# Toward a Comprehensive Molecular Design Framework for Reduced Hazard

Adelina M. Voutchkova,[†] Thomas G. Osimitz,[‡] and Paul T. Anastas*[,†]

*Center for Green Chemistry and Green Engineering, Yale University, New Haven, Connecticut 06511, and Science Strategies LLC, 600 East Water St., Charlottesville, VA 22902*

*Received September 11, 2009*

## Contents

1. Introduction 5845
2. Scope of This Work 5846
3. The Present State of Toxicology and the Need for a Framework for the Design of Safer Chemicals 5847
4. Toxicology Resources for Chemists 5847
   4.1. Explosion of New Information on Toxicity 5849
      4.1.1. Toxicogenomics 5849
      4.1.2. High-Throughput Screening (HTS) 5850
   4.2. Mechanistic Toxicological Considerations for the Design of Safer Chemicals 5850
      4.2.1. Toxicokinetics and Toxicodynamics 5850
5. Quantitative Structure−Activity (and Toxicity) Relationships 5862
   5.1. Historical Development of QSARs 5862
   5.2. Advancement of Whole-Organism QSTR Models 5863
   5.3. QSARs for General Versus Specific Toxicity 5863
6. In Silico Approaches 5863
   6.1. Estimation of Toxicity 5864
      6.1.1. Automated Rule Induction (ARI) Systems Approach 5864
      6.1.2. Knowledge-Based Systems (Expert Systems) 5865
   6.2. Prediction of Metabolism and Biotransformation 5866
7. New Perspectives: Toward Property-Based Design Guidelines 5867
   7.1. Toxicodynamic and Toxicokinetic Behavior and Chemical Properties 5867
      7.1.1. Example: Do Highly Toxic Chemicals Share Common Physicochemical Properties? 5868
      7.1.2. Structural Interventions That Reduce Absorption (Bioavailability) 5869
      7.1.3. Structural Interventions That Reduce Distribution 5871
      7.1.4. Structural Interventions That Can Reduce Bioactivation 5871
   7.2. Designing Molecules That Do Not Interfere with CYP Regulation Pathways 5871
   7.3. Structural and Property Modifications that Can Reduce Toxicity 5871
      7.3.1. Reducing CNS Activity 5871
      7.3.2. Reducing Carcinogenicity 5876
   7.4. Strategies for Reducing Hazard beyond Human Toxicity 5877
      7.4.1. Reducing Aquatic Toxicity 5877
      7.4.2. Enhancing Biodegradability 5877
      7.4.3. Minimizing Bioaccumulation 5879
8. Conclusion 5879
9. Acknowledgments 5880
10. References 5880

* To whom correspondence should be addressed. E-mail: Paul.Anastas@yale.edu.
[†] Yale University.
[‡] Science Strategies LLC.

## 1. Introduction

The history of chemistry has been one of understanding the properties and transformations of matter. Perhaps the most important aspect of this understanding is the properties that have an impact on human and environmental health and the transformations that take place in our bodies and in the biosphere. Only through a mastery of this understanding will chemistry be able to genuinely design molecules that perform their intended function (e.g., therapeutic or industrial) and are safer for humans and the environment.

Knowledge about the nature of toxic effects comes from the field of toxicology. Once primarily a descriptive science, relying to a large extent on whole-animal toxicology studies, the field has developed an extensive understanding of many of the mechanisms by which chemicals can exert toxicity.[1] Application of this knowledge has made it possible to develop correlations, equations, and models that relate chemical structure and properties to biological responses. This has led to an increasingly sophisticated in silico predictive aspect of toxicology[2] and provides the basis for current work being pursued in the development of a comprehensive design strategy for safer chemicals.

While there has been significant work in the field of chemistry in designing for various functions ranging from medicines to materials, there has been a lack of a comprehensive framework for designing molecules to have a reduced impact on human health and the environment. A framework for the design of safer chemicals was originally published by the noted medicinal chemist E. J. Ariens in 1980, titled appropriately *Domestication of Chemistry by Design of Safer Chemicals*[3] and later revised in 1985.[4] An ACS Symposium Series book published in 1996 on *Designing Safer Chemicals*[5] puts forth a framework that draws on a variety of sources and contains chapters that illustrate how the framework can be applied. In light of the tremendous advances in toxicology and molecular science in the 25 and 14 years since these prior perspectives were written, this review seeks to incorporate the new knowledge and tools available to today's chemist.

In the final measure, the ultimate success of deeply studying a problem is not simply to admire the problem but rather to solve it. This review provides an overview of the excellent research that has been done in the evolution of the

10.1021/cr9003105 
Published on Web 09/27/2010

Paul T. Anastas is the Teresa and H. John Heinz III Professor in the Practice of Chemistry for the Environment. He is the Professor in the Practice of Green Chemistry with appointments in the School of Forestry and Environmental Studies, Department of Chemistry, and Department of Chemical Engineering. In addition, Dr. Anastas serves as the Director of the Center for Green Chemistry and Green Engineering at Yale. From 2004−2006, Paul Anastas served as Director of the Green Chemistry Institute in Washington, D.C. He was previously the Assistant Director for the Environment in the White House Office of Science and Technology Policy where he worked from 1999−2004. Dr. Anastas received his Ph.D. in organic chemistry from Brandeis University. He is credited with establishing the field of green chemistry during his time working for the U.S. EPA as the Chief of the Industrial Chemistry Branch and as the Director of the U.S. Green Chemistry Program. Dr. Anastas has published widely on topics of science through sustainability; his most cited work is with co-author John Warner, entitled *Green Chemistry: Theory and Practice*. He is also Series Editor of the 12-volume series entitled *Handbook of Green Chemistry*, published by Wiley.

Dr. Adelina Voutchkova is a postdoctoral researcher at the Yale Center for Green Chemistry and Green Engineering under the advisement of Prof. Paul Anastas and Prof. Julie Zimmerman. Her current research interest is in the field of rational design of chemicals that are safer for humans and the environment. She received her Ph.D. in organometallic chemistry and catalysis from the group of Bob Crabtree at Yale, prior to which she completed her undergraduate work in Middlebury College, working with Prof. Sunhee Choi on the mechanism of platinum anticancer compounds. Together with Dr. Robert Boethling and Prof. Paul Anastas, she is currently coeditor of Volume 9 of the *Handbook of Green Chemistry* (Wiley), entitled "Designing Safer Chemicals".

molecular understanding of toxicity, from bioavailability to modes and mechanisms of toxic action to predictive toxicology (Figure 1). The information is presented through the perspective of providing the essential elements needed for the development of approaches for reducing the intrinsic hazards of chemicals—design rules for safer chemicals. The review seeks to provide the basis for a dialogue between two scientific communities that seldom interact—toxicologists and molecular designers—but whose conversation could be tremendously fruitful in the quest to reduce toxic threats in everyday life.

Dr. Osimitz is founder and President of Science Strategies LLC, a consulting firm specializing in toxicology and product development. He has over twenty-five years of experience in safety assessment and product development in commodities industry, including directing scientific programs to assess inhalation toxicology, developmental neurotoxicology, mechanistic toxicology, and endocrine disruption. Dr. Osimitz has a Ph.D. in Toxicology from The University of Michigan and a B.S. in Biology from The University of Minnesota. He is Board certified in Toxicology by the American Board of Toxicology.

## 2. Scope of This Work

The main aim of this review is to provide a relatively detailed understanding to chemists not trained in toxicology in (i) the current state of knowledge of the relationships between physical−chemical properties and toxicological hazard of chemicals and (ii) the available tools that chemists can apply to further define such relationships that may guide the design of safer chemicals.

It is important to note that the goal of this effort is to reduce the hazard or *intrinsic toxicity* of chemicals to humans and the environment. This is distinct from risk assessment, which seeks to characterize the probability that a specific exposure scenario will result in toxicity. Implicit in risk assessment is a thorough knowledge of the potential toxicity and associated dose−response relationships, as well as a reasonable estimate of the exposure that an organism will receive under certain circumstances (the external dose). While risk assessment is a useful tool in evaluating comparative risks of existing chemicals and the identification of risk management strategies when needed, we submit that the focus for new chemicals should be on the reduction of hazard.

The current understanding of how to design safer chemicals is still in its nascent stages. We therefore propose a

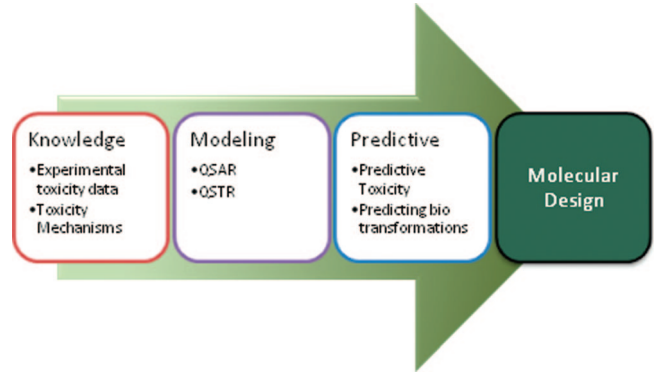



**Figure 1.** Outline of current review.

comprehensive framework for how this field can be further developed with the existing experimental and computational knowledge in toxicology and chemistry. We show that, with the existing knowledge of medicinal chemistry, we can already establish some ground rules for designing less toxic chemicals via incorporation of specific design features that, for example, block their access into organisms, thereby reducing or eliminating the internal dose of a chemical.

Medicinal chemists will find that many of the in silico tools presented are ones they commonly use in understanding how to achieve and maximize therapeutic activity through rational design and structural manipulations. As a result, numerous detailed reports can be found on any one of these tools from the medicinal chemistry perspective, such as quantitative structure−activity relationships (QSARs) and other predictive toxicology tools.[6] However, we have examined these tools from a novel perspective: how to minimize biological activity rather than maximize it.

If we consider the design of commercial chemicals that are benign not only to humans but also to the environment, we see that toxicity is not the only consideration. In fact, we can broadly segregate the hazards into three types—toxicological (human and animals), physical (such as explosivity, material corrosion, and flammability), and global (large-scale effects on our planet: influencing climate change and causing an increased loading of persistent and bioaccumulative chemicals). In this work, we have only addressed the toxicological hazards and touched on global hazards through environmental persistence as it relates to biodegradation. The other factors, equally important, will remain the subject of future work.

## 3. The Present State of Toxicology and the Need for a Framework for the Design of Safer Chemicals

As mentioned above, toxicology is now a well-developed discipline that identifies and assesses biological responses caused by chemicals at the molecular, biochemical, cellular, tissue, and system levels. Recent developments in the field of toxicology highlight the complexity of the interaction between xenobiotics and living systems. One example is the growing appreciation of the role that epigenetic changes may play in the safety assessment of chemicals. Evidence suggests that gene expression can be significantly changed by various mechanisms (such as the alteration by DNA methylation and functional changes in cell surface receptor molecules), resulting in changes in cellular behavior relevant for carcinogenesis, developmental toxicity, and other adverse effects. Moreover, toxicology has recognized that, in addition to understanding the mechanism of toxic action, the application of complementary systems biology approaches and an appreciation of toxicology pathways is essential to a full understanding of how inherent toxicity of a chemical is manifested in a complex organism. Those pathways consist of a dynamic set of complex biochemical interactions of genes, proteins, and small molecules that maintain normal cellular function, control communication between cells, and allow cells to adapt to environmental changes.

The increased knowledge in toxicology has created some controversies as well. Some researchers question the traditional dose-selection paradigm used for routine toxicology studies because several reports have suggested the presence of complex, nonmonotonic dose−response curves for some chemicals and responses.[7] Evidence also exists that some estrogenic chemicals are active at concentrations far below those currently being tested in toxicological studies.[8] Given these emerging complexities in toxicology, a sole reliance on assessing risk by controlling exposure (reducing the external dose) may not be prudent.[9]

In addition to the emerging biological complexities, there is the simple fact that the traditional animal-based toxicology testing is not likely to be able meet the demand for new data on the many chemicals for which the hazard has not been adequately characterized. There are currently about 82 000 chemicals included in the Toxic Substances Control Act that are permitted to be used in commerce in the United States. Approximately 700 new commercial chemicals are invented and introduced each year in the U.S. market.[10] Moreover, the emergence of advanced materials, such as nanomaterials, puts additional strain on the expectations of experimental toxicity testing. The cost of thorough toxicity testing of new commercial chemicals introduced to the market is likely to rise as new assays are developed and additional health concerns are identified.[11] It should be noted that the EPA currently does not require any toxicological testing for new commercial chemicals.The growing need for more thorough and sophisticated toxicology assessments often runs counter to the pressure to minimize animal use for such testing. Although in vitro and in silico alternatives to whole-organism toxicity testing are being developed and validated, many of the fundamental toxicology end points will likely need to be assessed with animal testing for some time.[12]

There is a growing consensus that feels it is necessary to have an increased emphasis on the intentional design of molecules that have reduced inherent toxicity (or hazard). Although this will not obliterate the need for toxicity testing, it will increase the probability that the chemicals being tested are safe. Success in this endeavor will mean that fewer resources will be needed to characterize the toxicity of such molecules and to assess and control exposure to humans and the environment. Molecular designers, skilled at designing compounds to have specific properties, have a crucial role in designing safer chemicals.[13]

## 4. Toxicology Resources for Chemists

Toxicologists have generated a large amount of experimental toxicity data on a wide variety of structurally diverse chemicals, which is an excellent source of information for chemists to extrapolate design rules for reducing toxicity. For this reason, we have briefly summarized information on some of the major toxicity databases available (Table 1).

Each database listed in Table 1 has specific strengths, but two are most notable. The first is EPA's National Center for Computational Toxicology ACToR (Aggregated Computational Toxicology Resource), a relatively new resource that was launched in December 2008. This database comprises a collection of data sources and is a very powerful source of experimental toxicity data. The second is GVK Bioscience's SAR and Mechanistic Based Toxicity Database, a commercial database that provides the mechanisms of toxicity and metabolism information, where available, for query chemicals. This feature can be exceptionally useful to molecular designers, as one can easily identify toxicophores in existing molecules from the mechanism and either eliminate it from the planned molecule or, if that is infeasible, incorporate structural features to reduce toxicity, such as by

**Table 1. List of Toxicity Databases Useful in Informing Molecular Design of Safer Chemicals**

| name | organization | no. of chemicals | details | web address |
|---|---|---|---|---|
| ACToR | EPA NCCT | 2 533 938 | Over 200 sources of publicly available data on chemicals of environmental interest have been brought together and are searchable by name, CAS number, or structure. Available data include chemical structure, physicochemical property values, in vitro assay data, and in vivo toxicology data, organized by phenotype. | http://actor.epa.gov/actor/faces/ACToRHome.jsp |
| Cancer Potency Database | Berkeley Laboratories | 1 547 | Provides results of chronic, long-term animal cancer tests on chemicals, including detailed information on the bioassay and carcinogenic potency (TD50) and its statistical significance. Data are downloadable in an easy-to-use format. | http://potency.berkeley.edu/ |
| ECOTOX | EPA | 8 400 | Aquatic and terrestrial data downloads with data on adverse effects of ecologically relevant species. It includes more than 400 000 test records covering 5 900 aquatic and terrestrial species and 8 400 chemicals. | http://cfpub.epa.gov/ecotox/ |
| Chemical Carcinogenesis Research Information System (CCRIS) | EPA | 8 000 | Carcinogenicity and mutagenicity test results, part of TOXNET. | http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS |
| Chemical Categories Report | EPA | 55 chemical classes | This database provides category statements that the EPA has developed about toxicity patterns for 55 chemical classes based on their wealth of experience in chemical assessment. Although the inclusion of a chemical of interest in the category statements provides insight into possible toxicity, lack of mention does not imply safety of the chemical. | http://www.epa.gov/oppt/sf/pubs/oncologic.htm |
| Hazardous Substances Data Bank (HSDB) | NLM | 5 000 | Comprehensive, peer-reviewed toxicology data, part of TOXNET. | http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB |
| Genetic Toxicology Data Bank (GENE-TOX) | EPA | 3 000 | Peer-reviewed genetic toxicology test data. | http://www.nlm.nih.gov/pubs/factsheets/genetxfs.html |
| ChemBank | Broad Institute's Chem.Biol. Program | 800 000 | Drug-like molecules with freely available in vitro bioassay data derived from small molecule screens. | http://chembank.broad.harvard.edu/ |
| SAR and Mechanistic Toxicity Database | GVK | | SAR database that contains the quantitative details of various in vitro and in vivo toxicity values. It also includes a mechanistic database of drug and drug-like substances and environmentally toxic substances. | http://www.gvkbio.com |
| Acute Aquatic Toxicity Database | USGS | 4 901 | Free online database summarizes the results from aquatic acute toxicity tests conducted by the USGS CERC. Database is downloadable. | http://www.cerc.usgs.gov/data/acute/multiselect.asp |
| DSSTox | EPA | | The DSSTox website provides a public forum for publishing downloadable, structure-searchable, standardized chemical structure files associated with toxicity data. | http://www.epa.gov/ncct/dsstox/index.html |
| TRI-CHIP | EPA | | Database containing peer-reviewed toxicity data on chemicals included in EPA's Toxic Inventory list. It offers many features for customized search capabilities. | http://www.epa.gov/tri/tri-chip/index.html |
| TOXNET | National Library of Medicine | | Collection of databases on toxicology, hazardous chemicals, environmental health, and toxic releases, includes the HSDB and CCRIS listed here, among others | http://toxnet.nlm.nih.gov |
| Integrated Risk Information System (IRIS) | EPA | | Human health assessment program that evaluates quantitative and qualitative risk information on effects that may result from exposure to environmental contaminants. IRIS was initially developed for EPA staff in response to a growing demand for consistent information on substances for use in risk assessments, decision making, and regulatory activities. The information in IRIS is intended for those without extensive training in toxicology but with some knowledge of health sciences. | http://www.epa.gov/iris/ |
| European Chemical Substances Information System (ESIS) | European Commission Joint Research Center | | Provides access to a collection of European regulatory inventories and to the data associated with them. The data are searchable by CAS number, chemical name, and molecular formula. For high production volume chemicals, ESIS stores more than 38 000 acute toxicity tests and about 3 000 tests for chronic toxicity. | http://ecb.jrc.ec.europa.eu/esis/ |

decreasing bioavailability. Although the disadvantage of the GSK database is that it is limited to therapeutic agents, it contains over 13 000 chemicals.

## 4.1. Explosion of New Information on Toxicity

### 4.1.1. Toxicogenomics

Toxicogenomics is defined by the U.S. National Center of Toxicogenomics as "The collection, interpretation, and storage of information about gene and protein activity in order to identify toxic substances in the environment, and to help treat people at the greatest risk of diseases caused by environmental pollutants or toxicants." Toxicogenomic, metabonomic, and proteomic studies (collectively known as Nomics or 'Omics) thus represent the study of the structure, function, and nucleotide sequences of component genes of the genome in order to determine how these genes interact and influence biological pathways, networks, and cellular physiology.[14] During the past two decades, and especially since the publication of the sequence of the human genome, rapid advances in molecular biology have been achieved, which has resulted in a dramatic increase of our knowledge in the field of genomics. The link between conventional toxicological research and functional genomics resulted in the emergence of toxicogenomics.[15]

Toxicogenomics studies the complex interactions between the structure and activity of the genome and adverse biological effects caused by exogenous agents, such as environmental stressors, drugs, and other chemicals. This term, in its broadest context, encompasses genetics and genome-scale mRNA expression (transcriptomics). Most of the work in this field thus far has focused on mammalian toxicology, although toxicogenomic studies have been conducted using genes from nonmammalian species as well.

Nomics can be combined with conventional toxicological approaches and high-throughput screening (discussed later) to form a very powerful tool for characterizing the interactions of chemicals with genes and their ultimate effect on cells and whole organisms. While traditional toxicology testing involves screening compounds through in vivo and in vitro tests with a focus on defined end points (e.g., neurotoxicity, developmental toxicity) or mechanisms of action (e.g., mutagenicity, cytotoxicity, regenerative hyperplasia), the use of microarray (cDNA or oligonucleotide) technology[16] permits the simultaneous analysis of the transcriptional expression level for *thousands* of individual genes, providing an expansive view of how a genome responds to a given chemical (toxic or otherwise). Seeing that at most a few hundred parameters may be measured in a typical comprehensive set of animal toxicology studies, the expansion of available data points afforded by the toxicogenomics technologies is staggering.[17] The challenges both in handling and interpreting the data presented by this data avalanche are discussed below.

**4.1.1.1. Use of Toxicogenomics to Classify Compounds.** Toxicologists are grappling with the task of interpreting toxicogenomic data for the purpose of classifying compounds in toxicity categories. The data collected after the application of microarray technology are codified by bioinformatics using advanced computing technologies in ways that facilitate data mining, i.e., the quick extraction of relevant parameters stored in a database.[16] The direct monitoring of patterns of cellular perturbations in specific pathways by characterizing gene expression resulting from pathological alterations within cells creates the potential to identify a characteristic gene expression profile (i.e., molecular signature or fingerprint) resulting from exposure to specific classes of chemicals.

Having defined molecular signatures that are diagnostic for certain specific forms of toxicity (or precursor events to toxicity), gene expression profiles induced by candidate compounds with unknown toxic properties in the same model systems can then be compared with the established and validated signatures. A positive correlation with an archived molecular signature provides an indication of the potential toxic effects of a new compound from the pattern of the altered gene expression it elicits in vivo or in vitro.[18] Compounds that belong to the same class reveal discernible molecular signatures with a greater similarity to each other than to the molecular signatures corresponding to exposure to a different chemical class. Additionally, intraclass gene expression profiles show greater similarity to each other than to profiles corresponding to compounds from different classes. Clearly, the potential for some of the knowledge derived from this kind of analysis to be applied to the design of safer chemicals is significant. This concept is perhaps best illustrated with the examples below.

**4.1.1.2. Example 1: Use of Toxicogenomics to Study Potential Liver Toxicants.** The liver may respond to chemical exposures in many ways, ranging from no response at all to adaptive responses, such as hypertrophy and induction of metabolic enzymes, peroxisome proliferation, cellular necrosis, and cirrhosis (a result of extensive cellular necrosis) and liver cancer. Toxicogenomics has been used to examine genomic responses to chemicals associated with some of these effects. This includes the evaluation of a series of both known and unknown compounds with respect to their ability to act as liver peroxisome proliferators or enzyme inducers.[19,20]

Over 100 compounds have been categorized into three classes of hepatotoxicants (macrophage activators, peroxisome proliferators, and oxidative stressors/reactive metabolites) in rat livers according to their transcriptional profiles using cDNA microarrays.[21,22]

Waring and colleagues showed that toxicogenomics could distinguish between a diverse set of liver toxicants based on their respective mechanisms of toxicity in either the whole rat or primary rat hepatocytes.[23,24] In the rat, there was a strong correlation between histopathology effects, clinical chemistry parameters, and gene expression signatures induced by the compounds used in the study.[23,24]

Hamadeh and others described the classification of samples based on gene expression changes and have shown a concordance between the level of toxicity identified by histopathological changes and the extent of gene expression alterations.[25] This is one of the first investigations to suggest that *gene expression changes are more sensitive indicators of adverse effects than classical indicators*, such as histopathology.[25] A similar observation was made by Zhou et al., who studied the low-dose effects of cadmium in mice.[26]

**4.1.1.3. Example 2: Use of Toxicogenomics to Study Potential Endocrine Disruptors.** Naciff, Daston, and co-workers[27,28] have conducted extensive research into the relationship between endocrine-related changes in standard in vivo toxicology studies and toxicogenomic changes. They administered well-defined estrogenic chemicals (17-ethynyl estradiol, bisphenol A, and genistein) to rats to characterize the effects on the developing reproductive system (uterus and ovaries) at both fetal and prepubertal

life stages and at the same time characterized gene expression using oligonucleotide-based microarray technology. They discovered that the exposure of female rats to chemicals with estrogenic activity caused an alteration in the expression of a selected number of genes from the uterus and ovaries in a developmental-stage specific manner.

Moreover, they concluded that

(1) Chemicals known to be estrogenic were associated with a specific gene expression signature.
(2) The signature can be used to better understand the mechanism of action of estrogenicity.
(3) The gene transcripts identified in these studies could be used as the basis for a screening assay for chemicals with estrogenic activity.

**4.1.1.4. Challenges to Overcome with the Use of Nomics.** Most promising to their ultimate application of the design of safer chemicals, data from toxicogenomics, proteomics, and metabonomics are beginning to be integrated. This integration should result in a more holistic understanding of the cellular response to toxic chemicals.[29−32]Nomics does have a tremendous promise to increase our understanding of chemical toxicity, but as discussed later, many challenges exist before the data can be reduced to the level at which they can inform the design and selection of safer chemicals.

Although a number of tools can be used in the interpretation of nomics data independently, tools that integrate these disparate data are lacking. In contrast to traditional measures of toxicity in which a persistent and easily observed end point is determined (e.g., liver necrosis), toxicogenomic responses are often subtle, dynamic, and subject to reversible changes. Therefore, capturing predictive profiles will be time-sensitive, and temporal toxicogenomic data will need to be collected and phenotypically correlated to established end points of toxicity, such as liver cell necrosis, before their utility can be maximized.

Comparison of nomics data will require sampling at multiple time points. Moreover, relating early toxicogenomic changes to complex organ-level effects is a difficult exercise due to the gaps in mechanistic understanding. Yet another challenge is determining whether short-term toxicogenomic responses can actually be used to predict toxic effects that would only be seen in whole animals after subchronic or chronic exposures. One of the biggest challenges to the interpretation of toxicogenetic data will be differentiating between adaptive and toxic responses and establishing thresholds beyond which a cascade of molecular responses results in a true adverse effect.

### *4.1.2. High-Throughput Screening (HTS)*

Regulatory agencies worldwide are faced with demands to evaluate the safety of a large number of compounds, far too many to evaluate through traditional end-point-based in vivo testing. Moreover, despite the discussion earlier about predictive toxicology, the current predictive models are not robust enough to handle many of the chemical classes of interest or all of the toxicological end points of concern. Fundamentally, there is a need to distinguish between compounds that are likely to be of little or no concern from those with the greatest likelihood of causing an adverse effect in the target species. Although currently the National Institutes of Health Chemical Genomics Center screens primarily drug candidates for biological activity, HTS could be a used in the future to derive useful data on the toxicity of commercial chemicals as well.

High-throughput screening methods using specific toxicologic end points are an important means of addressing this need. Chemicals with a high likelihood of toxicity can be given a high priority for further testing to better characterize the nature of the toxicity to be expected.

Over the past two decades, HTS has developed into a primary tool for drug discovery based upon bioactivity screening of the proteome.[33,34] On a more limited scale, HTS has also been adapted to agrochemical discovery for the analysis of target species and model organisms.[35,36] Recently, HTS applications to toxicology have been expanding as a useful complement to traditional toxicology.[37,38]

The NIH Chemical Genomics Center (NCGC) is using industrial-scale HTS technologies to collect data that are useful for developing small-molecule chemical probes for basic biological research.[39] The NIH Molecular Libraries & Imaging (MLI) Roadmap Initiative[40] has the potential to greatly enrich data resources and our basic understanding of chemical structure and biological activity. As a part of this effort, the Molecular Library Small Molecule Repository, a collection of more than 70 000 chemicals with a wide range of structures spanning large regions of chemical diversity space, will be put through initially hundreds, but ultimately thousands, of HTS assays. The resulting data will be made fully publicly available in PubChem.

