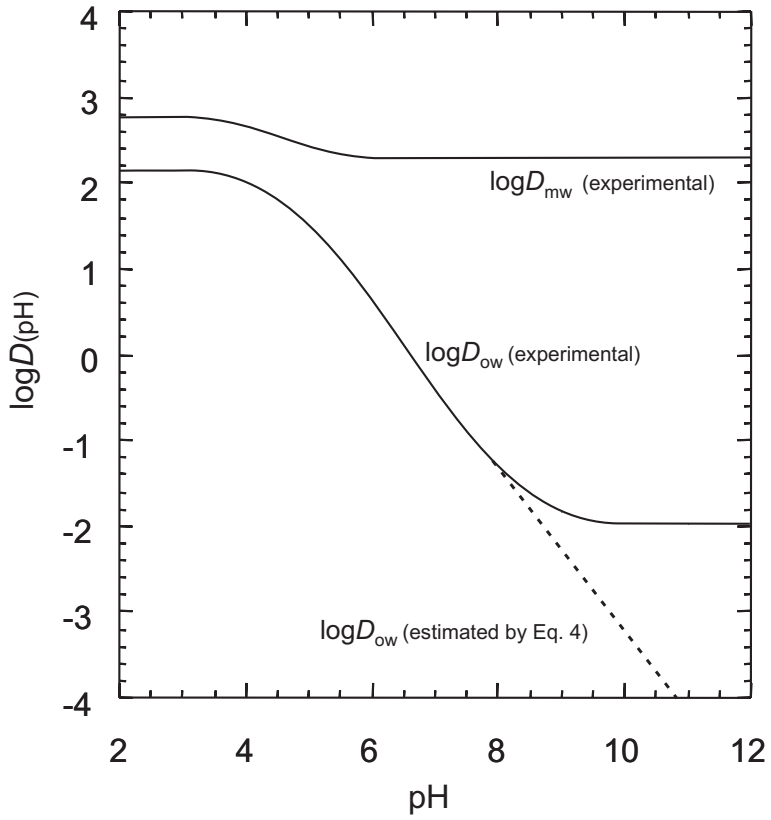


Fig. 4.4 Experimentally determined pH-dependence of log D_{ow} for 2,4-dinitro-o-cresol (pK_{a} = 4.3) and of log D_{mw} determined with liposomes (curve modeled by Equation 4.5). The curve for log D_{ow} is based on values measured by Jafvert *et al.* [22]. Both curves were measured at a $\mathrm{K^+}$-concentration of 0.01 mol·$\mathrm{L^{-1}}$. The dotted line shows the estimate for log D_{ow} by Equation 4.4.

For pH-values below 3, log D_{mw} is approximately equal to log $K_{\mathrm{mw,HA}}$ which means that the lipophilicity of the toxicant is approximately equal to the lipophilicity of the neutral species which is highly predominant in the membrane. For pH-values above 5.5, D_{mw} is approximately equal to $K_{\mathrm{mw,A^-}}$ and the lipophilicity of the toxicant is approximately equal to the lipophilicity of the charged species. For the present data set, the average difference between $K_{\mathrm{mw,HA}}$ and $K_{\mathrm{mw,A^-}}$ is one log unit. Thus, DNOC, the example in Figure 4.4, has a particularly small difference between log $K_{\mathrm{mw,HA}}$ and log $K_{\mathrm{mw,A^-}}$ with only 0.4 log units. Equation 4.2 and Equation 4.5 allow one to cover the uptake part depicted in Figure 4.2 and to accurately transform the aqueous effect concentration, $\mathrm{EC_w}$ to the intrinsic effect concentration, $\mathrm{EC_m}$, in the membrane. Note, that both $\mathrm{EC_w}$ and $\mathrm{EC_m}$ are defined as total concentrations of HA and $\mathrm{A^-}$ in water and the membrane phase, respectively. Any other definition, like Equation 4.4, sometimes used for other modes of toxic action, which considers only the uptake of HA does not make much sense in the case of uncouplers acting by the proton shuttle mechanism, as both HA and $\mathrm{A^-}$ have to be present in the membrane.

Description of the Permeability of Weak Acids: The intrinsic activity, $\mathrm{EC_m}$ describes how easily a compound, once it has been taken up *into* the membrane, can transport protons *through* the membrane. The bottleneck of this process is the energy required to move the charged species, $\mathrm{A^-}$, through the hydrophobic core of the membrane (cf. Figure 4.3). Kasianowicz and Benz calculated that the free energy difference between carbonyl cyanide *m*-chlorophenylhydrazone (CCCP) situated in the region of the polar head groups and CCCP in the core is 3.5 kcal/mol higher for the anion, $\mathrm{A^-}$, than for the neutral species, HA [19]. Efficient uncouplers like CCCP are compounds which have found a way to keep this energy as low as possible, or, in other words, to allow anions which are *per se* hydrophilic to show hydrophobic behavior. There seem to be two ways how a compound can achieve this. The first option is a high delocalization of the negative charge ideally accompanied with shielding of the negative charge [1] (Figure 4.5a). The second option is the delocalization of the negative charge accompanied by the formation of heterodimers, $\mathrm{AHA^-}$, i.e., one neutral species, HA associated with one charged species, $\mathrm{A^-}$ (Figure 4.5b) [26].

From studies in which the uncoupling activity was measured for different concentrations and at different pH-values, it was possible to develop a kinetic uncoupler model with experimentally determined input variables [27]. By solving the differential equations for the transport through diffusion (neutral species) and through migration (charged species) the

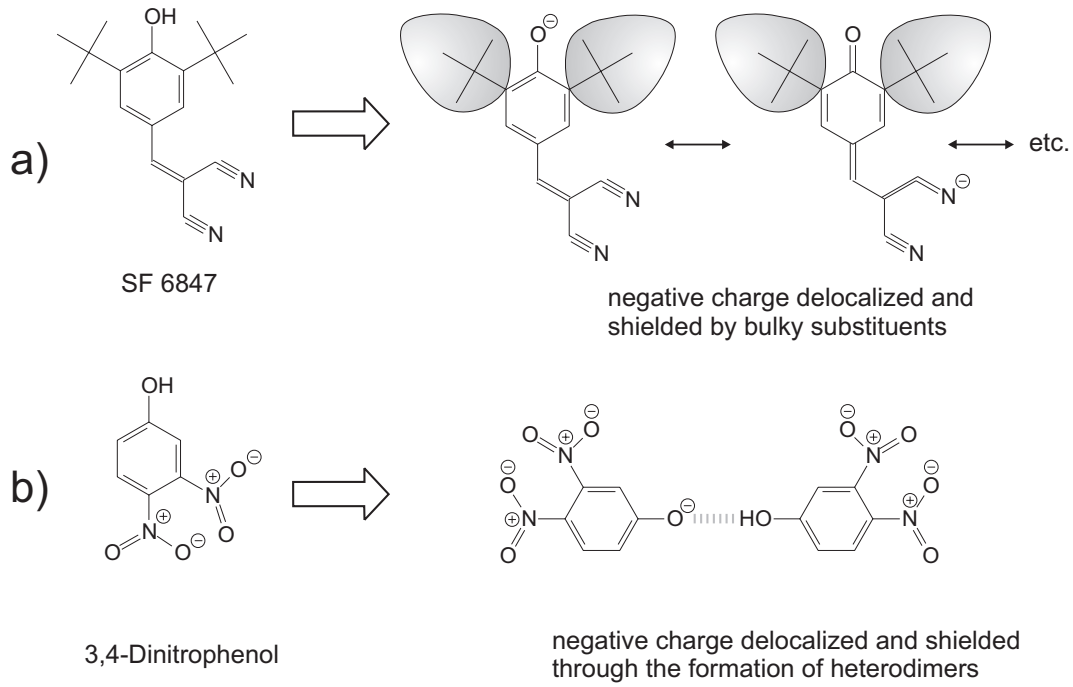


Fig. 4.5 Two typical structures with high intrinsic uncoupling activity. a) Class 1 uncouplers: The negative charge of the SF 6847 anion is shielded through the two bulky t-butyl-groups in ortho-position and additionally delocalized to the dicyanovinyl-substituent. b) Class 2 uncouplers: 3,4-dinitrophenolate has a high delocalization of the negative charge and its high activity suggests the formation of heterodimers [26] which was corroborated in kinetic models [27].

translocation rate constants of all involved species were determined, namely HA, A^- and AHA^-. The most important findings from Ref. [27] with regards to the present study can be summarized as follows: 1. Many compounds use both mechanisms shown in Figure 4.5, thus, have significant amounts of both A^- and AHA^- crossing the membrane 2. Compounds with bulky ortho-substituents or ortho-substituents carrying a large partial charge did not form AHA^- at all. 3. Although the free energy difference of the charged species is the energy bottleneck, the back-diffusion of the neutral species can have a significant contribution to the activity of some compounds. 4. The activity of an uncoupler shows an optimum at a given pH in the proximity of the pK_a value.

Unfortunately, the data required for fitting the full kinetic model, i.e., the measured effects for different concentrations *and* different pH-values, were available only for five compounds [27]. Nevertheless, the insights gained from these experimental kinetic studies could be used in the present study as a guide for the calculation and selection of descriptors.

4.2.2 Data Set

The data set consists of *in vitro* effect concentrations for 21 phenols (Table 4.1) with effect concentrations, $\mathrm{EC_w}$, ranging over four orders of magnitude. All compounds were measured in the same laboratory at the Swiss Federal Institute of Aquatic Science and Technology, EAWAG, using the Kinspec system. All $\mathrm{EC_w}$-values given in Table 4.1 were measured at a pH-value of 7. The compounds with low activities have a toxicity which comes close to nonspecific membrane toxicity (narcosis), but the activity was still higher than that of compounds with purely nonspecific membrane toxicity. Details about the classification of these compounds as weak uncouplers are given elsewhere [28]. For all 21 compounds also measured values for membrane-water partition coefficients of both HA and $\mathrm{A^-}$ and for pK_a values were available. This allowed us to derive values for the intrinsic activity, $\mathrm{EC_m}$. The logarithm of $\mathrm{EC_m}$ extends over two and a half orders of magnitude. Additionally, experimental log K_ow values were compiled from literature.

Table. 4.1 Measured values of uncoupling data set. Effective concentrations at pH 7 in Kinspec system. All values taken from [28], except N ° 21 [29] and N° 9 (taken as average value of [28] and two older unpublished measurements).

N°	Abbrev.	Name	CAS	pK_{a}	log $K_{\mathrm{mw,HA}}$ ($\mathrm{L \cdot kg_{lip}^{-1}}$)	log $K_{\mathrm{mw,A^-}}$ ($\mathrm{L \cdot kg_{lip}^{-1}}$)	log K_{ow}	$-\log \mathrm{EC_w}$ (M)	$-\log \mathrm{EC_m}$ ($\mathrm{mol \cdot kg_{lip}^{-1}}$)
1	24DCP	2,4-Dichlorophenol	120-83-2	7.85	3.59	2.69	3.23	4.63	1.08
2	34DCP	3,4-Dichlorophenol	95-77-2	8.59	3.76	2.85	3.05	4.78	1.03
3	35DCP	3,5-Dichlorophenol	591-35-5	8.26	3.76	2.85	3.62	5.40	1.66
4	245TriCP	2,4,5-Trichlorophenol	95-95-4	6.94	4.46	2.98	4.19	6.03	1.90
5	246TriCP	2,4,6-Trichlorophenol	88-06-2	6.15	3.99	2.50	3.72	4.33	1.15
6	345TriCP	3,4,5-Trichlorophenol	609-19-8	7.73	4.71	3.16	4.41	6.98	2.34
7	2345TeCP	2,3,4,5-Tetrachlorophenol	4901-51-3	6.35	4.76	3.90	4.87	7.15	2.92
8	2346TeCP	2,3,4,6-Tetrachlorophenol	58-90-2	5.40	4.46	3.46	4.42	5.29	1.74
9	PCP	Pentachlorophenol	87-86-5	4.75	5.10	4.35	5.24	7.07	2.70
10	35DBC	3,5-Dibromo-4-methylphenol	13979-81-2	8.28	4.51	3.18	5.44	6.41	1.92
11	Bromox	3,5-Dibromo-4-OH-benzonitrile	1689-84-5	4.09	3.16	2.10	2.97	4.90	2.79
12	4NP	4-Nitrophenol	100-02-7	7.08	2.72	0.95	1.91	4.00	1.54
13	24DNP	2,4-Dinitrophenol	51-28-5	3.94	2.64	1.90	1.67	4.14	2.24
14	26DNP	2,6-Dinitrophenol	573-56-8	3.70	2.03	1.86	1.22	3.09	1.23
15	34DNP	3,4-Dinitrophenol	577-71-9	5.48	3.17	1.90	1.79	4.84	2.76
16	DNOC	2-Methyl-4,6-dinitrophenol	534-52-1	4.31	2.76	2.35	2.12	4.75	2.40
17	DNPC	4-Methyl-2,6-dinitrophenol	609-93-8	4.06	2.34	2.26	1.88	3.42	1.16
18	Dinoseb	2-sec-Butyl-4,6-dinitrophenol	88-85-7	4.62	3.96	3.35	3.56	6.39	3.04
19	Dino2terb	2-tert-Butyl-4,6-dinitrophenol	1420-07-1	4.80	4.10	3.59	3.54	7.00	3.40
20	Dino4terb	4-tert-Butyl-2,6-dinitrophenol	4097-49-8	4.11	3.81	3.23	3.36	4.28	1.05
21	Triclosan	2,4,4’-Trichloro-2’-OH-diphenylether	3380-34-5	8.05	3.89	2.96	4.76	6.01	2.16

4.2.3 Calculation of Descriptors

The models and findings described in section 4.2.1 lead to the conclusion that a QSAR model for intrinsic uncoupling activity must cover the processes listed in Table 4.2. Some processes, e.g., mobility of the charged species A^- and AHA^- are influenced by several chemical mechanisms and, correspondingly, might need several descriptors in combination.

Table. 4.2 Factors of importance for uncoupling activity and descriptors tested. The abbreviations of the descriptors are explained in the text.

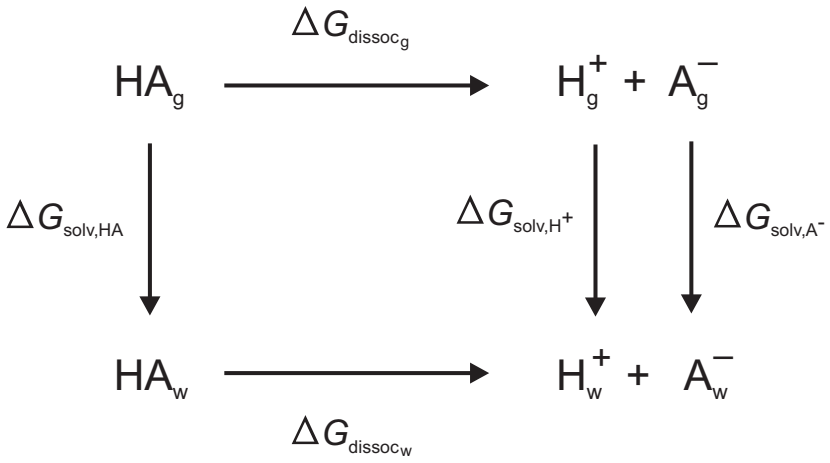
Process	Chemical mechanism	Descriptors tested
Mobility of charged species	Delocalization of charge Shielding of charge	D^-, $\Delta G_{\mathrm{dissoc_g}}$, $\Delta G_{\mathrm{solv,A^-}}$ $\Sigma\mathrm{ASA}^{\mathrm{ortho}}$
Formation of AHA^-	Hindrance of AHA^--formation	$\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$, $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$
Mobility of HA	Permeability	ASA
Speciation	Fraction of A^- present	$\mathrm{p}K_{\mathrm{a_w}}$, $\mathrm{p}K_{\mathrm{a_m}}$

Each descriptor will be briefly described in the following sections ordered according to the chemical mechanism listed in Table 4.2. The values of the descriptors are given in Table 4.3.

Delocalization of negative charge: The energy of mesomeric stabilization, D^-, was calculated using PETRA [30] (Keyword: B_NDELOC) which contains an empirical model derived from gas-phase acidities [31]. The gas-phase acidity, $\Delta G_{\mathrm{dissoc_g}}$, and the solvation free energy of the anion, $\Delta G_{\mathrm{solv,A^-}}$, were calculated using SPARC [32]. Both descriptors $\Delta G_{\mathrm{dissoc_g}}$ and $\Delta G_{\mathrm{solv,A^-}}$ are calculated with empirical methods incorporating monopole electrostatic interaction models calibrated with gas phase acidities [33]. The two descriptors are connected through the thermodynamic cycle shown in Scheme 2.

Shielding of negative charge: The atomic contribution to the approximate surface area, ASA [34] was calculated using ASA 1.0 [35] which is part of the MOSES-library developed over the last years [36, 37]. ASA is a topological descriptor calculated using a template of van der Waals radii [34]. In order to represent the shielding effect the sum of the ASAs of the ortho-substituents was calculated, $\Sigma\mathrm{ASA}^{\mathrm{ortho}}$. As an example, for 2-Methyl-4,6-dinitrophenol (DNOC) $\Sigma\mathrm{ASA}^{\mathrm{ortho}}$ is calculated as the sum of the surface contribution of three hydrogen atoms and one methyl-carbon atom at position 2 plus the sum of two nitro-oxygen atoms and one nitro-nitrogen atom at position 6.

Scheme 2. Thermodynamic cycle connecting the descriptors $\Delta G_{\mathrm{dissoc_g}}$ and $\Delta G_{\mathrm{solv,A^-}}$ with the superscript g denoting the dissociation in the gas phase and the subscript w denoting the dissociation in water.



Hindrance of $\mathbf{AHA^-}$-H-Bond-Formation: As a simple approach to calculating the strength of intermolecular hydrogen bonds we followed the hypothesis that in a highly hydrophobic environment, like the core of a membrane, any phenol would establish a hydrogen bond with a phenolate anion unless hindered by the ortho-substituents. Two alternative descriptors approximating this hindrance were calculated. The first one is based on a comparison of van der Waals radii, $\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$ and the second one on charged surfaces of the ortho-substituents, $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ as described below.

In order to calculate the overlap of the van der Waals radii of the phenolic oxygen atom with its substituents the distances, d_2 and d_6, of the phenolic oxygen atom to the closest atom of the ortho-substituent at position 2 and 6 were measured using the 3D-structure generated by CORINA [38]. As an example, for 2-Methyl-4,6-dinitrophenol d_2 is the distance from the phenolic oxygen atom to the closest hydrogen atom and d_6 the distance to the closest nitro-oxygen atom. The difference between d_i and the van der Waals radius of the phenolic oxygen atom plus the van der Waals radius of the closest substituent atom was calculated and summed up to the obtain the descriptor $\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$.