## 4.2. Mechanistic Toxicological Considerations for the Design of Safer Chemicals

Mechanistic toxicology aims to understand and identify the molecular events that lead from initial exposure to the chemical to the ultimate manifestation of toxic injury to the organism. Although mechanistic understanding of toxicology has progressed rapidly, much still remains to be elucidated. Most would agree that, if we had a complete understanding of all the biochemical steps involved in all the possible toxic responses, it would be possible to derive the rules that would incorporate safety at the design stage of a molecule. The science is far from having such an understanding of mechanistic toxicology for all of the classes of chemicals in commerce, but this goal can only be achieved by mapping out the current state of knowledge and identifying pressing gaps in our mechanistic understanding of toxic pathways. However, presently a significant body of mechanistic knowledge already exists that can be used by chemists to avoid some hazardous properties before the molecule is ever synthesized. Furthermore, it is likely that mechanisms of toxicity of many chemicals can be elucidated from an analysis of the existing data.

A useful categorization of toxic responses from a phenotypic standpoint is proposed by Boelsterli in the book *Mechanistic Toxicology*,[41] (Figure 2).

### *4.2.1. Toxicokinetics and Toxicodynamics*

In broad terms, there are two types of factors to be considered when determining the potential toxic effect of a chemical—toxicokinetic and toxicodynamic ones. These concepts are briefly defined here for the purposes of our discussion on toxicity, but the reader can find more sophisticated discussions in several leading texts.[41,42] It may be helpful to consider Boelsterli's pithy simplification of these two concepts, where toxicokinetics is said to encompass

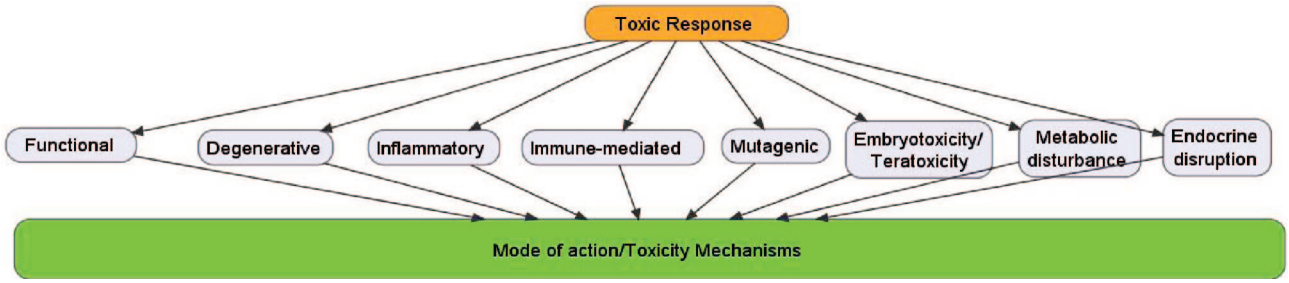



**Figure 2.** Classification of toxic responses based on phenotypic distinctions, modified from ref 41.

"what the body does to a chemical" while toxicodynamics is "what the chemical does to the body".

*Toxicokinetics* is best described as the uptake and fate of a chemical in an organism. Medicinal chemists and toxicologists characterize toxicokinetics by considering ADME (adsorption, distribution, metabolism, and excretion), because all four criteria influence the disposition of the chemical inside the body. Thus, the concentration flux of a chemical in the body is determined by (i) the extent and rate of uptake of the compound into the organism; (ii) how fast and where it is distributed; (iii) how fast it is converted to metabolites; and (iv) its rate of excretion. The kinetic behavior of each metabolite also has to be considered, because the toxicity of many compounds can be ascribed to the biological activity of their metabolites. Storage or accumulation of a compound in particular tissues due to compromised efflux mechanism, for example, can be a toxicokinetic cause of toxicity.

The second factor is *toxicodynamics*. This term is used to describe the interactions of a compound with a particular biological target and the consequences of these interactions. One common example of toxicodynamics is a covalent modification of a biological macromolecule, such as an enzyme, protein, or DNA by a xenobiotic. Such an interaction may trigger a functional or structural alteration of a cell and ultimately lead to an irreversible toxic effect. These interactions can be very precise (often involving receptors) and can account for the organ-specific toxicity that we observe with many xenobiotics. Or, they can be less specific, occurring when a xenobiotic interacts with a variety of macromolecules in the cell. In the latter case, if the damage to the target tissue or organ is great, and if it escapes or overwhelms repair mechanisms, irreversible cellular injury ensues.

In section 7 of this review, we highlight some of the design guidelines that have already been developed. It is worthy to note that these guidelines thus far have been derived almost exclusively from toxicokinetic ADME considerations. This is due to the inherently greater complexity of the entire set of possible in vivo toxicodynamics interactions of a xenobiotic.

**4.2.1.1. Example: Thalidomide.** A famous example of a chemical for which understanding the intricacies of the toxicokinetic and toxicodynamic interactions was crucial is Thalidomide. Thalidomide was introduced in 1956 as a "safe" drug for morning sickness in pregnant women. However, it was soon found that its use was associated with unusual increases in newborns bearing malformations of limbs, or sometimes missing limbs.[43] As a result, the drug was withdrawn from the market in 1962, and since, such malformations have become rare. Recently, Thalidomide has come into the limelight again as a potential drug for various diseases, including autoimmune diseases, AIDS, and some cancers, because the drug is deemed not toxic to adults.[44−46] The chemical was found to exert its teratogenic effect (induction of abnormalities by exogenous compounds during the process of organogenesis) during days 24−33 of pregnancy,[47,48] the time period during which organogenesis occurs. The toxicokinetic factors responsible for the embryo-selective effects of this drug include formation of toxic metabolites and accumulation of the drug or its metabolites in the embryonic tissue compartments.[49] These factors, however important, are not solely responsible for Thalidomide's teratogenic effects. For this, we must also consider possible toxicodynamic interactions. It is known that Thalidomide binds to DNA by intercolation and interferes with mechanisms of gene expression involved in several regulatory pathways, such as those for production of integrins (compounds that help intercellular adhesion)[50] and those responsible for angiogenesis (production of new blood vessels).[51,52]

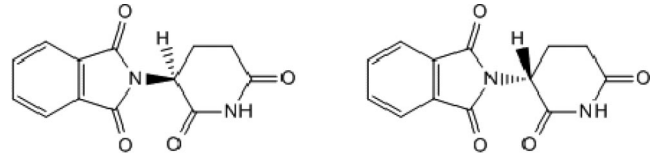



**Figure 3.** Enantiomers of Thalidomide: the *S*-enantiomer is able to intercolate in DNA at specific promoter regions and as a result is a teratogen, while the *R*-enantiomer is too sterically hindered to intercolate in DNA and is thus considered a safe drug for several diseases.

Notably, it was found that only the *S*-enantiomer of Thalidomide fits into the major groove of DNA,[53] while the *R*-enantiomer is unable to bind due to steric hindrance (Figure 3).[54] Athough these findings do not completely explain why there is such a large difference in toxicity of this drug from embryos to adults, it is clear that both the toxicodynamic and toxicokinetic interactions must be appreciated in order to understand how to avoid the molecular design features that are responsible for its toxicity.

**4.2.1.2. Toxicokinetic Considerations for Designing Safer Chemicals.** The terms "exposure" and "dose" are widely used in toxicology, and it is thus important to clarify their definitions for chemists. *External dose* or *applied dose* refers to the amount of the chemical at the interface between the environment and the organism. In the case of humans and other mammals, this includes the gastrointestinal tract, the respiratory tract, and the skin. Most efforts at risk assessment use the external dose as their measure of exposure. The *internal dose* refers to the amount of the chemical that has crossed the biological interface and is available for distribution in the organism and interaction with possible targets of toxicity. The amount that reaches the fluid, organ, or tissue of interest is referred to as the *delivered dose*. The *biologically effective dose* is relevant to toxic chemicals and is the amount that actually reaches the cellular or molecular targets of toxicity.


**Figure 4.** Phases of interaction of an intrinsically toxic chemical with a living system. Adapted from ref 55.

There are three fundamental requirements for chemical toxicity, as illustrated in Figure 4:

- There must be exposure to the chemical substance (external dose).
- The substance must be bioavailable (have a biologically effective dose).
- It must be capable of directly or indirectly causing an alteration in the normal cellular biochemistry and physiology inside the organism (toxicodynamics).

Since this discussion focuses on reducing intrinsic hazard through intentional molecular design, rather than through reducing the likelihood that a given exposure of the organism to the chemical will occur, we will assume that exposure is unavoidable and does happen, resulting in an external dose to the organism. Assuming an external dose is delivered and that the chemical is not corrosive to the skin, the next requirement is bioavailability—if a chemical cannot gain access into an organism, it will most likely not exert a toxic effect.

Our discussion will focus on the physicochemical properties that we, as molecular designers, can influence and the ways in which they can be manipulated to decrease bioavailability and reduce some modes of bioactivation.

*4.2.1.2.a. Absorption.* Absorption is often used interchangeably with the term bioavailability. While bioavailability is a direct function of the extent of absorption, the two terms are not synonymous. Bioavailability is a measure of the extent to which a chemical crosses a barrier of the external environment and enters the body to reach tissues in organ systems where it may interact with cellular macromolecules. It is often expressed as the letter $F$ in pharmacokinetic equations, as the fraction of external dose that enters the body to yield a delivered dose.[56] Numerous factors are known to influence bioavailability, which implies that bioavailability is more complex than just "internal dose". Some of these factors include, but are not limited to, intraspecies (individual) differences based on

- Age—In general, chemicals are metabolized more slowly in fetal, neonatal, and geriatric populations;
- Disease state, e.g., hepatic insufficiency, poor renal function, compromised skin;
- Levels of substrate efflux transporters (e.g., P-glycoprotein);
- Other phenotypic and individual metabolic differences, enterohepatic circulation, diet, gender, circadian differences, gastric emptying rate;
- Enzyme induction, causing increased rate of metabolism (e.g., Phenytoin (antiepileptic) induces CYP 450) and enzyme inhibition, or decreased rate of metabolism (e.g., grapefruit juice inhibits CYP 450);
- External dose—possible saturation of absorption mechanisms;
- Interactions with other xenobiotics (e.g., antacids, nicotine);
- Physicochemical properties of the toxicant (hydrophobicity, p$K_a$, solubility, polarity, vapor pressure, etc.).

Each of these factors exhibits inter- as well as intraindividual variation. The rate and extent of absorption of an ingested chemical will vary depending upon the amount of food the individual has consumed. This may alter first-pass metabolism, intestinal motility, and the degree of chemical degradation by intestinal microflora. Disease states affecting liver metabolism or gastrointestinal function will also have an effect.

As shown in Figure 4, blocking or reducing bioavailability is a fundamental means of reducing intrinsic toxicity of a chemical. Given the various sites and mechanisms of absorption, this can be a complicated process. However, reducing bioavailability of a chemical by molecular design is significantly more straightforward than influencing the distribution, metabolism, and elimination of the chemical once inside the body. Medicinal chemistry has made great strides in understanding the relationship between chemical structures and bioavailability in both humans, other mammals,[57] and ecologically relevant species, especially aquatic organisms.[5] The experimental understanding has also led to the development of in silico models that predict bioavailability in humans and other species.[58,59] Bioavailability has also been examined at length by the toxicology community with regard to organic[60−62] and inorganic[63] pollutants in the environment. Absorption of chemicals in mammals occurs primarily by the following three routes: the gastrointestinal tract, the respiratory tract, and the skin.

***Gastrointestinal Tract.*** Of the three possible sites for initial exposure in humans (gastrointestinal (GI) tract, respiratory tract, skin), the GI tract has the largest surface area (250 $m^2$ for the small intestine, the primary absorptive site) and the second-largest blood flow (20% of total cardiac output for the small intestine). Chemicals enter the GI tract by simple ingestion of a chemical present in food, drink, or other ingested particles, such as dust. For some chemicals, such as residential-use pesticides, incidental ingestion in children resulting from hand-to-mouth contact and dissolution may be a significant source of pesticide exposure. Airborne xenobiotics can be inhaled, either dissolved in or otherwise adhered to the mucous lining of the upper respiratory tract, and be swallowed, ending up in the GI tract.

Most xenobiotics pass the GI mucosa by passive diffusion, a process heavily dependent upon the proportion of the chemical that is nonionized or otherwise somewhat lipophilic. Brodie et al[64] proposed the pH−partition theory to explain the influence of GI pH and the chemical's p$K_a$ on the extent of chemical absorption into the systemic circulation and determined the ratio **D** of concentration of the chemical in the blood versus that in the GI tract, defined below. At physiological pH, most weak organic acids and bases will exist in varying proportions of un-ionized and ionized forms, depending on their p$K_a$ value (Figure 5).[64] The variation of pH over the length of the GI tract is large (in the small intestine, it is 3−7, and it reaches 8 in the large intestine).

**D** = ratio of total concentration of chemical in blood to total concentration in GI tract, i.e.,

$$\mathbf{D} = \frac{[U]_b + [I]_b}{[U]_g + [I]_g} \quad (1)$$

where $[U]_b$ = concentration in blood of un-ionized species, $[I]_b$ = concentration in blood of ionized species, $[U]_g$ = concentration in GI tract of un-ionized species, and $[I]_b$ = concentration in GI tract of ionized species. The ratio [U]/[I] is thus a function of the pH inside the intestine and the p$K_a$ of the molecule, as described by the Henderson−Hasselbach equation as follows.

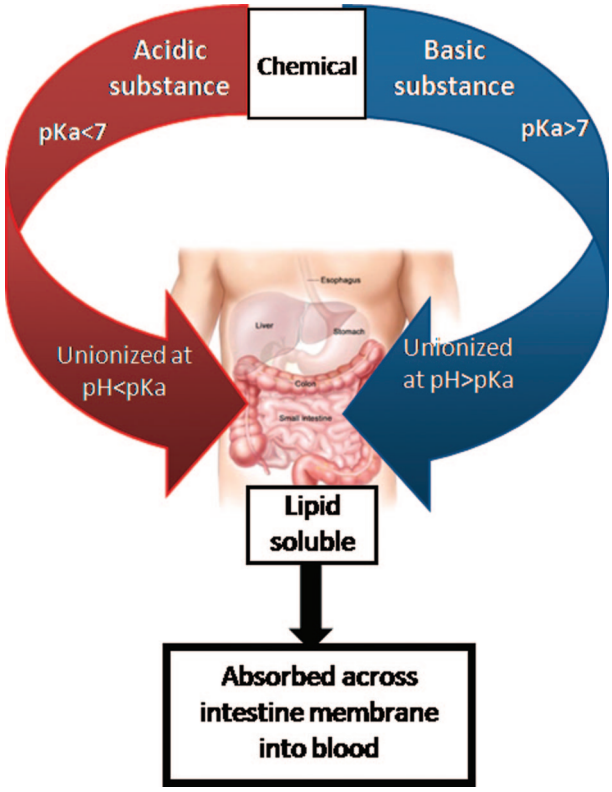



**Figure 5.** Influence of p$K_a$ on degree of ionization of a chemical, which in turn influences the extent of its absorption in the GI tract.

For weak acids:

$$pK_a - pH = \log \frac{[I]}{[U]} = \log \frac{[A^-]}{[HA]} \quad (2)$$

For weak bases:

$$pK_a - pH = \log \frac{[U]}{[I]} = \log \frac{[B]}{[BH^+]} \quad (3)$$

***Respiratory Tract.*** The lungs are just one portion of the respiratory tract, which also includes the nasopharyngeal region and the tracheobronchial region. The nasopharyngeal region consists of the nasal turbinates, glottis, epiglottis, and larynx. The tracheobronchial region contains primarily conducting airways that carry inspired air to the pulmonary region of the lungs where gas exchange takes place. While the nasopharyngeal and tracheobronchial regions are part of the respiratory tract, the external doses delivered there are dissolved in the mucous lining and are swallowed, resulting in a GI tract exposure. In contrast, true delivery of internal doses via inhalation exposure occurs in the lungs.

In humans, the lungs have the second-largest surface area (140 $m^2$), the greatest blood flow (receiving 100% of the cardiac output), and the thinnest absorption barrier (0.1−0.4 $\mu$m) of the three absorption routes, enabling rapid and efficient absorption of not only vital gases but also potentially hazardous chemicals.

Highly water-soluble and/or reactive gases and vapors (such as acetone, formaldehyde, methanol, and sulfur dioxide) tend to dissolve in mucus or react with tissue in the nasopharyngeal region and are removed from the air stream, resulting in GI tract exposure and sparing the pulmonary region. Less reactive and less water-soluble vapors and gases (such as methylene chloride) will penetrate to the pulmonary region and will be absorbed into the blood.

In contrast to the GI tract and skin, lipophilicity is not the primary determinant of the rate of absorption in the lungs. Because only a very thin alveolar membrane (typically 0.2−0.4 $\mu$m in humans) separates the outside of the body (in this case, the alveolar space in the deep lung) from systemic blood circulation, diffusion is the driving force. Thus, the blood-to-gas partition coefficient is the most important factor in the *rate* of uptake of a chemical from the inhaled air. Absorption from the alveoli and into the blood continues until equilibrium is obtained (equal concentrations in the airspace and the blood). Chemicals with a high blood-to-gas partition coefficient like chloroform (blood-to-gas partition coefficient = 15) will more rapidly be carried into the circulation than ethylene (partition coefficient = 0.14).

However, chemicals with a low blood-to-gas partition coefficient (such as ethylene = 0.14) will take a shorter time to equilibrate than chemicals with a higher blood-to-gas partition coefficient (such as chloroform = 15). Because of other important factors, such as tissue storage and metabolism, that affect the dynamic equilibrium in place, it is difficult to generalize about the net internal dose likely to be received based purely on consideration of the blood-to-gas partition coefficient.

For particles, the most important factors affecting the bioavailability are the particle size and the water solubility. The particle size is the determinant of the primary site of deposition of an inhaled particle:

- <1 $\mu$m penetrate to the pulmonary region. They are either absorbed into the blood across the alveoli (especially for ultrafine particles < 0.1 mm), scavenged by pulmonary macrophages (ultimately resulting in exposure via the lymphatic system), or dissolved and absorbed.
- 5 $\mu$m are deposited in the tracheobronchiolar regions of the lungs, from where they are mostly cleared by mucous layer and ultimately swallowed (clearance times, $t_{1/2}$ of 30−300 min), resulting in GI exposure.
- >5 $\mu$m are stopped in the nasopharyngeal region and swallowed, resulting in GI exposure.

The water solubility of a particle is especially important to clearance from the pulmonary region, as this is the primary means by which most particles are cleared from this region.

***Skin.*** The skin has a much smaller surface area (1.75 $m^2$) and blood flow than do the lungs and GI tract (~9% of total cardiac output) and offers a thicker barrier (100−1000 $\mu$m), but remains an important route of exposure for nonvolatile compounds. The rate-determining step for dermal permeability ($K_p$) is diffusion through the *stratum corneum* (uppermost layer of epidermis). There is substantial variability of permeability through skin for various chemicals (e.g., $K_p$ for octane is 5 pmol/$cm^2$/h,[65] while that for DMSO is 200 $\mu$mol/$cm^2$/h).[66] When considering $K_p$, it must be noted that some chemicals will degrade the integrity of the skin, which will render it inherently more permeable. In addition, the thickness and permeability of the skin in different regions of the human body varies by a factor of at least 20 (e.g., on the palm the stratum corneum is 400 $\mu$m thick, while the scrotum has 5 $\mu$m thickness). The chemical properties that are known to have the greatest influence on skin permeability are molecular size, water solubility, and the presence of solvents/carriers when exposure occurs. However, other physiological factors, such as the integrity of the stratum

corneum, its hydration state, and the ambient temperature can also have a significant effect on the xenobiotic absorption of chemicals through the skin.

*4.2.1.2.b. Distribution.* Distribution refers to the movement of a compound from its site of entry to other parts of the organism through the bloodstream. The rate of distribution is primarily determined by blood flow and the rate of diffusion out of the capillaries and into the tissues of a particular organ. Since cardiac output is approximately 4.8−6.4 L/min,[67] distribution is complete shortly after the chemical enters the blood. Well-perfused organs quickly take delivery of any foreign chemical that is in the blood. Where a substance is distributed, however, depends on its physicochemical and structural properties. The extent to which a chemical leaves the blood and distributes throughout the body to tissues (such as adipose tissue, bone, etc.) depends upon its relative physicochemical affinity for that tissue, as well as its ability to perfuse through the capillaries. Macromolecules, for example, such as heparin, cannot diffuse out of capillaries. Distribution is often nonuniform due to the differing nature of the various physiological compartments in the body (plasma, adipose tissue, bone, highly perfused tissues, etc).

A quantitative measure of the extent to which a chemical is, at least theoretically, distributed in the body is provided by the *apparent volume of distribution* ($V_D$). This is the theoretical volume of fluid into which the xenobiotic administered would have to be diluted to produce a certain observed concentration in the plasma. $V_D$ is thus given by

$$V_D = \frac{\text{total amount of xenobiotic in body}}{\text{observed concentration of xenobiotic in blood}}$$

$V_D$ is a descriptor of the equilibrium between the amount of chemical in plasma and that in the tissues. Compounds that have high affinity for tissues (such as 2,3,7,8-tetrachlorodibenzodioxin, a potent environmental toxicant that has a high affinity for adipose tissue[68]) are quickly distributed, and only a small fraction of the internal dose remains in blood. Their $V_D$ is therefore high. By contrast, compounds that remain primarily in the systemic circulation for long periods of time, such as the dye Evans Blue that binds to plasma proteins,[69] tend to have a lower $V_D$. Given the nature of this calculation, the value does not distinguish between a chemical that is truly widely distributed and one with a high affinity for tissue binding with restricted distribution. Nevertheless, $V_D$ is a useful descriptive parameter in pharmaco- and toxicokinetics.

Several predictive algorithms for $V_D$ have been developed.[70,71] These models indicate that the main physicochemical properties that affect $V_D$ are lipophilicity, H-bonding, size, and acid/base properties. Obach and co-workers have noted that basic compounds have higher $V_D$ values than neutral and zwitterionic ones, while acidic compounds have the lowest $V_D$.[72] This trend is likely to be related to the effect of p$K_a$ on the plasma protein binding of the chemicals. Also, since many organic acids have a p$K_a$ lower than the physiological pH of 7.4, they exist primarily in the dissociated form, and as such, tend to persist more in the blood than organic bases.

***Special Case: Blood−Brain Barrier.*** The blood−brain barrier (BBB) is a unique biological membrane that warrants special consideration. Outside of the brain, the endothelial cells lining the capillaries contain gap junctions that allow polar compounds in the blood to pass relatively freely from the bloodstream into surrounding tissues. In contrast, in the brain, the endothelial cells lining the capillaries are thicker and form tight junctions that prevent polar compounds from passing through the capillaries and into the cells of the brain.[73]

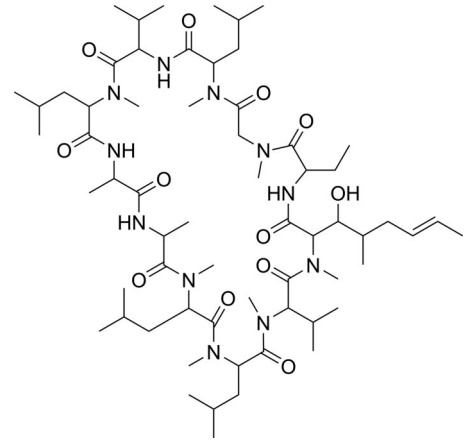


**Figure 6.** Cyclosporine enters endothelial cells but is efficiently removed by xenobiotic transporters.

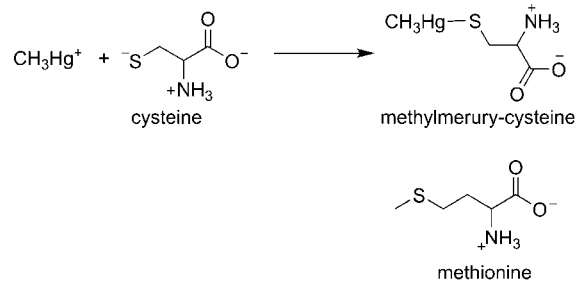



**Figure 7.** Methionine is mimicked by methylmercury−cysteine complex, which is thus allowed entry across the blood−brain barrier.

Upon entering endothelial cells, lipophilic compounds (such as cyclosporine,[74] shown in Figure 6) are subject to efflux by special xenobiotic transporters.[73] A few xenobiotics, such as cysteine-bound methylmercury, which reacts much like the endogenous substrate methionine, may cross the blood−brain barrier via active transport carrier-mediated processes (Figure 7).[75]

The blood−brain barrier is thus a natural defense designed to protect the brain and successfully block many chemicals from entering. Nonetheless, it is not a perfect one, and although highly polar and ionized molecules are blocked, lipid-soluble xenobiotics may still cross the blood−brain barrier. Molecular polarity is often quantified using polar surface area (PSA). Molecules with PSA of >140 Å$^2$ are usually believed to be poor at permeating cell membranes, while those with PSA of <60 Å$^2$ can usually pass through.[76] Size, however, also matters—even moderately polar small organic chemicals (e.g., ethanol) may cross the blood−brain barrier.[77]

*4.2.1.2.c. Accumulation and Storage.* Because of their physical and chemical characteristics, some chemicals are subject to accumulation and storage within the body. Stored chemicals can be released under certain circumstances at some point in time and cause toxic effects, specific examples of which will be discussed below. It should be noted that the site of accumulation is often not the target for toxicity and that stored toxicants are always in equilibrium with the

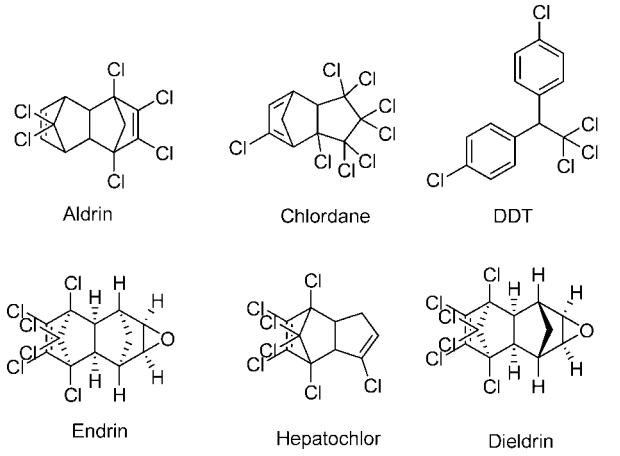



**Figure 8.** Highly lipophilic chemicals, which tend to be stored in the adipose tissues of humans and other animals.

toxicant in circulation. The four primary storage systems in the body are described below to raise the chemist's awareness of the structural motifs that are associated with chemicals stored by each system.

***Adipose Tissue.*** Highly lipophilic chemicals (ones with high log $P_{ow}$) are more likely to be stored in the body's fat cells than to be metabolized and excreted. Such chemicals also have a tendency to accumulate with continued exposure to the chemicals. Examples include persistent organic pollutants such asaldrin, chlordane, 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (DDT), dieldrin, endrin, heptachlor, mirex, toxaphene, polyhalogenated biphenyls, dioxins, and furans (some of which are shown in Figure 8). When fat stores are depleted, such as during fasting, mobilization of the stored chemicals results and causes redistribution of the xenobiotics throughout the body.