The charged approximate surface area, CASA, of an atom i, was calculated as the product of the approximate surface area of an atom i, ASA_i, with its total charge, q_{tot}. CASA is a descriptor which is related to CPSA, the charged polar surface area [39, 40]. Total charges, q_{tot}, were calculated as the sum of σ charge [41] and π-charge [42] as implemented in PE-

TRA. The sum of the products $q_{tot} \cdot \mathrm{ASA}_i$ for all ortho-substituents atoms, $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$, was used as a descriptor.

Mobility of HA: The approximate surface area, ASA, of a molecule was calculated by summing up the ASA_i contributions of all i atoms.

Speciation: This descriptor should describe the speciation, i.e., the fraction of anion present in the membrane. Measured values for the dissociation constants in water, $\mathrm{p}K_{\mathrm{a_w}}$, were used from publications as given in Table 4.1. The $\mathrm{p}K_{\mathrm{a}}$ value in the membrane, $\mathrm{p}K_{\mathrm{a_m}}$, is defined in terms of the proton activity in the aqueous phase, $a_{\mathrm{H_w^+}}$, assuming that there are no free protons in the membrane and that the protons of the acid-base reaction in the membrane are directly exchanged with the adjacent aqueous solution [27]. The mass law expression for $K_{\mathrm{a_m}}$ is

$$K_{\mathrm{a_m}} = \frac{C_{\mathrm{A_m^-}} \cdot a_{\mathrm{H_w^+}}}{C_{\mathrm{HA_m}}} = K_{\mathrm{a_w}} \cdot \frac{K_{\mathrm{mw,A^-}}}{K_{\mathrm{mw,HA}}} \tag{4.6}$$

and $\mathrm{p}K_{\mathrm{a_m}}$ is the negative decadic logarithm of $K_{\mathrm{a_m}}$.

Uptake: Once a good model for the intrinsic activity, $\mathrm{EC_m}$, has been established one needs to relate the predictions to the aqueous effect concentration, $\mathrm{EC_w}$. As a first step the predictions for $\mathrm{EC_m}$ were directly transformed using Equation 4.2 and the $\log D_{\mathrm{mw}}$ definition of Equation 4.5 without the use of a regression model. In a second approach, the measured values for membrane-water distribution coefficient of the neutral species, $K_{\mathrm{mw,HA}}$, was used as the additional descriptor modeling the uptake in the regression model. Alternatively, the octanol-water partition coefficient, K_{ow}, was used. In case the latter would result in a model of acceptable quality it would have the advantage that it can easily be calculated, e.g. by the XLOGP-method implemented in the MOSES-library [43, 44] or by the CLOGP-method [45, 46]. The values of all calculated descriptors are given in Table 4.3.

4.2.4 Statistical Methods and Model Validation

The negative logarithm of the intrinsic effect concentrations, $\mathrm{EC_m}$ was used as the response (Y-variable) in multiple linear regression models using ordinary least squares (OLS regression). Although Table 4.2 lists a total of nine descriptors (X-variables), no automated variable selection method (like stepwise exclusion) was applied, but instead the reasonings pre-

Table. 4.3 Calculated descriptors for 21 uncouplers. Values for log K_{ow} were calculated with CLOGP (C) and XLOGP (X), respectively.

N°	D^- (eV)	$\Delta G_{\mathrm{dissoc_g}}$ ($\mathrm{kcal \cdot mol^{-1}}$)	$\Delta G_{\mathrm{solv,A^-}}$ ($\mathrm{kcal \cdot mol^{-1}}$)	$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ (e·Å^2)	$\Sigma\mathrm{ASA}^{\mathrm{ortho}}$ (Å^2)	log $K_{\mathrm{ow-C}}$	log $K_{\mathrm{ow-X}}$
1	18.80	−331.80	−68.20	−3.50	50.90	2.80	2.80
2	16.00	−333.40	−69.70	1.50	23.50	2.90	2.80
3	13.20	−332.70	−69.20	1.60	23.50	2.80	2.80
4	18.90	−329.20	−67.00	−3.50	50.90	3.50	3.50
5	22.00	−330.40	−65.50	−8.60	78.30	3.40	3.50
6	16.00	−330.70	−68.40	1.60	23.50	3.50	3.50
7	18.90	−326.60	−65.80	−3.40	50.90	4.10	4.10
8	22.00	−328.00	−64.30	−8.50	78.30	4.10	4.10
9	21.90	−325.60	−63.20	−8.40	78.30	4.70	4.70
10	15.30	−332.80	−69.50	1.60	23.50	3.40	3.80
11	20.40	−321.10	−61.60	−7.30	91.90	3.00	3.40
12	17.80	−327.70	−69.10	1.60	23.50	1.40	1.90
13	22.60	−317.90	−62.70	−14.80	69.80	1.90	1.70
14	22.60	−310.70	−55.80	−31.30	116.00	2.30	1.70
15	18.90	−322.80	−67.70	1.60	23.50	1.30	1.70
16	25.90	−315.10	−59.40	−14.90	102.20	2.10	2.30
17	25.90	−310.90	−55.30	−31.30	116.00	2.70	2.30
18	25.90	−304.90	−47.10	−14.10	184.10	3.60	3.70
19	26.00	−304.50	−46.00	−14.60	190.40	3.60	3.60
20	25.90	−310.00	−53.40	−31.20	116.00	4.30	3.60
21	22.20	−317.90	−55.80	−1.90	22.80	4.70	4.70

sented in subsection 4.2.1 and in Table 4.2 were used for a focused search for appropriate variables. A look at the number of mechanisms listed in Table 4.2 shows that a maximum of five variables should do the job and cover all factors of influence. However, in the first attempt only the bottleneck of the process shall be modeled, namely the transfer of charged species through the membrane. This should be feasible with two or three variables, namely one for delocalization of negative charge, one for shielding of negative charge and one for the formation of heterodimers.

Following the arguments of Hawkins [47] the data set was not split into a training and a test set, because with such small data sets this leads to high variance in the estimate of the predictive power. Instead a five-fold cross-validation was used which results in estimates which are closer to the true error if conducted properly [47, 48]. The data set was split 19 times randomly into five subsets, i.e., the cross-validation was repeated 19 times in order to

obtain more balanced samples.

Statistical calculations were made with R 2.0 [49].

4.3 Results

4.3.1 Intrinsic Activity, $\mathrm{EC_m}$

The first goal of this study was to develop a model for intrinsic activity $\mathrm{EC_m}$. Once it is established, a suitable method to transform the predictions for intrinsic activity back to aqueous activity $\mathrm{EC_w}$ can be searched. The set of descriptors presented in Table 4.2 was tested starting with the lowest number of descriptors possible, namely one for the mobility of the charged species and one for the formation of $\mathrm{AHA^-}$. Table 4.3 gives an overview of the correlations between fitted and experimental values and the standard deviation of the residuals.

Table. 4.4 Possible combinations of descriptors for the two processes mobility of charged species and formation of $\mathrm{AHA^-}$.

Variable combination	R^2	sd
$\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$ + D^-	0.05	0.72
$\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{dissoc_g}}$	0.16	0.68
$\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{solv,A^-}}$	0.18	0.67
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + D^-	0.35	0.60
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{dissoc_g}}$	0.59	0.48
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{solv,A^-}}$	0.47	0.54

The descriptor $\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$ did not contribute to good correlations in any combination while $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ as the alternative descriptor for the formation of $\mathrm{AHA^-}$ shows better correlations. An interesting result was that the gas phase acidity $\Delta G_{\mathrm{dissoc_g}}$ resulted in a higher correlation than the solvation free energy of the anion $\Delta G_{\mathrm{solv,A^-}}$. In the thermodynamic cycle of Scheme 2 $\Delta G_{\mathrm{solv,A^-}}$ is the dominant term of $\Delta G_{\mathrm{dissoc_g}}$ and these two values are correlated with an r^2 of 0.90, but apparently the additional information contained in the other terms of Scheme 2 most of all the acidity in water, $\Delta G_{\mathrm{dissoc_w}}$ adds to the fit which is an indicator that the fraction of toxicant present as anion plays a role. Therefore, models for

$\mathrm{EC_m}$ with $\mathrm{p}K_{\mathrm{a_m}}$ as an additional descriptor describing the speciation were derived and the correlations of these models with $\mathrm{EC_m}$ are shown in Table 4.5.

Table. 4.5 Addition of the third descriptor $\mathrm{p}K_{\mathrm{a_m}}$ for the effect of speciation. The apparent R^2 and sd (using the full data set) and the equivalent numbers, R^2_{CV} and sd_{CV}, calculated for predictions from a 5-fold cross-validation are given.

Descriptor	R^2	sd	R^2_{CV}	sd_{CV}
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + D^- + $\mathrm{p}K_{\mathrm{a_m}}$	0.64	0.45	0.53	0.5
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{dissoc_g}}$ + $\mathrm{p}K_{\mathrm{a_m}}$	0.80	0.33	0.72	0.4
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{solv,A^-}}$ + $\mathrm{p}K_{\mathrm{a_m}}$	0.82	0.31	0.76	0.36
$\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ + $\Delta G_{\mathrm{solv,A^-}}$ + $\mathrm{p}K_{\mathrm{a_w}}$	0.80	0.33	0.73	0.39

The model using the solvation free energy of the anion, $\Delta G_{\mathrm{solv,A^-}}$, showed the best correlation and the lowest standard deviation of the residuals for both the apparent and also for the cross-validation estimates. One pair of the three descriptor was intercorrelated with an $r^2 > 0.5$ namely $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ and $\mathrm{p}K_{\mathrm{a_m}}$ with an r^2 of 0.72 while the other two pair combinations had intercorrelations below 0.5. The coefficients took the values given in Equation 4.7.

$$\log 1/\mathrm{EC_m} = 0.114(\pm 0.014)\Sigma\mathrm{CASA}^{\mathrm{ortho}} + 0.072(\pm 0.015)\Delta G_{\mathrm{solv,A^-}} - 0.428(\pm 0.073)\mathrm{p}K_{\mathrm{a_m}} + 10.436(\pm 1.039) \quad (4.7)$$

R^2 = 0.82, $F_{3,15}$ = 26.4, n = 21

with all descriptors having two sided p-values < 0.001 and, thus, high significance in the t-test.

Two other effects mentioned in Table 4.3 were tested, namely the shielding of the negative charge using the descriptor $\Sigma\mathrm{ASA}^{\mathrm{ortho}}$ and the mobility of the neutral species using ASA. Both were not significant in the equation and decreased the predictive power estimated with cross-validation. Figure 4.6 shows a plot of predicted versus experimental intrinsic activities.

The largest residuals have 3,4,5-trichlorophenol (0.61) and 2,3,4,5-tetrachlorophenol (0.69) which were predicted too low and 3,4-dichlorophenol (-0.5) whose toxicity was predicted too high. The same three compounds showed the highest deviations in the cross-validation, namely (0.71,0.75 and -0.58).

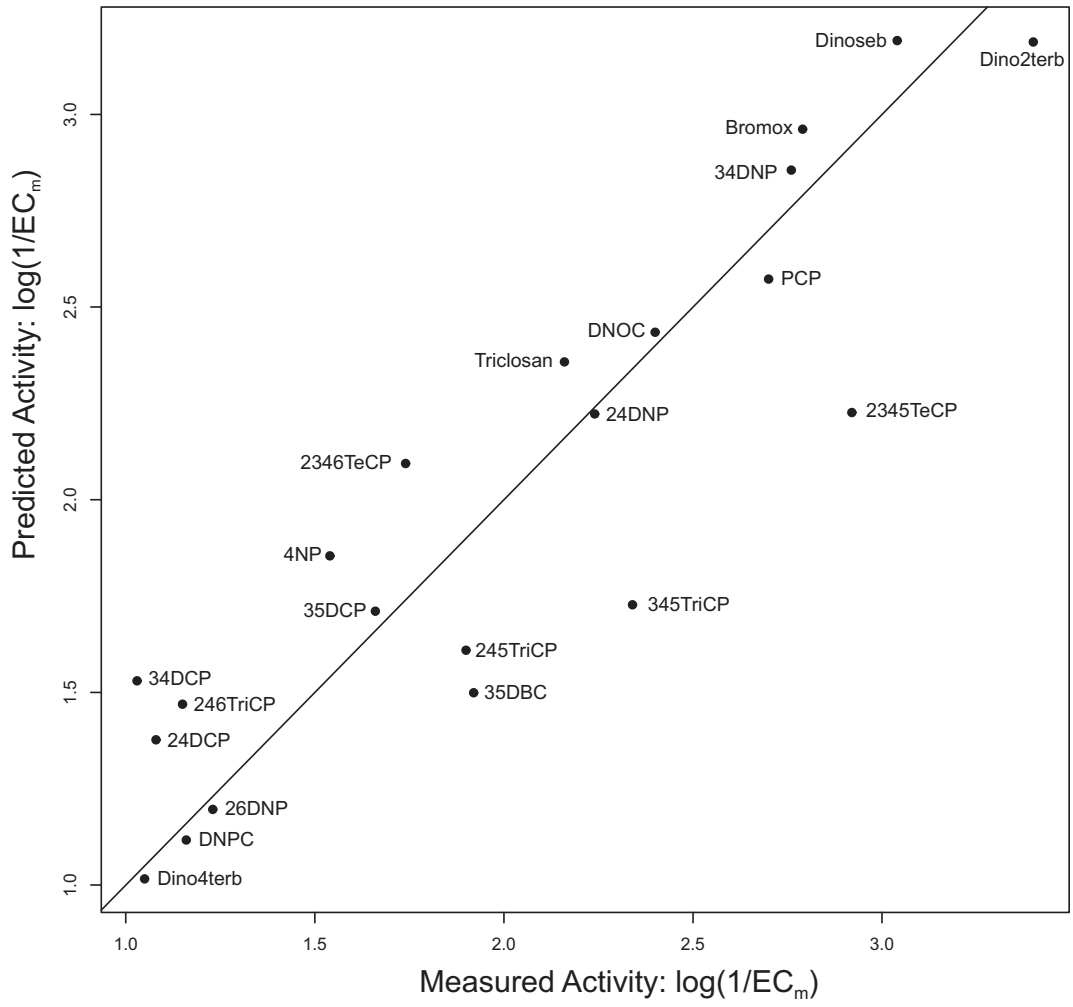


Fig. 4.6 Values predicted with Equation 4.7 versus measured values for the intrinsic uncoupling activity, $\mathrm{EC_m}$. The abbreviations for the names of the compounds are given in Table 4.1.

As $\mathrm{p}K_{\mathrm{a_m}}$, the $\mathrm{p}K_{\mathrm{a}}$ value in the membrane, is a descriptor which currently cannot easily be calculated with QSAR methods, a three variable model with $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$, $\Delta G_{\mathrm{solv,A^-}}$ and $\mathrm{p}K_{\mathrm{a_w}}$ instead of $\mathrm{p}K_{\mathrm{a_m}}$ was also fitted. The fit dropped only slightly with R^2 = 0.80 (0.82 with $\mathrm{p}K_{\mathrm{a_m}}$) as can be seen in Table 4.5.

4.3.2 Aqueous Toxicity, $\mathrm{EC_w}$

The three variable model for intrinsic activity with $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$, $\Delta G_{\mathrm{solv,A^-}}$ and $\mathrm{p}K_{\mathrm{a_m}}$ was combined with different approaches to modeling the uptake with the goal to predict the aqueous effect concentration $\mathrm{EC_w}$. The first approach is the rare ideal case that one has measured data for D_{mw} and can directly apply the rearrangement of Equation 4.2:

$$log1/\mathrm{EC_w} = log1/\mathrm{EC_m} + logD \tag{4.8}$$

The second still rare case is that only the membrane-water distribution of the neutral species $K_{\mathrm{mw,HA}}$ is available. The third case is that measured values of the octanol-water partitioning coefficient $K_{\mathrm{ow,HA}}$ are available and the fourth that only calculated $K_{\mathrm{ow,HA}}$ can be used.

The correlations of the regression models based on the three descriptors used in Equation Equation 4.7 combined with the different descriptors for the uptake are shown in Table 4.6.

Table. 4.6 Different models for aqueous effect concentrations. The apparent R^2 and sd and the equivalent figures calculated with predictions from a 5-fold cross-validation are given.

Descriptor	R^2	sd	R^2_{CV}	sd_{CV}
log D_{mw}	0.94	0.33	0.92	0.40
log $K_{\mathrm{mw,HA}}$	0.88	0.42	0.81	0.54
log $K_{\mathrm{ow,HA}}$-experimental	0.85	0.48	0.76	0.62
log $K_{\mathrm{ow,HA}}$-calculated (XLOGP)	0.76	0.60	0.60	0.81
log $K_{\mathrm{ow,HA}}$-calculated (CLOGP)	0.75	0.62	0.58	0.83

In order to test whether the decrease of predictive power from experimental log $K_{\mathrm{ow,HA}}$ to calculated log $K_{\mathrm{ow,HA}}$ was a specific problem of the atom-based method XLOGP, log $K_{\mathrm{ow,HA}}$ values were also calculated with the fragment-based method CLOGP [45]. With an R^2 dropping slightly to 0.75 and R^2_{CV} = 0.58 the problem could not be resolved.

The pattern of the compounds with large residuals was preserved when Equation 4.8 was used to transform the predictions for $\mathrm{EC_m}$ to $\mathrm{EC_w}$ with the same three compounds (3,4,5-trichlorophenol, 2,3,4,5-tetrachlorophenol and 3,4-dichlorophenol) having residuals above half a log unit. For the regression model using log $K_{\mathrm{mw,HA}}$ again these three compounds were among the top 3 residuals. With transition to models with experimental log $K_{\mathrm{ow,HA}}$ these compounds still caused high residuals but there were also some compounds which had very small residuals before now causing large residuals, namely Triclosan (-0.78,-1.16), Bromox (-0.73,-0.94) and Dino2terb (0.60,0.98). The first value in the parentheses is the prediction with the full data set, the second one is the prediction in the cross-validation with negative values indicating too high predictions and positive too low ones. The model with calculated log $K_{\mathrm{ow,HA}}$ showed a similar trend as the model with experimental log $K_{\mathrm{ow,HA}}$ values, but the residuals became higher.

For the regression models based on experimental log $K_{\mathrm{ow,HA}}$ the significance of the two descriptors $\Delta G_{\mathrm{solv,A^-}}$ and $\mathrm{p}K_{\mathrm{a_m}}$ took quite low values with P-values of 0.01 and 0.04, respectively. In the case of calculated log $K_{\mathrm{ow,HA}}$ the P-values were 0.11 and 0.10 and thus, not significant at 0.05-level. However, removing these variables decreased the R^2_{CV} for both the model with experimental log $K_{\mathrm{ow,HA}}$ and the model with calculated log $K_{\mathrm{ow,HA}}$.