***Bone.*** Compounds containing fluoride, as well as heavy metals (e.g., lead and strontium, among others), may be incorporated and stored in the bone matrix by an equilibrium process. The mobilization of toxicants from bone can cause severe toxic effects and is often hormonally controlled (e.g., an increase in osteolytic activity that typically occurs during pregnancy[78]). Children chronically exposed to high concentrations of lead have been reported to have high lead concentrations in the long bone.[79] It is often observed that, after exposure has ceased and blood lead concentrations have been reduced through chelation therapy, lead starts being released from the stores in the long bones back into the blood.[80,81]

***Liver and Kidneys.*** The liver and kidneys also have a high capacity for storing chemicals. Toxic metals, such as lead and cadmium, are often stored in these organs, where they are bound and sequestered by various metallothioneins, which prevents their excretion. Other proteins, such as α2u-globulin, sequester hydrocarbons like D-limonene. Binding of hydrocarbons to α2u-globulins interferes with their clearance from renal tubular cells, and chronic exposure to such hydrocarbons has been shown to lead to nephrotoxicity and carcinogenicity in the male rat.[82]

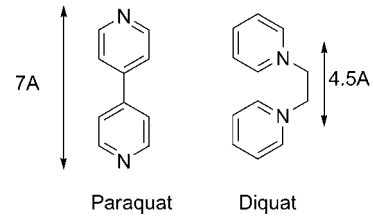
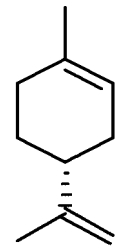



**Figure 9.** Paraquat (highly toxic), Diquat (much less toxic), and the native PAT polyamine.



D-limonene

***Example: Toxicity through Selective Accumulation in Lung Tissue—Diquat and Paraquat.*** Paraquat is a nonselective contact herbicide of the diphenyl class that was introduced in the 1960s and became used worldwide. However, it was soon recognized that Paraquat causes selective pulmonary toxicity to humans (damage to alveolar cells type I and II, causing edema, inflammation, hemorrhage, and lung fibrosis) and in laboratory animals such as dogs.[83]

A vital step in its toxicity mechanism was found to involve selective accumulation. After oral ingestion, a very small fraction of the compound is absorbed, but the blood levels remain constant for many hours because Paraquat is not metabolized by the liver. Instead, it accumulates selectively in the lung and is retained there even after blood concentrations decrease.[84] The reason for this accumulation was found to involve entry into alveolar cells via carrier-mediated transmembrane transport, specifically the polyamine transport (PAT) system.[85,86] Alveolar cells I and II have much more efficient amine transport systems than any cells in other major organs. Once inside the cells, the positively charged Paraquat exerted strong electrostatic interactions with DNA and RNA. A new generation of this herbicide, Diquat, was found to be just as efficacious but much less toxic to mammals (Figure 9).[87]

The reason for this was that Diquat does not accumulate in the alveolar cells like its predecessor, Paraquat. This is due to the rigorous substrate specificity of the PAT system. One of the native amines PAT transports is putrescine, which has four methylene groups separating the two positively charged nitrogens (it is known that PAT has affinity for substances with 4−7 methylene groups separating the two N's) and minimal steric hindrance.[88] Because Diquat is more sterically hindered than Paraquat and has only two methylene groups between the N's, PAT systems show a low affinity toward it, and it is thus not accumulated in the alveolar cells (Figure 10).

***Plasma Proteins.*** Binding to plasma proteins is a common storage depot for endogenous chemicals and xenobiotics. Of the plasma proteins, albumin is the major protein that binds a variety of xenobiotics, but several others are also important (Table 2). Plasma binding decreases both the effective (i.e., free) concentration of the xenobiotic in the system and its

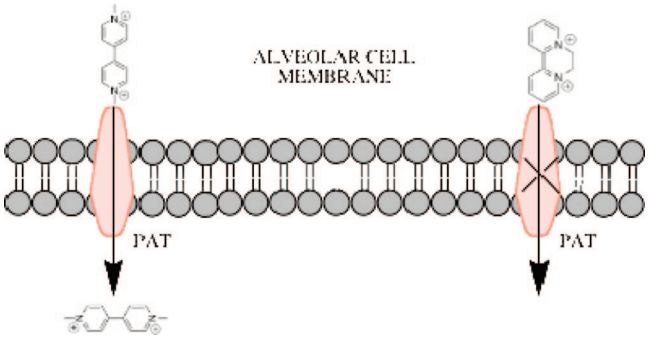



**Figure 10.** Polyamine transporters (PATs) responsible for Paraquat accumulation into cells do not have affinity for Diquat, which is much less toxic than its predecessor Paraquat.

clearance rate, extending the residency time of the chemical in the body.[73] Depending upon relative binding affinity, it may displace other xenobiotics bound to proteins, such as pharamaceuticals, which can in turn cause toxicity. Plasma protein binding is often reflected in the apparent volume of distribution ($V_D$), as discussed previously. Many acidic drugs (e.g., warfarin, aspirin) are highly protein-bound and thus have small $V_D$ values. Basic drugs, on the other hand (e.g., amphetamine, meperidine), leave the blood compartment and are taken up by tissues to yield an apparent volume of distribution larger than the volume of the entire body.[72]

*4.2.1.2.d. Metabolism and Elimination.* Two major pathways are available to a hydrophobic xenobiotic after it enters the body:

(1) It can partition into lipophilic compartments of the body, such as adipose tissue.
(1) It can be converted to a more hydrophilic compound and excreted.

Both pathways result in a reduction in the amount of free xenobiotic available for interaction with potential target tissues. As a general rule, xenobiotics with a log $D_{7.4} > 0$ require biotransformation to facilitate their elimination.[89] Some highly lipophilic compounds, like polyhalogenated biphenyls and chlorinated hydrocarbons, are resistant to significant biotransformation. They will tend to accumulate in the body upon repeated exposure as the alternative methods for excretion (mammary, bile, and intestinal excretion) are of limited capacity.

Metabolism can be thought of as the body's attempt to make lipophilic compounds more hydrophilic to facilitate excretion through the kidneys. The process generally involves enzymatic oxidation, reduction, or hydrolysis (phase I metabolism) and enzymatic conjugation with polar endogenous substrates (phase II metabolism). Numerous enzyme systems are involved: phase I is dominated by the cytochrome P450 (CYP) superfamily, whereas glutathione transferase, sulfotransferase, and UGP-glucuronosyltransferase are some of the primary phase II enzymes.

**4.2.1.3. Toxicodynamic Considerations for Designing Safer Chemicals.** If we look back to the three fundamental requirements for chemical toxicity, as illustrated in Figure 4, the third requirement is the direct ability to cause an alteration in the normal cellular biochemistry and physiology inside the organism through biomolecular interactions with cellular macromolecules. We do not have the scope to cover the numerous toxicodynamic mechanisms in this review, but full discussion of the current state of knowledge on these mechanisms can be found in excellent texts.[41,90]

We would like to devote some special consideration to mechanisms of bioactivation and detoxification, which are a consequence of metabolism. Metabolism can lead to bioactivation (or toxication) or detoxication depending on the nature of the chemical and the enzyme(s) involved in metabolism. Some decades ago, Brodie suggested[91] that chemically inert organic compounds may cause tissue lesions by the formation of covalent linkages between an activated metabolite and various cellular macromolecules. He supported his theory with experiments that showed that prior treatment of animals with phenobarbital (which induces cytochrome P450 activity) markedly increased the centrolobular hepatotoxicity of a number of aromatic hydrocarbons.[92] A similar mechanism has been invoked to account for the carcinogenicity activity of chemically inert substances, such as dialkylnitrosamines, azodyes, *N*-acetylaminofluorene, and polycyclic hydrocarbons.[93] For example, the insertion of an oxygen atom across a carbon-to-carbon double bond to form an epoxide typically makes the chemical electrophilic and increases the likelihood that it will react with cellular macromolecules and cause toxicity. The formation of free radicals, such as the one that occurs in peroxidase-catalyzed reactions with phenols and aromatic amines, can result in significant cellular toxicity. The phase II enzymes play a major role in the conjugation and detoxification of nucleophiles (by sulfation or glucuronidation) and electrophiles (by reaction with glutathione), many of which may themselves be the product of phase I metabolism. Superoxide dismutase plays a significant role in the elimination of highly reactive free radicals. Ultimately, it is the balance between activation and detoxification that determines whether a chemical will be toxic. Such a balance is dynamic and influenced by many factors.

We will focus on the Cytochrome P450 family of enzymes because it is responsible for the majority of oxidative reactions of xenobiotics. Keep in mind that some of the reactions result in bioactivation and others result in detoxification.

*4.2.1.3.a. Cytochrome P450 Metabolic Transformations.* CYP enzymes are responsible for about 90% of phase I metabolism.[94]Although there are 57 human CYPs,[95] fewer than a dozen seem to play a significant role in xenobiotic metabolism. One of the most prevalent CYP transformations, monooxygenation, involves the staged fission of a dioxygen molecule and subsequent oxygen atom insertion into a substrate (RH) to form an oxygenated metabolite (ROH) and a water molecule, according to the following scheme:

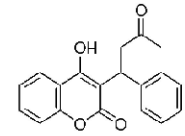
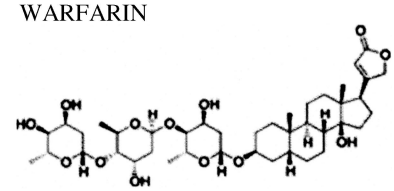
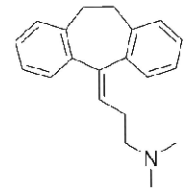
The two reducing equivalents (electrons) are supplied by either NADH or NADPH, involving either a flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN)-containing oxidoreductase or iron−sulfur redoxin, depending on the type of CYP system involved.

In order to design a chemical that either will not be bioactivated by CYP or will be metabolized to less toxic substances (detoxified), one must understand the following:

- The physicochemical parameters that influence the overall metabolic (or clearance) rate, expressed as $k_{cat}/K_m$.
- The selectivity that governs CYP-mediated hydrogen-atom abstraction (i.e., which substrate hydrogen atom is most likely to be abstracted by CYP), which will provide insight into the potential structure and thus toxicity of the metabolite.
- A xenobiotic-specific distinction between the CYP-mediated pathways responsible for detoxication and bioactivation.

We will first discuss the current state of knowledge in understanding the physicochemical properties that guide the relative reactivity of different substrates toward CYP and

**Table 2. Major Plasma Proteins in Humans and The Characteristic Compounds That They Bind**

| Plasma protein | Blood conc. g/100 mL[73] | Bind compounds with characteristics | Examples |
|---|---|---|---|
| Albumin | 50-65 | Acidic compounds | Salicylic acid, phenylbutazone,  WARFARIN  DIGITOXIN |
| α1-acid glycoprotein | 0.5-1.5 | Basic compounds |  AMITRIPTYLINE |
| Globulins (e.g. transferrin) | (γ, 3-6.5 β&$\alpha_2$) | Steroids, metal ions | $Fe^{3+}$, $Bi^{3+}$, $Ga^{3+}$, $In^{3+}$, $Sc^{3+}$, $Al^{3+}$, $Gd^{3+}$, $Th^{5+}$ |
| Ceruloplasmin | 5-9.5 | Metals | copper |
| α and β-lipoproteins | | Lipid-soluble compounds | vitamins, cholesterol, steroids |

bind.

more briefly address efforts to predict the selectivity of CYP oxidations using computed and experimental properties. Because the distinction between bioactivation and detoxification is derived from the CYP chemoselectivity toward the particular chemical, we will not have a direct discussion on how to make this distinction.

The reactivity of compounds with CYP is quite broad and, as many QSAR studies indicate, is governed by a complex combination of physicochemical and structural properties.[96] In order to clarify how some of these properties influence the monooxygenase reactivity, the process can be broken down into four steps,[97] as illustrated in Figure 11 below:

(1) The substrate's desolvation in the cellular matrix;
(2) The affinity of the substrate for the P450 active site, which is controlled by the substrate lipophilicity and its compatibility with the CYP active site structure;
(3) Its efficient turnover or intrinsic reactivity of individual C−H bonds in the substrate, which is largely determined by the C−H bond strength; and
(4) The constraints imposed by the active site on the oxidation reaction by orienting the substrate relative to the iron-bound oxidizing species and by restricting its mobility.

Naturally each of these steps is favored by different physicochemical and structural properties, and we will briefly explore them in sequence.

***Desolvation and Partitioning.*** The rate of desolvation, for example, is controlled by electrostatic interactions, polarity, H-bonding, $\pi-\pi$ stacking (hydrophobic interactions), and loss of substrate translational and rotational degrees of freedom upon solvation. The desolvation energy, $\Delta G_{desolv}$, and partitioning energy, $\Delta G_{part}$, are related quantities as they both deal with solubility in the cellular environment. $\Delta G_{desolv}$

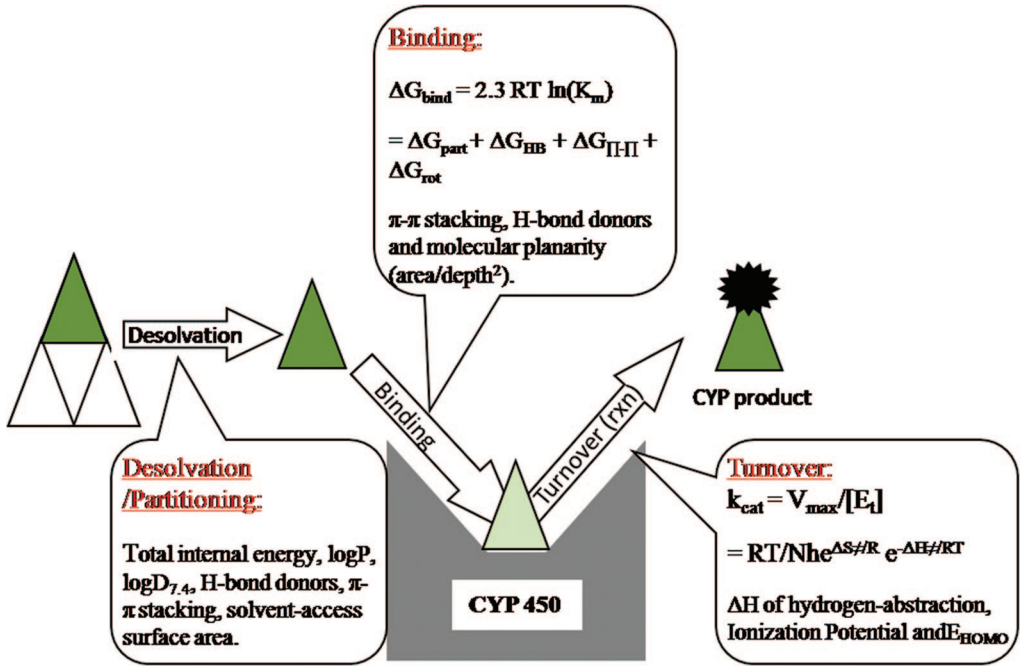



**Figure 11.** Influence of physicochemical and structural properties of substrate on reactivity with CYP.

**Table 3. Properties Used in Estimation of Substrate−CYP Binding Affinity ($\Delta G_{bind}$)**

| property | estimated average value in biological systems | coefficient from QSAR obtained for 90 substrates, $r^2 = 0.9737$[98] |
|---|---|---|
| hydrogen bonding energy $\Delta G_{HB}$ | −2 kcal/mol | $-1.925 \pm 0.039$ |
| $\pi-\pi$ stacking interaction energy $\Delta G_{\pi-\pi}$ | −0.9 kcal/mol for interactions involving 6-membered aromatic rings, less negative value for 5-membered ring systems[99] | $-1.027 \pm 0.043$ |
| loss of bond rotational energy $\Delta G_{rot}$ | +0.6 kcal/mol for each rotatable bond | $+0.600 \pm 0.016$ |
| ionic interaction energy $\Delta G_{io}$ | −4 kcal/mol | |
| loss of translational/rotational energy $\Delta G_{TR}$ | negligible with the exception of large CYP substrates, such as cyclosporine | 0 |

can be expressed in terms of the solvent-accessible surface area (SASA), which can be easily calculated[98] as shown in eq 4.

$$\Delta G_{desolv} = -0.025\text{SASA} \quad (2)$$

where SASA has units of Å$^2$.

The partition energy can be understood as desolvation of the active site that accompanies substrate binding and is strongly related to the octanol−water partition coefficient, $P_{ow}$, via the relation

$$\Delta G_{part} = RT \ln P_{ow} \quad (3)$$

where $R$ = gas constant and $T$ = absolute temperature (usually taken as 310 K). According to QSAR data,[98] the ln $P_{ow}$ term has an overall negative contribution to the total binding energy $\Delta G_{bind}$, in which $\Delta G_{part}$ is also included.

***Binding.*** As proposed by Lewis,[98] the contributions to overall binding affinity, $\Delta G_{bind}$, can be estimated using the following properties of CYP substrates, listed in Table 3.

Lewis includes desolvation energy in his treatment of binding, but for our purposes, we will treat it as a separate term. The loss of translational/rotational energy $\Delta G_{TR}$ is generally negligible with the exception of large CYP substrates, such as cyclosporin. A similar argument is made for the ionic interactions, as the majority of CYP substrates do not rely heavily on ionic interactions for binding. The experimental binding energy, $\Delta G_{bind}$, can also be thermodynamically related to the enzyme−substrate dissociation constant $K_D$ by the expression

$$\Delta G_{bind} = RT \ln K_D \quad (4)$$

Although not identical to $K_D$, under certain conditions of low substrate concentrations, $K_D$ can be approximated as the $K_m$ value (the apparent Michaelis constant),[100] and it is found that the equivalence between $K_m$ and $K_D$ is quite satisfactory for most substrates. Consequently, we substitute $K_m$ in eq 5

$$\Delta G_{bind} = RT \ln K_m \quad (5)$$

to obtain

As $K_m$ values tend to correlate well with log $P_{ow}$ (or desolvation parameters) in most cases, it is likely that the favorable entropic change associated with the substrate-mediated exclusion of water from the CYP P450 heme environment makes a major contribution to overall substrate binding energy; this is related to $K_m$, as has been described above.

Detailed investigations have also been made on the possible relationships between lipophilicity, in the form of either log $P_{ow}$ or as the ionization-corrected log $P_{ow}$ value at pH 7.4 (log $D_{7.4}$), and binding to human CYP enzymes that are associated with the metabolism of drugs and other

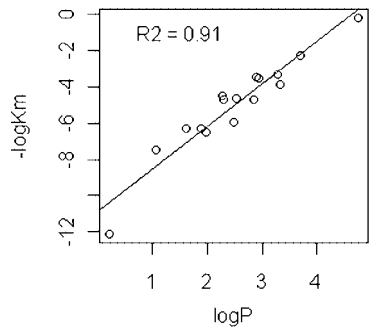
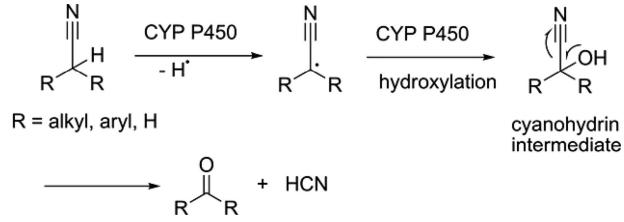



**Figure 12.** Plot of this correlation of log $P_{ow}$ and $-\log K_m$ for 16 ***CYP2B6*** substrates.[102] Linear equation: $\log K_m = 0.881 \log P_{ow} (\pm 0.058) + 1.676$, $R^2 = 0.91$. The $K_m$ values have been taken from a review of the relevant literature, while the log $P_{ow}$ data are either experimental or, where not available, calculated using the Pallas software (CompuDrug). Reprinted with permission from ref 102. Copyright 2004 Elsevier.

**Scheme 1**

$$RH + O_2 + 2H^+ + 2e^- \rightarrow ROH + H_2O$$

**Scheme 2. Proposed Mechanism for Cyanide Toxicity, Mediated by α-Hydrogen Atom Abstraction by CYP, Which Causes Cyanide Release**




xenobiotics.[101] A direct correlation of lipophilicity with $K_d$ has been repeatedly observed and has been incorporated into numerous Hansch relationships for individual series of compounds.[102] There have been several reports in the literature that point to the relevance of lipophilicity to CYP enzyme binding and inhibition.[103,104] log $P_{ow}$ and log $D_{7.4}$ both have also been correlated linearly to $K_i$, the inhibition constant of CYP.[105,106]

It has been shown that substrate binding affinities obtained from $K_m$ or $K_D$ data exhibit linear correlations with log $P_{ow}$, either expressed as the corresponding partitioning energy or as the raw log $P_{ow}$ values versus $-\log K_m$. For example, Lewis has shown that, when considering a set of 16 structurally diverse substrates of CYP2B6, a simple linear relationship between $-\log K_m$ and log $P_{ow}$ can be derived, as shown in Figure 12.[102]

Lipophilicity can thus be a useful indicator of CYP selectivity. Particular ranges of log $P_{ow}$ values for individual CYP isoforms can be rationalized in terms of variations in hydrophobicity and local dielectric constant of each CYP active site, and the slope of the lipophilicity relationship for substrates of a particular CYP can provide a measure of the degree of hydrophobicity in the environment of the heme.[101-103]

***Example: Nitriles.*** Although nitriles have broad commercial utility, many are known to produce symptoms similar to those of cyanide poisoning,[107] suggesting that cyanide release could be responsible for their acute toxicity (Scheme 2). DeVito and co-workers have developed a mechanism-based model for prediction of acute nitrile toxicity using octanol/water partition coefficients (log $P_{ow}$) and estimated rates of α-hydrogen atom abstraction by CYP as variables.[108] They have further shown that, for a series of nitriles, the higher rate constants for α-hydrogen abstraction by CYP are associated with increased toxicity (lower $LD_{50}$ in mice, Table 4).[109] This can be rationalized by understanding the stability of the α-radical formed after hydrogen abstraction. As expected, secondary nitriles are more toxic than primary nitriles, as 3° radical formed in this case is more stable than the 2° one formed on primary nitriles (Scheme 2).

**Table 4. Predicted Rate Constants for α-Hydrogen Atom Abstraction by CYP for a Series of Nitriles and Their Acute Toxicity Values for Mice ($LD_{50}$ in mmol/kg)[109]**

| αC | nitrile name | $k(\alpha)$ Cyt P450 | $LD_{50}$ mice (mmol/kg) |
|---|---|---|---|
| 3° | isobutyronitrile | 8288.5 | 0.37 |
| 2° | propionitrile | 251 | 0.65 |
| 2° | butyronitrile | 58.4 | 0.57 |
| 2° $sp^2$ | malonitrile | 5.4 | 1.8 |
| 1° | acetonitrile | 4 | 6.55 |
| 2° | 3-methylbutyronitrile | 0.98 | 2.8 |

*4.2.1.3.b. Turnover/Reactivity.* A key set of parameters associated with substrate oxidation by the substrate−CYP complex are the frontier orbital energies and the $\Delta H$ or $\Delta H^{\ddagger}$ for hydrogen-atom abstraction from the molecule. In fact, frontier orbital energies, especially the $E_{HOMO}$ and $E_{(HOMO-LUMO)}$, i.e., $\Delta E$ or excitation energy, have been identified as important electronic descriptors for determining the relative rates of CYP-mediated oxygenations.[110]

***Example: Nitrosamines.*** The tobacco-related nitrosamines, known to be potent carcinogens in rats, mice, and humans,[111] are believed to be activated by CYP2E1[112] via α-hydroxylation, as shown in Scheme 3.[111,113,114]

For a series of symmetric dialkylnitrosamine congeners, Lewis found that $TD_{50}$ correlates linearly ($R = 0.95$) with excitation energy, $\Delta E$, computed with CNDO/2 method, as shown in Figure 13.[111,115]

This result suggests that compounds with larger HOMO−LUMO gaps, i.e., $\Delta E$ values, will be activated at a faster rate by CYP. This data can also be rationalized mechanistically as an ability to form toxic intermediates, such as most carbonium ions, which is related to the electron donor/acceptor characteristics or activation energy of the molecule, that is, the $\Delta E$. This $\Delta E$ has previously been shown to be a contributing factor in biological activity and carcinogenicity.[117] In the case of the short-chain dialkyl nitrosamines studied in this report,[118] the greater the value of $\Delta E$ for symmetric dialkylnitrosamines, the greater is the carcinogenic potential ($R^2 = 0.90$) of the compound.

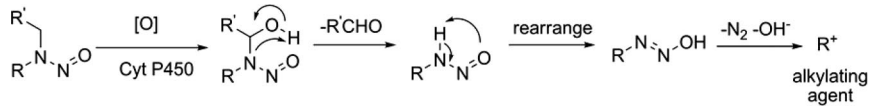
***Example: p-Subtituted Toluenes.*** Similarly, for a series of eight *p*-substituted toluenes, the CYP-mediated hydroxylation rates and binding affinities can be correlated exceptionally well with two properties: ionization potential (which is the negative of AM1-calculated $E_{HOMO}$) and calculated activation energies ($\Delta H^{\ddagger}$) for the formation of the hydrogen-abstracted species, represented below.[119]

$$\log k = 19.97 - 0.95\mathrm{IP} - 0.02H^{\ddagger} \qquad (6)$$

where $n = 8$, $R = 0.99$, and $k$ = rate constant for P450-catalyzed hydroxylation in rabbit liver microsomes.

While the above examples illustrate that correlations between electronic properties and CYP reactivity do exist, few or no studies have focused on understanding what the values of the property limits associated with high reactivity

**Scheme 3**




are across a larger range of compounds. However, evidence suggests that such limits can be defined.

For example, parabolic relationships between CYP activity and electronic properties of substrates have been reported. AM1 HOMO energies of a series of monosubstituted benzenes show a parabolic correlation with the relative rates of CYP-mediated aromatic nitration (Figure 14). The parabolic relationship indicates that the highest reactivity can be associated with a fairly narrow range of HOMO energies of −9.2 to −8.7 eV.[116]

Another example of parabolic behavior has been reported for the relationship between the electron reuptake rate constant ($k_e$) and $\Delta E$. The relative electrophilicity, which is related to carcinogenic potency, is measured as the ability to pick up solvated electrons.[119] The electron reuptake rate constant $k_e$ was correlated to the $\Delta E$ of 64 miscellaneous chemicals, as shown by the parabolic curve of Figure 15.

A relationship similar to that shown in Figure 15 was found between bacterial mutagenicity of 20 nitrogenous compounds formed during the cooking of meat products and their corresponding $\Delta E$ values (Figure 16). This similarity suggests that, as expected, there are commonalities in mechanism between mutagenicity and carcinogenicity. It is possible that these quadratic expressions in $\Delta E$ correspond to a Marcus-type relationship governing electron-transfer processes.[120] If so, it is to be expected that the combined parameter $\Delta E$ may be promising as a descriptor for mutagenicity, although additional factors, such as lipophilicity and nitrene stability in the case of nitrogenous compounds, are also likely to be play a role in mutagenic potential, particularly as far as nitrogenous (hetero) aromatic bases are concerned.

Figure 16 illustrates that when establishing such relationships it is important to examine compounds with a large range of $\Delta E$ values. For example, if either compounds with low $\Delta E$ values or high $\Delta E$ values were excluded from this analysis, the parabolic relationship would appear linear.

In addition to frontier orbital energies, predicted bond strengths in CYP substrates have also proven to be invaluable tools for predicting CYP selectivity and reactivity. This is because the intrinsic reactivity of a given C−H bond can be defined by its bond strength, which is in turn defined as the energy required to break the C−H bond homolytically. Because the energy of the hydrogen radical is constant, the bond strength is determined by the stability of the carbon radical that is generated in the reaction.[97] As shown by the bond strengths listed in Table 5, this value approximation predicts the following order of C−H bond reactivity: benzylic or allylic > tertiary > secondary > primary. Vinylic hydrogens have too high a bond strength to undergo direct hydroxylation. This general order is observed when C−H bond oxidation is controlled by intrinsic reactivity rather than by steric constraints or positioning of the substrate within the active site.