4.4 Discussion

Given the complexity of the different processes taking place in energy-transducing membranes the data set used in this study is quite small. However, the availability of experimental data for both activity and for descriptors helps to overcome the problems and the dangers of small data sets, because it allows one to sequentially replace measured parameters with calculated parameters. The insights gained in experimental studies with Kinspec in the past are also of central importance to determine which descriptors are needed and to keep the number of descriptor candidates low.

Three descriptors for delocalization and two for $\mathrm{AHA^-}$-formation were defined in Table 4.2 which resulted in the six combinations listed in Table 4.4. Apparently, the descriptor $\Sigma\Delta\mathrm{vdWR}^{\mathrm{ortho}}$ limited to geometric aspects cannot fully describe the process. Thus, only combinations with $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ remained. The fact that $\Delta G_{\mathrm{dissoc_g}}$ and $\Delta G_{\mathrm{solv,A^-}}$ resulted in substantially better correlations than the delocalization descriptor D^- shows that the processes that need to be described are not delocalization, as suggested in the central review on the topic [1], but solvation energies particularly of the charged species. The addition of $\mathrm{p}K_{\mathrm{a_m}}$ as a descriptor for the speciation had a great impact on the fit of the models especially for the combination with $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ and $\Delta G_{\mathrm{solv,A^-}}$ where the R^2 rose from 0.47 to 0.82. This can be explained by the fact that for toxicants, which are present in the membrane mainly as anions, the energy required to move through the membrane is less important than for those that exist predominantly as neutral species. This explains also the coefficients in Equation 4.7. The coefficient for $\mathrm{p}K_{\mathrm{a_m}}$ is negative which means that high values of $\mathrm{p}K_{\mathrm{a_m}}$ reduce the activity, i.e., a toxicant whose anion could pass the membrane quite easily based on its $\Delta G_{\mathrm{solv,A^-}}$ value, but has, at pH 7, only a fraction of 1% of $\mathrm{A^-}$ in the membrane, will be less active than one with a fraction of 50% (i.e., a lower $\mathrm{p}K_{\mathrm{a_m}}$).

The coefficients of the other two descriptors are both positive. For $\Delta G_{\mathrm{solv,A^-}}$ with nega-

tive energies ranging from -46 to -70 kcal/mol this means that the lower the solvation energy the more active the compound. For $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$, which takes values ranging from +1.6 to -31.2, the coefficient expresses that ortho-substituents with negative partial charges taking a large surface and thus a large negative value for $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ the activity decreases.

The interpretation of $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ as local descriptor around a reaction center differs from the original charged polar surface area (CPSA) descriptor introduced by Stanton and Jurs [50] who intended to capture information about the features of molecules responsible for polar intermolecular interactions, but later used it for a wide variety of properties as reviewed recently [40]. The idea to use $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ for describing the tendency to form heterodimers was stimulated by the fact that hydrogen bonds to anions are very short. With distances ranging from 1.2 to 1.6 Å and O–O-distances ranging from 2.4 to 2.6 Å [51] there would be significant overlap of the van der Waals-Radii of phenolate and phenol for ortho-substituents like chlorine especially in the case of 2,6-substitution. So the idea behind choosing $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ as a descriptor was that it reflects the potential of the ortho-substituents to "disturb" the formation of heterodimers. The topologic nature of $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ might also explain the consistent pattern of observed residuals that 2,6-chlorosubstituted compounds are predicted too high and compounds with only one chloro-substitutent in ortho-position are predicted too low (cf. Figure 4.6). However, as the data set is very small the interpretation given in this subsection also has to be taken with some caution, because although the choice of this descriptor was motivated by mechanistic considerations it might cover other aspects of these structures like intramolecular hydrogen-bonding in the case of nitro-groups in ortho-position. Concerning the existence of heterodimers there are two experimental studies [52, 53] which favor the theory of hydrogen bonding as illustrated in Figure 4.5 and also confirm their primary occurrence in very hydrophobic environments. Due to the introduction of the concept of heterodimers the proposed model can also reproduce the low values of 2,4,6-trichlorophenol and Dinoseb which were outliers in other studies [7].

With the current set of descriptors the model is still limited to phenols because of the definition of $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$, but it should be possible to extend this descriptor also to other chemical classes known to cause uncoupling like benzimidazoles [54] or salicylanilides [11]. As the other descriptors $\Delta G_{\mathrm{solv,A^-}}$ and $\mathrm{p}K_{\mathrm{a}}$ are independent of the chemical class, this approach might allow a generalization to all weak acids causing uncoupling of oxidative phosphorylation.

A limitation of the model is that no term for the back-diffusion of the neutral species was found yet. This is either due to an inappropriate descriptor choice or to the low number of compounds in this data set for which the back-diffusion is of relevance. The process is important for compounds with pK_{a_m} that are much lower than the pH-value of the system. Note that pK_{a_m} is on average about a log unit higher than pK_{a_w} and therefore in the present data set there are only five compounds with $pK_{a_m} < 5$. As a consequence compounds with very low pK_{a_m} values would be predicted as highly active with the current model. In reality efficient uncouplers require both species to be present in significant concentrations and thus, the linearity of the descriptor pK_{a_m} probably ends at one or two units below the pH of 7 at which the effect concentrations were determined. The problem can, however, be eased if the present model is constrained to compounds with pK_a values known to allow uncoupling. This pK_a-window was observed to lie between 3.5 and 8.7 for Kinspec data which is in excellent agreement with an alternative uncoupling test system based on tobacco pollen growth inhibition [8] where a pK_a-window of 3.8 to 8.5 has been observed.

The backtransformation of EC_m to the aqueous concentration resulted in predictions with a high R^2 of 0.94 for the full data set and one of 0.92 during cross-validation. This is very high, but partially also a consequence that the uptake terms were previously used to define EC_m and thus, do not add any error to the model.

The drop of fit for the second model in Table 4.6 using experimental log $K_{mw,HA}$ shows that it is relevant to know both log $K_{mw,HA}$ and log K_{mw,A^-} to obtain the full picture of the uptake. This is corroborated by the fact that the highest residuals are also compounds which have the largest differences between $K_{mw,HA}$ and K_{mw,A^-}. An additional smaller drop of fit is observed when log $K_{mw,HA}$ is replaced by experimental log $K_{ow,HA}$. The drop of fit is mainly caused for the same reason, i.e., the neglect of differences between $K_{mw,HA}$ and K_{mw,A^-}. Given the enormous differences between log D_{ow} and log D_{mw} shown in Figure 4.4 it is clear why in the past no sucessful models based on log D_{ow} could be established. The error of assuming that only the neutral species is sorbed into the membrane is much higher than the error of setting the uptake equal to log K_{ow}. However, the decrease of the correlation coefficient from an R^2 of 0.94 of the model using log D_{mw} to an R^2 of 0.85, when log K_{ow} is used, must be attributed to this simplification.

The drop from experimental log $K_{ow,HA}$ (R^2 = 0.85) to calculated log $K_{ow,HA}$ (R^2 = 0.76) is remarkable. The comparison of experimental and calculated log $K_{ow,HA}$ values

showed high deviations ranging from -0.51 for 26DNP (predicted too low) to 1.6 units for 35DBC for CLOGP and even more pronounced for XLOGP where deviations ranged from -1.1 for 26DNP to 2.2 for 35DBC. Apparently, predicting log $K_{\mathrm{ow,HA}}$ still is a difficult task for compounds with large π-systems and also for ortho-substituted phenols. The fact that the two descriptors $\Delta G_{\mathrm{solv,A^-}}$ and $\mathrm{p}K_{\mathrm{a_m}}$ were not even significant in the models with calculated log $K_{\mathrm{ow,HA}}$ shows how important the choice of descriptor for lipophilicity is. This work shows also the importance of further QSAR studies on membrane-water partitioning [55, 56] when more data become available.

4.5 Conclusion

The basic approach of this study is the clear separation of uptake and intrinsic activity at the target site. This allowed the derivation of a model for the intrinsic activity based on simple empirical descriptors, which can explain the basic behavior of uncouplers of oxidative phosphorylation. Two new types of descriptors which have not been used for modeling of uncouplers so far have been introduced, namely solvation free energies of anions and an estimate for the formation of heterodimers. The general approach in this study to sequentially replace measured parameters with calculated parameters proved valuable as it facilitates the detection of potential problems and possibilities for improvement. Regarding the uptake of uncouplers into the membrane the use of log D has advantages, but only if it is based on an appropriate organic phase like liposomes. If log D is based on the bulk organic solvent octanol as the organic phase the uptake of the charged species is strongly underestimated. Using log K_{ow} instead of liposomal log D leads to a lower but still acceptable quality of the model. The presented approach is to a large extent independent of the chemical class and, thus, might allow a generalization to all weak acids causing uncoupling of oxidative phosphorylation.

Acknowledgements

The authors would like to thank Said Hilal of the United States EPA for highly valuable discussions and support in making the calculations. We would also like to thank Eric Pellegrini, Dimitar Hristozov and Thomas Kleinöder for their support, all of the Computer-Chemie-Centrum of the University of Erlangen. Simon Spycher would like to thank the Bundesministerium für Bildung und Forschung (BMBF) for partial funding of this work

(Systems Biology, project no. FZJ 0313080 C)

References

[1] H. Terada, "Uncouplers of oxidative phosphorylation", *Environ. Health Perspect.* **1990,** *87,* 213–218.

[2] K. B. Wallace, A. A. Starkov, "Mitochondrial targets of drug toxicity", *Annu. Rev. Pharmacol. Toxicol* **2000,** *40,* 353–388.

[3] S. G. A. McLaughlin, J. P. Dilger, "Transport of protons across membranes by weak acids", *Physiol. Rev.* **1980,** *60,* 825–863.

[4] J. Gasteiger, "A Hierarchy of Structure Representations", in "Handbook of Chemoinformatics", J. Gasteiger (Ed.), Wiley-VCH, **2003**, Volume 3, 1034–1061.

[5] C. Hansch, K. Kiehs, G. L. Lawrence, "The role of substituents in the hydrophobic bonding of phenols by serum and mitochondrial proteins", *J. Am. Chem. Soc.* **1965,** *87,* 5770–5773.

[6] H. Miyoshi, T. Fujita, "Quantitative analyses of the uncoupling activity of substituted phenols with mitochondria from flight muscles of house flies", *Biochim. Biophys. Acta* **1988,** *935,* 312–321.

[7] H. Miyoshi, H. Tsujishita, N. Tokutake, T. Fujita, "Quantitative analysis of uncoupling activity of substituted phenols with a physicochemical substituent and molecular parameters", *Biochim. Biophys. Acta* **1990,** *1016,* 99–106.

[8] G. Schüürmann, R. K. Somashekar, U. Kristen, "Structure-activity relationships for chloro- and nitrophenol toxicity in the pollen tube growth test", *Environ. Toxicol. Chem.* **1996,** *15,* 1702–1708.

[9] E. Argese, C. Bettiol, G. Giurin, P. Miana, "Quantitative structure-activity relationships for the toxicity of chlorophenols to mammalian submitochondrial particles", *Chemosphere* **1999,** *38,* 2281–2292.

[10] Z. J. Guo, H. Miyoshi, T. Komyoji, T. Haga, T. Fujita, "Quantitative analysis with physicochemical substituent and molecular parameters of uncoupling activity of substituted diarylamines", *Biochim. Biophys. Acta* **1991,** *1059,* 91–98.

[11] H. Terada, S. Goto, K. Yamamoto, I. Takeuchi, Y. Hamada, K. Miyake, "Structural requirements of salicylanilides for uncoupling activity in mitochondria: quantitative analysis of structure-uncoupling relationships", *Biochim. Biophys. Acta* **1988,** *936,* 504–512.

[12] D. M. Gange, S. Donovan, R. J. Lopata, K. Henegar, "The QSAR of insecticidal uncouplers", *ACS Symp. Ser. – Classical and Three-Dimensional QSAR in Agrochemistry* **1995,** *606,* 199–212.

[13] G. Schüürmann, "Ecotoxic modes of action of chemical substances", in "Ecotoxicology", John Wiley and Spektrum Akademischer Verlag, New York, **1998**, Ecotoxicology, 665–749.

[14] M. T. D. Cronin, Y. H. Zhao, R. L. Yu, "pH-dependence and QSAR analysis of the toxicity of phenols and anilines to Daphnia magna", *Environ. Toxicol.* **2000,** *15,* 140–148.

[15] P. Mitchell, "Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism", *Nature* **1961,** *191,* 144–148.

[16] B. I. Escher, M. Snozzi, K. Häberli, R. P. Schwarzenbach, "A new method for simultaneous quantification of uncoupling and inhibitory activity of organic pollutants in energy-transducing membranes", *Environ. Toxicol. Chem.* **1997,** *16,* 405–414.

[17] B. I. Escher, M. Snozzi, R. P. Schwarzenbach, "Uptake, Speciation, and Uncoupling Activity of Substituted Phenols in Energy Transducing Membranes", *Environ. Sci. Technol.* **1996,** *30,* 3071–3079.

[18] B. I. Escher, R. I. L. Eggen, U. Schreiber, Z. Schreiber, E. Vye, B. Wisner, R. P. Schwarzenbach, "Baseline Toxicity (Narcosis) of Organic Chemicals Determined by In Vitro Membrane Potential Measurements in Energy-Transducing Membranes", *Environ. Sci. Technol.* **2002,** *36,* 1971–1979.

[19] J. Kasianowicz, R. Benz, S. McLaughlin, “The kinetic mechanism by which CCCP (carbonyl cyanide m-chlorophenylhydrazone) transports protons across membranes”, *J. Membrane Biol.* **1984,** *82,* 179–190.

[20] J. G. Lessard, M. Fragata, “Micropolarities of lipid bilayers and micelles. 3. Effect of monovalent ions on the dielectric constant of the water-membrane interface of unilamellar phosphatidylcholine vesicles”, *J. Phys. Chem.* **1986,** *90,* 811–817.

[21] B. I. Escher, L. Sigg, “Chemical speciation of organics and of metals at biological interphases”, in “Physicochemical Kinetics and Transport at Biointerfaces”, J. Buffle, H. P. van Leuwen (Eds.), John Wiley, Chichester, **2004**, Volume 9 of *IUPAC Series on Analytical and Physical Chemistry of Environmental Systems*, 205–269.

[22] C. T. Jafvert, J. C. Westall, E. Grieder, R. P. Schwarzenbach, “Distribution of hydrophobic ionogenic organic compounds between octanol and water: organic acids”, *Environ. Sci. Technol.* **1990,** *24,* 1795–1803.

[23] P. Smejtek, S. Wang, “Distribution of hydrophobic ionizable xenobiotics between water and lipid membranes: pentachlorophenol and pentachlorophenate. A comparison with octanol-water partition”, *Arch. Environ. Contam. Toxicol* **1993,** *25,* 394–404.

[24] B. I. Escher, R. P. Schwarzenbach, “Partitioning of Substituted Phenols in Liposome-Water, Biomembrane-Water, and Octanol-Water Systems”, *Environ. Sci. Technol.* **1996,** *30,* 260–270.

[25] A. Avdeef, K. J. Box, J. E. A. Comer, C. Hibbert, K. Y. Tam, “pH-metric logP 10. Determination of liposomal membrane-water partition coefficients of ionizable drugs”, *Pharm. Res.* **1998,** *15,* 209–215.

[26] A. Finkelstein, “Weak-acid uncouplers of oxidative phosphorylation. Mechanism of action on thin lipid membranes”, *Biochim. Biophys. Acta* **1970,** *205,* 1–6.

[27] B. I. Escher, R. Hunziker, R. P. Schwarzenbach, J. C. Westall, “Kinetic Model To Describe the Intrinsic Uncoupling Activity of Substituted Phenols in Energy Transducing Membranes”, *Environ. Sci. Technol.* **1999,** *33,* 560–570.

[28] B. I. Escher, R. P. Schwarzenbach, “Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms”, *Aquat. Sci.* **2002,** *64,* 20–35.

[29] R. Hunziker, B. Escher, “Unpublished Measurements, EAWAG, Switzerland”, , **1999**.

[30] “PETRA - Parameter Estimation for the Treatment of Reactivity Applications”, Version 3.1, MolNet GmbH, Erlangen, **2002**.
`http://www.mol-net.com`
`http://www2.chemie.uni-erlangen.de/services/petra/index.html`

[31] A. Fröhlich, “Modelle zur Beschreibung elektronischer Effekte in π-Systemen und ihr Einsatz in der Modellierung der elektrophilen aromatischen Substitution”, PhD-Thesis, Technische Universität München, **1993**.

[32] S. H. Hilal, S. W. Karickhoff, L. A. Carreira, “SPARC - Performs Automated Reasoning in Chemistry”, University of Georgia, Athens, GA.
`http://ibmlc2.chem.uga.edu/sparc/`

[33] S. H. Hilal, S. W. Karickhoff, L. A. Carreira, “Prediction of Chemical Reactivity Parameters and Physical Properties of Organic Compounds from Molecular Structure using SPARC”, Report: EPA/600/R-03/030, **2003**.
`http://www.epa.gov/athens/publications/reports/EPA_600_R03_030.pdf`

[34] P. Labute, “A widely applicable set of descriptors”, *J. Mol. Graphics Mod.* **2000,** *18,* 464–477.

[35] D. P. Hristozov, “ASA”, Version 1.0, Calculator of in-house chemoinformatics library, **2005**.

[36] A. Herwig, K. Thomas, J. Marusczyck, L. Terfloth, J. Gasteiger, “MOSES - Molecular Structure Encoding System”, In-house chemoinformatics library, **2002**.

[37] A. Herwig, “Development of an integrated framework for chemoinformatics applications”, PhD-Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, **2004**.

[38] "CORINA", Version 2.4, MolNet GmbH, Erlangen.
`http://www.mol-net.com`
`http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html`

[39] P. C. Jurs, "Quantitative Structure-Property Relationships", in "Handbook of Chemoinformatics", J. Gasteiger (Ed.), Wiley-VCH Verlag, Weinheim, **2003**, Volume 3, 1314–1335.

[40] D. T. Stanton, S. Dimitrov, V. Grancharov, O. G. Mekenyan, "Charged partial surface area (CPSA) descriptors QSAR applications", *SAR QSAR Environ. Res.* **2002,** *13,* 341–351.

[41] J. Gasteiger, H. Saller, "Calculation of the charge distribution in conjugated systems by quantification of the mesomerism concept", *Angew. Chem.* **1985,** *97,* 699–701.