Bond strengths have been shown to be useful predictors of CYP selectivity. By comparison to aromatic oxidation,

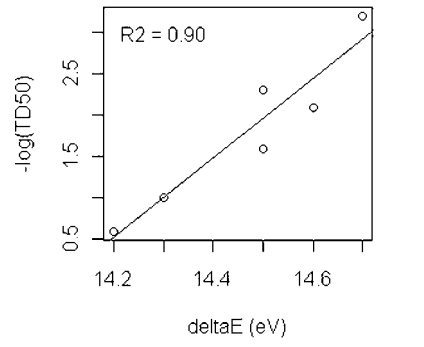



**Figure 13.** Dependence of carcinogenicity (−log $TD_{50}$) on $\Delta E$ (in eV) for six symmetric dialkyl nitrosamines, $R^2 = 0.90$.[111] Reprinted with permission from ref 116. Copyright 2004 Elsevier.

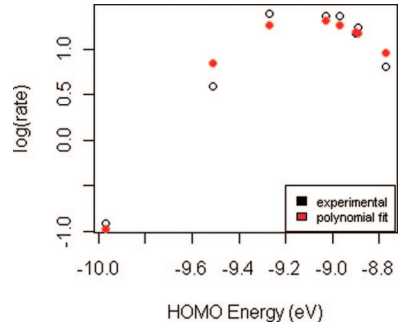



**Figure 14.** Plot of the parabolic relationship between log(relative rate) of aromatic substitution and AM1 $E_{HOMO}$ for eight monosubstituted benzene derivatives, $R^2 = 0.971$.[116]

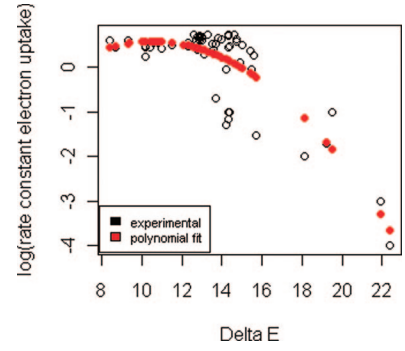



**Figure 15.** Plot of log(electron uptake rate constant) ($k_e$) versus $\Delta E$ for 64 miscellaneous chemicals. log $k_e = 0.64\Delta E$ (±0.15) − $0.03\Delta E^2$ − 2.77 (±0.005) where $n = 64$ and $R = 0.84$. $\Delta E = E_{LUMO} - E_{HOMO}$, where $E_{LUMO}$ and $E_{HOMO}$ are the CNDO/2-calculated frontier orbital energies (eV). $k_e$ = rate constant for electron reuptake.[116]

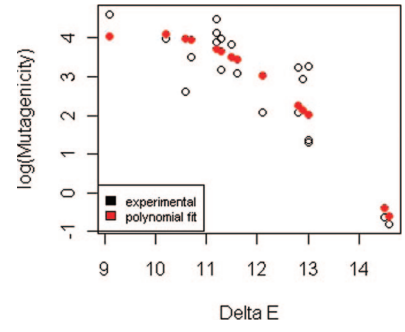



**Figure 16.** Plot of log(bacterial mutagenicity) versus $\Delta E$ for 20 cooked food mutagens.[116,121]

**Table 5. Relative Energies and C−H and O−H Bond Lengths in the Transition State for Hydrogen Abstraction, Ordered from Highest to Lowest (Reproduced with permission from ref 110. Copyright 2006 American Chemical Society.)**

| entry | substrate[a] | energy[b] (kJ/mol) | C−H (Å) |
|---|---|---|---|
| 1 | methane | 86.7 | 1.39 |
| 2 | propane(1) | 73.9 | 1.35 |
| 3 | propane(2) | 62.0 | 1.32 |
| 4 | isobutene | 59.7 | 1.31 |
| 5 | propene | 53.9 | 1.31 |
| 6 | propionaldehyde | 47.9 | 1.35 |
| 7 | toluene | 54.6 | 1.31 |
| 8 | ethylbenzene(2) | 50.6 | 1.29 |
| 9 | 1-methylethylbenzene | 55.8 | 1.30 |
| 10 | dimethylether | 50.9 | 1.31 |
| 11 | dimethylsulfane | 45.9 | 1.33 |
| 12 | methyl(phenyl)sulfane | 45.4 | 1.34 |
| 13 | dimethylamine | 31.9 | 1.27 |
| 14 | trimethylamine | 27.9 | 1.27 |
| 15 | fluoroethane(2) | 77.2 | 1.38 |
| 16 | fluoroethane(1) | 61.6 | 1.34 |
| 17 | ethylbenzene(1) | 72.2 | 1.35 |
| 18 | 2-fluoroprop-1-ene | 55.2 | 1.33 |
| 19 | prop-1-en-2-ol | 49.1 | 1.32 |
| 20 | *p*-xylene | 53.0 | 1.31 |
| 21 | 1-methyl-4-nitrosobenzene | 49.5 | 1.33 |
| 22 | methoxybenzene | 54.5 | 1.34 |
| 23 | *N*-methylaniline | 31.9 | 1.30 |
| 24 | *N*,*N*-dimethylaniline | 28.9 | 1.29 |

[a] Numbers in parentheses for the substrates indicate the position from which the hydrogen is abstracted. [b] The energies were determined at B3LYP/6-311++G(2d,2p)//B3LYP/6-31G(d) level, with the zero-point vibrational energy at the B3LYP/6-31G(d) level included. All energies are relative to the sum of the energies of the isolated substrate and the compound I model.

aliphatic oxidation can be predicted more easily both in terms of selectivity and reactivity. For example, the Olsen group has studied 24 aliphatic hydrogen-abstraction substrates using the Fe(porphine)($SCH_3$)O model of CYP compound I, the heme-iron(IV)-oxo porphyrin radical cation formed in peroxidase and catalase enzymes by reaction with hydrogen peroxide.[110]As shown in Table 5, the amines were found to have the lowest activation energies for the hydrogen abstraction, 28−32 kJ/mol, followed by ethers and thioethers (45−55 kJ/mol), which were followed by compounds whose reactive carbon was next to $sp^2$-hybridized carbons (48−56 kJ/mol range). Substrates with only $sp^3$-hybridized atoms have the highest barriers. Abstractions from secondary and tertiary carbon atoms or from carbons bound to F atoms have activation energies of 60−62 kJ/mol, whereas abstractions from primary carbon atoms (entries 1, 2, 15, and 17) have even larger activation energies (>72 kJ/mol). Most importantly, these computed activation energies follow the qualitative stability of the radicals, as has been noted before.[122]

***Selectivity.*** The selectivity of CYP substrate oxidation is determined by many factors, such as the nature of the rate-limiting step, docking-related steric effects, and intrinsic electronic reactivity.[123] A classic example is oxidation of substituted benzene. Benzene ingestion is known to cause hematopoietic toxicity, such as bone marrow damage and leukemia.[124]To circumvent the initial bioactivation step, a more readily oxidizable C−H bond can be incorporated in the molecule, such as a benzylic methyl group (i.e., toluene). It has been proposed that the bioconversion of benzenes to toxic intermediates proceeds through formation of epoxides[125] that can covalently bind nucleophiles, such as DNA or protein residues, irreversibly modifying vital cellular macromolecules and triggering biochemical pathways leading to cell death, as shown in Scheme 4. In addition, the phenols can get further oxidized to bisphenols and quinines, which can cause serious oxidative damage to cells through radical pathways or can alkylate N- or S-nucleophiles, such as glutathione and glycine (Scheme 4).

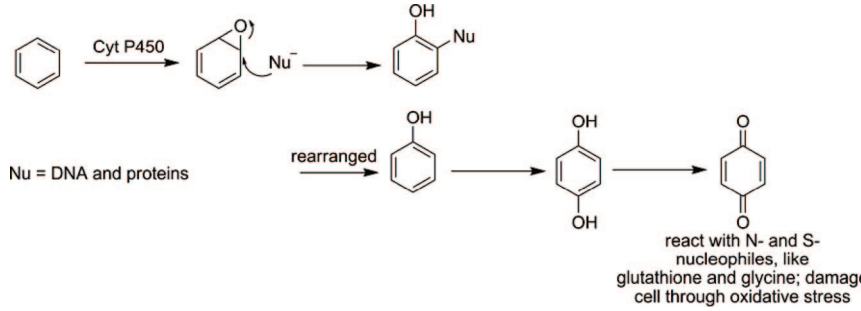
To understand selectivity, much effort has been devoted to computationally probing the mechanism of hydroxylation and specifically the nature of the radical or cationic intermediates. Aliphatic CYP-mediated hydroxylation appears in general to have been studied more than aromatic hydroxylation and is known to proceed via radical intermediates. The mechanism and selectivity of aromatic hydroxylation has been explored computationally using B3LYP DFT, and it was shown that the rate-determining step of the reaction between the enzyme and an aromatic carbon atom proceeds via a transition state with partial radical and cationic character. Reactivity is further shown to depend strongly on ring substituents, with both electron-withdrawing and -donating groups strongly decreasing the barrier for addition to the para position, thus enhancing the reaction.[126]

The difficulty in predicting aromatic regioselectivities is illustrated by the study of the hydroxylation of fluorobenzenes. In vivo metabolites of fluorobenzenes were studied by identifying various phenols in the urine of rats. The regioselectivities observed for the aromatic hydroxylation were shown to correlate with the values predicted for regioselectivity on the basis of HOMO/HOMO-1 frontier orbital density with only 6% accuracy. Regioselectivities predicted on the basis of a CYP P450 hydroxylation proceeding through (i) an initial nucleophilic attack on the benzene's LUMO/LUMO+1 or (ii) an initial electron abstraction followed by an attack of the $(Fe0)^{2+}$ on the benzene cation radical singly occupied molecular orbital (SOMO) cannot correctly predict the regioselectivities of all five fluorobenzenes tested.[127]

More recently, Olsen and co-workers have been able to obtain good correlations for the relative density functional theory (DFT)-calculated activation energies of aromatic oxidation of substituted benzenes with experimental rates and to predict the selectivities (product ratios) of methyl-substituted 3-fluoroanilines (Figure 17). Their results further show that the site of oxidation is in general determined by two factors: the solvent accessibility and the intrinsic reactivity (the activation energy) of the site. For the methyl-substituted F-anilines, substituted phenols, and the three druglike molecules, all but one of the reactions occur for carbon atoms with SASA > 18 $Å^2$, and among these, the reactions with the lowest activation energies are also observed experimentally. In fact, the expression $E^{\ddagger}_{react} - 3SASA$ is shown to be able to correctly predict the site of metabolism in all tested systems.[128]

Above are just a few examples of the numerous structure/property studies on CYP that can be used to inform chemists of potential molecular property ranges that are most likely to be associated with minimal CYP interaction. Attempts to carry out more complex analyses that can help us identify characteristics of a broad range of chemical classes that disfavor CYP transformations have also been carried out[129] using machine learning techniques, such as support vector machine (SVM), random forest, decision tree, and kappa nearest neighbors ($\kappa$NN).

**Scheme 4**




## 5. Quantitative Structure−Activity (and Toxicity) Relationships

### 5.1. Historical Development of QSARs

The relationship between chemical structure and biological activity has drawn the attention of many investigators since the end of the 19th century. As early as 1893, Richet had stated that "the more soluble [alcohols and ethers] are, the less toxic they are".[130] Several years later, Meyer[131] and Overton[132] proposed to use oil−water partition coefficients to explain the difference in narcotic activity of several substances. It has been recognized for decades that partition coefficients can be good predictors of biological activity.[133] Ferguson later delivered a thermodynamic interpretation of the observed data, connecting baseline toxicity with chemical activity, which is in turn proportional to the lethal body burden.[134] In his opinion, it was not the ultimate concentration of a substance in the fish that is most important, but its "chemical potential" (a thermodynamically defined quantity), which can be measured outside of the organism, in an equilibrium situation. This approach was restricted to substances with limited chemical reactivity, substances with a so-called "physical action". Soon after, McGowan recognized the importance of the molecular volume and hydrogen bonding to biological activity,[135] and Mullins expanded the thermodynamic treatment.[136] Hansch and co-workers introduced the partition coefficient between octanol and water ($K_{ow}$ or $P_{ow}$) as an additive parameter describing hydrophobicity and extended the QSAR equations into linear free energy relationships.[137,138] The $P_{ow}$ has since then remained the most frequently used parameter in QSAR studies.[139]

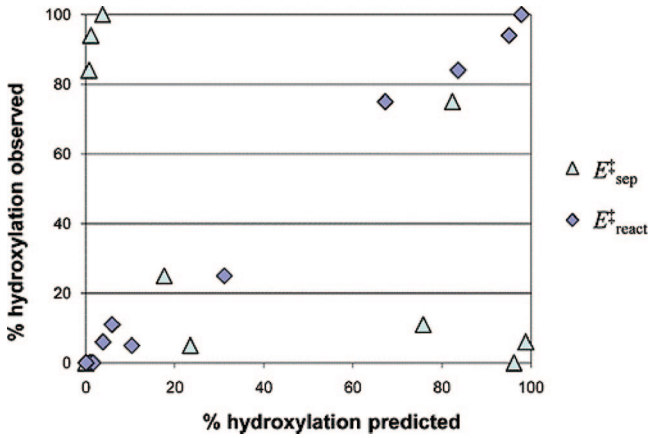


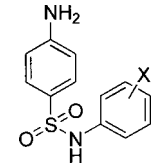

**Figure 17.** Predicted product ratios for methyl-substituted 3-fluoroaniline compounds based on the activation energies calculated from the B3LYP methoxy-radical model. It is assumed that the product ratios are proportional to the ratios of the rate constants ($k$) and that the rate constants are related to the activation energies according to $\Delta E^{\ddagger} = -RT \ln k$ ($T = 310.15$ K). (Reproduced with permission from ref 128. Copyright 2008 American Chemical Society.)

Hansch's general equation states:

$$\log 1/C = k_1 \log \mathrm{P} - k_2(\log \mathrm{P})^2 + k_3 \mathrm{p}K_a + k_4 E_s + \cdots + k_n \quad (7)$$

where $C$ = concentration of a substance required to produce a certain biological response, e.g., lethality in 50% of the exposed population (i.e., $LC_{50}$; $LD_{50}$); P is a partition coefficient (usually in the $n$-octanol/water system); $K_a$ is the acid dissociation constant; $E_s$ is a steric parameter; and $k_1-k_n$ = coefficients obtained by fitting the equation to the experimental data.

An early example of this application of linear free energy relationships connected Hammett's electronic descriptors to QSARs relating the inhibition of bacterial growth by a series of sulfonamides,




where X represents various substituents.[140] A QSAR was developed based on the $\sigma$ values of the substituents (eq 8).

$$\log\left(\frac{1}{c}\right) = 1.05\sigma - 1.28 \quad (8)$$

Most QSAR applications lie in the field of drug and pesticide research, but for a number of years, QSARs have also been applied to predictive aquatic toxicology.[141] It is generally considered that QSARs can be applied only to groups of chemicals that share the same mechanism of biological action.[142] The QSAR or quantitative structure−toxicity relationship (QSTR) models that have greatest applicability to developing heuristic rules for designing safer chemicals will be addressed here.

QSAR models, for the purpose of this discussion, can be divided into two groups—models that have been put forward to directly describe unicellular or multicellular whole-organism toxicity of specific classes of compounds and models that describe activity in specific biochemical pathways associated with toxicity, such as a single enzyme. Since the focus is the derivation of design rules that encompass whole-organism toxicity, we will limit our discussion to this class of QSARs. Excellent reviews are available elsewhere[143,144] that discuss various biochemical pathways.

It should be noted, however, that although QSARs are extremely useful mathematical tools, their application to a problem must be preceded by a qualitative understanding of the factors that influence the biological response under investigation. This qualitative understanding facilitates the interpretation of the models obtained from QSARs and alerts to nonsensical results.

## 5.2. Advancement of Whole-Organism QSTR Models

Whole-organism quantitative structure−toxicity relationship analysis has become a commonly used tool in toxicology modeling, especially in ecotoxicological risk assessments. A successful QSAR or QSTR model is largely determined by the identification of effective descriptors that are related to the chemicals' ability to cause an adverse biological effect and by the use of appropriate modeling methods. The biological parameter, the dependent variable used in describing whole-organism toxicity, can be expressed as $\log(1/C)$, where $C$ is the molar concentration or dose of compound that induces a particular biological effect, e.g., $LC_{50}$ (molar concentration that induces 50% lethality), $LD_{50}$ (molar dose which induces 50% lethality), $EC_{50}$ (molar concentration that impacts a defined biological response by 50%), or $IC_{50}$ (molar concentration that causes 50% inhibition in growth of cells).[145] The dominating parameters in whole-organism toxicity models tend to be solubility and partitioning descriptors, which strongly influence the bioavailability and bioaccumulation.[146] Recently, much effort has been directed toward increased understanding of ecotoxicity and human toxicity using global QSAR models. Konemann showed that aquatic toxicity in *Poeciliareticulata* (guppies) can be modeled by QSARs that use only $P_{ow}$ as a physical parameter for about 70 hydrophobic industrial pollutants.[147] $P_{ow}$, therefore, accounts for a significant part of the variation in toxicity.

Recently, attempts have been made to correlate whole-organism acute toxicity with computationally derived molecular descriptors. Group contribution methods represent one such approach and have been used by Martin and Young to correlate the acute toxicity (96-h $LC_{50}$) to the fathead minnow (*Pimephales promelas*) for 397 organic chemicals.[148] Using multilinear regressions and computational neural networks (CNNs) for model building, they have achieved a fairly good data correlation of $R^2 > 0.9$. Sophisticated statistical methods, such as neural and fuzzy-neural networks, are also increasingly being applied to correlating toxicity directly to a diverse set of molecular descriptors and facilitating effective modeling of toxicity.[149] Numerous other useful QSAR models have been developed for compounds grouped mostly by functional properties, such as substituted benzenes,[150] carboxylic acids,[151]and alcohols,[152] among others. In addition, extensive computerized databases of QSAR data are available[139] that can be used both as lateral validation of existing QSARs and also to derive new QSAR models for the derivation of heuristic rules that describe toxicity in terms of physicochemical descriptors.

It is thus clear that medicinal chemistry tools, such as existing and emerging QSARs, will provide a rich resource for the extrapolation of rules governing human and ecotoxicity. However, it is the analysis of this available data and its morphing into explicit design rules that will prove challenging for several reasons.

Many 2D or 3D QSAR models rely on multiple variables (properties or attributes) to build the model, and this makes them difficult to qualitatively interpret as design guidelines. Many QSARs use relatively small training sets of structurally similar compounds. While this often allows a robust linear model to be built, it does not facilitate design guidelines that can be applied to broad classes of chemicals.

This means that using existing QSAR models to develop design rules is likely to be challenging. It is proposed, however, that new QSAR-like analyses of chemicals from a range of chemical classes should be pursued as a means of deriving molecular design guidelines. Some of the novel statistical techniques that are being applied to the field of drug discovery, such as partial least-squares (PLS), naïve Bayes classifier (NBC), $\kappa$-nearest neighbor ($\kappa$NN), self-organizing map (SOM), recursive partition (RP), artificial neutral network (ANN), support vector machine (SVM), and machine learning are likely to also find great utility in deriving design rules. Excellent descriptions of these methods are available.[96]

## 5.3. QSARs for General Versus Specific Toxicity

To forward our discussion of the available medicinal chemistry tools that can be applied to molecular design, it is necessary to distinguish between general toxicity (or narcosis) and specific (or reactive) toxicity. General toxicity, narcosis, or baseline toxicity is considered the minimum toxicity exhibited by organic compounds and occurs by nonspecific disruption of the functioning of the cell membrane. Careful examination of the different symptoms caused by a wide variety of narcotic chemicals, however, suggests the possibility that there are numerous specific mechanisms of narcosis.[153]

Nonspecific toxicity can be thought of as a reversible state of arrested activity of protoplasmic structures caused by a wide variety of organic chemicals and is characterized by progressive lethargy, unconsciousness, and death without any specific sustained symptoms such as hyperventilation, erratic or convulsive movement, or hemorrhage.

Specific toxicity, on the other hand, refers to chemicals that interact with or disrupt the function of a defined receptor site, in addition to disrupting the cell membrane. Examples of chemicals that act by specific mechanisms of toxicity include nucleophiles, electrophiles, respiratory inhibitors, acetylcholinesterase inhibitors, and central nervous system seizure agents.[42] The reason this distinction is important in understanding QSARs is that, for nonreactive chemicals, the baseline toxicity is expected to be broadly additive, implying that the QSARs can be applied to a diverse set of chemical compounds. Reactive chemicals that cause specific toxicity, on the other hand, have to be treated more carefully on a case-by-case basis, as their toxicity is often related to parameters that are most relevant to the mechanism of toxicity.

# 6. In Silico Approaches

There is little doubt that there is a dire need for reliable computer-based (or in silico) predictive toxicology tools that would eventually replace whole-organism testing. Not only is animal testing prohibitively expensive both for government and industry, but the current protocols do not necessarily address all of the toxic end points and exposure scenarios that are of interest. Moreover, numerous cases show the disparity between rat, mice, and dog toxicity profiles and

those in humans. Finally, animal welfare concerns have led to a reconsideration of the value of a toxicology assessment regime primarily built upon extensive animal testing. For the above reasons, development of in silico methods of estimation of toxicity from chemical structures began in the 1980s and has since made considerable strides, as we will show here. For an in-depth analysis, the reader is directed to an excellent resource by Cronin and Livingstone entitled "Predicting Chemical Toxicity and Fate".[154]

Owing to the complexity inherent in toxicological assessments, however, the development of computational models for various toxicities has long been considered highly complicated, and the resulting models had limited utility for whole-organism toxicity prediction. Nevertheless, toxicity models have been implemented that compute a wide range of toxicological end points with varying accuracy. In addition to the complexity of toxicology, the limited availability of toxicological data in a structured, computer-readable format has also added to the problem.[155] This challenge has partially been met with new technologies that have emerged in the past two decades, such as high-throughput screening (HTS) and high-content screening (HCS), both of which have yielded large amounts of chemical and biological data.[156]

Microarray experiments allow the quantification of changes in expression levels of all genes in an organism resulting from the administration of a chemical agent. New interdisciplinary databases have emerged that hold a plethora of data—from protein assays to whole organisms.[157]

It is the combination of all these complementary types of data that is needed to build more realistic models of the potential toxic effects of a compound on an organism. The main driving force for computational toxicology is, therefore, the increasing amount of quantitative data held in regulatory, public, and industry archives on the effects of chemicals on biological systems that have become available over the past decade.[158] We must, however, stress that although the fields of predictive toxicology and a priori molecular design for safety are far from identical, they share a common foundation and driving force—increasing amount of available quantitative toxicity data—and as such there is a lot we can learn from predictive toxicology that can be directly applied to molecular design.

## 6.1. Estimation of Toxicity

Several types of methods have been developed for estimation of toxicity, but they are mostly directly or indirectly statistics based. The in silico approaches to toxicity predictions have been classified by some as QSAR-based and so-called "expert systems",[2] with the latter technically defined as a program that mimics the judgment of experts by following sets of knowledge rules. These "knowledge rules" are derived based on studies of toxicity mechanisms in animals and humans. Others[159] have classified them as knowledge-based systems (KBS) versus automated rule induction (ARI) systems, based on whether the system is fed human knowledge (KBS) or whether it derives its own prediction rules based on data-derived patterns. While some of the in silico toxicity prediction tools fall clearly in one category (e.g., TOPKAT), others, especially newer models that use combinations of these approaches, are more ambiguous (e.g., MULTICASE). Nevertheless, for the purpose of organizing the presented information and highlighting the aspects important to designing safer chemicals, we will utilize the second classification—KBS versus ARI systems.

### 6.1.1. Automated Rule Induction (ARI) Systems Approach

Chemicals of known toxicity are used to form training sets for particular biological end points by fragmentation into all possible atom pairs. Pattern-recognition techniques and statistical (QSAR) analyses are then used to compare the frequency of occurrence of specific structural features in sets of active and inactive compounds. From this analysis, the features most essential to biological activity are identified. This allows the identification of toxicophores (fragments directly responsible for toxicity) in novel molecules. It should be noted that the QSAR-derived relationships are sometimes not validated with up-to-date data, and systematic improvements to the systems are often rather challenging. This makes ARI's good for specific models that have rich data sources but not as good for general models.

**TOPKAT** (toxicity prediction by (k)computer assisted technology), now part of Accelrys Inc., was originally developed by Enslein and Craig in the 1970s.[160,161] TOPKAT quantifies electronic, steric, and shape attributes of a structure in terms of electrotopological state (E-state) values for all possible two-atom fragments from 2D molecular structures. Atomic size-adjusted E-states are computed from the rescaled count of valence electrons, molecular weight, topological shape indices, and symmetry indices. This methodology can be considered an extension of the classic QSARs, yielding robust QSTR models for assessing specific toxicological end points. A stepwise regression and discriminant analysis are used to search out mathematical relationships between the structural descriptors and toxicity. Molecular descriptors were related to the oral $LD_{50}$ of the compound in rats, resulting in the following relation,

$$(1000 \cdot MW)/LD50 = \sum [coeffn*FG] \qquad (9)$$

where coeffn = calculated coefficients for each functional group, FG = value of 0 or 1 depending on whether a particular functional group is present in the molecules, MW = molecular weight of molecule, and $LD_{50}$ = dose that causes 50% mortality in rats.

The rules are devised by integration of toxicological knowledge, expert judgment, QSAR models, and neural network logic. HazardExpert is similar to TOPKAT, except it derives the parameters it uses from the entered molecular structure. It uses a program called METABOLEXPERT to examine possibility of toxic metabolites. The utility of HazardExpert to the molecular designer is that it also provides estimates for bioavailability and bioaccumulation. Reducing bioavailability, as discussed previously, is one of the primary goals in the design of safer chemicals.

**TOXBOXES** is available as a new stand-alone program or as a free online tool that predicts three basic toxicity end points: acute toxicity, genotoxicity, and organ-specific health effects. Its sister application, ADME Boxes, is also available from the same location and gives insightful property analysis relating to bioavailability. Toxic effects of molecules are predicted by ToxBoxes solely from the chemical structure by using validated databases and QSAR models. A reference compound dictionary, which provides the available experimental data of related compounds, is also quite useful, enabling the user to compare the query structure with related compounds. ToxBoxes is one of the fastest tools that any chemist can freely access for quick toxicity predictions. ToxBoxes has already been used in scientific publications

**Table 6. Selected Common Toxicity Predicting Tools**

| program | type | originators | now licensed by | link |
|---|---|---|---|---|
| TOPKAT | ARI | Enslein, 1970s | Accelrys Inc. | http://accelrys.com/products/discovery-studio/ |
| LAZAR | ARI | Helma, 2006 | In silico Toxicology (Free) | http://www.predictive-toxicology.org and http://lazar.in-silico.de |
| DEREK | KBS | Schering Agrochemicals | Lhasa Ltd. (nonprofit) | http://www.lhasalimited.org |
| ADAPT | KBS | | Compudrug | http://www.compudrug.com |
| HazardExpert | KBS | | | |
| OncoLogic | KBS | | EPA (Free) | http://www.epa.gov/oppt/newchems/ |
| MULTICASE | ARI/KBS | Klopman, Case Western Reserve Univ., 1980s | Multicase | http://www.multicase.com |
| ToxBoxes | ARI | | Pharma Algorithms Inc. (Free web-based version available) | http://www.pharma-algorithms.com/webboxes/ |
| TOXTREE | ARI | Ideaconsult Ltd. (EU REACH) | European Commssion | http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE |

for risk analysis,[162] but the community has not had a chance yet to thoroughly test and validate its predictive capabilities.