[42] J. Gasteiger, M. Marsili, "Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges", *Tetrahedron* **1980,** *36,* 3219–3222.

[43] B. Schmidt, "XLOGP", Calculator of in-house chemoinformatics library, **2004**.

[44] R. Wang, Y. Gao, L. Lai, "Calculating partition coefficient by atom-additive method", *Perspect. Drug Discovery Des.* **2000,** *19,* 47–66.

[45] "KowWin", Syracuse Research Corporation, Syracuse, NY, **2005**.
`http://www.syrres.com/esc/kowdemo.htm`

[46] W. M. Meylan, P. H. Howard, "Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients", *J. Pharm. Sci.* **1995,** *84,* 83–92.

[47] D. M. Hawkins, "The problem of overfitting", *J. Chem. Inf. Comput. Sci.* **2004,** *44,* 1–12.

[48] S. Spycher, M. Nendza, J. Gasteiger, "Comparison of different classification methods applied to a mode of toxic action data set", *QSAR Comb. Sci.* **2004,** *23,* 779–791.

[49] "R: A language and environment for statistical computing", Version 2.0, R Foundation for Statistical Computing, Vienna, Austria, **2003**.
`http://www.r-project.org`

[50] D. T. Stanton, P. C. Jurs, "Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies", *Anal. Chem.* **1990,** *62,* 2323–2329.

[51] G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford, **1997**.

[52] A. W. Barstad, D. H. Peyton, P. Smejtek, "AHA- heterodimer of a class-2 uncoupler: pentachlorophenol", *Biochim. Biophys. Acta* **1993,** *1140,* 262–270.

[53] M. Siam, G. Reiter, R. Hunziker, B. Escher, A. Karpfen, A. Simperler, D. Baurecht, U. P. Fringeli, "Evidence for heterodimers of 2,4,5-trichlorophenol on planar lipid layers. A FTIR-ATR investigation", *Biochim. Biophys. Acta* **2004,** *1664,* 88–99.

[54] J. Ilivicky, J. E. Casida, "Uncoupling action of 2,4-dinitrophenols, 2-trifluoromethylbenzimidazoles, and certain other pesticide chemicals upon mitochondria from different sources and its relation to toxicity", *Biochem. Pharmacol.* **1969,** *18,* 1389–1401.

[55] W. H. J. Vaes, E. U. Ramos, C. Hamwijk, I. van Holsteijn, B. J. Blaauboer, W. Seinen, H. J. M. Verhaar, J. L. M. Hermens, "Solid Phase Microextraction as a Tool To Determine Membrane/Water Partition Coefficients and Bioavailable Concentrations in Vitro Systems", *Chem. Res. Toxicol.* **1997,** *10,* 1067–1072.

[56] H. Patel, T. W. Schultz, M. T. D. Cronin, "Physico-chemical interpretation and prediction of the dimyristoyl phosphatidyl choline-water partition coefficient", *THEOCHEM* **2002,** *593,* 9–18.

4.6 Further Discussion of Toxicological Aspects and Mechanistic Descriptors

The study presented in this chapter introduced new descriptors which, to the knowledge of the authors, have not been used yet in QSAR studies. Both the solvation free energies of anions and the estimate for the formation of heterodimers fit well into the current picture of the uncouplers acting by the proton shuttle mechanism. However, with a small data set of 21 compounds there is always a danger that one has found descriptors which merely reflect the composition of the data set and not of the underlying mechanisms.

Therefore, in a future study alternative methods to calculate these descriptors should be tested. The solvation free energy can be calculated using implicit solvent models such as the finite difference Poisson-Boltzmann method (FDPB-method) [57], the generalized born model (GBSA) [58], or the conductor-like screening model for real solvents COSMO-RS [59,60]. The FDPB-method can also be used with empirical or semi-empirical charges which would allow the rapid screening of large data bases. If such alternatively calculated descriptors result in similar correlations, it would present a strong confirmation of the model.

An additional new aspect for a QSAR model of uncoupling is the use of log D. The crucial point for the successful use of log D was the use of experimental membrane-water distributions, $K_{\mathrm{mw,HA}}$ and $K_{\mathrm{mw,A^-}}$ for the neutral and the charged species, respectively.

However, this is also a drawback for future predictions as there are very few QSAR models to predict $K_{\mathrm{mw,HA}}$ and none to predict $K_{\mathrm{mw,A^-}}$. If $K_{\mathrm{mw,HA}}$ is substituted with predicted octanol-water distributions coefficients from well-established programs like CLOGP the quality of the model drops significantly as can be seen in Table 4.6. However, Analytical companies like Sirius, Inc. have compiled large databases on liposome-water distribution coefficients and hopefully, they will be available for the QSAR community in the future.

Concerning the data analysis aspect of QSAR, the model presented in this chapter is very simple as it is based on ordinary least squares regression. For a data set of only 21 compounds it makes sense to use a simple method, as neural networks show their advantages rather with large data sets.

As mentioned in the discussion, the model is still limited to phenols because of the definition of the steric descriptor $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$. Thus, to come back to the three structures shown in Figure 1.2 of the Introduction of this thesis the uncoupler FCCP can not (yet) be predicted.

However, as the other two descriptors pK_{a} and $\Delta G_{\mathrm{solv,A^-}}$ are independent of the chemical class it should be feasible to extend this model to other chemical classes with uncoupling activity.

The other challenge illustrated by Figure 1.2 was whether the uncoupler 2,3,4,5-tetrachlorophenol can be successfully discriminated from the 2,6-dichlorophenol which shows no uncoupling activity. This issue will be addressed in the next chapter using the insights gained in chapter 4.

Additional References:

[57] M. K. Gilson, B. Honig, "Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis", *Proteins* **1988,** *4,* 7–18.

[58] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics", *J. Am. Chem. Soc.* **1990,** *112,* 6127–6129.

[59] A. Klamt, G. Schüürmann, "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient", *J. Chem. Soc. Perkin Trans.* **1993,** *2,* 799–805.

[60] A. Klamt, F. Eckert, M. Hornig, "COSMO-RS: a novel view to physiological solvation and partition questions", *J. Comput. Aided Mol. Des.* **2001,** *15,* 355–365.

Chapter 5

Classification Study for the Uncoupling Activity Model

In Chapter 4 a regression model for the prediction of uncoupling activity was derived. However, the prerequisite for using this model for the screening of data bases is a classification method which can pre-filter potential candidates for uncoupling activity from compounds with little to no likelihood to be active. Therefore, a simple and pragmatic approach for the classification of uncouplers will be presented here. In section 5.2 the descriptors used for the regression model and for the classification model will be tested with an external test set.

5.1 Derivation of a classification method for uncoupling activity

5.1.1 Toxic Mechanism

In section 4.2.1 the factors determining the activity of uncouplers of oxidative and photophosphorylation were introduced. The relevant question for a classification study is which molecular features allow a molecule to be an uncoupler at all and which ones result in the complete loss of uncoupling activity. These features are not necessarily the same as those for the quantitative model, e.g., a descriptor for lipophilicity like log K_{ow} is central for developing a quantitative model for the activity, but does not necessarily help to develop a classification model for the discrimination of uncouplers from other toxicants.

The central common feature of uncouplers acting by the proton shuttle mechanism is first

and foremost to have an acidic group. The replacement of the acidic group by a non-acidic group results in a complete loss of uncoupling activity [1]. Additionally, it is known from previous experimental studies that there are upper and lower limits for the pK_{a}-values of uncouplers. This pK_{a}-window was observed to lie between pK_{a}-values of 3.8 to 8.5 [2]. Thus, as a first filter rule, taking only compounds which are within this pK_{a}-window would already filter large portions of most data bases.

However, some weak acids with pK_{a}-values within the range of 3.8 to 8.5 are known not to show uncoupling activity, e.g. some aliphatic carboxylic acids were measured with the *in vitro* assay Kinspec without any of them showing uncoupling activity [3]. Therefore, the pK_{a}-filter rule needs to be extended with other criteria.

5.1.2 Data

The data set with 21 active compounds used to derive the regression model was extended with 18 nonactive compounds shown in Appendix B (compounds 32-49). These compounds were measured with Kinspec, but showed no activity albeit being structurally similar to known uncouplers.

Additionally, four catechols determined to be uncouplers with the Kinspec were added [4]. They could not be used for the quantitative study of Chapter 4 as the catechols seem to degrade rather fast [5] which would distort the value of the effect concentrations, but they could be used for this qualitative study (compounds 22-25 in Appendix B). An additional six active uncouplers taken from literature were added (compounds 26–31). These six compounds are all experimentally well-studied uncouplers and were included in order to increase the diversity of the active compounds and consisted of the two carbonylcyanid phenylhydrazones (FCCP and CCCP) [6], one benzimidazoles (TTFB) [6], one phenylpyridineamine (Fluazinam) [7], one salicylanilide (S-13) [8] and SF 6847, i.e., 2,6-di-t-butyl-4-(2,2-dicyanoovinyl)phenol, which is the most active uncoupler known [1]. The first five of these six additional uncouplers have a remarkable feature: their acidic group is an NH-group. More than half of all publications on uncoupling activity consist of compounds with acidic NH-groups, hence these examples are of relevance [7, 9–11].

5.1.3 Results

Three descriptors were tested with the goal to refine the pK_{a}-filter rule: the solvation free energy of the anion, $\Delta G_{\mathrm{solv,A^-}}$, the membrane-water distribution coefficient of the neutral species, log $K_{\mathrm{mw,HA}}$, and the octanol-water distribution coefficient of the neutral species, log K_{ow}, as an easy to calculate alternative. For $\Delta G_{\mathrm{solv,A^-}}$ values calculated by SPARC [12] and for the other two descriptors measured values were taken. The values of the descriptors are given in Table B.1 in Appendix B.

Figure 5.1 shows the position of actives and nonactives with regard to pK_{a} and log $K_{\mathrm{mw,HA}}$.

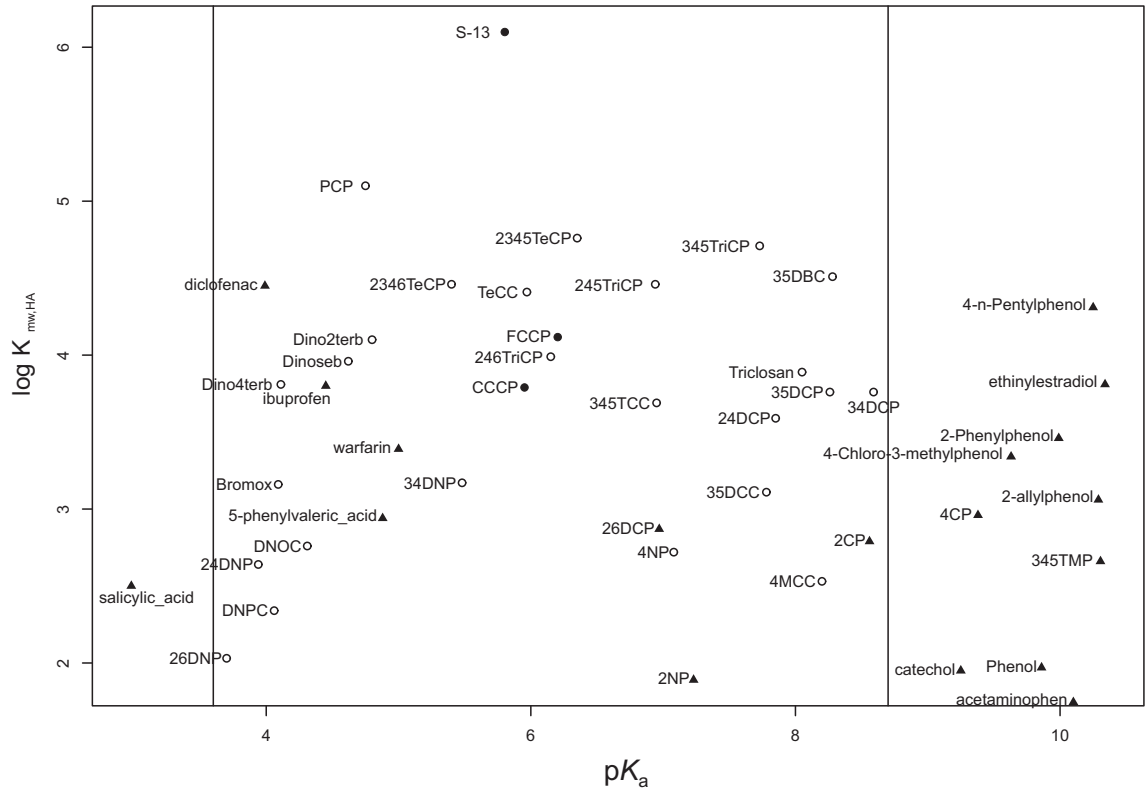


Fig. 5.1 Position of compounds on the two axes pK_{a} and log $K_{\mathrm{mw,HA}}$. Uncouplers represented with a ○ (phenols) or ● (acidic NH-groups) and other MOAs with ▲. The vertical lines indicate the currently known pK_{a}-window of active uncouplers.

This simple plot first shows the central importance of pK_{a} as a means for separating actives from nonactives. The lowest pK_{a} of an active compound was 3.7 and the highest 8.6. Almost the same values were determined in an earlier study with a test system based on tobacco pollen growth [2]. Ten of the 18 nonactives clearly fall outside of the pK_{a}-window,

which in Figure 5.1 is marked by a lower pK_{a}-value of 3.5 and an upper limit of 8.8. Secondly, Figure 5.1 shows that lipophilicity described by measured $K_{\mathrm{mw,HA}}$-values, although an indispensable descriptor for quantitative models of uncouplers, does not contribute at all to the classification of uncouplers. The lipophilicity of uncouplers extends over more than four orders of magnitude with $K_{\mathrm{mw,HA}}$-values ranging from 2.0 (2,6-dinitrophenol) to 6.1 (S-13). A similar picture was obtained with the descriptor log K_{ow}, which also cannot help to separate actives from nonactives either.

Figure 5.2 shows the same type of plot with pK_{a} and the solvation free energy of the anions $\Delta G_{\mathrm{solv,A^-}}$.

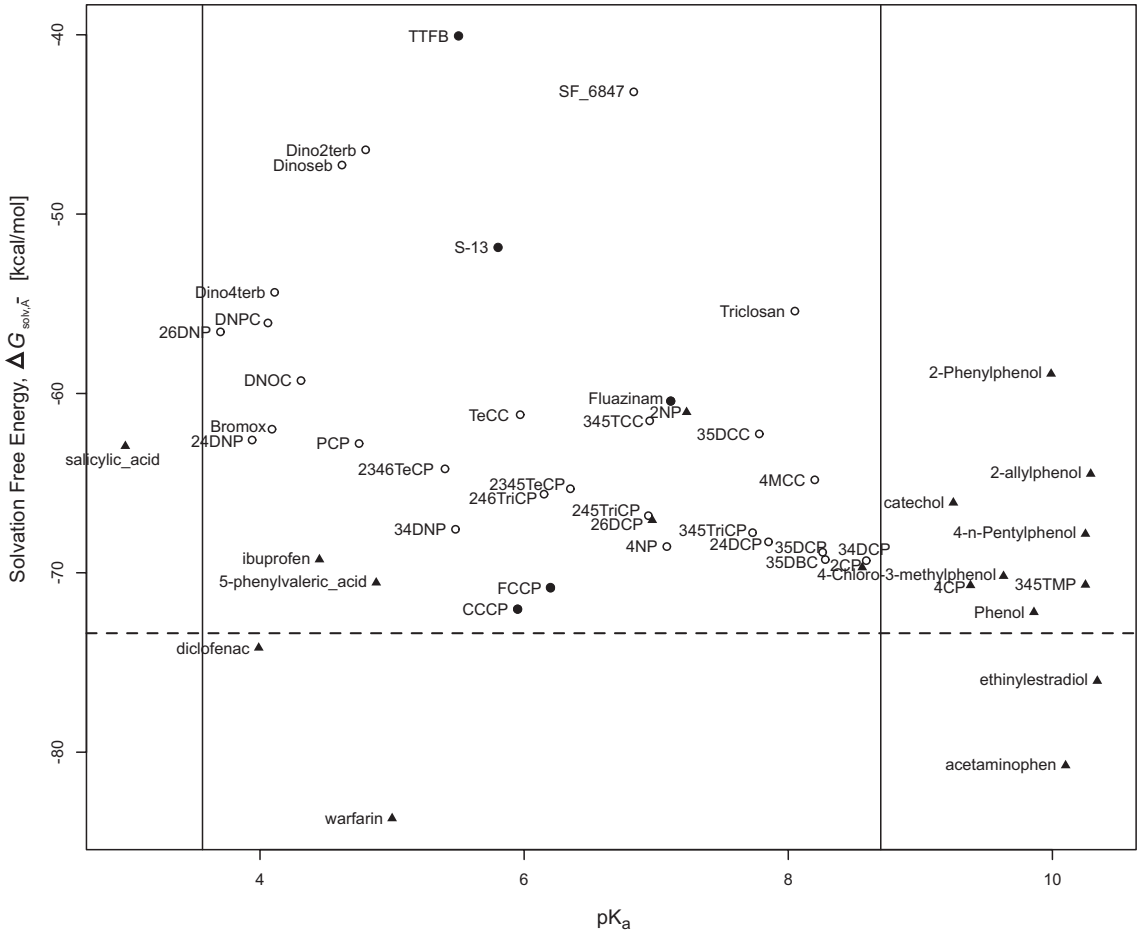


Fig. 5.2 Position of compounds on the two axes pK_{a} and $\Delta G_{\mathrm{solv,A^-}}$. Uncouplers represented with a ○ (phenols) or ● (acidic NH-groups) and other MOAs with a ▲.

The descriptor $\Delta G_{\mathrm{solv,A^-}}$ partially helps to separate the nonactive from active compounds. The nonactive carboxylic acid diclofenac and especially the nonactive warfarin seem to have much larger solvation free energies then all active compounds. In the quantita-

tive model developed in Chapter 4 (Equation 4.7) the coefficient for $\Delta G_{\mathrm{solv,A^-}}$ was positive. The values for $\Delta G_{\mathrm{solv,A^-}}$ were given as negative values. This means that compounds with very large (negative) values for $\Delta G_{\mathrm{solv,A^-}}$ would have a very low activity or, as in the case of diclofenac and warfarin, be inactive. Apparently, these carboxylic acids lack the ability to delocalize the negative charge to the same extent as the known phenolic uncouplers. The consequence is that the free energy of solvation of the nonactive compounds is too large to allow them to move efficiently enough through energy-transducing membranes and hence they are not active.

Thus, the second filter rule would be that $\Delta G_{\mathrm{solv,A^-}}$ should not be too strongly negative. With only two "interesting" nonactives, i.e. nonactives within the pK_{a}-window, it is not possible to determine precisely at which level $\Delta G_{\mathrm{solv,A^-}}$ gets too negative, but some value between –70 and –75 kcal/mol indicated by the dashed line in Figure 5.2 appears reasonable. However, there is a strong need to measure additional nonactives within the pK_{a}-window of uncoupler. With only two compounds with too strongly negative $\Delta G_{\mathrm{solv,A^-}}$ it would not make sense to define an exact limit.

The five uncouplers with an acidic NH-group taken from literature had $\Delta G_{\mathrm{solv,A^-}}$-values comparable to the phenolic uncouplers. This is an indicator that it is potentially possible to develop a QSAR equation for uncouplers belonging to completely different chemical classes. However, the strongly negative values of $\Delta G_{\mathrm{solv,A^-}}$ of FCCP and CCCP are unexpected. These two compounds are very potent uncouplers and therefore should have much less negative $\Delta G_{\mathrm{solv,A^-}}$-values.

Five nonactives had similar $\Delta G_{\mathrm{solv,A^-}}$-values than active uncouplers and thus would be wrongly classified by the proposed classification filter: ibuprofen, 5-phenylvaleric acid, 2-nitrophenol, 2,6-dichlorophenol and 2-chlorophenol. The relatively less negative $\Delta G_{\mathrm{solv,A^-}}$ values of the two nonactives ibuprofen and 5-phenylvaleric acid are conflicts require further investigation. The other three compounds, 2-nitrophenol, 2,6-dichlorophenol and 2-chlorophenol, can be classified correctly by applying a third rule.

In order to discriminate them from active uncouplers, they were predicted using the regression model of Equation 4.7 of the previous chapter. If their predicted activities are very low this would mean that they potentially are uncouplers, but their activity is so low that it does not differ from the baseline toxicity (called narcosis by other authors) which each organic compound exhibits by default. Figure 5.3 shows the experimental and the predicted

aqueous activity of the three compounds 2,6-dichlorophenol (26DCP), 2-nitrophenol (2NP) and 2-chlorophenol (2CP).

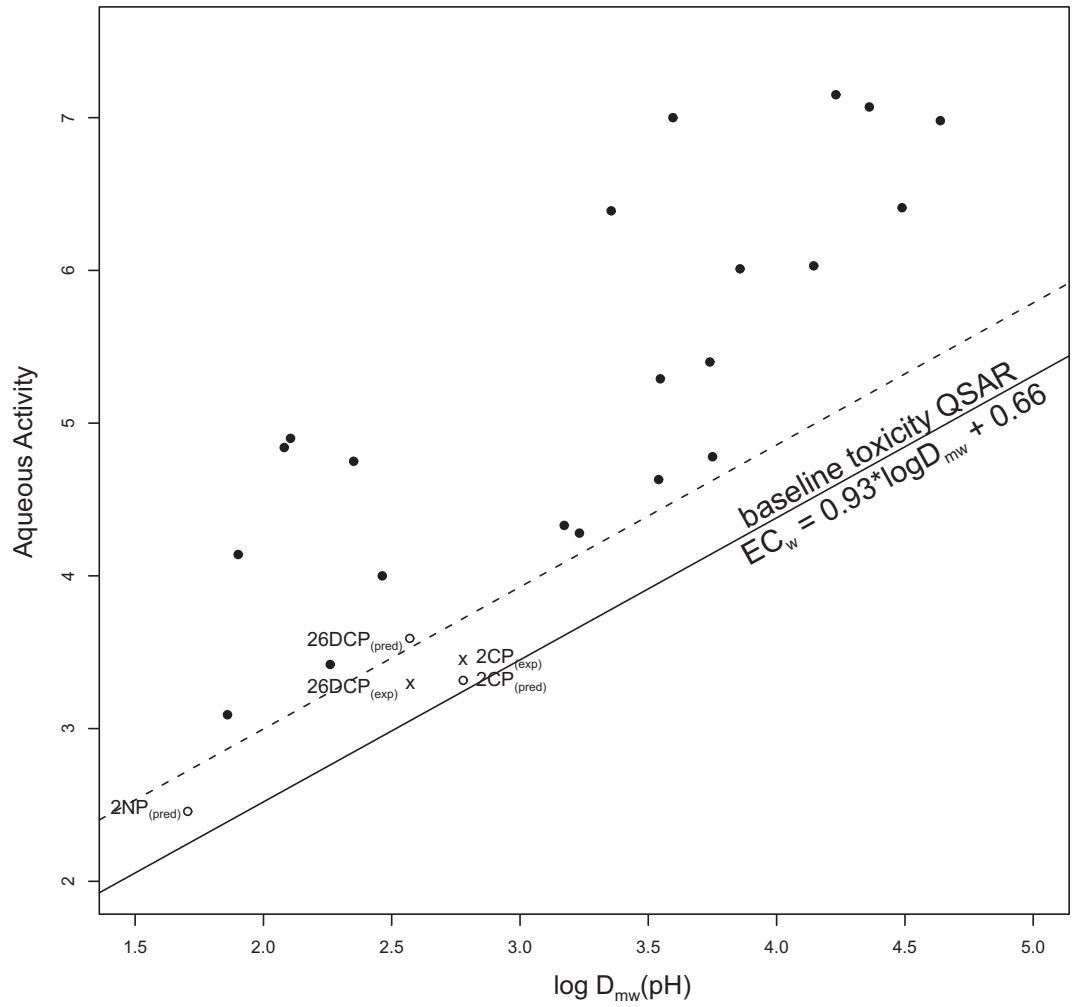


Fig. 5.3 Predicted uncoupling activity of three known nonactive compounds (○) and measured baseline toxicity (×) (the experimental value for 2-nitrophenol is missing). The straight line indicates the baseline toxicity of organic compounds which in Kinspec can be expressed as a function of log D_{mw} [13]. The experimental uncoupling activity of the 21 uncouplers (●) used to derive the regression model is plotted for comparison.

The dashed line in Figure 3 indicates the level of activity above the baseline which in Kinspec is considered as uncoupling. Like most other test systems, Kinspec cannot distinguish between nonspecific membrane disturbances, i.e. baseline toxicity, and the more specific MOA of uncoupling. Therefore, in the experimental MOA determination method compounds were defined to have uncoupling activity if their activity is three times more

active than baseline toxicity [13]. Thus, the third rule emulates this experimental MOA assignment. For the three compounds predicted, only 2,6-dichlorophenol would be wrongly assigned as an uncoupler, however as a very weak one.

To summarize, the proposed classification filter would consist of three steps:

1. Exclusion of compounds not falling into a pK_{a}-window of 3.2 to 9.1 (adding a safety margin of 0.5 log units to the minimal and maximal uncoupler-pK_{a}-value observed so far)

2. Exclusion of compounds with too strongly negative $\Delta G_{\mathrm{solv,A^-}}$

3. Prediction of the remaining compounds with the regression model and comparison of predicted uncoupling activity with predicted baseline activity. Compounds whose uncoupling activity is not predicted three times higher than their baseline toxicity are not considered as active with regards to the MOA of uncoupling.

The problem remains to understand the conflicts caused by the $\Delta G_{\mathrm{solv,A^-}}$-values of ibuprofen and 5-phenylvaleric acid and the unexpectedly large $\Delta G_{\mathrm{solv,A^-}}$-values of FCCP and CCCP. Therefore, in section 5.2 the quality of $\Delta G_{\mathrm{solv,A^-}}$ will be investigated with external data.

5.2 Evaluation of Descriptor Quality

One of the valuable features of the quantitative model proposed in Chapter 4 is that it is possible to study the influence of the descriptor quality as has been done by the comparison of experimental and calculated logK_{ow}-values. In Equation 4.7 of Chapter 4, the standard errors, σ, of the regression coefficients, β, for the intrinsic activity is given. In order rank the descriptors according to the largest uncertainty to the model, the mean value of each descriptor was multiplied with the regression coefficient $\beta \pm \sigma$ which gives the maximum and minimum contribution of each descriptor to the regression model. The difference between the thusly obtained maximum and minimum was taken as an indicator for the uncertainty of the descriptor estimate and took the following values: $\Sigma\mathrm{CASA}^{\mathrm{ortho}}$ (0.24), $\Delta G_{\mathrm{solv,A^-}}$ (1.88) and pK_a^m (1.00). Therefore, evaluating the quality of the $\Delta G_{\mathrm{solv,A^-}}$-calculations has the highest importance. However, only few experimental data for $\Delta G_{\mathrm{solv,A^-}}$ of anions were found. Therefore, data for the free energy of dissociation in the gas phase (gas phase acidity),

$\Delta G_{\mathrm{dissoc_g}}$, were searched. As the method used to predict $\Delta G_{\mathrm{dissoc_g}}$ and the method used to predict $\Delta G_{\mathrm{solv,A^-}}$ are calculated using the same concepts, [14] this seems to be a valid approach.

5.2.1 Gas Phase Acidity

The NIST ionization energy data base was searched for gas phase acidity data on phenols. The data base currently contains data on the free energy of dissociation in gas phase, $\Delta G_{\mathrm{dissoc_g}}$, for over 1000 compounds which is a highly valuable source to test the quality of the calculation methods used. About 45 $\Delta G_{\mathrm{dissoc_g}}$-entries for phenols were found. As a large part of known uncouplers of oxidative phosphorylation are compounds with acidic NH-groups a literature search for such compounds was made. Five $\Delta G_{\mathrm{dissoc_g}}$-values for compounds with NH-acidity were found [15]. The correlation between the gas phase acidities predicted with SPARC [12] and the experimental data of NIST and from literature, respectively, allow one to judge how much error is introduced by the estimation method.

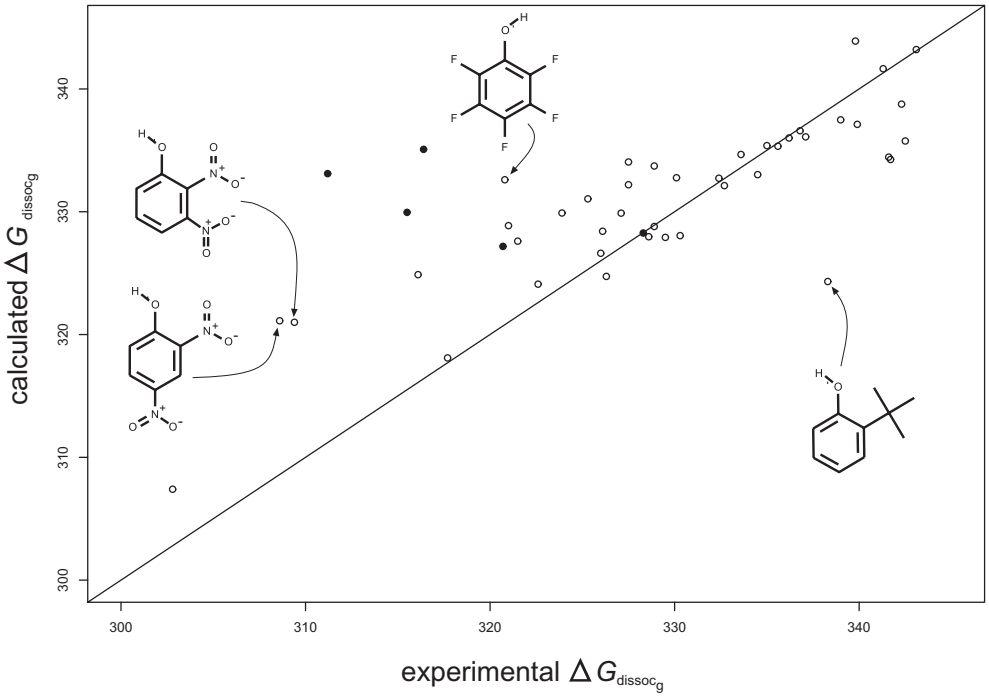


Fig. 5.4 Predicted versus experimental $\Delta G_{\mathrm{dissoc_g}}$ from the NIST ionization data base (given as positive values). Compounds with acidic NH-group are indicated as • and phenols as ○. The structure of the four phenols with deviations large than 10 kcal/mol are drawn on the plot.

The correlation between the experimental values and the predicted values for the phenols

was r^2 = 0.72 and the standard deviation of the differences was surprisingly large with 5.27 kcal/mol which corresponds approximately to four pKa-units. For the five compounds with NH-acidity there was no correlation between experimental and calculated values and a large standard deviation of 9.0 kcal/mol. Thus, the use of $\Delta G_{\mathrm{solv,A^-}}$ which is strongly related to $\Delta G_{\mathrm{dissoc_g}}$ brings a considerable amount of error into the model which explains the large standard error,σ, of the regression coefficient for $\Delta G_{\mathrm{solv,A^-}}$ in Equation 4.7. The uncertainty of the predictions might also be the reason for the misclassification of ibuprofen and 5-phenylvaleric acid in the classification rule presented above.

5.3 Discussion

The proposed classification filter is rather crude, but has the advantage of being pragmatic and transparent.

However, the classification data set composed of all *in vitro* data yet measured with the Kinspec system is still limited. The majority of the 18 nonactives can be separated from the actives by a single descriptor, namely the pK_{a}-value. Therefore, there is a strong need for further experimental data of potential uncouplers, i.e., with a pK_{a} value between 3.2 and 9.1, which show no activity. Only if the data base is enlarged it makes sense to work with more refined statistical or neural network methods for classification. With an enlarged database it would also be possible to obtain a clearer picture at which level the solvation free energy of the anion, $\Delta G_{\mathrm{solv,A^-}}$, becomes too negative for a compound to be active as an uncoupler.

Concerning the uncoupling activity of carboxylic acids there also seems to be a disagreement among different experimental methods. Masabuchi *et al.* studied several anti-inflammatory drugs using rat liver mitochondria and found several carboxylic acids with uncoupling activity [16]. One compound, diclofenac, was also used in this study and no activity was observed in the Kinspec-system. Considering that mitochondria are very complex organelles with numerous possible sites of attack it might be that carboxylic acids do not act acting by the proton shuttle mechanism, but interfere with some other mitochondrial target.

Concerning the chemical domain of toxicity SAR and QSAR, the evaluation of the descriptor quality also showed that alternative $\Delta G_{\mathrm{solv,A^-}}$ calculation methods are necessary. First of all, because this descriptor adds the largest error to the quantitative model and might also be responsible for the misclassifications in the classification model. Secondly, because

no predictive power was observed for compounds with acidic NH-groups. Apparently, predicting $\Delta G_{\mathrm{dissoc_g}}$ and $\Delta G_{\mathrm{solv,A^-}}$ with empirical methods is very difficult and only few data are available. However, if the present model should be extended to other chemical classes, compounds with acidic NH-groups need to be predicted with sufficient quality, as this type of compound is very frequent among the currently known uncouplers of oxidative and photophosphorylation.

The fact that the models of chapter 4 and 5 are based on data which consist not only of measured biological activity but partially also of measured descriptors, is remarkable. With other models based on descriptors, which cannot be directly compared to experimental data, it is much more difficult to judge a model's strength and weaknesses. Thus, this can be seen as a general conclusion from the Chapters 4 and 5, that models based on descriptors which can also be determined experimentally are more transparent. In the case of the regression model of Equation 4.7 this was particularly impressive. In section 4.3.2 different lipophilicity terms for the final prediction of $\mathrm{EC_w}$ were compared. When calculated octanol-water distribution coefficients were used instead of the experimentally determined log D values, two of the three previously highly significant descriptors were not significant any more. This means that if only the biological activity would have been available, which is the normal case in QSAR modeling, these descriptors would have been rejected.

In this light, the experiences from the past two chapters result in the recommendation that more data sets should be established containing not only a biological activity, but also other experimental values like K_{ow} or D_{mw}. This would give the modelers more opportunities to "trace" the errors and make improvements at the right place.

On the data analysis level the move from classification systems with many classes to a simple two-class system offers advantages. In the classification study presented here the MOA of uncouplers was discriminated from compounds of other MOAs mainly baseline toxicants (narcotics). As could be seen in Chapter 2 it would be desirable to have classification systems for multiple MOA. The compounds determined to be uncouplers with the proposed filters, might also be determined active for another MOA using another similar filter. As an example, the chlorocatecholes were determined by the classification filter to be uncouplers. In the data set of Chapter 3 the same compound would have been a proelectrophile. If in a future study someone develops a filter for proelectrophiles the compound can then be assigned to uncouplers and also to proelectrophiles. The crucial point for the toxic effect is

then for which MOA the quantitative prediction is higher.

5.4 Conclusion

Given the few data which can be used for establishing a classification model the simple filter method presented here seems an appropriate approach. It resulted in very small number of misclassifications.

With a good tool for the prediction of pK_{a} values and one for the prediction of $\Delta G_{\mathrm{solv,A^-}}$ it should be possible to screen a large data base. This would result in a list of potential candidates which then can be predicted with the regression model. Compounds whose predicted uncoupling activity is substantially higher than the predicted baseline activity would then be considered as uncouplers.

References

[1] H. Terada, "Uncouplers of oxidative phosphorylation", *Environ. Health Perspect.* **1990,** *87,* 213–218.

[2] G. Schüürmann, R. K. Somashekar, U. Kristen, "Structure-activity relationships for chloro- and nitrophenol toxicity in the pollen tube growth test", *Environ. Toxicol. Chem.* **1996,** *15,* 1702–1708.

[3] B. I. Escher, R. I. L. Eggen, U. Schreiber, Z. Schreiber, E. Vye, B. Wisner, R. P. Schwarzenbach, "Baseline Toxicity (Narcosis) of Organic Chemicals Determined by In Vitro Membrane Potential Measurements in Energy-Transducing Membranes", *Environ. Sci. Technol.* **2002,** *36,* 1971–1979.

[4] N. Schweigert, R. W. Hunziker, B. I. Escher, R. I. L. Eggen, "Acute toxicity of (chloro-)catechols and (chloro-)catechol-copper combinations in Escherichia coli corresponds to their membrane toxicity in vitro", *Environ. Toxicol. Chem.* **2001,** *20,* 239–247.

[5] B. I. Escher, Personal Communication, **2005**.

[6] J. Ilivicky, J. E. Casida, "Uncoupling action of 2,4-dinitrophenols, 2-trifluoromethylbenzimidazoles, and certain other pesticide chemicals upon mitochondria from different sources and its relation to toxicity", *Biochem. Pharmacol.* **1969,** *18,* 1389–1401.

[7] Z. J. Guo, H. Miyoshi, T. Komyoji, T. Haga, T. Fujita, "Quantitative analysis with physicochemical substituent and molecular parameters of uncoupling activity of substituted diarylamines", *Biochim. Biophys. Acta* **1991,** *1059,* 91–98.