### 6.1.2. Knowledge-Based Systems (Expert Systems)

These systems can be distinguished from the ARIs by focusing on structural alerts in molecules that indicate toxicophores. The accuracy of the system is determined to a large extent by the quality of the toxicity data used to build the rules. The advantage of these systems is that there is some human judgment built in, and the rules are backed-up by an understanding of the mechanism of toxicity in most cases. Some systems even provide references and predicted mechanisms of toxicity. The consequent disadvantage is that these systems are often not quantitative but qualitative, often predicting the toxicity type and fate of the compound. The most popular systems that fit into this category are described below.

**ADAPT** is a statistical-based methodology developed in academia for the correlation of chemical structure with molecular properties. It uses more modern statistical manipulations than TOPKAT, such as pattern recognition and neural networks. The structures are entered in the form of connection tables (2D structures). The program generates structural descriptors, physicochemical parameters, elemental composition, and molecular shape. On the basis of these descriptors, toxicity classifiers are produced. The program has an open-source code in Fortran.

**DEREK** (deductive estimate of risk from existing knowledge) is a rule-based expert system that assimilates a chemical structure, decomposes it into substructural fragments, and assesses the toxic potential of each fragment and, hence, to the complete molecule.[163] It addresses many toxicity end points including carcinogenicity, mutagenicity, genotoxicity, teratogenicity, hepatotoxicity, and neurotoxicity, among others. Its originators, Sanderson and Earnshaw, stated DEREK's premise as follows: "if chemical structure is present, then specific toxic action is a possibility".[163] Thus, DEREK indicates whether a specific toxic response may occur and does not provide a quantitative estimate for the prediction. Because substituents can exist in a variety of molecular contexts, the rules are not chemical-specific but rather serve as broad generalizations with regard to the chemical structure. DEREK's fragment rule base is open-ended—many of the rules represent a distillation of much toxicological experience, such as FDA structure alerts for carcinogenicity, while others are more statistical. When the program encounters a structure that is inadequately covered by its rules, however, it does not return a prediction. The fact that the tool recognizes its limitations can be viewed as an advantageous feature.

**TOXTREE**, a tool for predicting different types of toxicological hazards and modes of action by applying decision tree approaches, can be used for the grouping and hazard profiling of chemicals.[164,165] Toxtree identifies the types (and in some cases the levels) of potential toxicological effects. For example, to support human health risk assessments, Toxtree identifies the so-called Cramer class, which represents the degree of potential oral systemic toxicity. The Cramer classification scheme,[166] which is based mostly on structural rules, is perhaps the best known structure-based approach for estimating thresholds of concern.

**MCASE** (multiple computer automated structure evaluation) is now a combination of several modules marketed by the company Multicase. The original was called CASE, and its successor MCASE/MC4PC incorporates features for organizing toxicological data obtained from evaluation of diverse chemicals. This program is somewhat unique in that, although classified under KBS, it is a statistical analysis-based hybrid program, as it combines 2D-QSAR with an expert-based program structure. The module decomposes chemical structures into 2−5 carbon fragments, and discriminant analysis followed by linear regressions is used to quantitate properties such as $LD_{50}$ values. The result of toxicity prediction is also presented in CASE units, from 10−100, with higher values indicating increasing activity. These CASE units are essentially sums of the activities of the molecular fragments. The output of the program gives a designation of active or inactive fragments,[167] which is a useful learning tool for molecular designers. One of the strengths of this model over KBS is its ability to predict toxicity profiles for chemicals whose mechanism of action is still unknown.

**OncoLogic** is a powerful EPA-developed expert system[168] that is capable of dealing with a variety of materials beyond organics—metals, metalloids, inorganics, fibers, and polymers. OncoLogic is thus composed of 4 modules—metals, polymers, fibers, and 48 classes of organics. The code is a logic-based decision tree structure, where at each level a compound must be categorized. Compounds in particular subgroups are then subjected to expert rules, which gives rise to a carcinogenicity rating based upon the most current knowledge available. This is one of the most highly regarded systems for mutagenicity prediction and has been thoroughly validated.

The links to the programs discussed are listed in Table 6.

Chemical grouping methods, such as TOXTREE, are sometimes classified separately but can also be considered to be a type of expert system. Irrespective of the size of the group, chemicals are selected based on the hypothesis that they will show common physicochemical properties, as well as toxicological (human health/ecotoxicity) effects or environmental fate properties. The formation of chemical categories provides significant efficiencies and benefits when identifying and filling data gaps.

Apart from the list of tools listed in Table 6, several impressive efforts provide online searchable tools for predictive toxicity based on QSARs. One is the European Comission's Computational Toxicology Group's QSAR resources (available at http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/). Another, from the U.S. EPA, also provides a similar resource (EPISUITE, available at http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm).

It is clear that, although numerous predictive tools and expert systems are available, no single system has proven to be able to replace in vivo testing. All have their strengths and weaknesses, with varying sensitivity (ratio of correctly predicted toxic molecules to total number of toxic molecules), specificity (ratio of correctly predicted nontoxic molecules to total number of nontoxic molecules), and accuracy. These programs have been evaluated both by industry and government, and a maximum accuracy of 65−70% was attained for noncarcinogenic chemicals in most of the expert systems, which, although encouraging, is still unable to replace in vitro and in vivo testing in most cases.[161] In addition, the applicability domains of the ARIs require that the test compound be as similar as possible to the training set, which is always descriptor-dependent for each specific case. Thus, the descriptor must be mechanistically related to the predicted end point. We can and should, however, still use these programs as useful tools to guide molecular design.

The expert "rules" of KBS systems can be particularly useful for deriving design rules. Revealing these expert rules would give chemists a priori knowledge of what the most obvious red flag structural features are so they can incorporate them into molecular design.

## 6.2. Prediction of Metabolism and Biotransformation

It is important to keep in mind that the majority of the toxicity prediction modules do not explicitly address the toxicity of the metabolites of the compound in question. Several software developers have addressed this issue by providing separate, stand-alone applications that predict metabolites, such as METAPC by Multicase and METEOR by Lhasa.

Prediction of metabolic fate is, like that of toxicity, highly challenging. CYP, the primary enzyme superfamily responsible for phase I metabolism, consists of numerous isoforms. In addition, most CYP isoforms have broad and overlapping substrate specificities, and the interactions between CYPs and their substrates in vitro are complex and often dose-dependent. Non-Michaelis−Menten kinetics are often observed, and inhibition of CYP-mediated conversions may be substrate-dependent. Furthermore, data from various studies indicate the possibility that the active site may simultaneously accommodate several ligands. On the phase II side, no single enzyme family predominates, but uridine diphosphate-dependent glucuronosyltransferase (UGT), sulfotransferases, and glutathione-*S*-transferase (GST) are the major players. Of these three, glucuronidation of small lipophilic molecules by UGTs is arguably the most important phase II process for the clearance of xenobiotics.[169] Prediction of UGT activity is complicated by its 18 isoforms, most of which show overlapping substrate selectivities, and the numerous factors known to influence UGT activity in vivo. The kinetics of all the biotransformation reactions also have to be considered together, as they are linked to metabolic stability of the chemicals in question. This alone is a significant challenge.

From a computational viewpoint, these challenges have to be met with modeling of the enzyme systems. The ideal computational system has to answer three primary questions for each enzyme family:

(1) Can the molecule in question act as an inducer or inhibitor to any of the enzymes involved in biotransformation?
(2) Can it be a potential substrate? If so, will it be a suicide substrate (e.g., one that covalently modifies the enzyme and inactivates it permanently)?
(3) What are the most likely metabolites (with some consideration to rate of production)?

These goals have been tackled with three types of computational methods—rule-based, protein-based, and ligand-based systems[158]—as shown in Figure 18.

*Protein-based* methods rely on X-ray crystal structures of specific proteins that are used to examine ligand-binding interactions by automated docking approaches that attempt to place molecules in enzyme binding site and assess the binding affinity with scoring functions.[170] Proteins can be modeled quantum mechanically (e.g. DivCon, ONETEP), classically (with molecular mechanics) or by a combination

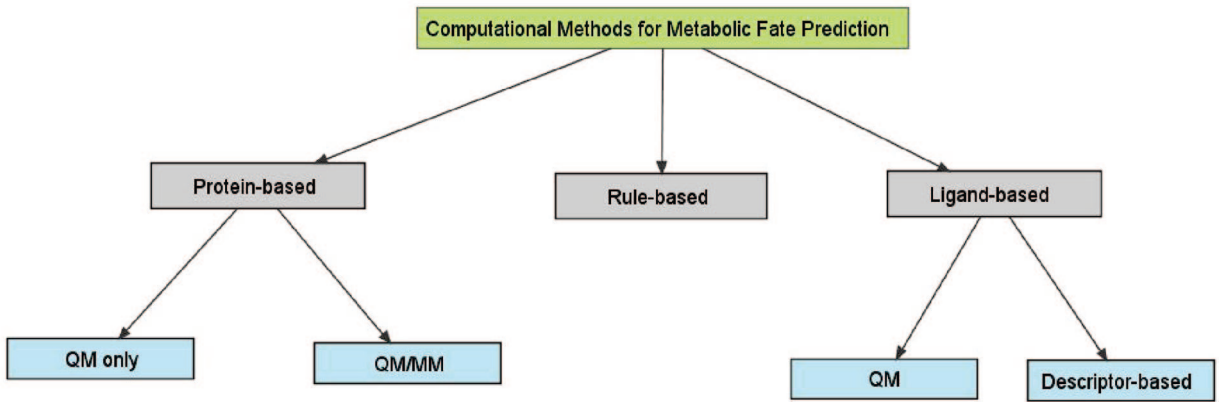
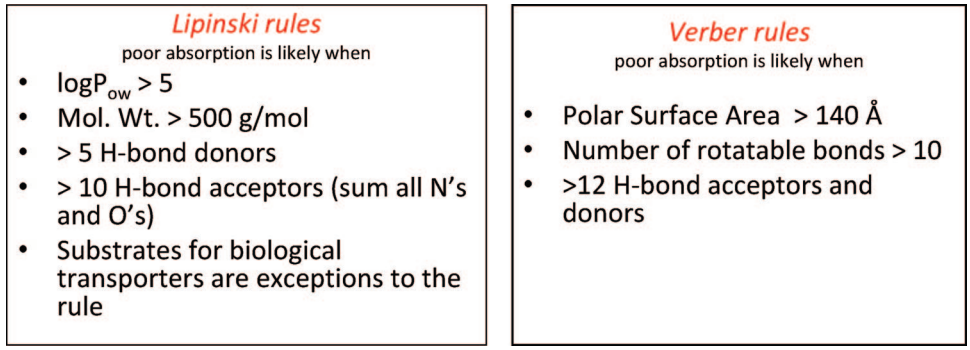



**Figure 18.** Classification of computational methods for prediction of metabolic fate.

**Scheme 5**




of the two (QM/MM).[158] QM/MM techniques have become quite popular in the past decade and found wide application in various enzyme types.[171−172]

The inverse of this is the *ligand-based* approach, where, as opposed to asking the question "what ligands fit this enzyme?", one asks "what enzymes can bind this ligand?" Ligand approaches either use QM methods to describe the electronic structure of ligands and the energies of various species along the enzyme−substrate reaction path or focus on physicochemical descriptors. These properties are calculated from the structrure of the ligand, and a statistical relationship between these descriptors and experimental activity is established using linear and nonlinear regressions.[173] These data could also be used to generate 3D-QSAR models to correlate the descriptors to activity.[174]The disadvantage of this method is its reliance on a sufficiently large and high-quality data set to act as a training set. As with the ARI predictive methods described earlier, if the query compound is far outside the range of the training set, the prediction accuracy may be compromised. Furthermore, since enzymes have very specific binding sites, if the training set is too broad, subtle differences governing protein−ligand interactions might not be captured in the descriptors. Yet, if it is too narrow, its performance on structurally different compounds might be less than satisfactory.

Finally, the *rule-based* methods depend on data-mining techniques on large databases of metabolic data to extract generalized rules and determine which molecular fragment undergoes metabolic transformation and the metabolites that result.

Although great strides have been made in our predictive capabilities for metabolic fate, significant challenges remain to be addressed before we can apply these techniques with confidence and replace in vitro and in vivo testing. There is, however, still a great deal to be learned by a chemist from these predictive approaches, especially as they pertain to furthering our understanding of toxicity mechanisms. Imperfect as our current biotransformation predictions might be, they would still be very informative in alerting the chemist to consider the toxicity of the suggested metabolites. Ultimately, it would be desirable to attain a clear understanding of an updated rule base that can be used by chemists upfront in the design stage.

## *7. New Perspectives: Toward Property-Based Design Guidelines*

In previous sections of this paper, we described how physicochemical properties can be used to manipulate bioavailability, distribution, absorption, and excretion (ADME). Here, we show how this knowledge, together with data on toxicity and physicochemical properties, can be used to develop property-based design criteria for hazard reduction. Chemists can then directly apply such guidelines during the molecular design process to minimize risks of particular types of hazards. We will show that some of these guidelines have already been derived, while others can be derived from existing property and toxicity data.

Although the complexity of toxicity mechanisms is vast and broad, the feasibility of identifying a set of property criteria that are met by the majority of compounds exhibiting particular types of toxicity is not implausible. In fact, medicinal chemists have met similar challenges in identifying the property criteria met by biologically active compounds. In 1997, Lipinski and colleagues formulated the "Rule of five" of druglikeness.[175] These simple "rules" described the predominant property value ranges of molecular weight, log $P_{ow}$, and number of hydrogen bond donors and acceptors of ∼900 commercial pharmaceutical compounds. Since then, the Lipinski rules, as well as rules proposed by Veber et al.,[176] have become accepted as standards by the pharmaceutical industry for evaluating the feasibility of active lead compounds (Scheme 5). The properties Lipinski and Veber et al. described are mechanistically associated with the potential for favorable ADME (adsorption, distribution, metabolism, and excretion) properties of the lead compounds.

In the commercial chemical industry, any biological activity of chemicals is unintentional and should be avoided. Thus, what we seek is a set of Lipinski-like guidelines for compounds that are biologically inactive. One way to achieve lack of activity is to ensure the chemical is not bioavailable. If that is not possible, one can also aim to reduce distribution, reduce bioactivation (or increase deactivation), accelerate excretion, and finally eliminate any potential toxicodynamic interactions responsible for specific toxicity. These aims are illustrated in Figure 19.

### 7.1. Toxicodynamic and Toxicokinetic Behavior and Chemical Properties

The biological interaction with xenobiotics, as mentioned earlier, can be described by the physicochemical characteristics of the chemical and the internal dose that reaches the target tissue. Both toxicokinetic and toxicodynamic interactions can be mechanistically accounted for by the physicochemical properties, or structural motifs, of a chemical.

From a toxicokinetic perspective, we are concerned with the properties that affect ADME. To be absorbed into the systemic circulation, a xenobiotic must pass through several biological lipid bilayer membranes. It must therefore have

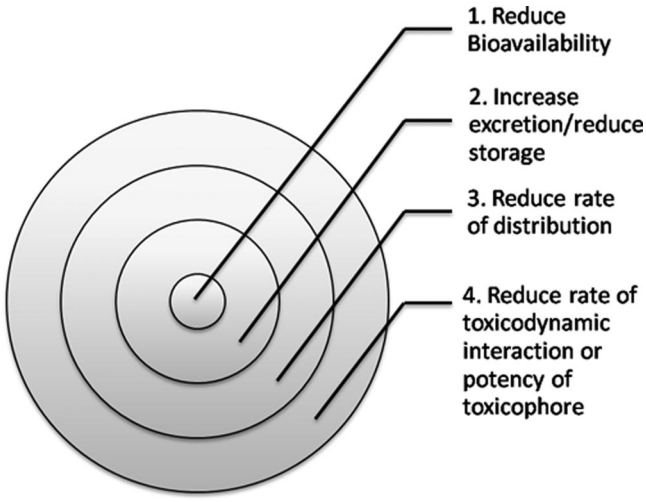



**Figure 19.** Mechanistic layers of reducing toxicological hazard.

properties that facilitate such transport. Most simply, its log $P_{ow}$, molecular weight and number of hydrogen bond acceptors/donors, polar surface area, and number of freely rotatable bonds must be within certain ranges (Scheme 5).[175]Similarly, in order to be eliminated efficiently, a chemical must have particular ranges of water solubility, polarity, p$K_a$ and size have plasma protein affinity.

Physicochemical properties and structural descriptors also dictate toxicodynamic interactions. Most covalent interactions that result in direct toxicity (e.g., suicide inhibition of vital enzymes, binding and modification of DNA) require that the toxicant have HOMO and/or LUMO energies within particular ranges, which facilitate the reaction with biochemical targets. For example, empirical relationships have been found between HOMO−LUMO energies and carcinogenicity (Figure 13—the dependence of carcinogenicity on $\Delta E$ for six symmetric dialkyl nitrosamines[116]). It would thus be highly useful to identify, where possible, the properties (or combination of properties) that are associated with specific toxicity end points.

### *7.1.1. Example: Do Highly Toxic Chemicals Share Common Physicochemical Properties?*

A Lipinski-like analysis of the properties of toxic, EPA-regulated compounds was recently reported by our group.[179] This study analyzed the properties of 546 compounds with established human toxicity, comprising the U.S. EPA's Toxic Release Inventory (TRI) list of toxic chemicals.[177] The compounds on this list include commercial chemicals released into the environment or otherwise managed as waste by facilities from various industry sectors, including manufacturing, metal and coal mining, electric utilities, and hazardous waste treatment, among others.[178] The authors report the distributions of the calculated physicochemical properties of these compounds, predicted using Schrodinger's QikProp software, and compare them to those of a randomly selected group of compounds from the ZINC database, an extensive collection of all commercially available compounds, which was selected to represent "chemical space". The resulting property distributions indicated that, for almost all of the physicochemical property descriptors predicted, the toxic group (TRI chemicals) showed statistically different property distributions from those of the ZINC group. The statistics summarizing these distributions for the toxic chemicals are shown in Table 7. Although the statistical differences may not be obvious from a first glance at the ranges of property values, the means of each distribution are more indicative.

These results should be viewed with some caution, as the toxicity of the chemicals in the ZINC group is completely unknown.

Particle or molecular size is closely related to human dermal, pulmonary, and oral absorption. Compounds with molecular weight > 400 Da, for example, are generally poorly absorbed through the skin.[180] Similarly, those with MW > 500 Da cannot typically cross the GI tract and enter the bloodstream.[181,182] The molecular weight distribution of the TRI chemicals was significantly different from that of the control ZINC group (2.5 and 97.5% quantiles we observe of the distribution of molecular weights of the TRI compounds fall within 43−461 amu, Table 7). These results indicate that the weights of >92% of the TRI compounds fall well within the molecular weight range for likely absorption. It is possible that outliers may include compounds that contain hydrolyzable linkages, which may degrade in the stomach to smaller compounds that can be absorbed in the gut.

Polar surface area (PSA) is considered to be an important property in the prediction of oral bioavailability of drugs.[183] Veber et al. proposed that drugs with PSA > 140 Å$^2$ are poorly absorbed.[176] Others have also suggested that membrane permeability can better be correlated to PSA than molecular weight.[183−187] The predicted property data for TRI chemicals (Table 7, entry 3) shows that the majority of these compounds have low PSA, with an upper 97.5% limit at 122 Å$^2$, well below the proposed maximum for oral bioavailability of 140 Å$^2$.

**Table 7. Summary Statistics of Selected Predicted Physicochemical Property Distributions of the 546 Organic Compounds from the EPA Toxic Release Inventory (Reproduced (Modified) with Permission from Ref 179. Copyright 2010 Elsevier.)**

| property | TRI mean (2.5%−97.5% quantiles) | ZINC mean (2.5%−97.5% quantiles) |
|---|---|---|
| molecular weight (amu) | 206 (43−461) | 395 (234−504) |
| log $P_{oct/water}$ | 2.03 (−2.14−6.99) | 3.59 (0.56−6.28) |
| polar surface area: N, O, and X−H surface area (Å$^2$) | 37 (0−122) | 77 (29−150) |
| globularity = $4\pi(r^2)$/SASA, where $r$ is radius of sphere whose volume is equal to the molecular volume | 0.91 (0.80−0.99) | 0.74−0.89 |
| H-bond donors: estimated number of HBs that would be donated by the solute to water | 0.62 (0−3) | 1.07 (0−3) |
| H-bond acceptors: number of HBs that would be accepted by the solute from water in aqueous solution | 2.68 (0−9) | 6.46 (2.5−11) |
| polarizability (A$^3$) | 18.3 (2.6−43.5) | 40.8 (23.8−53.5) |
| electron affinity (eV) | 0.41 (−3.25−3.46) | 0.83 (−0.22−1.89) |
| ionization potential (IP) | 9.44 (5.98−11.81) | 8.85 (7.61−9.83) |
| water solubility log (mol/L) | −2.31 (−9.11−1.79) | −5.30 (−6.5−0.5) |

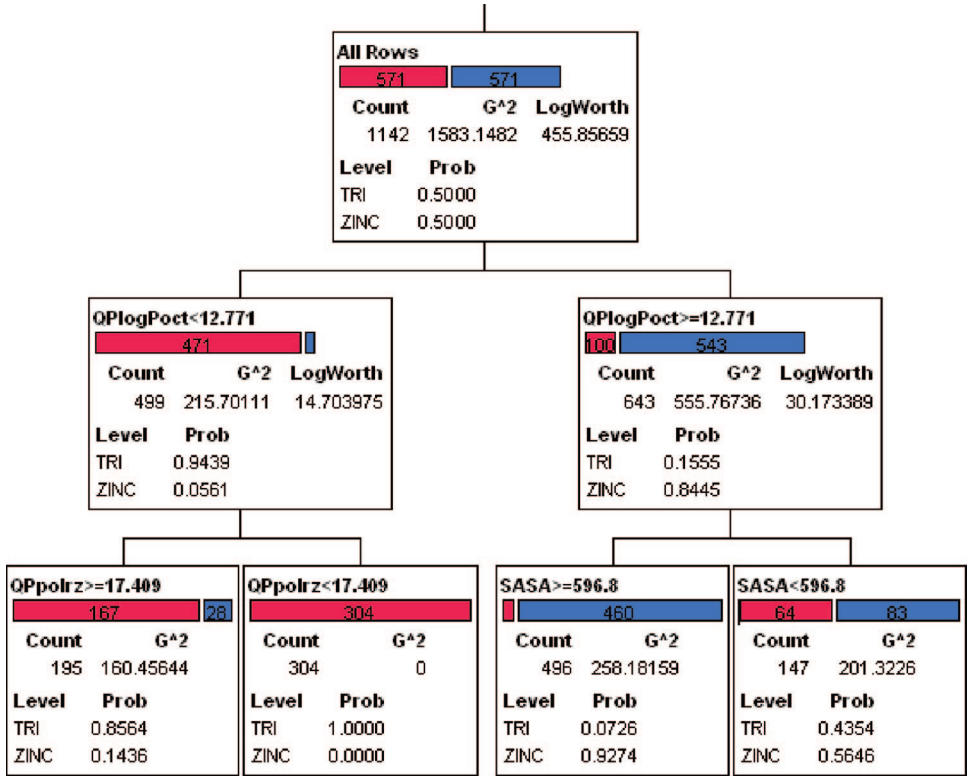



**Figure 20.** Decision tree and splitting diagram of partition analysis of TRI data set and an equal randomly selected sample from the ZINC database. Properties represented (all calculated by QikProp): QPlogPoct = log of octanol/gas partition coefficient; QPpolrz = polarizability; SASA = solvent accessible surface area). Reproduced with permission from ref 179. Copyright 2010 Elsevier.

Although single properties did show differences between the TRI and control group, a multivariate method was sought to gain a more holistic understanding of whether the physicochemical properties studied can be used to segregate the compounds into two groups. Partitioning, or decision tree analysis, was applied, which showed that compounds with log $P_{octanol/gas}$ < 12.771 were significantly more likely to belong to the TRI group of toxic chemicals than to the contol set. Figure 20 illustrates the tree diagram and partition graph representing the best splitting.

The implication of this analysis is that a classification of highly toxic chemicals could indeed be possible with as few as three physicochemical properties—in this case, octanol/gas partition coefficient, surface area, and polarizability. It must be noted, however, that the control group does not represent compounds that are likely to be safe but merely the diversity of all commercial chemicals, or what is known as "chemical space". In addition, organometallic compounds, charged organic species, and those with odd numbers of electrons were not included in this study because their properties could not be predicted.[188]

### *7.1.2. Structural Interventions That Reduce Absorption (Bioavailability)*

A reduction in the amount of a chemical that is able to enter the body will limit the internal dose and reduce the likelihood of toxicity. Thus, an understanding of the properties that impact absorption and uptake is essential to the design of safer chemicals. Anatomical considerations are also important, including surface area of the organism that is exposed, thickness of membrane or tissue, and regional blood flow. Details for the primary sites of absorption in humans (GI tract, respiratory tract, and skin) and physicochemical properties or design criteria that decrease their absorption through each site are presented below.

**7.1.2.1. Gastrointestinal Tract.** The physicochemical properties that specifically affect absorption from the GI can be summarized as follows:

- log $P_{ow}$ is a key factor in determining the extent of absorption. Generally more lipophilic compounds are absorbed better through passive diffusion, the main permeability mechanism in the GI tract.
- Physical state: liquids or solutions absorbed better than solids.
- Particle size: smaller particle size and bigger surface area/volume ratio results in faster absorption.
- Dissociation constant (p$K_a$) and ionization: organic salts are absorbed better than neutral organics;[55] neutral molecules are better absorbed by passive diffusion.
- Molecular weight and size: MW < 300 Da = well-absorbed, 300−500 Da = less well-absorbed, >500 Da = poorly absorbed.

From this, we extrapolate that, for reduced gastrointestinal bioavailability, a chemical should possess the properties listed in Table 8.

If we consider nonhuman mammals, many variations of these rules can exist due to interspecies differences (e.g., the drug Nadolol is absorbed orally significantly more by dogs than humans and rats). Interspecies differences among mammals are often due to differences in pH of the GI tract, as well as differences in number and nature of microbes present and their distribution.

**7.1.2.2. Respiratory Tract.** Most notably, due to the thin alveolar membrane, absorption from the lungs differs from intestinal and dermal absorption in that lipid solubility is less important than water solubility, which directly affects the blood-to-gas partition coefficient.

**Table 8. Desirable Physicochemical Properties for Decreased GI Absorption and Bioavailability of Chemicals in Humans**

| physicochemical property | decreased oral absorption favored by | desired value |
|---|---|---|
| particle size | Increased particle size (for nanoparticles) | ideally > 100 nm |
| ionization | Keeping substance in un-ionized form (ionization increases solubility) is favorable. An exception to this rule is made for nanoparticles, as ionized nanoparticles show decreased absorption. Incorporation of substituents that remain ionized at pH 2, such as $-SO_3^-$, would make the chemical too polar to cross intestinal membrane | un-ionized or ionized at pH 2 (e.g., $-SO_3^-$) |
| log $P_{ow}$ | High log $P_{ow}$ implies chemical is too lipid-soluble to dissolve in GI tract; conversely, if log $P_{ow}$ is too low, increased water solubility will limit absorption and enhance elimination | log $P_{ow}$ < 0 or > 5 |
| molecular weight | Increased molecular weight decreases chance of absorption in GI tract | >500 Da |
| melting point | Liquids are absorbed faster than the corresponding solids, so solids at body temperature (37 °C) are preferred | melting point >150 °C |
| hydrogen bonds | Increased number of hydrogen bond donors and acceptors limits absorption (unless transported by specific active transport carrier protein, such as erythromycin or methotrexate). | >5 H-bond donors or >10 H-bond acceptors |
| hydrolyzable linkages | Ability to be hydrolyzed by acidic conditions of stomach (pH 1−3) or biotransformed by intestinal enzymes or bacterial intestinal flora | avoiding hydrolyzable ester and amide linkages |

**Table 9. Desirable Physicochemical Properties for Decreased Respiratory Tract Absorption and Bioavailability of Chemicals in Humans**

| physicochemical property | decreased respiratory absorption favored by | desired value |
|---|---|---|
| gases | Lower blood-to-gas partition coefficient will slow the rate of absorption across the alveoli | <1 |
| vapor pressure | Lower vapor pressure decreases chance of delivery of external dose to respiratory tract | <0.001 mmHg |
| molecular weight | Increased molecular weight decreases vapor pressure and lessens chance of delivery of external dose to respiratory tract | >400 Da |
| particle size | Larger particles deposit higher in the respiratory tract, are removed in the mucous layer, and are swallowed, resulting in a GItract exposure | MMAD[a] > 5 $\mu$m |

[a] MMAD = Mass median aerodynamic diameter.

Although consideration of water solubility and the related blood-to-gas partition coefficient can result in significant differences in the rate and extent to which a vapor/gas will cross the alveoli and enter systemic circulation, the property with the single greatest impact on bioavailability through the lungs for liquids and gases is vapor pressure. Simply put, chemicals that will not readily become airborne are less likely to be inhaled and have the potential to be absorbed via respiratory tract exposure.

Although respiratory absorption and toxicity of particles depends on multiple factors such as morphology, composition and size, several studies have shown that larger particles (with mass median aerodynamic diameter, MMAD > 5100 $\mu$m) are less likely to become airborne and be inhaled.[189] Thus, to reduce bioavailability via the respiratory tract, a chemical should be designed to possess the properties described in Table 9.

**7.1.2.3. Skin.** The predominant properties that affect skin permeability are as follows:

- Physical state: liquids absorbed better than solids;
- Melting point: nonionic solids with MP < 125 °C have higher probability of being absorbed;
- Ionization: ionic solids or highly polar substances are generally not well-absorbed;
- log $P_{ow}$: lipophilic substances absorbed better; higher log $P_{ow}$ = better skin absorption, up to log $P_{ow}$ of 6, when they become too lipophilic;
- Molecular weight: compounds >400 Da are generally poorly absorbed.

**Table 10. Desirable Physicochemical Properties for Decreased Dermal Absorption in Humans**

| physicochemical property | decreased dermal absorption favored by | desired value |
|---|---|---|
| physical state | solid must first be dissolved to permeate skin, while liquids may be absorbed directly | solid |
| ionization | increased polarity means increased water solubility, which decreases skin absorption | polar or ionized (salt) |
| log $P_{ow}$ | low log P means higher water solubility and lower lipophilicity | 0 < log $P_{ow}$ < 6 |
| molecular weight and particle size | increased molecular weight decreases rate and extent of absorption through skin | >400 Da |

Therefore, to design chemicals with decreased potential for dermal absorption and bioavailability, we can refer to the properties in Table 10.

It should be noted that there is a wide variation in skin absorption among species. This is due to differences in the thickness of the outer skin layer and in the number of sweat glands. For this reason, experimental dermal absorption data from laboratory animals, such as rats, must be used with caution when estimating potential dermal absorption in humans. As discussed previously (section 4.2.1.2.a), since the thickness and permeability of skin in different body regions in human varies by as much as a factor of 20, the dermal absorption also varies significantly depending on the location of exposure.

### 7.1.3. Structural Interventions That Reduce Distribution

Designing molecules so that they have decreased distribution is a challenge, because once absorbed (yielding an internal dose), chemicals are usually distributed rather readily. Depending on the route of exposure and absorption, decreasing water solubility may decrease absorption. However, decreased aqueous solubility can also hinder excretion. Polarity and size also appear to be important—polar molecules of molecular weight > 200 Da have more difficulty getting into cells without active or facilitated transport, so increased polarity and molecular weight are also key.

Other factors, such as extent of binding to plasma proteins (listed in Table 2) and storage, will also influence distribution. Although general categorical physicochemical criteria for these processes are still to be established, reports have indicated that a significant correlation between the logarithm of the partition coefficient and the plasma protein binding exist for some classes of compounds, such as bulky organic cations.[190]

The same principles that guide diffusion of chemicals across other cells in the body apply to the blood−brain barrier: lipophilicity and a nonionized state favor passage. It should be noted, however, that the BBB is a highly complex series of specialized capillary membranes in the brain and consideration of numerous biological and chemical factors is often necessary.

### 7.1.4. Structural Interventions That Can Reduce Bioactivation

The central role of metabolic enzymes, such as CYP, in the mechanisms of toxicity of numerous chemicals has been reiterated in numerous literature reports. Table 11 highlights the mechanisms of toxicity of several classes of chemicals and illustrates how structural modifications can be used to reduce the undesired biological activity. This information can serve as a guide for informing design of molecules and will undoubtedly be continuously expanded.

## 7.2. Designing Molecules That Do Not Interfere with CYP Regulation Pathways

In addition to being concerned about the molecular structure of CYP substrates and their possible enzyme-mediated transformations to reactive metabolites, one also has to appreciate the complexity of regulation of the CYP family of enzymes, as enhanced enzyme expression can greatly augment the bioactivation or detoxication rates of xenobiotics. The mechanisms of induction of CYPs primarily involve activation of receptors involved in transcription, such as the aryl hydrocarbon (Ah) receptor for the 1A CYP family, the constitutive androstane receptor (CAR) for the 2B family, and the pregnane X receptor (PXR), glucocorticoid receptor (GR), and vitamin D receptor (VDR) for the 3A family.[204]

Although we are perhaps still faced with an incomplete scheme of all the interactions required for enzyme or transporter induction to occur, experimental and theoretical approaches are being used to make predictions about the ultimate in vivo response based on an understanding of the binding pockets of the above-stated receptors. Table 12 summarizes the currently available information on the five receptors and provides a succinct description of the physical parameters that describe the active bonding site, which can guide the design of molecules that lack a good geometric (steric) or electronic fit. For example, the aryl hydrocarbon receptor (Ah) has a constricted binding pocket of 6.8 × 13.7 Å planar rectangle, and it has been shown that molecules that are nonplanar and/or do not fit in these dimensions are generally poor inducers of CYP via this receptor.

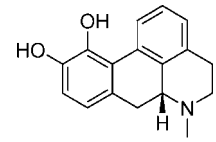
A few of the most powerful inducers of CYPs in humans are phenobarbital, acetaminophen, extract from St John's wart (hyperforin), progesterone, estrogen, corticosterone, rifampicin, nifedipine, clotrimazole, and metyrapone. Since several distinct biochemical pathways are responsible for CYP regulation, a rather broad class of compounds can be potential inducers. Most of them, however, fall into the following classes of compounds: (i) steroid hormones and their metabolites, (ii) polyaromatic hydrocarbons (PAHs), (iii) polychlorinated biphenyls and dioxins, and (iv) furans. Structures of common CYP inducers are shown in Scheme 6.

Inhibition of CYP can also be important to toxicity. The most striking examples of irreversible (suicide) inhibitors are highly electrophilic halomethanes such as $CCl_4$, $HCCl_3$, and $HCClBr_2$. In addition, compounds with terminal alkynes are also suicide inhibitors for CYP. Data on safer analogues of these highly toxic compounds does not yet exist, thus chemists should aim to avoid these functionalities in molecular design. A complete list of commonly encountered CYP substrates, inducers, and inhibitors is available.[217]

## 7.3. Structural and Property Modifications that Can Reduce Toxicity

### 7.3.1. Reducing CNS Activity

CNS active compounds are known to share several topological or electronic characteristics.[218] For example, Seeman[219] has proposed that there are structural requirements for agonists of dopamine receptors (such as (−)-apomorphine) on the basis of extensive binding data of semirigid chemicals.



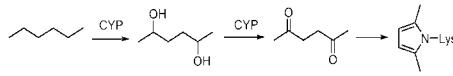
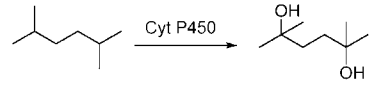
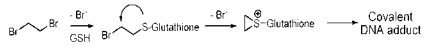
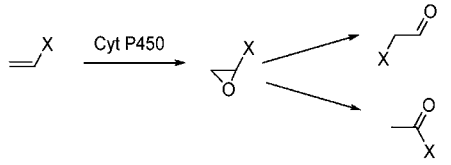
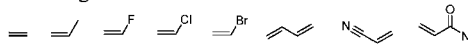
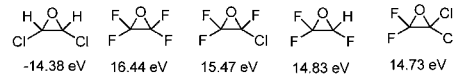
(-)-apomorphine

These structural requirements can be summarized as follows:

- A hydrogen bonding group (OH or NH) corresponding to the 3-hydroxy group of dopamine;
- A nitrogen atom positioned ∼0.6 Å from the plane of the aromatic ring;
- A distance <7.3 A between the nitrogen and the hydroxyl group;
- High lipid solubility; and
- Steric hindrance.

Similarly, many other classes of CNS-active compounds are known to have common pharmacophores. The most common class of antidepressants, for example, contains a tricyclic motif (e.g., impramine and amitriptyline). These tricyclic antidepressants (TCAs) are thought to act by blocking the uptake of the neurotransmitters noradrenaline and serotonin, which causes an increase in the concentration of these transmitters in the synapse and thus counteracts the deficiency-associated depression.[220]

**Table 11. Select Metabolic Activation Pathways and Structural Modifications That Can Be Used to Reduce Toxic Effect**

| Chemical Group | Subgroup | Enzyme | Mechanism (s) | Structural Modifications to decrease toxicity | Toxicity endpoint |
|---|---|---|---|---|---|
| Akanes  | Hexanes  | CYP |  The diol is further oxidized to the 2,5-dione, which reacts quickly with the epsilon amino groups of Lys within proteins of axonal nerve fibers to form pyrrole adducts.[191] These adducts lead to the protein cross-linking which eventually causes the nerve destruction and neurotoxicty. | Substituting the 2 and 5 positions of hexane with methyl groups prevents the dione formation, and is expected to not be neurotoxic in human studies.[192]  | neurotoxicity |
| | 1,2-di-haloalkanes | GST | The SH group of GSH attacks the first and second carbon, releasing two halides and forming a ring-shaped, highly electrophilic episulfonium ion, which can covalently bind DNA and cause mutations. Stable GS-conjugates are formed hepatically, transported in the bile and re-absorbed in the gut, and finally transported to the kidneys to be degraded and excreted. Thus the kidneys often become the target organ of accumulation and toxicity.  | Substituting the alpha bromide with a chloride or fluoride should decrease toxicity as these are worse leaving groups than bromides. | |
| Alkene  | Vinyl halides  | CYP |  It was found that mutagenicity and carcinogenicity of these compounds was associated with the two-center bond energies of the corresponding epoxides. Thus when this energy falls between -14.1eV and -12.9 eV, the epoxides formed are highly toxic. The most common parent olefins which fall in that region are listed below, while the epoxide of isobutene was predicted and experimentally proven to not be mutagenic in-vitro.[193]  | The two-center C-O bond energies should fall outside of the range or -14.1 eV to -12.9 eV. NB: This QSAR has possibly not been verified by a big enough data set. The following epoxides fall outside of that range, and were tested in-vitro to be non-oncolgenic. However, it should be noted that this model does not take into account the rate of epoxide formation by CYP. Other predictive models have attempted to take this into account also.[194, 195]  -14.38 eV 16.44 eV 15.47 eV 14.83 eV 14.73 eV | mutagenicity |
| Arene  | benzene | CYP | Oxidation to the epoxide occurs via a tetrahedral intermediate, which can form either an epoxide or a phenol directly (Scheme below). The epoxide can covalently bind nucleophiles, such as DNA or proteins, to open up the epoxide to a phenol and make toxic covalent adducts. The phenols can get further oxidized to bis-phenols, which can form quinones. Quinones can cause serious oxidative damage to cells through radical pathways, or can alkylate N- or S-nucleophiles, such as glutathione and glycine. | To circumvent this bioactivation pathway, a more easily oxidizable C-H bond can be included in the molecule, such as a benzylic methyl group. Toluene, as a result, is significantly less toxic than benzene since its major CYP metabolite is benzyl alcohol, which ALDH converts to benzoic acid. Benzoic acid is conjugated with glycine and eliminated in the urine as hippuric acid, which is much less toxic than the metabolites of benzene. | carcinogenicity |

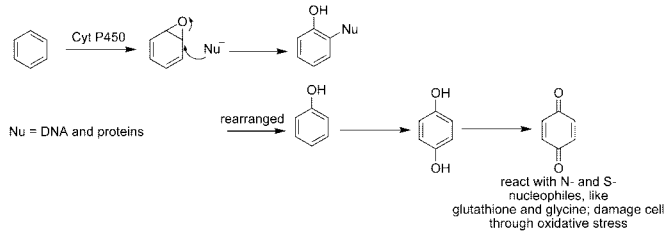


**Table 11. Continued**

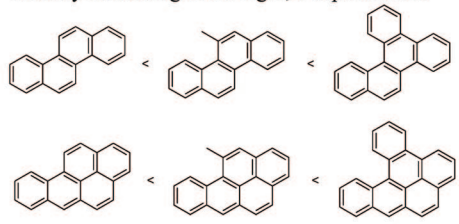
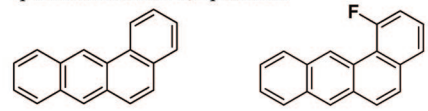
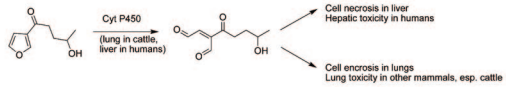
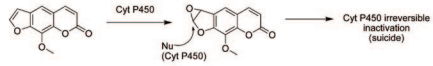
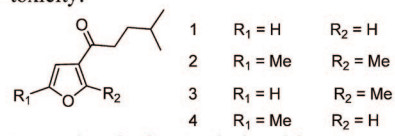
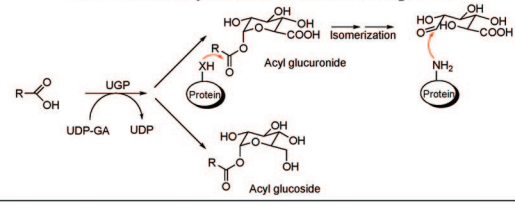
| Chemical Group | Subgroup | Enzyme | Mechanism (s) | Structural Modifications to decrease toxicity | Toxicity endpoint |
|---|---|---|---|---|---|
| | Polycyclic aromatic hydrcarbons | CYP | Polycyclic aromatic hydrcarbons can also be activated twice by CYP to dihydrodioleperoxides. It was found that carcinogenicity is affected by both the number of rings and the methyl substitution, so that the toxicity pattern for the series below is as follows:<br>Toxicity increasing to the right, 2 separate sets.<br>  | It has also been noted that isosteric replacement with fluorine is an effective way to decrease carcinogenicity of polyaromatic hydrocarbons. For example, 7-methyl-benzo[a]anthracene is highly carcinogenic, but 1-fluoro analog is not, as the presence of fluoride prevents CYP epoxidation at the 1,2 position.<br>  <br>Carcinogenic Non-carcinogenic<br>It is also known that carcinogenicity is affected by both the number of rings and the methyl substitution, so that the toxicity pattern for the series below is shown to the left.<br>Using QSAR data, molecular weight, volume and log P (all of which are related) of unsubstituted PAHs have been found to correlate well with toxicity (lower molecular volume of unsubstituted PAHs – lower toxicity), but seem to underestimate the toxicity of substituted PAH. Using volume appears to give stronger correlations than using weight. | |
| | Furans | CYP | Furans are epoxidized by CYP to yield a highly electrophilic species which rearrange to dialdehydes, ultimately leading to cell death (necrosis) in the liver or lungs (species-dependent).[222]<br>  <br>Furanocoumarins, such as 8-methoxypsoralen, are epoxidized by CYP. The epoxide directly reacts with an amino acid residue of CYP itself and thus inactivates the enzyme irreversibly (thus it is also in the class of suicide inhibitors).<br>  | Methylation and chlorination of furans affects the ability of CYP to form epoxides. Although methylation does not prevent oxidation, it prevents epoxidation, giving rise to the much more stable hydroxyl product.<br>In the case of 4-ipomeanol (IPO) and its analogs, methylation has a vast effect on toxicity.[196]<br>  <br>$LD_{50}$ values for four 3-substituted furans:<br>Compound $LD_{50}$<br>1 30 ± 1.8<br>2 2238±1242<br>3 193 ± 15<br>4 8806±3455 | Hepatic toxicity or lung toxicity |
| Acids | | UDT | Glucouronidation of carboxylic acids results in unstable acyl glucuronides, which are prone to hydrolysis, isomerization and covalent binding to proteins and amino acids. Isomerization refers to translocation of the acyl group to the 2, 3 and 4-position of the GA molecule. These isomers can transiently undergo chain-opening, which exposes reactive aldehydes to cellular nucleophiles.[197]<br>  | The currently available QSAR data are divided on glucouronidation being a detoxification or bioactivation mechanism, depending on the precise structure. For NSAIDS, it appears to be mostly a detoxification mechanism, but the toxicity of acyl glucuronide and glucoside adducts must be considered (in terms of isomerization and protein adduct formation) when designing carboxylic acids.[198] | Ulceration |

**Table 11. Continued**

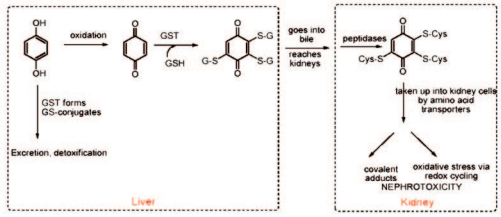
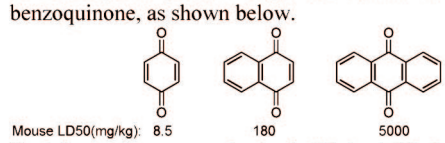
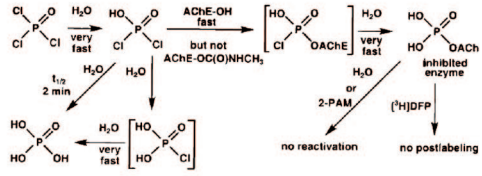
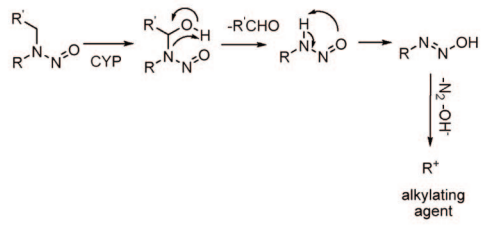
| Chemical Group | Subgroup | Enzyme | Mechanism (s) | Structural Modifications to decrease toxicity | Toxicity endpoint |
|---|---|---|---|---|---|
| Benzoquinone | Also hydrobenzoquinone | GST | Benzoquinone (either direct exposure, or exposure to benzohydroquinone, such as from cigarette smoke, which is partially metabolized to benzoquinone) is conjugated to GSH by GST to form mono, bis, tris or tetra GS conjugates. The mono and tetra are non-toxic, but the bis and especially tris are highly toxic in the kidneys, as once they reach the kidneys from the bile, peptidases cleave two amino acids off the GS, leaving cysteine conjugates. These are transported into the proximal tubular epithelia cells, where they are oxidized to the substituted benzoquinone. These compounds are highly electrophilic and also undergo redox cycling, translating to oxidative stress for the cell.   | A decrease of acute toxicity to mice is observed as subsequent phenyl rings are added to benzoquinone, as shown below.   Toxicity values are obtained EPA ACToR database (NLM dataset). | Nephrotoxicity |
| Organophosphorus (OP) | | AChE | Inhibition of AChE to 70-90% is lethal in mammals, insects and nematodes. To some extent they can be hydrolyzed by A-esterases (paraoxonases or lactonases) but B-esterases are inhibited (carboxyesterases and cholinesterases). $POCl_3$ is a selective inhibitor of AChE compared with other serine hydrolases. The average $IC_{50}$s for $POCl_3$ observed here are 28 μM for four AChEs, 740 μM for five BuChEs, 140 μM for the carboxylesterase, and >1000 μM for three other hydrolases.1 The active agent for selective phosphorylation of AChE is proposed to be $HOP(O)Cl_2$ rather than $POCl_3$ itself. [199]   | The lower reactivity of thiono analogues of OP compounds as AChE inhibitors in comparison to their oxono analogues has been well known, and is believed to be due to the fact that since O is more electronegative than S, the electrophilicity of the P in oxono analogues is increased, which makes it more reactive with nucleophiles, such as the active site serine of AChE. The inhibitory potency of organophosphorus on AChE is dependent also upon hydrophobicity, steric and electronic parameters.[200] | Neurotoxicity (lethal) |
| Nitrosamine | dimethylnitrosamine | CYP | The toxicity and carcinogenicity of N-alkyl-N-nitroso compounds depend on the formation in vivo of one or more species of alkylating agent, i.e., diazoalkanes, carbonium ions, or both[93] They are believed to be activated by CYP2A6 via alpha-hydroxylation, as shown below.   | In a series of nitrosamines congeners, it is found that carcinogenicity correlates linearly ($R = 0.95$) with excitation energy, $\Delta E$.[111] | carcinogenicity |

**Table 11. Continued**

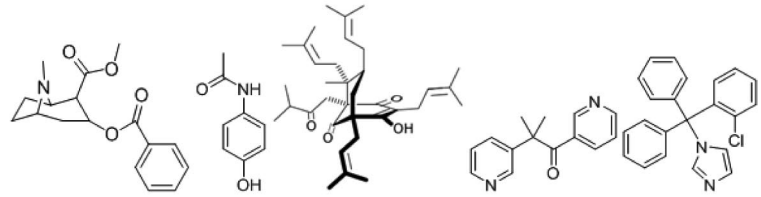
| Chemical Group | Subgroup | Enzyme | Mechanism (s) | Structural Modifications to decrease toxicity | Toxicity endpoint |
|---|---|---|---|---|---|
| Amines | Aromatic amines | CYP, GST, ST and others | Mechenism:<br>- N-hydroxylation and/or acetylation<br>- O-acylation (where acyl can also be sulfonyl or phosphonyl<br>- N or O glucouridation to form acyloxylamines and glucuronides<br>Factors known to affect carcinogenicity are:<br>- Nature of amine-generating group<br>- Position (and other ring substituents) of amine-generatng group<br>- Size, shape and planarity of molecule.[203] | Properties associated with lower carcinogenic potency are: [203,204]<br>- More fused phenyl rings lead to higher conjugation, and higher carcinogenic potency. Non-aromatic amines are generally less carcinogenic<br>- Alkylation of the N with bulky groups decreases carcinogenic potential by hindering delakylation<br>- Ortho substituents provide steric hindrance slow down amine activation<br>- Distorting the planarity of the molecule by introducing bulky groups ortho to the inter-cyclic linkages decreases ability to intercalate DNA and makes it a poor substrate for activation enzymes<br>- Decrease resonance stabilization of intermediates by altering the position of amine or replacing electron-conducting intercyclic linkages (eg CH=CH or -$CH_2$-) with electron-insulating ones (-C(O) -$CH_2$- or (-$CH_2$-)$_n$ where n>1)<br>- Render the molecule more water soluble by introducing hydrophilic substituents (eg sulfonyl), thus making it less bioavailable. | |

**Table 12. Description of CYP Induction Receptors and Physicochemical Parameters That Describe the Active Binding Site**

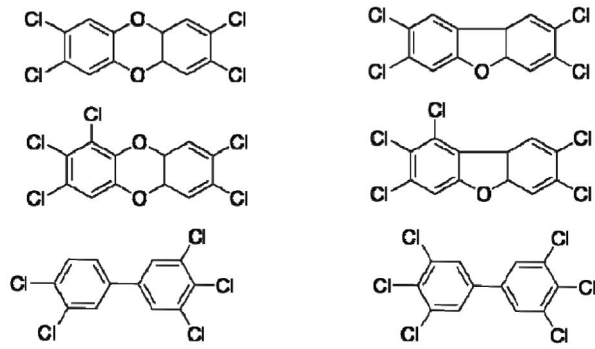
| CYP inducer receptor | binding site characteristics | known strong-binding ligands | properties for strong binding |
|---|---|---|---|
| aryl hydrocarbon receptor (Ah) | hydrophobic pocket | indigo | van der Waals volume < 40 $nm^3/mol$[205,206] |
| | | indirubicin | high lateral polarizability, lateral chlorine substitution[207] |
| | | ITE | compounds with Van der Waals dimensions 14 × 12 × 5 Å are favored[208] |
| | | TCCD | |
| pregnane X receptor (PXR) | large and flexible hydrophobic binding domain with few polar residues | bile acids | four hydrophobic features and at least one hydrogen bonding[209] |
| | upon binding, the ligand can change the protein conformation, but could be ligand-dependent | statins | binding cavity volume (1150 $Å^3$)[210] |
| | very promiscuous | hyperforin | higher affinity for larger molecules than smaller planar ones[211] |
| | | HIV protease inhibitors | |
| | | calcium channel modifiers | |
| | | steroids | |
| | | plasticizer monomers | |
| | | estradiol | |
| | | exteinascidin | |
| constitutive active/androstane receptor (CAR) | | androstane | high affinity for the rigid repressors and androstane metabolites[212] |
| | | | three hydrophobes and one hydrogen bond acceptor[212] |
| | | | binding cavity volume (1170 $Å^3$)[213] |
| glucocorticoid receptor (GR) | control hepatic expression of PXR and CAR | dexamethasone | hydrogen bonding key[214] |
| | extensive H-bonding network | | molecular planarity and rectangularity[215] |
| vit D receptor | binding pocket changes little while the ligand adopts a conformation in order to fit | bile acids such as lithocholic acid | hydrophobic, electrostatic interactions and hydrogen bonds[216] |
| | less promiscuous | vitamin D | |

**Scheme 6. Cyt P450 Inducers**

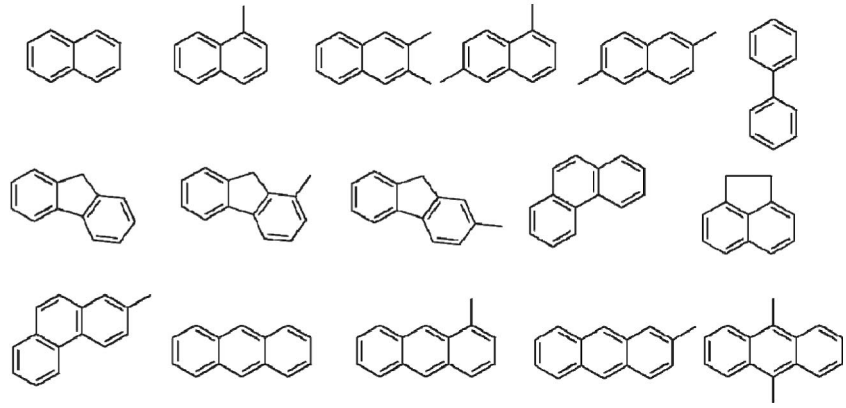
(a) drugs

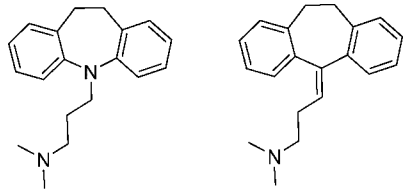



(b) Chlorinated PAHs, furans and dioxins



(c) Polyaromatic hydrocarbons (PAHs)






A study by Grunewald et al.[221] has demonstrated that binding of TCAs to noradrenaline uptake sites requires a steric fit, largely determined by the semirigid tricyclic system, together with some hydrophobic interactions. Thus, the reduction of one aromatic group resulted in a 10-fold decrease in activity, while reduction of both rings produced no further significant change.