[8] H. Terada, S. Goto, K. Yamamoto, I. Takeuchi, Y. Hamada, K. Miyake, "Structural requirements of salicylanilides for uncoupling activity in mitochondria: quantitative analysis of structure-uncoupling relationships", *Biochim. Biophys. Acta* **1988,** *936,* 504–512.

[9] D. M. Gange, S. Donovan, R. J. Lopata, K. Henegar, "The QSAR of insecticidal uncouplers", *ACS Symp. Ser. – Classical and Three-Dimensional QSAR in Agrochemistry* **1995,** *606,* 199–212.

[10] J. Kasianowicz, R. Benz, S. McLaughlin, "The kinetic mechanism by which CCCP (carbonyl cyanide m-chlorophenylhydrazone) transports protons across membranes", *J. Membrane Biol.* **1984,** *82,* 179–190.

[11] D. A. Hunt, "2-arylpyrroles: novel uncouplers of oxidative phosphorylation", *Special Publication - Royal Society of Chemistry, Advances in the Chemistry of Insect Control III* **1994,** *147,* 127–140.

[12] S. H. Hilal, S. W. Karickhoff, L. A. Carreira, "SPARC - Performs Automated Reasoning in Chemistry", University of Georgia, Athens, GA.
`http://ibmlc2.chem.uga.edu/sparc/`

[13] B. I. Escher, R. P. Schwarzenbach, "Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms", *Aquat. Sci.* **2002,** *64,* 20–35.

[14] S. H. Hilal, S. W. Karickhoff, L. A. Carreira, "Prediction of Chemical Reactivity Parameters and Physical Properties of Organic Compounds from Molecular Structure using

SPARC", Report: EPA/600/R-03/030, **2003**.
http://www.epa.gov/athens/publications/reports/EPA_600_R03_030.pdf

[15] I. Koppel, J. Koppel, P. C. Maria, J. F. Gal, R. Notario, V. M. Vlasov, R. W. Taft, "Comparison of Bronsted acidities of neutral NH-acids in gas phase, dimethyl sulfoxide and water", *Int. J. Mass Spectrom. Ion Proc.* **1998,** *175,* 61–69.

[16] Y. Masubuchi, S. Yamada, T. Horie, "Diphenylamine as an important structure of non-steroidal anti-inflammatory drugs to uncouple mitochondrial oxidative phosphorylation", *Biochem. Pharmacol.* **1999,** *58,* 861–865.

Chapter 6

Conclusion and Outlook

The three studies presented in this work used a wide range of different data analysis methods and of different chemical descriptors to model quite different types of modes of action. Thus, the question is when is which approach appropriate, especially when should which type of descriptors be chosen. Figure 6.1 can serve as an orientation guide for the different types of descriptors to classify MOAs and for QSAR in general.

Two criteria guided the decisions to use the different approaches in this work. The first one was the available knowledge. If little is known about the processes leading to a toxic effect it is hard to find the underlying molecular or atomic properties of relevance. This was the case in Chapter 3 for a data set of 220 phenols. Although the data set consisted just of phenols their toxicity and metabolism is quite complex [1]. Therefore, a global structure representation based on autocorrelation vectors (AC) was used. As could be shown in Chapter 3 such descriptors require a certain size of the data set. Therefore, the methods involving structure descriptors (Radial Distribution Functions (RDF) and AC) where put in the area of large data sets and little knowledge about mechanisms.

Quite the opposite was true for the data sets used in Chapters 4 with 21 phenols and in Chapter 5 with 49 weak acids composed of phenols and compounds with acidic NH-groups. This data set is small, but less complex as it focuses only on one MOA. Additionally, it turned out that its mechanisms are well understood. Thus, it was appropriate to model the activity and the classification model for uncouplers with molecular descriptors.

The data set used in Chapter 2 with 104 highly diverse compounds and 7 MOAs is hard to position on Figure 6.1. The low correct classification rates obtained with different methods might suggest that its size is too small for its complexity. However, as Figure 6.1 suggests

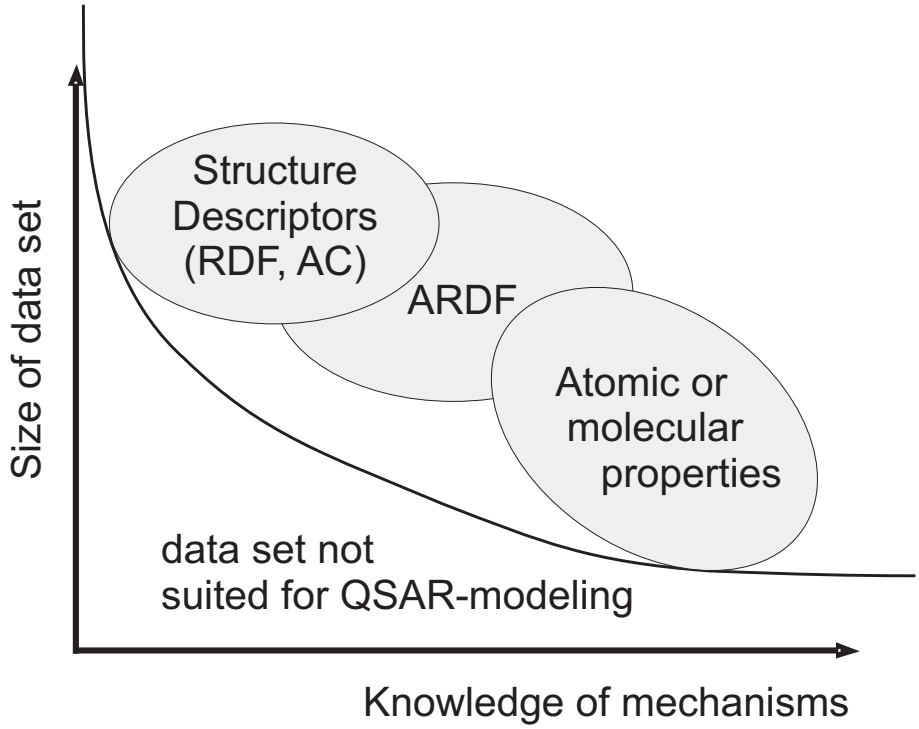


Fig. 6.1 Appropriateness of the different descriptor types with the criteria size of data set and knowledge about mechanisms of data set. To the left of the curve are data sets which are not appropriate for establishing QSAR and SAR models. Abbreviations: Autocorrelation vectors (AC), Radial Distribution Function vectors (RDF), Atomic Radial Distribution Function vectors (ARDF).

new insights into the mechanisms can move this data set along the X-Axis and thereby the problem of the small size might be overcome.

An intermediate situation depicted in Figure 6.1 is the case when a reaction center is known, but not much knowledge is available about the reaction mechanisms. In such a case an atom centered radial distribution is an appropriate structure representation as it uses the available knowledge and requires less data then the global RDF.

In this light, the availability of data is clearly a limiting factor for MOA classification models. As an example even the largest MOA data set, the US-EPA fathead minnow data base with approximately 280 compounds with high confidence MOA assignments, nevertheless contains rather few compounds with specific MOA. With only nine neurotoxicants, 11 uncouplers, 17 inhibitors of AChE there is a clear need to enlarge the database, in order to profit from the full power of the methods of chemoinformatics.

Regarding the contribution of the three disciplines required to establish MOA-based toxicity QSAR and SAR the following experiences were made in this work:

- **Data Analysis**: A large number of pattern recognition methods is available today. Neural networks have become established methods available in major data analysis packages. Given the small size of toxicity data sets neural networks (currently) offer less potential in toxicity QSAR than in other fields.

  A central issue is the validation of data sets. In Chapter 2 the applicability of cross-validation and other resampling methods was discussed at length. This was not just an academic exercise, because if applicable resampling methods offer also an economic advantage. Efron and Tibshirani estimated in simulations that the gain in efficiency of their resampling method corresponds to a large increase in the size of the training set [2]. However, it must be admitted that the applicability of cross-validation and other resampling methods will remain limited by the practice of developing QSAR and SAR models as was discussed in detail in the comments to Chapter 2.

  The determination of applicability domain is of high importance for future application of QSAR as a predictive tool. A recent report identifies still major problems in this field to be solved [3].

- **Chemistry:** In the field of chemistry a wide range of approaches can be used today. Similarity oriented methods like autocorrelation are especially suited for poorly understood problems [4], while whole molecule and atomic descriptors appear to suit mechanistic models best. Both approaches have their limits. A mechanistic approach can fail when a unexpected factor comes into play, e.g., the dimer formation of phenols which affects uncoupling activity might be described accurately by empirical mechanistic descriptors, but it might completely fail if new compounds like a catechol should be predicted which might form other types of dimers. On the other, hand similarity oriented descriptors might fail if the training set is not highly balanced as observed in Chapter 3. It could be shown that for classification methods the training set needs to be balanced in a way that it contains representative compounds for *each* class.

- **Toxicology:** The field of toxicology is probably the most important one for further advances in toxicity QSAR and SAR. Toxicologists can give the modeler the most

valuable leads in which direction to search. There is also a strong need for further harmonization of the MOA-assignment, as was shown in Figure 1.3. Indeed in this thesis in each chapter compounds formerly called narcotics were called differently: nonspecific nonreactives, polar (and nonpolar) narcosis and baseline toxicants. The reason is an ongoing discussion among toxicologists whether the distinction of polar and nonpolar narcosis is justified or whether it is an artifact [5–7]. Each data set used in this study came from adherents of another view in this discussion and such issues cannot be resolved by chemoinformaticians.

The studies of Chapters 4 and 5 are an example of how a chemoinformaticians can benefit from toxicological science. However, it must also be noted that the MOA of uncouplers with its comparable simple mechanism and its clearly defined target is probably a quite simple case. The modeling of other MOA like reactive chemicals which can affect numerous targets seem far more difficult as was impressively demonstrated in past predictive toxicology challenges for the classification of carcinogens [8].

Concerning the general confidence one can have in toxicity SAR and QSAR it is clear that with the limited data available for establishing models an individual chemical may still be hard to predict reliably. However, as Benigni and Richards pointed out, modeling systems show their greatest advantage at the statistical level, i.e. when SAR and QSAR methods are used to screen large databases [9]. This is currently already the case in drug design and combinatorial chemistry and will probably be the case in environmental risk assessment in the near future. In such a case, even a small shift from random assignment to a significant enrichment constitutes a remarkable advantage.

References

[1] I. M. C. M. Rietjens, C. den Besten, R. P. Hanzlik, P. J. van Bladeren, “Cytochrome P450-Catalyzed Oxidation of Halobenzene Derivatives”, *Chem. Res. Toxicol.* **1997,** *10,* 629–635.

[2] B. Efron, R. Tibshirani, “Improvements on Cross-Validation: The .632+ Bootstrap Method”, *JASA* **1997,** *92,* 548–560.

[3] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. van de Sandt, W. Tong, G. Veith, C. Yang, "Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure Activity Relationships – The Report and Recommendations of ECVAM Workshop 52", *ATLA* **2005,** *33,* 1–19.

[4] A. Bender, R. C. Glen, "Molecular similarity: a key technique in molecular informatics", *Org. Biomol. Chem.* **2004,** *2,* 3204–3218.

[5] W. H. J. Vaes, E. U. Ramos, H. J. M. Verhaar, J. L. M. Hermens, "Acute toxicity of nonpolar versus polar narcosis: is there a difference?", *Environ. Toxicol. Chem.* **1998,** *17,* 1380–1384.

[6] B. I. Escher, J. L. M. Hermens, "Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects", *Environ. Sci. Technol.* **2002,** *36,* 4201–4217.

[7] D. W. Roberts, J. F. Costello, "Mechanisms of action for general and polar narcosis: A difference in dimension", *QSAR Comb. Sci.* **2003,** *22,* 226–233.

[8] R. Benigni, A. Giuliani, "Putting the Predictive Toxicology Challenge into perspective: reflections on the results", *Bioinformatics* **2003,** *19,* 1194–1200.

[9] R. Benigni, A. M. Richard, "Quantitative structure-based modeling applied to characterization and prediction of chemical toxicity", *Methods (Orlando, Florida)* **1998,** *14,* 264–276.

Chapter 7

Summary

The prediction of biological activity from chemical structure is a central domain of chemoinformatics and has evolved to a large research field over the last decades.

Traditionally, quantitative structure-activity relationships (QSAR) for toxic effects are derived from a series of compounds belonging to the same chemical class. However, structural similarity does not necessarily lead to similar toxic effects. As a consequence, considerable effort has been made to move toxicity QSAR from a class-based perspective to one that is consistent with modes of toxic action (MOA). The prerequisite for screening large databases of compounds with MOA-based QSARs is the successful classification of compounds into the different MOAs. The goal of this work is therefore to improve MOA classification methods.

Toxicity QSAR is based on three main disciplines: data analysis, chemistry and toxicology. Therefore this work is structured around these three disciplines with each chapter consisting of a study on MOA classification but with a main focus on one of these disciplines. Two of the studies have already been published, the third one has been submitted for publication.

Data analysis methods: Data analysis methods contribute in several ways to improved MOA classification. First of all, they provide classification methods. These can be used to establish a functional relationship between the calculated characteristics of a structure and its class, i.e., between the descriptors and the MOA. Using a classification method for a given data set then, allows one to establish a classification model. Secondly, data analysis methods are necessary to estimate the predictive power of classification models and, thirdly, they are

required to determine the confidence one can have in new predictions of a given classification model.

In the first part of this thesis, the predictive power of different classification models was compared. A data set of 104 compounds with seven different MOAs was used to build classification models using linear discriminant analysis (LDA), multinomial logistic regression (multinom), partial least squares (PLS) and counter-propagation neural networks (CPG NN).

The predictive power varied surprisingly little between the methods and ranged from 52% to 59% overall correct classification when appropriate parameters and descriptor selection methods were chosen. The fact that no approach could further improve the predictive power was attributed to the combination of high diversity, small size of the data set and the large number of classes. Concerning the validation methods, descriptor selection appeared to be one of the crucial processes of too optimistic estimates of predictive power. However, a methodology to correct this optimistic bias in cross-validation methods was developed.

Thus, the recommendations resulting from the first study can be summarized as follows: first, it is not possible to say *a priori* which classification methods works best for a given data set, therefore it is advisable to use both a linear and a nonlinear method for each data set. Secondly, cross-validation requires that all model building steps have to be reproduced, a requirement which in QSAR model building is not easy to meet. However, if this requirement is met it offers advantages over splitting the data set into training and test set, as cross-validation and other resampling methods allow the extract more information from a given data set.

Chemistry: Progress in computational chemistry and especially in chemoinformatics contribute to advances of toxicity prediction in two ways. First of all, it allows the calculation of an increasing number of molecular and atomic descriptors and, secondly, it provides methods for describing molecular similarity in the context of biological activity.

In this second study, a data set of 220 phenols with exhibiting four MOAs was used. The structures were described by a representation previously developed in drug design. The structure representation used consists of vectors based on empirical atomic π-charge and σ-electronegativity encoded by topological autocorrelation and of vectors based on hydrogen bonding potential encoded by surface autocorrelation. CPG NN and multinom were used as classification methods. This led to two classification models with an estimated predictive power of 92% and 93% overall correct classification, respectively.

The performance of autocorrelation descriptors proved to be equal to the results obtained in the evaluation of previously published quantum-chemical descriptors. Thus, the greatest advantage of the proposed structure representation is the ease and speed of calculation. This advantage was utilized to screen all 3142 monocyclic phenols contained in the open NCI database in an exploratory study. Subsequently, a simple method for determining the prediction space was applied. It could be shown that the training set and the structure representation used allow one to safely predict a fair share of the NCI phenols, but also that a relatively large part of the predictions consist of extrapolations.

Toxicology: If mechanistic knowledge of a toxic effect is available, this knowledge can simplify both descriptor selection and data analysis. As a third study a mechanistic QSAR model of uncouplers of oxidative and photo-phosphorylation was developed based on insights gained in kinetic experimental studies. The model was based on a data set of 21 phenols with measured biological activity, and additionally with experimental membrane-water distribution coefficients. This allowed the calculation of effect concentrations at the site of action, which is an advantage compared with the traditional approach of estimating the uptake with octanol-water coefficients.

A linear regression model of good quality could be established using the following descriptors: solvation free energies of the anions, an estimate for heterodimer formation, pK_{a} values and a term for lipophilicity. The study showed deficiencies of octanol-water partitioning coefficients as a descriptor of lipophilicity in the case of charged molecules. Additionally the study indicated a potential to establish MOA based QSARs containing data from different chemical classes which is one of the central goals of MOA based QSAR.

In a last study, a pragmatic approach for a classification model for uncouplers was presented. It is based on a data set of 31 uncouplers and 18 compounds showing no activity in this MOA and could classify 46 of the 49 compounds correctly.

The experiences made in this work can help to guide future studies. The size of data set and the knowledge of toxic mechanisms are taken as the two criteria which should guide through the QSAR model building process. Similarity oriented descriptors like autocorrelation vectors are most suited for complex processes and for large data sets. If mechanistic insights are available, which still is rarely the case, then a more specific description of the processes is possible and smaller data sets become suitable to develop QSAR models.

Concerning the state of MOA classification in general one must say that this field is still in an early stage. However, given that tens of thousands of compounds will be tested over the next years, even models with intermediate or modest predictive power used for the prioritization of testing needs are of remarkable practical value.

Chapter 8

Zusammenfassung

Die Vorhersage der biologischen Aktivität aufgrund der chemischen Struktur ist ein zentrales Gebiet der Chemoinformatik und hat sich in den vergangenen Jahrzehnten zu einem grossen Forschungsgebiet entwickelt.

Die meisten quantitativen Struktur-Wirkungsbeziehungen (quantitative structure-activity relationships: QSAR) für toxische Effekte basieren auf Strukturen der gleichen chemischen Klasse. Strukturelle Ähnlichkeit führt jedoch nicht notwendigerweise zu ähnlichen toxischen Wirktypen. Aus diesem Grund wird in den letzten Jahren vermehrt versucht, QSARs von der Basis der chemischen Klasse weg zu bringen und auf der Basis toxischer Wirktypen (mode of toxic action: MOA) zu arbeiten.

Die Voraussetzung, um grosse Datenbanken mit MOA-basierten QSARs nach problematischen Chemikalien zu durchsuchen, ist eine erfolgreiche Klassifikation der Chemikalien in ihre MOAs. Ziel dieser Arbeit ist deshalb, die Untersuchung und Verbesserung von MOA-Klassifikationsmodellen.

QSARs zur Toxizität basieren hauptsächlich auf drei Disziplinen: Datenanalyse, Chemie und Toxikologie. Diese Arbeit wurde deshalb auch anhand dieser drei Disziplinen gegliedert, wobei in jedem Kapitel ein MOA-Datensatz mit Hauptbezug auf eine dieser Disziplinen untersucht wurde. Zwei dieser Studien wurden bereits publiziert, eine dritte wurde vor kurzem zur Publikation eingereicht.

Datenanalyse: Methoden der Datenanalyse tragen in verschiedenen Bereichen zu Verbesserungen der MOA-Klassifikation bei. Den ersten Bereich bilden Klassifikationsmethoden. Diese stellen den funktionellen Zusammenhang zwischen den berechneten strukturellen

Eigenschaften und der Klassenzugehörigkeit her, also zwischen Deskriptoren und MOAs. Wird eine Klassifikationsmethode auf die Deskriptoren eines gegebenen Datensatz angewandt, lässt sich damit ein Klassifikationsmodell erstellen. Den zweiten Bereich der Datenanalyse bilden die Methoden zur Bestimmung der Vorhersagekraft eines Klassifikationsmodells und den dritten Bereich die Methodik, mit der für ein gegebenes Klassifikationsmodell das Vertrauen in die Vorhersagen neuer Verbindungen bestimmt wird.

Anhand eines MOA-Datensatzes mit 104 Verbindungen und sieben MOAs wurde die Vorhersagekraft verschiedener Klassifikationsmodelle getestet. Folgende Klassifikationsmethoden wurden dafür herangezogen: Lineare Diskriminanzanalyse (LDA), multinomiale logistische Regression (multinom), partial least squares (PLS) und Counter-Propagation Neuronale Netze (CPG NN).

Die Unterschiede in der Vorhersagekraft der Modelle waren relativ gering und reichten mit optimierter Parameterwahl und mit Deskriptorselektionsmethoden von 52% zu 59% insgesamt korrekt vorhergesagten Verbindungen. Die Tatsache, dass sich die Vorhersagekraft mit keinem Ansatz weiter erhöhen liess, wurde auf die sehr hohe molekulare Diversität des Datensatzes und auf dessen geringe Grösse des im Verhältnis zur Anzahl Klassen, zurückgeführt. In weiteren Untersuchungen konnte die Deskriptorselektion als eine der Hauptursachen für die Überschätzung der Vorhersagekraft identifiziert werden. Deshalb wurde eine Methodologie, um solche Überschätzungen in Kreuzvalidierungsmethoden zu korrigieren, erarbeitet.

Aus der Studie liessen sich die folgenden Empfehlungen für das weitere Vorgehen ableiten: Erstens lässt es sich nicht *a priori* sagen, welche Klassifikationsmethoden für einen bestimmten Datensatz am besten funktionieren und deshalb sollten jeweils lineare und nichtlineare Methoden ergänzend eingesetzt werden. Zweitens ist es bei der Verwendung der Kreuzvalidierung unerlässlich, dass alle Schritte der Modellbildung in den Validierungsprozess eingebunden werden, was in der QSAR-Modellierung häufig schwierig zu realisieren ist. Ist dies jedoch möglich, dann bieten Kreuzvalidierungs- und andere Resamplingmethoden erhebliche Vorteile gegenüber Validierungsmethoden, die auf dem Aufteilen in Trainings- und Testdatensatz beruhen, weil sich mit geringerer Datensatzgrösse gleichwertige Schlüsse ziehen lassen.

Chemie: Fortschritte in der theoretischen Chemie, der Computer-Chemie und der Chemoinformatik können in zweifacher Weise zu Verbesserungen von Toxizitätsvorhersagen beitra-

gen. Erstens erlauben sie die zunehmend genaue Berechnung einer Vielzahl von molekularen und chemischen Deskriptoren, und zweitens stellen sie Methoden zur Beschreibung der molekularen Ähnlichkeit im Kontext der biologischen Aktivität zur Verfügung.

Die zweite Studie beruht auf einem Datensatz von 220 Phenolen mit vier MOAs. Eine für die Wirkstoffforschung entwickelte Strukturrepräsentation wurde dabei verwendet. Die Strukturrepräsentation basiert auf empirisch berechneten atomaren π-Ladungen und σ-Elektronegativitäten, welche mit Autokorrelationsvektoren codiert wurden. Zusätzlich wurden Wasserstoffbrückenbindungspotentiale mit Oberflächenautokorrelation codiert. Mit CPG NN und multinom wurden Klassifikationsmodelle gebildet, deren Vorhersagekraft auf 92% beziehungsweise 93% insgesamt korrekt vorhergesagter Verbindungen bestimmt wurde.

Die Vorhersagekraft war vergleichbar mit denjenigen, die in publizierten Arbeiten mit quanten-chemischen Deskriptoren erreicht wurde, wobei die hier vorgeschlagenen Deskriptoren erheblich einfacher und schneller zu berechnen sind. Dieser Vorteil wurde genutzt für eine explorative Analyse der 3142 monocyclischen Phenole in der open NCI Database, in der eine einfache Methode zur Bestimmung des Vorhersageraums verwendet wurde. Dabei zeigte sich, dass ein grosse Zahl dieser Phenole in den Vertrauensbereich des entwickelten Klassifikationsmodell fällt, aber auch schon ein erheblicher Teil der Vorhersagen aus Extrapolationen besteht.

Toxikologie: Ist mechanistisches Wissen über toxische Effekte verfügbar, kann es die Wahl der Deskriptoren und die Datenanalyse vereinfachen. In einer dritten Studie wurde ein mechanistisches QSAR-Modell für Entkoppler der oxidativen und Photo-Phosphorylierung entwickelt, das auf Ergebnissen experimentell hergeleiteter kinetischer Modelle basiert. Ein Datensatz von 21 Phenolen mit gemessener Aktivität und zusätzlich mit gemessenen Membran--Wasser-Verteilungskoeffizienten wurde an dieser Stelle verwendet. Daraus liess sich die Konzentration der Verbindungen am Wirkort berechnen, was ein Vorteil im Vergleich zu Studien darstellt, in denen die Aufnahme mit berechneten Oktanol-Wasser-Verteilungskoeffizienten quantifiziert wird.

Basierend auf den folgenden Deskriptoren liess sich ein lineares Regressionsmodell guter Qualität bilden: freie Solvatationsenergie des Anions, einer Abschätzung für die Heterodimerbildung, der Säuredissoziationskonstante und einem Term für die Lipophilie. In der Studie zeigten sich die Schwächen von Oktanol-Wasser-Verteilungskoeffizienten als Deskriptor der Lipophilie im Falle geladener Moleküle. Die Wahl der Deskriptoren weist ausserdem auf ein

Potential hin, auch Verbindungen anderer chemischer Klassen in das Modell zu integrieren, welches ein zentrales Ziel des MOA-basierten Ansatzes darstellt.

In einer zusätzlichen Studie wurde ein pragmatischer Ansatz zur Klassifikation von Entkopplern entwickelt. Er basiert auf einem Datensatz von 31 aktiven Entkopplern und 18 nicht aktiven Verbindungen und erlaubte 46 der 49 Verbindungen korrekt zu klassifizieren.

Aus den Erfahrungen mit den unterschiedlichen Datensätzen lassen sich zwei Kriterien zur Eignung der verschiedenen Ansätze definieren: Mechanistisches Wissen und Grösse des Datensatzes. Deskriptoren wie Autokorrelationsvektoren, welche eher molekulare Ähnlichkeit beschreiben, eignen sich vor allem für komplexe Prozesse und grosse Datensätze. Ist mechanistisches Wissen vorhanden, was noch immer eher selten zutrifft, ist eine spezifischere Beschreibung der Prozesse möglich, und auch kleinere Datensätze können dadurch für die QSAR-Modellbildung verwendet werden.

Was den Stand der MOA-Klassifikationsmethoden im Allgemeinen betrifft, muss gesagt werden, dass sich dieses Gebiet noch in einem frühen Stadium befindet. Die Tatsache, dass in den nächsten Jahren, zehntausende von Chemikalien getestet werden sollen, bedeutet aber auch, dass schon Modelle mit relativ bescheidener Vorhersagekraft von grossem praktischen Nutzen sein können.

Appendix A

Full Equations used in Chapter 4

The quantitative data of the *in vitro* assay for uncoupling activity used in chapter 4 works with nominal concentrations. This means, one has only data of the total concentration $\mathrm{EC_{tot}}$ in the system and not measured aqueous concentrations, $\mathrm{EC_w}$ nor measured concentrations in the membrane, $\mathrm{EC_m}$. $\mathrm{EC_w}$ and $\mathrm{EC_{tot}}$ are in most cases well correlated, i.e., for the 21 compounds used for the quantitative model of chapter 4 $\mathrm{EC_w}$ and $\mathrm{EC_{tot}}$ correlate with an r^2 of 0.89. Nevertheless this would be an additional source of error which can easily be avoided using the formulas derived below.

Therefore the full derivation used to calculate $\mathrm{EC_m}$ is given in this section. As $\mathrm{EC_w}$ is just one specific concentration, i.e., the one where a defined toxic effect occurs, the general notation of concentrations will be use here. For the concentration of a compound in water this will be C_{tot_w} (instead of $\mathrm{EC_w}$) and for the concentration in the membrane C_{tot_m} (instead of $\mathrm{EC_m}$).

Speciation in the Aqueous Phase (w): The dissociation of an acid in aqueous solution can be written as

$$\mathrm{HA_w} \rightleftharpoons \mathrm{H^+} + \mathrm{A_w^-} \tag{1.1}$$

where $\mathrm{HA_w}$ presents the neutral form, $\mathrm{A_w^-}$ the acid anion and $\mathrm{H^+}$ the aqueous hydrogen ion. Following the notation used in similar studies [1] the representation of the positive charge of the proton and the negative charge of the acid anion will be omitted in this section from now on.

The mass law expression for the reaction in water is

$$K_{a_w} = \frac{C_{A_w} \cdot a_H}{C_{HA_w}} \tag{1.2}$$

where C_{A_w} and C_{HA_w} $[\mathrm{mol} \cdot \mathrm{L}^{-1}]$ represent the concentrations of A and HA in the aqueous

phase and a_H the hydrogen activity. The total concentration of toxicant in the aqueous phase C_{tot_w} is defined by

$$C_{tot_w} = C_{HA_w} + C_{A_w} \tag{1.3}$$

Concentration and Speciation in the Membrane (m): Both species HA and A partition between the aqueous phase and the membrane phase. This effect can be described by the membrane-water partition coefficients of the neutral species and the anion, $K_{mw,HA}$ and $K_{mw,A}$ and the mass law equations for the partitioning processes are

$$K_{mw,HA} = \frac{C_{HA_m}}{C_{HA_w}} \tag{1.4}$$

$$K_{mw,A} = \frac{C_{A_m}}{C_{A_w}} \tag{1.5}$$

where C_{HA_m} and C_{A_m} are the concentrations of neutral species and of anion, respectively, in the membrane phase. As in the aqueous phase C_{tot_m} is defined by

$$C_{tot_m} = C_{HA_m} + C_{A_m} \tag{1.6}$$

Mass balance: In a system containing water and a given mass of membranes the total concentration of a weak acid is

$$C_{tot} = C_{tot_m} \cdot [m] + C_{tot_w} \tag{1.7}$$

where [m] is the ratio of membrane lipid to aqueous phase. In the case of Table 4.1 membrane concentrations are given in units of mol per kg phospholipid in the membrane and [m] takes the unit of kg phospholipid per liter. For a given pH and a given total concentration in the system it is now possible to calculate the concentration of each species in each compartment using the mass law expressions of Equations 1.2, 1.4, and 1.5 and the mass balance of Equation 1.7. In other words it is necessary to know – or in the case of a QSAR model to estimate – a compounds pK_a, the membrane-water partition coefficient of both neutral species and anion, $K_{mw,HA}$ and $K_{mw,A}$.

As an example the equation for the total concentration of acid in the membrane, C_{tot_m} (or $\mathrm{EC_m}$ in the case of the toxic effect concentration) is subsequently derived. Equation 1.6

is transformed into an expression containing only C_{HA_w} by combining Equations 1.4, 1.5, and 1.2

$$C_{tot_m} = C_{HA_m} + C_{A_m} = K_{mw,HA} \cdot C_{HA_w} + K_{mw,A} \cdot C_{A_w} = K_{mw,HA} \cdot C_{HA_w} + K_{mw,A} \cdot K_{a_w} \frac{C_{HA_w}}{a_H} \tag{1.8}$$

In order to arrange the formulas in a clearer form the notion of fractions of the different species present in the aqueous phase and inside the membrane, $\alpha_{\mathrm{i_w}}$ and $\alpha_{\mathrm{i_m}}$, respectively, at a given pH (here for the example of the neutral species, HA) is used

$$\alpha_{HA_w} = \frac{C_{HA_w}}{C_{tot_w}} = \frac{C_{HA_w}}{C_{HA_w} + C_{A_w}} = \frac{1}{1 + 10^{pH - pK_{a_w}}} \tag{1.9}$$

$$\alpha_{HA_m} = \frac{C_{HA_m}}{C_{tot_m}} = \frac{C_{HA_m}}{C_{HA_m} + C_{A_m}} = \frac{1}{1 + 10^{pH - pK_{a_m}}} \tag{1.10}$$

Note that the acidity constant in the membrane phase K_{a_m} is operationally defined in terms of the proton activity in the aqueous phase, assuming that there are no free protons in the membrane and that the protons of the acid-base reaction in the membrane are directly exchanged with the adjacent aqueous solution. The mass law expression for K_{a_m} is

$$K_{a_m} = \frac{C_{A_m} \cdot a_H}{C_{HA_m}} = \frac{K_{a_w} \cdot K_{mw,A}}{K_{mw,HA}} \tag{1.11}$$

Using the notion of species fractions Equation 1.8 can then be rewritten as

$$C_{tot_m} = C_{HA_w} \cdot (K_{mw,HA} + K_{mw,A} \cdot (\frac{1}{\alpha_{HA_w}} - 1)) \tag{1.12}$$

Then, the mass balance of Equation 1.7 is used to substitute C_{HA_w} with the total concentration of toxicant in the system, C_{tot} using the relationship

$$C_{HA_w} = \frac{C_{tot}}{[m] \cdot (K_{mw,HA} + K_{mw,A} \frac{\alpha_{A_w}}{\alpha_{HA_w}}) + \frac{1}{\alpha_{HA_w}}} \tag{1.13}$$

and putting it into the Equation 1.12 yields the final pH-dependent expression for C_{tot_m}

$$C_{tot_m} = \frac{C_{tot}}{[m] + \frac{1}{\alpha_{HA_w} \cdot K_{mw,HA} + \alpha_{A_w} \cdot K_{mw,A}}} \tag{1.14}$$

The formula for C_{tot_w} and hence $\mathrm{EC_w}$ is obtained using the analogous approach.

An additional source of error of all test systems working with nominal concentrations is that compounds might be eliminated from the system, e.g. by hydrolysis or transfer to the gas phase. As the *in vitro* assay for uncoupling activity is very fast this source of error should be negligible for the type of compounds studied in Chapters 4 and 5.

References:

[1] B. I. Escher, R. Hunziker, R. P. Schwarzenbach, J. C. Westall, "Kinetic Model To Describe the Intrinsic Uncoupling Activity of Substituted Phenols in Energy Transducing Membranes", *Environ. Sci. Technol.* **1999,** *33,* 560–570.

Appendix B

Data set used for the classification study of Chapter 5

On the next pages the strucutes of the 49 compounds used for the classification study are shown. The first 21 structures were used for the regression and the classification model (2,4-chlorophenol to triclosan) and are active uncouplers. An additional 28 compounds have been compiled for the classification data set. Compounds 22 to 31 are well known active uncouplers (4-monochlorocatecol to CCCP). Structures 32 (2-chloropenol) to 49 (catechol) were compounds without uncoupling activity, but potential activity in other MOAs. The values of the descriptors used in Chapter 5 are given in Table B.1, except for the first 21 compounds whose descriptor values have already been given in Table 4.1.

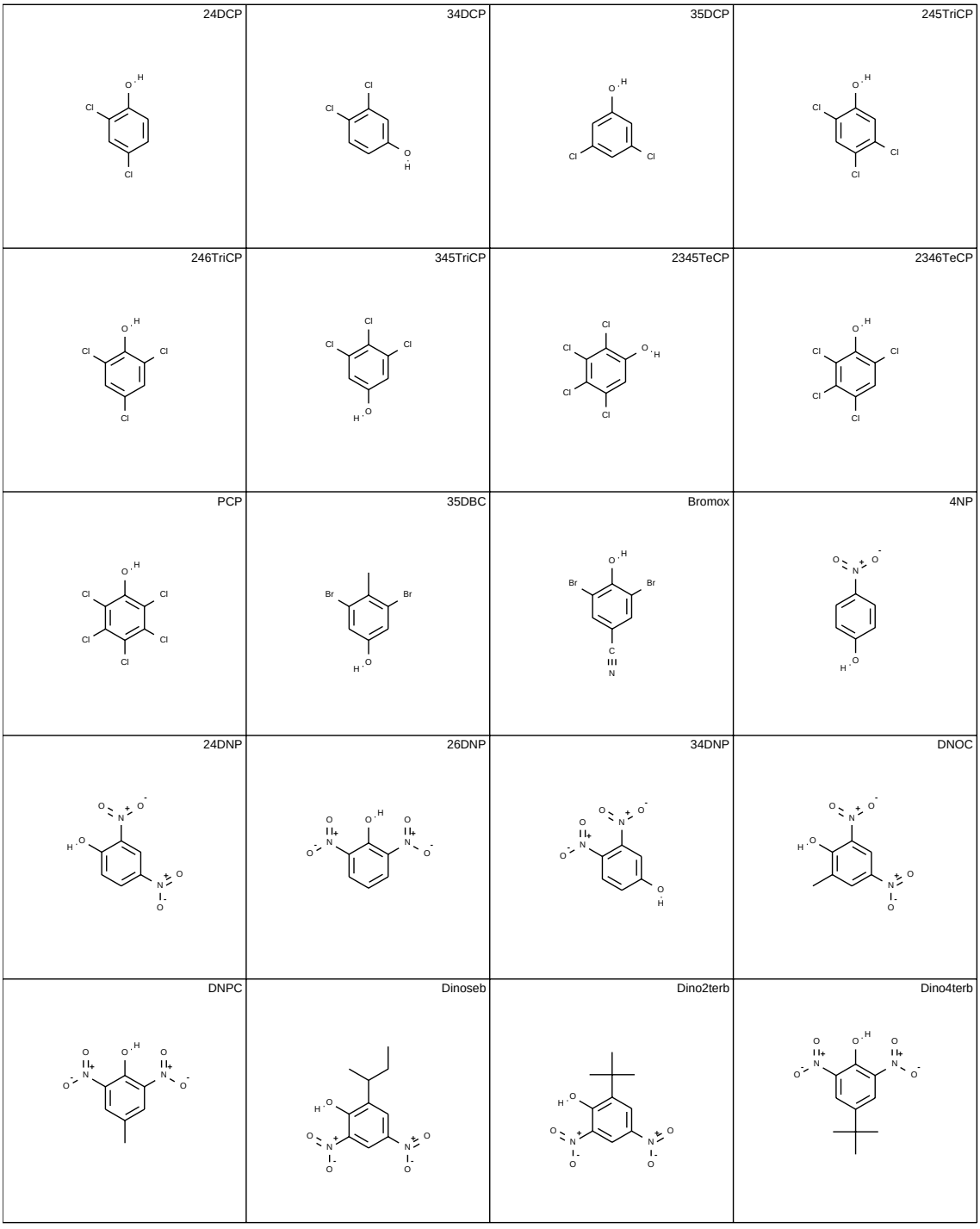


Fig. B.1 Active uncouplers 1-20 of the classification study.