### *7.3.2. Reducing Carcinogenicity*

Extensive research into the molecular characteristics of mammalian carcinogens over the last century has led to some important general principles that can inform the design of safer chemicals. In many cases, chemical carcinogens are bioactivated to electrophiles that bind covalently to nucleophilic cellular macromolecules (such as DNA), leading to mutagenicity and ultimately carcinogenesis. Structural modifications that reduce or eliminate bioactivation to electrophiles would likely reduce or eliminate the carcinogenic potential (see section 7.3).

Giving rise to electrophilic metabolites is not, however, the only governing factor of carcinogenic activity; it also depends on a host of other factors. As noted by Lai et al.,[203] some of the physicochemical properties or molecular parameters known to affect carcinogenicity of chemicals include the following:

- Molecular size and shape: The size of the most potent carcinogens (like PAHs) lies within a certain range. Almost all PAHs with highly elongated shape tend to be inactive.
- Substituent effects: Ring substitution with methyl groups at detoxification sites (e.g., L-region) of PAHs tends to increase carcinogenicity, while substitution with bulky or hydrophilic substituents decreases it, especially in the pro-electrophilic bay region.[223]
- Molecular flexibility (number of freely rotatable bonds): Epoxides on cycloaliphatic rings tend to be less active than

**Table 13. Changes in Physicochemical Properties to Favor Reduced Aquatic Toxicity**

| | |
|---|---|
| molecular size and weight | Generally, as molecular weight increases, bioavailability and toxicity decrease. At MW > 1000 Da,[230] bioavailability is negligible, as often is toxicity. Caution must, however, be taken to consider possible breakdown products that may have MW < 1000 Da and exert toxicity. |
| octanol−water partition coefficient (log $P_{ow}$) | Lipophilicity (log $P_{ow}$) usually correlates well with acute aquatic toxicity. For nonionic organic chemicals that operate though narcosis, acute and chronic toxicity increases exponentially with increases in log $P_{ow}$ up to a value of ∼5. For those whose log $P_{ow}$ > 5, bioavailability decreases along with acute toxicity, but bioaccumulation also increases. Minimal toxicity is likely with log $P_{ow}$ < 1.[225] |
| water solubility | Generally, increases in log $P_{ow}$ are associated with a decrease in water solubility. Very poorly water-soluble chemicals (<1 ppb) generally have low bioavailability and are thus less toxic. |
| LUMO energy | LUMO energies > 2 eV have shown to be associated with chemicals that are not toxic to some aquatic species (see example below). This is rationalized by the reduced electrophilicity of chemicals in this group—the higher the LUMO energy of a chemical, the less likely it is to be a strong electrophile. |

open-chain acyclic ones. Therefore, increased flexibility might be associated with increased potential for carcinogenicity.

• Polyfunctionality and distance between reactive groups: Diepoxides are significantly more toxic than monoepoxides, especially if the two epoxides are far enough apart to enable cross-linking.[224]

The example of aromatic amines, known for their high carcinogenic potency, has been thoroughly studied. Chemists now understand how to design aromatic amines with lower carcinogenic potency though manipulation of the molecular geometry and physicochemical properties of the molecule.[203] In summary, for aromatic amines, these structure and property manipulations are as follows (for more detail, see Table 11, section 7.b.iii, metabolism):

• Fewer fused phenyl rings and, thus, less conjugation results in lower carcinogenic potency; nonaromatic amines are generally less carcinogenic than their aromatic analogues.

• Alkylation of the N with bulky groups decreases carcinogenic potential by hindering dealkylation.

• Introducing ortho-substituents to provide steric hindrance slows amine bioactivation.

• Distorting the planarity of the molecule by introducing bulky groups ortho to the intercyclic linkages decreases ability to intercalate DNA and makes it a poor substrate for activation enzymes.

• Decreasing resonance stabilization of intermediates by altering the position of the amine or replacing electron-conducting intercyclic linkages (e.g., CH═CH or $-CH_2-$) with electron-insulating ones ($-C(O)CH_2-$ or $(-CH_2-)_n$ where $n > 1$).

• Rendering the molecule more water-soluble by introducing highly polar substituents (e.g., sulfonate), thus making it less bioavailable.

## 7.4. Strategies for Reducing Hazard beyond Human Toxicity

### *7.4.1. Reducing Aquatic Toxicity*

As discussed previously, aquatic toxicity is usually due to either nonspecific mechanisms (necrosis) or specific toxic interactions. Most chemicals cause toxicity to aquatic species by simple necrosis (either nonpolar or polar), resulting in a perturbation of cellular functions. Chlorinated hydrocarbons, alcohols, ethers, ketones, weak organic acids and bases, and simple nitro-compounds are among the chemicals that act by such a nonspecific mechanism. The mechanisms of specific aquatic toxicity include electrophilic attack of macromolecules, CNS seizure, AChE inhibition, neurodepression, respiratory blocking, and uncoupling of oxidative phosphirylation. The following physicochemical properties, listed in Table 13, are known to favor reduced toxicity.

**7.4.1.1. Example: Deriving Rules for Safer Aquatic Chemicals to Fathead Minnow (*Pimephales promelas*).** Data from the EPA's assay on aquatic toxicity using the fathead minnow[226] has been used to derive some physicochemical property design guidelines for aquatic toxicity.[225] Predicted properties were calculated using Schrodinger QikProp from optimized 3D molecular structures for the 617 compounds in the EPA assay, 571 of which were neutral organic compounds and, thus, could be analyzed.

A selection of the top three properties most relevant to toxicity yielded log $P_{ow}$, LUMO energy values, and globularity. The relationship between these three properties and toxicity is explored using a 3D scatterplot with color-coded confidence ellipsoids in Figure 21. This figure shows that the cluster of green ("no concern") compounds is spatially distinct from that of the red (highest concern) compounds and that these areas do not overlap.

The three property limits for log $P_{ow}$, LUMO, and globularity proposed as design rules are listed in Table 14. Of the 41 compounds in the EPA fathead minnow (FHM) data set that meet these design rules, 40 had $LC_{50}$ values greater than 200 mg/L (i.e. low or no concern for aquatic toxicity). The mean $LC_{50}$ value for these 45 compounds is 6 466 mg/L. The mean $LC_{50}$ for the entire FHM data set is 971 mg/L. There are also 146 compounds that do not meet any of the three criteria, i.e., have a log $P_{ow}$ of <1, LUMO energy < 2 eV, and globularity < 0.90. The mean $LC_{50}$ value for these compounds is 22.9 mg/L, indicating that they are of much greater concern for acute aquatic toxicity than the average FHM compound.

### *7.4.2. Enhancing Biodegradability*

Increased biodegradability to nontoxic or less toxic substances, like reduced toxicity, is integral to chemical design. Chemicals that resist biodegradation continue to exert toxic effects (if present) on the environment, and ones that are bioaccumulative are of even greater concern because levels may be achieved in organisms that appear safe on the

**Table 14. Design Rules for Reduced Aquatic Toxicity to the Fathead Minnow**

| property | minimum limit | maximum limit |
|---|---|---|
| log P(o/w) | −4[a] | 1 |
| LUMO (AM1, in eV) | 2 | 8[a] |
| globularity | 0.90 | 1[a] |

[a] These limits were guided by the range of the data set but can potentially be extended. For example, the range of log P(o/w) was −3.7 to 6.6, and thus the lower limit is set at −4, but it is possible that compounds with a log $P_{ow} < -4$ would also meet these criteria and be considered to have no concern for acute aquatic toxicity.

basis of a single daily exposure, but because the actual dose effectively accumulates overtime, the result may be unforeseen toxic effects. Producing chemicals that are subject to biodegradation and thus are removed from the environment quickly is of major importance. Much work has been done over the past 50 years in the development and application of methods to measure biodegradability of chemicals, especially detergent chemicals and pesticides—two classes of chemicals that are frequently discharged or used in the environment. By the late 1970s, standardized biodegradation protocols were emerging and have since been applied to thousands of chemicals and further developed.

On the basis of a review of the significant body of biodegradability data, a series of generalizations about the influence of chemical structure on biodegradability have been articulated by Boethling et al.[227] These "rules of thumb" include generalizations about the effects of various substituent groups or substructures, the number of times a given substituent appears in a molecule, and substituent position. As generalizations, these have qualitative value at the screening level of detail and are useful starting points for a more rigorous examination of biodegradability of a given chemical.

It must be noted that a compound's biodegradability is affected by broad categories of factors—the molecular structure and the exposure conditions, which broadly can

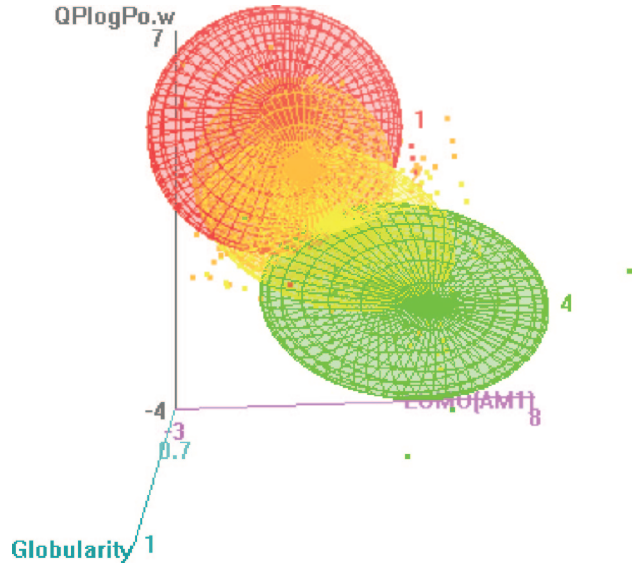
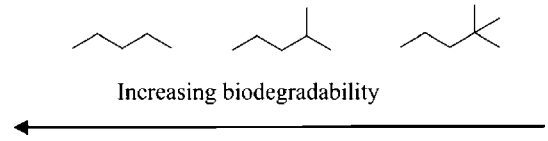


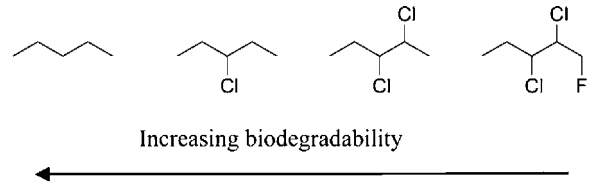

**Figure 21.** 3D scatter plot of the three properties (log $P_{ow}$, LUMO energy, and globularity) found to provide the greatest information about level of acute aquatic toxicity concern. Compounds with high, medium, low, and no concern for acute aquatic toxicity are represented by red, orange, yellow, and green dots, respectively. The 80% ellipsoids for each group are indicated in the same color.

mean the environmental conditions (like pH), the nature of micoroorganisms likely to attack the chemical, and waste treatment. At the molecular level, some of the reasons for a substrate to be recalcitrant to microbial attack include, but are not limited to, lack of transport mechanism into the cell, inability to be a substrate for available enzymes, toxicity to the microorganism, or inability to be metabolized into useful feedstocks.[228] On the basis of this analysis, the following molecular features, listed below, generally decrease the rate of aerobic biodegradation. It should certainly be noted that only a small number of generalizations are acceptable even for qualitative use and that there are many exceptions. The following features should thus be avoided to decrease chance of recalcitrance to biodegradation.

Chain branching, if extensive, impedes biodegradability. Quaternary carbons have the greatest impact, but there are exceptions—some naturally occurring compounds, such as vitamin A, cholesterol, and pantothenic acid have a quaternary carbon yet are relatively biodegradable.

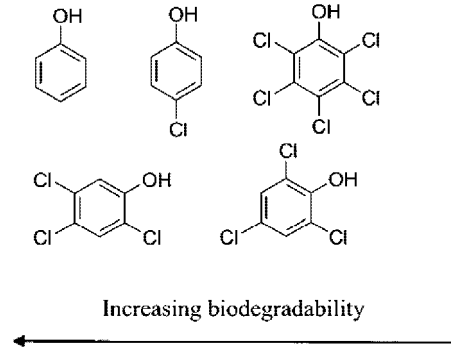



Strongly electron-withdrawing substituents, such as halogens (chlorine and fluorine especially), greatly impede biodegradability. This effect is magnified if there are more than three in a small molecule.




Heterocyclic residues (e.g., imidazole and aliphatic ether bonds, except in ethoxylates) also decrease biodegradability.

Presence of a chlorine atom on a phenyl ring makes the ring less susceptible to attack by oxygenase enzymes. The substituent position also matters (see example below), but no broadly applicable generalizations can be made about substituent position. Each class of compounds needs to be examined individually.




The substituent position also matters (see example below), but no broadly applicable generalizations can be made about substituent position, and thus each class of compounds must be examined individually: tertiary amine, nitro, nitroso, azo, and arylamino groups.

For polycyclic aromatics, more rings, especially with more than three fused rings, decrease biodegradability.

**Table 15. Effects on Biodegradability of Chemical Structure Modifications Recommended for Reducing Toxicity[a]**

| structural modification | toxicity end point/objective | effect on biodegradability |
|---|---|---|
| increase MW to >1000 | lower aquatic toxicity | decrease |
| reduce water solubility to <1 $\mu$g/L | lower aquatic toxicity | decreases availability to biodegradation enzymes |
| increase steric hindrance at active site | lower aquatic toxicity | decreases availability to biodegradation enzymes |
| add bulky ortho groups | reduce oncogenicity concern for aromatic amines | decreases accessibility to biodegradation enzymes |
| add hydrophilic groups (sulfonate or COOH) | reduce oncogenicity concern (enhance excretion) | may increase or decrease depending on group |

[a] Reproduced with permission from ref 227. Copyright 2007 American Chemical Society.

The presence of the following in a compound will generally increase aerobic biodegradability:

Groups susceptible to enzymatic hydrolysis, especially esters (including phosphate esters) and to a lesser extent amides, have a very favorable effect on biodegradability. The enzymatic hydrolysis of esters is a ubiquitous and crucial step in the degradation of chemicals in the environment. Esterases are widespread in the biota and have broad substrate specificity.

Oxygen atoms in the form of hydroxyl, aldehyde, ketones, or carboxylic acid groups help biodegradation. Ethers are not part of this group, with the exception of ethoxylate groups. The microbial-mediated oxygenation of chemicals is usually the first and often rate-limiting step in the biodegradation of many compounds without ester groups. Small molecules that contain oxygen biodegrade more readily than do the same molecules that do not. For example, phenol degrades more readily in a bacterial incubato than benzene.

Cyclohexanol and cyclohexanone degrade more easily than cyclohexane, and alcohols and carboxylic acids degrade more readily than the corresponding aliphatic hydrocarbons.

Unsubstituted linear alkyl chains (especially with 4 or more carbons) and phenyl rings are especially susceptible to biodegradation because these are easily acted upon oxygenases, resulting in the insertion of an oxygen molecule. As mentioned above, the presence of oxygen enhances the rate of biodegradability.

As duly noted by Boethling et al., the enhancement of biodegradability and reduction of toxicity by structural modifications can be divergent in some cases. For example, compounds with MW > 1000 that are poorly bioavailable are, in general, not easily biodegraded. Thus, the challenge to design molecules that are nontoxic to humans and other species in the environment but biodegradable is indeed formidable. Some examples of structural modifications that decrease both toxicity and biodegradability are reproduced from Boethling et al. in Table 15.

Moreover, it is important to consider the toxicity and fate of possible breakdown products of chemicals in which biodegradability has been intentionally designed. Although biodegradation generally results in less persistent and less toxic chemicals, this is not always the case, e.g., the degradation of the nonionic surfactant nonylphenolethoxylates to the more toxic and bioaccumulative nonylphenols. Thus, care must be taken to avoid unintended consequences when designing chemical to be favorable to biodegradation.

### 7.4.3. Minimizing Bioaccumulation

Bioaccumulation occurs in organisms when the rate of uptake of a chemical (by all routes of exposure—air, food, soil/sediment, and water) exceeds the rate of elimination. The bioaccumulation of chemicals is a concern that first gained public attention with DDT in the 1960s and continues today for chemicals such as the perfluorinated hydrocarbons, where the resulting increased body burden of the chemicals makes the possibility of adverse effects greater. Clearly, bioaccumulation is a phenomenon that should be minimized when designing safer chemicals.

The bioaccumulation factor (BAF) is the ratio between the concentration of the chemical in the biota and its associated exposure compartments (air, food, soil/sediment, and water) at steady state and is an estimate of the tendency of a chemical to bioacumulate in the food chain.

The term bioconcentration is a more narrowly defined subset of the process of bioaccumulation and refers to the uptake and concentration of chemicals from water into aquatic organisms. Thus, the bioconcentration factor (BCF) is the ratio between the concentration of the chemical in biota and the concentration *in water* at steady state. The BCF can also be calculated by the ratio of the first-order uptake and elimination rate constants, a method that does not require equilibrium conditions. The BCF can be measured experimentally directly and is generally considered important to obtain for chemicals that have a log $P_{ow} > 3$. Numerous QSAR models have been reported for the prediction of BCF, most of which are based on log $P_{ow}$ and apply well to neutral organic substances that are not easily metabolized.[229] The BCF−log $P_{ow}$ relationship applies generally to neutral organic substances with log $P_{ow}$ from 3−8. One must be careful applying this relationship. Poorly lipid-soluble chemicals, those that are highly lipophillic (log $P_{ow} > 8$), or chemicals with a molecular weight > 700 Da will generally not bioconcentrate as sometimes predicted by the QSAR model. Moreover, the role of metabolism (microbial, plant, or animal) must be considered as it can render a predicted bioaccumulative chemical to be relatively short-lived in the biota.

The fact that bioaccumulation is a result of an inbalance between uptake and elimination presents added complexity when establishing a framework with reduced hazard. For example, concerns over a high predicted BAF may be mitigated by designing for phase I and phase II metabolism. These and other considerations are needed for the optimum design of chemicals with reduced hazard and must be part of a new comprehensive molecular design framework.

## 8. Conclusion

This review has tracked the essential research advances that not only have advanced the current state of the knowledge of molecular toxicology but have also provided the building blocks to the emerging area of molecular design for reduced hazard. The efforts in this nascent area of research demonstrate that, as shown in drug development, it is possible to translate the knowledge base of molecular toxicology into insights for the molecular designer in reducing the potential for a new molecule to manifest adverse biological activity. Certainly, the early-stage work deals primarily with the relatively less complex aspects of toxicology related to bioavailability. It must be fully appreciated that the deeper the emerging understanding of system biology

becomes, the more complicated and difficult the task of the molecular designer will likely be. Future research both in the short and much longer term will need to address such issues as metabolic products of parent compounds, synergistic effects, bioactivation, epigenetics, and multiple competing mechanisms of action, to name just a few. While the research path ahead is certainly a long and likely torturous one, the positive aspect is that there is a path.

## 9. Acknowledgments

We acknowledge Prof. Julie Zimmerman for collaboration and fruitful discussions, Prof. William Jorgensen and Dr. Julian Tirado-Rives for assistance with use of QikProp software, and Dr. Stephen DeVito for many insightful discussions and provision of the U.S. EPA's Toxic Release Inventory list of chemicals. For funding, we thank the Johnson Family Foundation and the Kendeda Fund.

## 10. References

(1) Kent, C. 1−3 Role of the toxicologist. In *Basics of Toxicology (Preserving the Legacy)*; Wiley, John & Sons: New York, 1998.
(2) Dearden, J. C. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 119.
(3) Ariens, E. J. I.; Medicinal Chemistry, A. S. o. M. V., Design of Safer Chemicals. In *Drug Design*; Ariens, E. J., Ed.; Academic Press: New York, 1980; Vol. IX, pp 1−46.
(4) Ariens, E. J. *Drug Metab. Rev.* **1985**, *15*, 425.
(5) DeVito, S.; Garrett, R. *Designing Safer Chemicals: Green Chemistry for Pollution Prevention*; American Chemical Society: Washington, DC, 1996; Vol. 640.
(6) Topliss, J. G. *Quantitative structure−activity relationships of drugs*; Academic Press: New York, 1983; Vol. xii, p 519.
(7) Palanza, P.; Parmigiani, S.; vom Saal, F. S. In *Joint Meeting of the 6th International Conference on Hormones, Brain, and Behavior/ Society-for-Behavioral-Neuroendocrinology*; Academic Press Inc.: Madrid, Spain, 2000; pp 252−265.
(8) Welshons, W. V.; Thayer, K. A.; Judy, B. M.; Taylor, J. A.; Curran, E. M.; vom Saal, F. S. *Environ. Health Perspect.* **2003**, *111*, 994.
(9) Shrader-Frechette, K. *Hum. Exp. Toxicol.* **2008**, *27*, 647.
(10) General Acccounting Office. Options Exist to Improve EPA's Ability to Assess Health Risks and Manage Its Chemical Review Program. In *Chemical Regulation*; U.S. Government Accounting Office: Washington, DC, 2005; Vol. 2009.
(11) Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J.; *Toxicol. Sci.* **2007**; *95*, 5−12.
(12) Elmore, E. *In Vitro Cell. Dev. Biol.: Anim.* **2008**, *44*, S7.
(13) Horvath, I. T. *Molecular Design: Chemical Structure Generation from the Properties of Pure Organic Compounds*; Elsevier Science Ltd.: Oxford, 1992.
(14) Pennie, W. D.; Woodyatt, N. J.; Aldridge, T. C.; Orphanides, G. *Toxicol. Lett.* **2001**, *120*, 353.
(15) Urushidani, T.; Nagao, T. Toxicogenomics: Japanese initiative. In *Handbook of Toxicogenomics: Strategies and Applications*; Borlak, J., Ed.; Wiley-VCH: Weinheim, Germany, 2005.
(16) Nuwaysir, E. F.; Bittner, M.; Trent, J.; Barrett, J. C.; Afshari, C. A. *Mol. Carcinog.* **1999**, *24*, 153.
(17) Farr, S.; Dunn, R. T. *Toxicol. Sci.* **1999**, *50*, 1.
(18) Harries, H. M.; Fletcher, S. T.; Duggan, C. M.; Baker, V. A. *Toxicol. in Vitro* **2001**, *15*, 399.
(19) Bushel, P. R.; Bennett, L.; Hamadeh, H.; Green, J.; Ableson, A.; Misener, S.; Paules, R.; Afshari, C. *Funct. Monit. Drug-Tissue.* **2002**, *4623*, 85.
(20) Hamadeh, H. K.; Bushel, P. R.; Jayadev, S.; Martin, K.; DiSorbo, O.; Sieber, S.; Bennett, L.; Tennant, R.; Stoll, R.; Barrett, J. C.; Blanchard, K.; Paules, R. S.; Afshari, C. A. *Toxicol. Sci.* **2002**, *67*, 219.
(21) Al-Shaiba, R.; McMillan, D. C.; Angerson, W. J.; McArdle, C. S.; Horgan, P. *Br. J. Cancer* **2004**, *91*, 205.
(22) Barber, M. D.; Preston, T.; McMillan, D. C.; Slater, C.; Ross, J. A.; Fearon, K. C. *Clin. Sci. (Lond)* **2004**, *106*, 359.
(23) Waring, J. F.; Ciurlionis, R.; Jolly, R. A.; Heindel, M.; Ulrich, R. G. *Toxicol. Lett.* **2001**, *120*, 359.
(24) Waring, J. F.; Jolly, R. A.; Ciurlionis, R.; Lum, P. Y.; Praestgaard, J. T.; Morfitt, D. C.; Buratto, B.; Roberts, C.; Schadt, E.; Ulrich, R. G. *Toxicol. Appl. Pharmacol.* **2001**, *175*, 28.
(25) Hamadeh, H. K.; Knight, B. L.; Haugen, A. C.; Sieber, S.; Amin, R. P.; Bushel, P. R.; Stoll, R.; Blanchard, K.; Jayadev, S.; Tennant, R. W.; Cunningham, M. L.; Afshari, C. A.; Paules, R. S. *Toxicol. Pathol.* **2002**, *30*, 470.
(26) Zhou, T.; Jia, X.; Chapin, R. E.; Maronpot, R. R.; Harris, M. W.; Liu, J.; Waalkes, M. P.; Eddy, E. M. *Toxicol. Lett.* **2004**, *154*, 191.
(27) Naciff, J. M.; Overmann, G. J.; Torontali, S. M.; Carr, G. J.; Tiesman, J. P.; Daston, G. P. *Environ. Health Perspect.* **2004**, *112*, 1519.
(28) Naciff, J. M.; Daston, G. P. *Toxicol. Pathol.* **2004**, *32* (Suppl 2), 59.
(29) Ruepp, S. U.; Tonge, R. P.; Shaw, J.; Wallis, N.; Pognan, F. *Toxicol. Sci.* **2002**, *65*, 135.
(30) Coen, M.; Ruepp, S. U.; Lindon, J. C.; Nicholson, J. K.; Pognan, F.; Lenz, E. M.; Wilson, I. D. *J. Pharm. Biomed. Anal.* **2004**, *35*, 93.
(31) Kleno, T. G.; Kiehr, B.; Baunsgaard, D.; Sidelmann, U. G. *Biomarkers* **2004**, *9*, 116.
(32) Craig, A.; Sidaway, J.; Holmes, E.; Orton, T.; Jackson, D.; Rowlinson, R.; Nickson, J.; Tonge, R.; Wilson, I.; Nicholson, J. *J. Proteome Res.* **2006**, *5*, 1586.
(33) Fliri, A. F.; Loging, W. T.; Thadeio, P. F.; Volkmann, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 261.
(34) Janzen, W. P.; Hodge, C. N. *Chem. Biol. Drug Des.* **2006**, *67*, 85.
(35) Smith, S. C.; Delaney, J. S.; Robinson, M. P.; Rice, M. J. *Comb. Chem. High Throughput Screen.* **2005**, *8*, 577.
(36) Tietjen, K.; Drewes, M.; Stenzel, K. *Comb. Chem. High Throughput Screen.* **2005**, *8*, 589.
(37) Bhogal, N.; Grindon, C.; Combes, R.; Balls, M. *Trends Biotechnol.* **2005**, *23*, 299.
(38) Xiang, A. B.; Kanematsu, M.; Mitamura, M.; Kikkawa, H.; Asano, S.; Kinoshita, M. *Invest. Radiol.* **2006**, *41*, 704.
(39) Austin, C. P. *Annu. Rev. Med.* **2004**, *55*, 1.
(40) NIH. Molecular Libraries Program. 2007; http://mli.nih.gov/mli/ (accessed March 15, 2009).
(41) Boelsterli, U. A. *Mechanistic Toxicology: The Molecular Basis of How Chemicals Disrupt Biological Targets*; Taylor & Francis: New York, 2003.
(42) Klaassen, C. D. E. *Casarett & Doull's Toxicology: The Basic Science of Poisons*, 7th ed.; McGraw Hill: New York, 2008.
(43) Gordon, G. *Dev. Med. Child Neurol.* **1966**, *8*, 761.
(44) Bower, M.; Howard, M.; Gracie, F.; Phillips, R.; Fife, K. *JAIDS, J. Acquired Immune Defic. Syndr.* **1997**, *14*, 76.
(45) Riflkin, I. R.; Mchugh, S. M.; Bradley, J. R.; Thiru, S.; Ewan, P. W.; Lockwood, C. M. *Kidney Int.* **1995**, *47*, 672.
(46) D'Amato, R. J.; Lentzsch, S.; Anderson, K. C.; Rogers, M. S. *Semin. Oncol.* **2001**, *28*, 597.
(47) Fabro, S. *Acta Vitamin. Enzymol.* **1967**, *21*, 123.
(48) Fabro, S.; Shull, G.; Dixon, R. *Pharmacologist* **1976**, *18*, 231.
(49) Mcbride, W. G.; Vardy, P.; Stokes, P. A. *Teratology* **1982**, *25*, A61.
(50) Neubert, R.; Hinz, N.; Thiel, R.; Neubert, D. *Life Sci.* **1995**, *58*, 295.
(51) Shull, G. E. *J. Theor. Biol.* **1984**, *110*, 461.
(52) Stephens, T. D.; Bunde, C. J. W.; Fillmore, B. J. *Biochem. Pharmacol.* **2000**, *59*, 1489.
(53) Koch, H. P.; Czejka, M. J. *Z. Naturforsch., C: J. Biosci.* **1986**, *41*, 1057.
(54) Shoji, A.; Kuwahara, M.; Ozaki, H.; Sawai, H. *J. Am. Chem. Soc.* **2007**, *129*, 1456.
(55) DeVito, S. C. In *Designing Safer Chemicals*; DeVito, S., Garrett, R., Eds.; American Chemical Society: 1996; Vol. 640, 16.
(56) Shargel, L.; Yu, A. B. *Applied biopharmaceutics & pharmacokinetics*, 4th ed.; McGraw-Hill: New York, 1999.
(57) Hurst, S.; Loi, C. M.; Brodfuehrer, J.; El-Kattan, A. *Exp. Opin. Drug Metab. Toxicol.* **2007**, *3*, 469.
(58) Kesisoglou, F.; Wu, Y. H. *AAPS J.* **2008**, *10*, 516.
(59) Golovenko, N. Y.; Borisyuk, I. Y. *Biomed. Khim.* **2008**, *54*, 392.
(60) Oleszczuk, P. *Biotechnologia (Poznan)* **2007**, 9.
(61) Oleszczuk, P. *Biotechnologia (Poznan)* **2007**, 26.
(62) Hermens, J. L. M.; Heringa, M. B.; ter Laak, T. L. *J. Toxicol. Environ. Health, Part A* **2007**, *70*, 727.
(63) Slaveykova, V. I.; Wilkinson, K. J. *Environ. Chem.* **2005**, *2*, 9.
(64) Shore, P. A.; Brodie, B. B.; Hogben, C. A. M. *J. Pharmacol. Exp. Ther.* **1957**, *119*, 361.
(65) Tsuruta, H. *Ind. Health* **1989**, *27*, 37.
(66) Ursin, C.; Hansen, C. M.; Vandyk, J. W.; Jensen, P. O.; Christensen, I. J.; Ebbehoej, J. *Am. Ind. Hyg. Assoc. J.* **1995**, *56*, 651.
(67) Olsson, B.; Pool, J.; Vandermo, P.; Varnausk, E; Wassen, R. *Cardiology* **1970**, *55*, 136.
(68) Pohjanvirta, R.; Laitinen, J. T.; Vakkuri, O.; Linden, J.; Kokkola, T.; Unkila, M.; Tuomisto, J. *Toxicology* **1996**, *107*, 85.
(69) Huggins, R. A.; Smith, E. L.; Deavers, S. *Am. J. Physiol.* **1963**, *205*, 351.
(70) Avram, M. J.; Krejcie, T. C. *Clin. Pharmacol. Ther.* **2002**, *71*, P99.
(71) Sui, X. F.; Sun, J.; Wu, X.; Li, H. Y.; Liu, J. F.; He, Z. G. *Curr. Drug Metab.* **2008**, *9*, 574.
(72) Obach, R. S.; Lombardo, F.; Waters, N. J. *Drug Metab. Dispos.* **2008**, *36*, 1385.