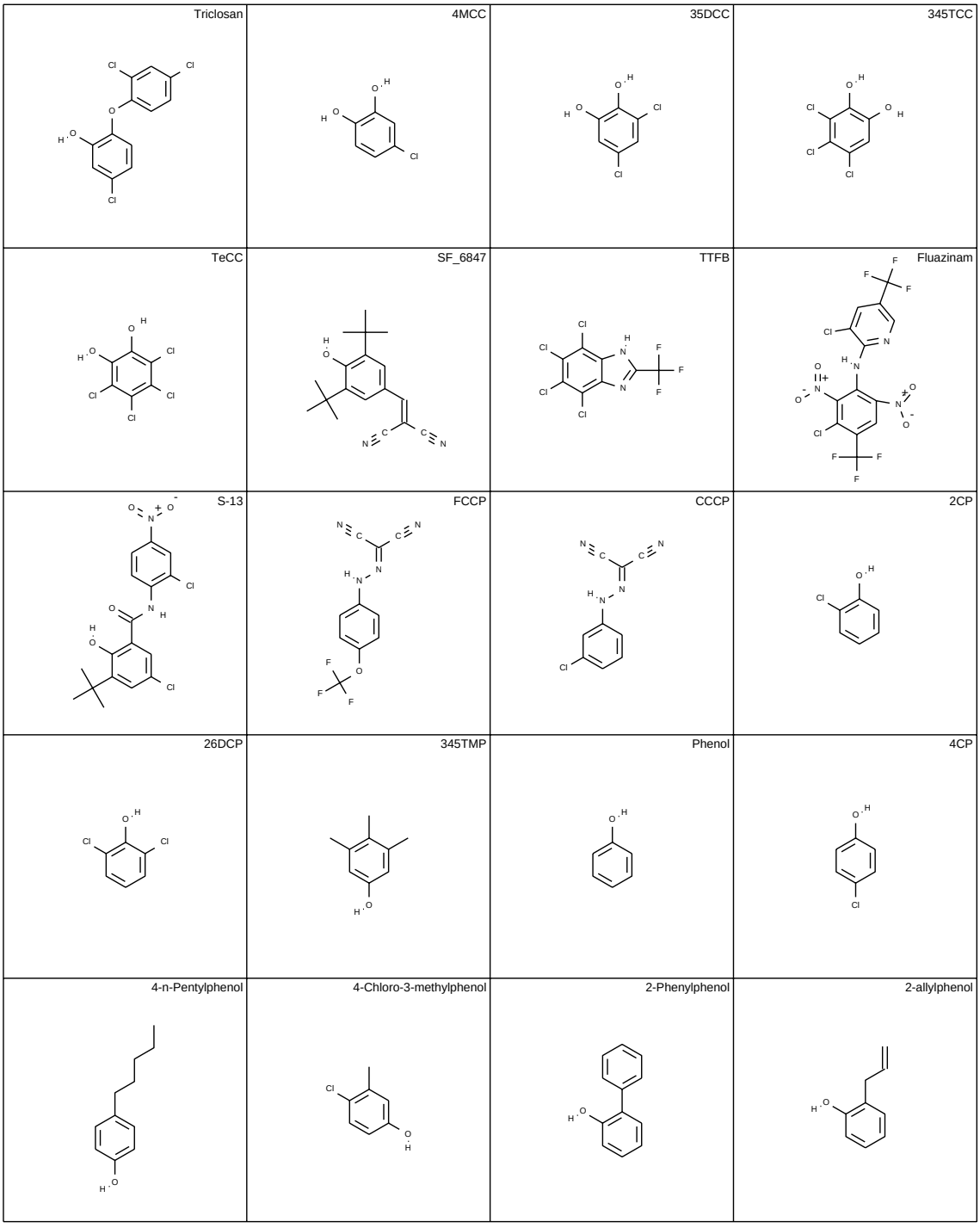


Fig. B.2 Active uncouplers 21-31 and nonactive compounds 32-40 of the classification study.

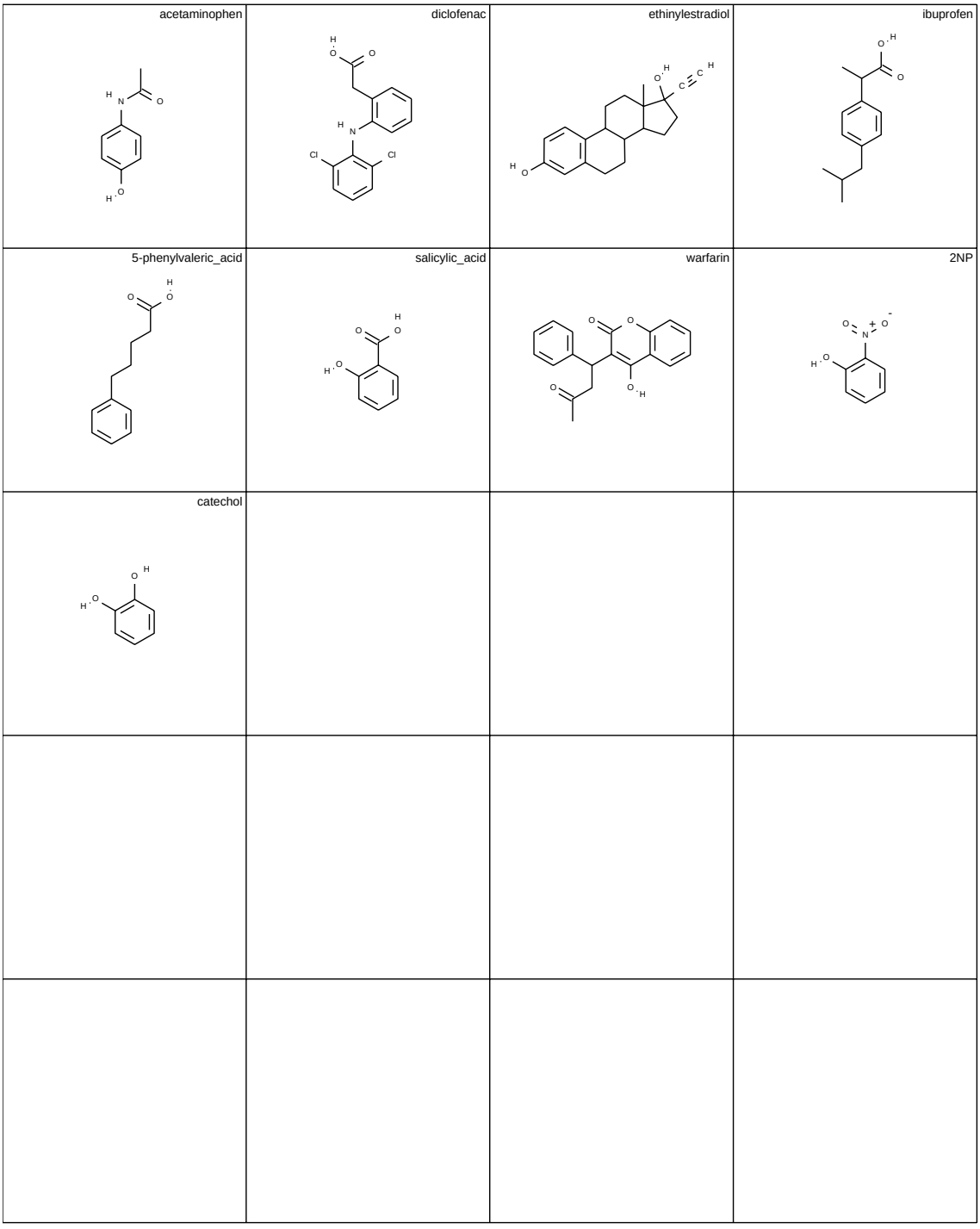


Fig. B.3 Nonactive compounds 41-49 of the classification study.

Table. B.1 Descriptor values of classification data set. The MOA is either uncoupler (u) or other MOAs (o). The references of the data sources for the acid dissociation constant, pK_{a}, and the liposome-water distribution coefficients, log $K_{\mathrm{mw,HA}}$ are given on the next page. $\Delta G_{\mathrm{solv,A^-}}$-values have been calculated with SPARC[l].

N°	Abbr	Compoundname	CAS	MOA	pK_{a}	log $K_{\mathrm{mw,HA}}$	$\Delta G_{\mathrm{solv,A^-}}$
22	4MCC	4-chlorocatechol	2138-22-9	u	8.20[f]	2.53[g]	-64.82
23	35DCC	3,5-dichlorocatechol	13673-92-2	u	7.78[f]	3.11[g]	-62.26
24	345TCC	3,4,5-trichlorocatechol	56961-20-7	u	6.95[f]	3.69[g]	-61.52
25	TeCC	tetrachlorocatechol	1198-55-6	u	5.97[f]	4.41[f]	-61.19
26	SF 6847	2,6-di-tert-butyl-4-(2,2-dicyanovinyl)phenol	10537-47-0	u	6.84[h]	n.d.	-43.19
27	TTFB	4,5,6,7-tetrachloro-2-(CF_3)-1H-benzimidazole	2338-29-6	u	5.50[i]	n.d.	-40.07
28	fluazinam	3-Cl-N-[3-Cl-2,6-dinitro-4-(CF_3)phenyl]-5-(CF_3)-2-pyridinamine	79622-59-6	u	7.11[j]	n.d.	-60.43
29	S-13	5-chloro-3-tert-butyl-2'-chloro-4'nitrosalicylanilide	16128-96-4	u	5.80[k]	6.10[k]	-51.85
30	FCCP	carbonyl cyanide p-methoxyphenylhydrazone	370-86-5	u	6.20[e]	4.12[e]	-70.85
31	CCCP	carbonyl cyanide m-chlorophenylhydrazone	555-60-2	u	5.95[e]	3.79[e]	-72.02
32	2CP	2-Chlorophenol	95-57-8	o	8.56[d]	2.79[d]	-69.70
33	26DCP	2,6-Dichlorophenol	87-65-0	o	6.97[d]	2.87[d]	-67.06
34	345TMP	3,4,5-Trimethylphenol	527-54-8	o	10.25[d]	2.66[d]	-70.67
35	phenol	phenol	108-95-2	o	9.86[l]	1.97[d]	-72.19
36	4CP	4-chlorophenol	106-48-9	o	9.38[a]	2.96[a]	-70.70
37	4-n-Pentylphenol	4-n-pentylphenol	14938-35-3	o	10.29[m]	4.31[n]	-67.83
38	4-Cl-3-methylphenol	4-chloro-3-methylphenol	59-50-7	o	9.55[l]	3.34[n]	-70.19
39	2-Phenylphenol	2-phenylphenol	90-43-7	o	10.01[l]	3.46[n]	-58.91
40	2-allylphenol	2-allylphenol	1745-81-9	o	10.29[l]	3.06[n]	-64.48
41	acetaminophen	acetaminophen	103-90-2	o	10.10[o]	n.d.[o]	-80.74
42	diclofenac	diclofenac	15307-86-5	o	3.99[o]	4.45[o]	-74.18
43	ethinylestradiol	ethinylestradiol	57-63-6	o	10.38[m]	3.81[o]	-76.03
44	ibuprofen	ibuprofen	15687-27-1	o	4.45[o]	3.80[o]	-69.26
45	5-phenylvaleric acid	5-phenylvaleric acid	2270-20-4	o	4.88[e]	2.94[e]	-70.55
46	salicylic acid	salicylic acid	69-72-7	o	2.98[e]	2.50[e]	-62.94
47	warfarin	warfarin	81-81-2	o	5.00[e]	3.39[e]	-83.69
48	2NP	2-nitrophenol	88-75-5	o	7.23[c]	2.09[c]	-61.05
49	catechol	catechol	120-80-9	o	9.25[f]	1.95[f]	-66.09

References:

[a] B. I. Escher, R. P. Schwarzenbach, J. C. Westall, "Evaluation of Liposome-Water Partitioning of Organic Acids and Bases. 1. Development of a Sorption Model", *Environ. Sci. Technol.* **2000,** *34,* 3954–3961.

[b] B. I. Escher, R. W. Hunziker, R. P. Schwarzenbach, "Interaction of Phenolic Uncouplers in Binary Mixtures: Concentration-Additive and Synergistic Effects", *Environ. Sci. Technol.* **2001,** *35,* 3905–3914.

[c] B. I. Escher, R. P. Schwarzenbach, "Partitioning of Substituted Phenols in Liposome-Water, Biomembrane-Water, and Octanol-Water Systems", *Environ. Sci. Technol.* **1996,** *30,* 260–270.

[d] B. I. Escher, R. P. Schwarzenbach, "Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms", *Aquat. Sci.* **2002,** *64,* 20–35.

[e] Escher, B. I., Sigg, L., "Chemical speciation of organics and of metals at biological interphases", in "Physicochemical Kinetics and Transport at Biointerfaces", Buffle, J. and van Leuwen, H. P. (Eds.), *IUPAC Series on Analytical and Physical Chemistry of Environmental Systems*, John Wiley, Chichester, **2004,** Vol. 9, pp 205–269.

[f] N. Schweigert, R. W. Hunziker, B. I. Escher, R. I. L. Eggen, "Acute toxicity of (chloro-)catechols and (chloro-)catechol-copper combinations in Escherichia coli corresponds to their membrane toxicity in vitro", *Environ. Toxicol. Chem.* **2001,** *20,* 239–247.

[g] from Ref. [f] but estimated value.

[h] H. Miyoshi, T. Fujita, "Quantitative analyses of the uncoupling activity of substituted phenols with mitochondria from flight muscles of house flies", *Biochim. Biophys. Acta* **1988,** *935,* 312–321.

[i] H. Terada, "Uncouplers of oxidative phosphorylation", *Environ. Health Perspect.* **1990,** *87,* 213–218.

[j] Z. J. Guo, H. Miyoshi, T. Komyoji, T. Haga, T. Fujita, "Quantitative analysis with physicochemical substituent and molecular parameters of uncoupling activity of substituted diarylamines", *Biochim. Biophys. Acta* **1991,** *1059,* 91–98.

[k] J. Kasianowicz, R. Benz, S. McLaughlin, "How do protons cross the membrane-solution interface? Kinetic studies on bilayer membranes exposed to the protonophore S-13 (5-chloro-3-tert-butyl-2'-chloro-4'nitrosalicylanilide)", *J. Membrane Biol.* **1987,** *95,* 73–89.

[l] S. H. Hilal, S. W. Karickhoff, L. A. Carreira, "SPARC - pKa-Database", University of Georgia, Athens, GA. http://ibmlc2.chem.uga.edu/sparc/

[m] from Ref. [l] but estimated value

[n] W. H. J. Vaes, E. U. Ramos, H. J. M. Verhaar, C. J. Cramer, J. L. M. Hermens, "Understanding and Estimating Membrane/Water Partition Coefficients: Approaches To Derive Quantitative Structure Property Relationships", *Chem. Res. Toxicol.* **1998,** *11,* 847–854.

[o] B. I. Escher, R. I. L. Eggen, U. Schreiber, Z. Schreiber, E. Vye, B. Wisner, R. P. Schwarzenbach, "Baseline Toxicity (Narcosis) of Organic Chemicals Determined by In Vitro Membrane Potential Measurements in Energy-Transducing Membranes", *Environ. Sci. Technol.* **2002,** *36,* 1971–1979.

Appendix C

Publications

- J. Gasteiger, A. Teckentrup, L. Terfloth, S. Spycher
  "Neural networks as data mining tools in drug design"
  J. Phys. Org. Chem. **2003**, *16*, 232–245.

- T. Kleinöder, A. Yan, S. Spycher
  "Prediction of Properties of Compounds",
  in "Chemoinformatics - A Textbook" (eds: J. Gasteiger, T. Engel),
  Wiley-VCH, Weinheim, **2003**, pp. 487-514.

- S. Spycher, M. Nendza, J. Gasteiger
  "Comparison of different classification methods applied to a mode of toxic action data set"
  QSAR & Comb. Sci. **2004**, *23*, 779–791.

- S. Spycher, E. Pellegrini, J. Gasteiger
  "Use of Structure Descriptors to Discriminate between Modes of Toxic Action of Phenols"
  J. Chem. Inf. Comput. Sci. **2005**, *45*, 200–208.

- S. Spycher, B. I. Escher, J. Gasteiger
  "A QSAR Model for the Intrinsic Activity of Uncouplers of Oxidative Phosphorylation"
  (submitted to *Chem. Res. Toxicol.*).

Appendix D

Lebenslauf

Name	Simon Lukas Spycher
Geburtsdatum	17. 01. 1972
Geburtsort	Grabs (Schweiz)
Staatsangehörigkeit	Schweizer
Eltern	Alfred Emil and Hella Annerose Spycher, geb. Paitz
Schulbildung	
08/1978 - 03/1984	Primarschule in Scuol, Kanton Graubünden
04/1984 - 06/1991	Gymnasium Alpines Institut Ftan, Kanton Graubünden Abschluß: Matura Typus B (Latein)
Hochschulausbildung	
10/1992 - 10/1994	Grundstudium der Umweltnaturwissenschaften an der ETH-Zürich
11/1994 - 05/1999	Hauptstudium der Umweltnaturwissenschaften mit Vertiefung Chemie und Bodenkunde an der ETH-Zürich Diplomarbeit in der Gruppe Umwelt- und Sicherheitstechnologie bei Prof. Dr. K. Hungerbühler Thema: „Pestizide im Kaffeeanbaut – Bewertung der Ökotoxizität mit Critical Surface Time 95”
05/1999	Abschluss als Dipl. Umweltnaturwissenschafter ETH
seit 02/2000	Anfertigung der Dissertation bei Prof. Dr. J. Gasteiger am Computer-Chemie-Centrum und Institut für Organische Chemie der Friederich-Alexander Universität Erlangen Nürnberg
Auslandsaufenthalt	
11/1994 - 08/1995	Praktikum bei Inyo County Environmental Health Services, USA