(73) Timbrell, J. A. *Principles of Biochemical Toxicology*, 4th ed.; Informa Healthcare: New York, 2009.
(74) Tsuji, A.; Tamai, I.; Sakata, A.; Tenda, Y.; Terasaki, T. *Biochem. Pharmacol.* **1993**, *46*, 1096.
(75) Aschner, M.; Aschner, J. L. *Neurosci. Biobehav. Rev.* **1990**, *14*, 169.
(76) Kelder, J.; Grootenhuis, P. D. J.; Bayada, D. M.; Delbressine, L. P. C.; Ploemen, J. P. *Pharm. Res.* **1999**, *16*, 1514.
(77) Strazielle, N.; Ghersi-Egea, J. F. *Rev. Med. Virol.* **2005**, *15*, 105.
(78) Hertz-Picciotto, I.; Schramm, M.; Watt-Morse, M.; Chantala, K.; Anderson, J.; Osterloh, J. *Am. J. Epidemiol.* **2000**, *152*, 829.
(79) Arruda, J. D. T.; de Oliveira, M. C. C.; Sarkis, J. E. S.; Bordini, P.; Manso-Guevara, M. V.; Garcia, F.; Prado, G. R.; Krug, F. J.; Mesa, J.; Bittencourt-Oliveira, M. C.; Garcia, C.; Rodrigues, T. E.; Shtejer, K.; Genofre, G. C. *Environ. Int.* **2009**, *35*, 614.
(80) Sanin, L. H.; Gonzalez-Cossio, T.; Romieu, I.; Hernandez-Avila, M. *Salud Publ. Mexico* **1998**, *40*, 359.
(81) Silbergeld, E. K.; Sauk, J.; Somerman, M.; Todd, A.; Mcneill, F.; Fowler, B.; Fontaine, A.; Vanburen, J. *Neurotoxicology* **1993**, *14*, 225.
(82) Dietrich, D. R.; Swenberg, J. A. *Cancer Res.* **1991**, *51*, 3512.
(83) Pond, S. M. *Med. J. Aust.* **1990**, *152*, 256.
(84) Dinis-Oliveira, R. J.; Duarte, J. A.; Sanchez-Navarro, A.; Remiao, F.; Bastos, M. L.; Carvalho, F. *Crit. Rev. Toxicol.* **2008**, *38*, 13.
(85) Groves, C. E.; Morales, M. N.; Gandolfi, A. J.; Dantzler, W. H.; Wright, S. H. *J. Pharmacol. Exp. Ther.* **1995**, *275*, 926.
(86) Toursarkissian, B.; Endean, E. D.; Aziz, S. M. *J. Surg. Res.* **1994**, *57*, 401.
(87) Chung, K. W.; Chandler, A. R.; Key, P. B. *J. Environ. Sci. Health, Part B* **2008**, *43*, 293.
(88) Osullivan, M. C.; Golding, B. T.; Smith, L. L.; Wyatt, I. *Biochem. Pharmacol.* **1991**, *41*, 1839.
(89) Williams, J. A.; Hurst, S. I.; Bauman, J. *Curr. Drug Metab.* **2003**, *4*, 527.
(90) Casarett, L. J.; Klaassen, C. D. *Casarett and Doull's toxicology: the basic science of poisons*, 7th ed.; McGraw-Hill Medical: New York, 2008; Vol. xv, 1310 pp.
(91) Brodie, B. B. In *Ciba Foundation Symposium on Drug Responses in Man*; Wolstenholme, G. E. W., Porter, R., Eds.; J. & A. Churchill: London, 1967; p 188.
(92) Brodie, B. B.; Reid, W. D.; Cho, A. K.; Sipes, G.; Krishna, G.; Gillette, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 160.
(93) Miller, E. C.; Miller, J. A. *Pharmacol. Rev.* **1966**, *18*, 805.
(94) Evans, W. E.; Relling, M. V. *Science* **1999**, *286*, 487.
(95) Lewis, D. F. V. *Pharmacogenomics* **2004**, *5*, 305.
(96) Li, H. Y.; Sun, J.; Fan, X. W.; Sui, X. F.; Zhang, L.; Wang, Y. J.; He, Z. G. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 843.
(97) de Montellano, P. R. O. *Chem. Rev.* **2010**, *110*, 932.
(98) Lewis, D. F. V. *Curr. Drug Metab.* **2003**, *4*, 331.
(99) Lewis, D. F. V.; Eddershaw, P. J.; Dickens, M.; Tarbit, M. H.; Goldfarb, P. S. *Chem.-Biol. Interact.* **1998**, *115*, 175.
(100) Bauer, C.; Osman, A. M.; Cercignani, G.; Gialluca, N.; Paolini, M. *Biochem. Pharmacol.* **2001**, (61), 1049.
(101) Lewis, D. F. V.; Dickins, M. *Drug Discov. Today* **2002**, *7*, 918.
(102) Lewis, D. F. V.; Jacobs, M. N.; Dickins, M. *Drug Discov. Today* **2004**, *9*, 530.
(103) Algailany, K. A. S.; Houston, J. B.; Bridges, J. W. *Biochem. Pharmacol.* **1978**, *27*, 783.
(104) Hansch, C.; Zhang, L. T. *Drug Metab. Rev.* **1993**, *25*, 1.
(105) Ishigami, M.; Honda, T.; Takasaki, W.; Ikeda, T.; Komai, T.; Ito, K.; Sugiyama, Y. *Drug Metab. Dispos.* **2001**, *29*, 282.
(106) Riley, R. J.; Parker, A. J.; Trigg, S.; Manners, C. N. *Pharm. Res.* **2001**, *18*, 652.
(107) Willhite, C. C.; Smith, R. P.; Smith, L. *Pharmacologist* **1979**, *21*, 187.
(108) Grogan, J.; DeVito, S. C.; Pearlman, R. S.; Korzekwa, K. R. *Chem. Res. Toxicol.* **1992**, *5*, 548.
(109) DeVito, S. C. Designing safer nitriles. In *Designing Safer Chemicals*; DeVito, S. C., Garrett, R., Eds.; American Chemical Society: Washington DC, 1996; Vol. 640, pp 194−223.
(110) Olsen, L.; Rydberg, P.; Rod, T. H.; Ryde, U. *J. Med. Chem.* **2006**, *49*, 6489.
(111) Lewis, D. F. V.; Brantom, P. G.; Ioannides, C.; Walker, R.; Parke, D. V. *Drug Metab. Rev.* **1997**, *29*, 1055.
(112) Yang, C. S.; Yoo, J. S. H.; Ishizaki, H.; Hong, J. Y. *Drug Metab. Rev.* **1990**, *22*, 147.
(113) Poulsen, M.; Spanger, D.; Loew, G. H. *Mol. Toxicol.* **1987**, *I*, 35.
(114) Kamataki, T.; Fujita, K.; Nakayama, K.; Yamazaki, Y.; Miyamoto, M.; Ariyoshi, N. *Drug Metab. Rev.* **2002**, *34*, 667.
(115) Wishnok, J. S.; Archer, M. C.; Edelman, A. S.; Rand, W. M. *Chem.-Biol. Interact.* **1978**, *20*, 43.
(116) Lewis, D. F. V. *Drug Metab. Rev.* **1999**, *31*, 755.
(117) Peto, R.; Gray, R.; Brantom, P.; Grasso, P. *Cancer Res.* **1991**, *51*, 6415.
(118) Lewis, D. F. V. *Drug Metab. Rev.* **1997**, *29*, 589.
(119) Lewis, D. F. V.; Pratt, J. M. *Drug Metab. Rev.* **1998**, *30*, 739.
(120) Benigni, R.; Andreoli, C.; Giuliani, A. *Carcinogenesis* **1989**, *10*, 55.
(121) Lewis, D. F. V.; Ioannides, C.; Parke, D. V.; Walker, R. *Food Addit. Contam.* **1995**, *12*, 715.
(122) Korzekwa, K. R.; Jones, J. P.; Gillette, J. R. *J. Am. Chem. Soc.* **1990**, *112*, 7042.
(123) White, R. E.; Mccarthy, M. B.; Egeberg, K. D.; Sligar, S. G. *Arch. Biochem. Biophys.* **1984**, *228*, 493.
(124) Agency for Toxic Substances and Disease Registry, D. o. H. a. H. S. ToxFAQs for Benzene. http://www.atsdr.cdc.gov/tfacts3.html, accessed 2007.
(125) Korzekwa, K. R.; Swinney, D. C.; Trager, W. F. *Biochemistry* **1989**, *28*, 9019.
(126) Bathelt, C. M.; Ridder, L.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2003**, *125*, 15004.
(127) Rietjens, I. M. C. M.; Soffers, A. E. M. F.; Veeger, C.; Vervoort, J. *Biochemistry* **1993**, *32*, 4801.
(128) Rydberg, P.; Ryde, U.; Olsen, L. *J. Phys. Chem. A* **2008**, *112*, 13058.
(129) Vasanthanathan, P.; Taboureau, O.; Oostenbrink, C.; Vermeulen, N. P. E.; Olsen, L.; Jorgensen, F. S. *Drug Metab. Rev.* **2008**, *40*, 36.
(130) Richet, C. *C. R. Soc. Biol.* **1893**, *54*, 775.
(131) Meyer, H. H. *Arch. Exp. Pathol. Pharmakol.* **1899**, *42*, 109.
(132) Overton, E. *Vierteljahresschr. Naturforsch. Ges. (Ziirich)* **1899**, *44*, 88.
(133) Kubinyi, H. *J. Med. Chem.* **1977**, *20*, 625.
(134) Ferguson, J. *Proc. R. Soc. London B* **1939**, *127*, 387.
(135) McGowan, J. C. *J. Appl. Chem. (London)* **1952**, *2*, 323.
(136) Mullins, L. J. *Chem. Rev.* **1954**, *54*, 289.
(137) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. *Nature* **1962**, *194*, 178.
(138) Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
(139) Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C. D. *Chem. Rev.* **2002**, *102*, 783.
(140) Seydel, J. K. *Mol. Pharmacol.* **1966**, *2*, 259.
(141) LeBlanc, G. A.; Kaiser, K. L. E. *QSAR in environmental toxicology. Proceedings of the workshop on quantitative structure−activity relationships in environmental toxicology held at McMaster University, Hamilton, Ontario, Canada, August 16−18, 1983* **1984**, 235.
(142) Gao, C.; Govind, R.; Tabak, H. H. *Environ. Toxicol. Chem.* **1992**, *11*, 631.
(143) Roy, K.; Roy, P. P. *Eur. J. Med. Chem.* **2009**, *44*, 1941.
(144) Costescu, A.; Moldovan, C.; Katona, G.; Diudea, M. *J. Math. Chem.* **2009**, *45*, 287.
(145) Selassie, C. D.; Garg, R.; Kapur, S.; Kurup, A.; Verma, R.; Mekapati, S. B.; Hansch, C. *Chem. Rev.* **2002**, *102*, 2585.
(146) Neely, W. B.; Branson, D. R.; Blau, G. E. *Environ. Sci. Technol.* **1974**, *8*, 1113.
(147) Konemann, H. *Toxicology* **1981**, *19*, 209.
(148) Martin, T. M.; Young, D. M. *Chem. Res. Toxicol.* **2001**, *14*, 1378.
(149) Mazzatorta, P.; Benfenati, E.; Neagu, C. D.; Gini, G. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 513.
(150) Gong, Z. G.; Xia, B. B.; Zhang, R. S.; Zhang, X. Y.; Fan, B. T. *QSAR Comb. Sci.* **2008**, *27*, 967.
(151) Kompany-Zareh, M. *Med. Chem. Res.* **2009**, *18*, 143.
(152) Singh, R. K.; Khan, A. K. R.; Sahu, V. K.; Singh, P. P. *Int. J. Quantum Chem.* **2009**, *109*, 185.
(153) Veith, G. D.; Broderius, S. J. *Environ. Health Perspect.* **1990**, *87*, 207.
(154) Cronin, M. T. D.; Livingstone, D. *Predicting chemical toxicity and fate*; CRC Press: Boca Raton, FL, 2004; 445 pp.
(155) Richard, A. M.; Yang, C.; Judson, R. S. *Toxicol. Mech. Meth.* **2008**, *18*, 103.
(156) Liu, B.; Li, S.; Hu, J. *Am. J. PharmacoGenomics* **2004**, *4*, 263.
(157) Database, B. I. o. H. a. M. C. http://chembank.broad.harvard.edu/, accessed 2008.
(158) Czodrowski, P.; Kriegl, J. M.; Scheuerer, S.; Fox, T. *Exp. Opin. Drug Metab. Toxicol.* **2009**, *5*, 15.
(159) Mohan, C. G.; Gandhi, T.; Garg, D.; Shinde, R. *Mini-Rev. Med. Chem.* **2007**, *7*, 499.
(160) Enslein, K.; Craig, P. N. *J. Environ. Pathol. Toxicol.* **1978**, *2*, 115.
(161) Enslein, K. *Pharmacol. Rev.* **1984**, *36*, 131S.
(162) Carlsen, L.; Kenessov, B. N.; Batyrbekova, S. Y. *Environ. Toxicol. Pharmacol.* **2009**, *27*, 415.
(163) Sanderson, D. M.; Earnshaw, C. *Hum. Exp. Toxicol.* **1991**, *10*.
(164) Ideaconsult. *Toxtree v1.51*; http://ecb.jrc.it/qsar/qsar-tools/index-.php?c1/4TOXTREE, Accessed March 15, 2010.
(165) Pavan, M.; Worth, A. P. *SAR QSAR Environ. Res.* **2008**, *19*, 785.
(166) Cramer, G. M.; Ford, R. A.; Hall, R. L. *Food Cosmet. Toxicol.* **1978**, *16*, 255.
(167) Benigni, R. *Drug Discovery Today: Technol.* **2004**, *1*.
(168) Woo, Y. T.; Lai, D. Y.; Argus, M. F.; Arcos, J. C. *Toxicol. Lett.* **1995**, *79*, 219.

(169) Mackenzie, P. I.; Owens, I. S.; Burchell, B.; Bock, K. W.; Bairoch, A.; Belanger, A.; FournelGigleux, S.; Green, M.; Hum, D. W.; Iyanagi, T.; Lancet, D.; Louisot, P.; Magdalou, J.; Chowdhury, J. R.; Ritter, J. K.; Schachter, H.; Tephly, T. R.; Tipton, K. F.; Nebert, D. W. *Pharmacogenetics* **1997**, *7*, 255.
(170) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. *Proteins* **2005**, *58*, 836.
(171) Tu, Y.; Laaksonen, A. *Advances in Quantum Chemistry, Vol 59* **2010**, *59*, 1.
(172) Ranaghan, K. E.; Mulholland, A. J. *International Reviews in Physical Chemistry* **2010**, *29*, 65.
(173) Fox, T.; Krieg, J. M. *Curr. Top. Med. Chem.* **2006**, *6*, 1579.
(174) Ekins, S.; De Groot, M. J.; Jones, J. P. *Drug Metab. Dispos.* **2001**, *29*, 936.
(175) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
(176) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615.
(177) Program, U. S. EPA TRI. http://www.epa.gov/TRI/, accessed November, 2009.
(178) Program, U. S. EPA TRI. Toxics Release Inventory (TRI) Program Fact Sheet. http://www.epa.gov/TRI/triprogram/tri_program_fact_sheet.htm, accessed November, 2009.
(179) Voutchkova, A. M.; Ferris, L. A.; Zimmerman, J. B.; Anastas, P. T. *Tetrahedron* **2010**, *66*, 1031.
(180) Schneider, T.; Vermeulen, R.; Brouwer, D. H.; Cherrie, J. W.; Kromhout, H.; Fogh, C. L. *Occup. Environ. Med.* **1999**, *56*, 765.
(181) Jackson, M. J. Drug Transport across Gastrointestinal Epithelia. In *Physiology of the Gastrointestinal Tract*, 2nd ed.; Johnson, L. R., Ed.; Raven Press: New York, 1987; pp 1597−1621.
(182) Zmuidinavicius, D.; Didziapetris, R.; Japertas, P.; Avdeef, A.; Petrauskas, A. *J. Pharm. Sci.* **2003**, *92*, 621.
(183) Lu, J. J.; Crimin, K.; Goodwin, J. T.; Crivori, P.; Orrenius, C.; Xing, L.; Tandler, P. J.; Vidmar, T. J.; Amore, B. M.; Wilson, A. G.; Stouten, P. F.; Burton, P. S. *J. Med. Chem.* **2004**, *47*, 6104.
(184) Dressman, J. B.; Thelen, K.; Jantratid, E. *Clin. Pharmacokinet.* **2008**, *47*, 655.
(185) Refsgaard, H. H. F.; Jensen, B. F.; Brockhoff, P. B.; Padkjaer, S. B.; Guldbrandt, M.; Christensen, M. S. *J. Med. Chem.* **2005**, *48*, 805.
(186) Faassen, F.; Kelder, J.; Lenders, J.; Onderwater, R.; Vromans, H. *Pharm. Res.* **2003**, *20*, 177.
(187) Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. *Pharm. Res.* **1997**, *14*, 568.
(188) Jorgensen, W. *QikProp version 2.3*; Schrodinger, LLC: New York, 2003.
(189) Brown, J. H.; Cook, K. M.; Ney, F. G.; Hatch, T. *Am. J. Public Health* **1950**, *40*, 450.
(190) Proost, J. H.; Roggeveld, J.; Wierda, J. M. K. H.; Meijer, D. K. F. *J. Pharmacol. Exp. Ther.* **1997**, *282*, 715.
(191) DeCaprio, A. P. *Neurotoxicology* **1987**, *8*, 199.
(192) Serve, M. P.; Bombick, D. D.; Roberts, J.; McDonald, G. A.; Mattie, D. R.; Yu, K. O. *Chemosphere* **1991**, *22*, 77.
(193) Jones, R. B.; Mackrodt, W. C. *Biochem. Pharmacol.* **1983**, *32*, 2359.
(194) Csanady, G. A.; Laib, R. J.; Filser, J. G. *Toxicol. Lett.* **1995**, *75*, 217.
(195) Eriksson, L.; Verhaar, H. J. M.; Hermens, J. L. M. *Environ. Toxicol. Chem.* **1994**, *13*, 683.
(196) Keeler, R. F.; Tu, A. T. *Toxicology of plant and fungal compounds*; Marcel Decker: New York, 1991; Vol. 6.
(197) Bailey, M. J.; Dickinson, R. G. *Chem.-Biol. Interact.* **2003**, *145*, 117.
(198) Siraki, A. G.; Chevaldina, T.; O'Brien, P. J. *Chem.-Biol. Interact.* **2005**, *152*, 177.
(199) Quistad, G. B.; Zhang, N.; Sparks, S.; Casida, J. *Chem. Res. Toxicol.* **2000**, *13*, 652.
(200) Metcalf, R. L.; Francis, B. M.; Metcalf, R. A.; Farage-Elawar, M.; Hansen, L. G. *Pestic. Biochem. Physiol.* **1988**, *30*, 46.
(201) Berenblum, I.; Ben-Ishai, D.; Haran-Ghera, N.; Lapidot, A.; Simon, E.; Trainin, N. *Biochem. Pharmacol.* **1959**, *2*, 168.
(202) Lai, D. Y.; Woo, Y.; Acros, J.; Argus, M. *Toxicologist* **1994**, *14*, 134.
(203) Lai, D. Y.; Woo, Y. T.; Argus, M. F.; Acros, J. C. In *Designing Safer Chemicals*; DeVito, S., Garrett, R., Eds.; American Chemical Society: 1996; Vol. 640, 62.
(204) Mankowski, D. C.; Ekins, S. *Curr. Drug Metab.* **2003**, *4*, 381.
(205) Gasiewicz, T. A.; Kende, A. S.; Rucci, G.; Whitney, B.; Jeff Willey, J. *Biochem. Pharmacol.* **1996**, *52*, 1787.
(206) Fujii-Kuriyama, Y.; Mimura, J. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 311.
(207) Mekenyan, O. G.; Veith, G. D.; Call, D. J.; Ankley, G. T. *Environ. Health Perspect.* **1996**, *104*, 1302.
(208) Ashida, H.; Fukuda, I.; Yamashita, T.; Kanazawa, K. *FEBS Lett.* **2000**, *476*, 213.
(209) Ekins, S.; Erickson, J. A. *Drug Metab. Dispos.* **2002**, *30*, 96.
(210) Watkins, R. E.; Wisely, G. B.; Moore, L. B.; Collins, J. L.; Lambert, M. H.; Williams, S. P.; Willson, T. M.; Kliewer, S. A.; Redinbo, M. R. *Science* **2001**, *292*, 2329.
(211) Kliewer, S. A.; Goodwin, B.; M., W. T. *Endocr. Rev.* **2002**, *23*, 687.
(212) Ekins, S.; Mirny, L.; Schuetz, E. G. *Pharm. Res.* **2002**, *19*, 1788.
(213) Windshügel, B.; Jyrkkärinne, J.; Poso, A.; Honkakoski, P.; Sippl, W. *J. Mol. Model.* **2005**, *11*, 69.
(214) Lewis, D. F. V.; Ogg, M. S.; Goldfarb, P. S.; Gibson, G. G. *J. Steroid Biochem. Mol. Biol.* **2002**, *82*, 195.
(215) Lewis, D. F. V.; Jacobs, M. N.; Dickins, M.; Lake, B. G. *Toxicology* **2002**, *176*, 51.
(216) Janowski, B. A.; Grogan, M. J.; Jones, S. A.; Wisely, G. B.; Kliewer, S. A.; Corey, E. J.; Mangelsdorf, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 266.
(217) Flockhart, D. A. Drug Interactions: Cytochrome P450 Drug Interaction Table, 2007. http://medicine.iupui.edu/clinpharm/ddis/table.asp (accessed June 5, 2009).
(218) Andres, P. R.; Lloyd, E. J. *Med. Res. Rev.* **1982**, *2*, 355.
(219) Seeman, P. *Pharmacol. Rev.* **1980**, *32*, 229.
(220) Lahti, R. A. *Naturwissenschaften* **1979**, *66*, 403.
(221) Grunewald, G. L.; Reitz, T. J.; Ruth, J. A.; Vollmer, S.; Eiden, L. E.; Rutledge, C. O. *Biochem. Pharmacol.* **1979**, *28*, 417.
(222) Verschoyle, R. D.; Philpot, R. M.; Wolf, C. R.; Dinsdale, D. *Toxicol. Appl. Pharmacol.* **1993**, *123*, 193.
(223) Acros, J.; Argus, M. *Chemical Induction of Cancer*; Academic Press: New York, 1974; Vol. IIA.
(224) Acros, J. *Environ. Pathol. Toxicol.* **1978**, *1*, 433.
(225) Voutchkova, A. M.; Kostal, J.; Long, J. J.; Steinfeld, J. B.; Emerson, J. W.; Zimmerman, J. B. *Environ. Sci. Technol.* 2010, submitted for publication.
(226) Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. *Environ. Toxicol. Chem.* **1997**, *16*, 948.
(227) Boethling, R. S.; Sommer, E.; DiFiore, D. *Chem. Rev.* **2007**, *107*, 2207.
(228) Fewson, C. A. *Trends Biotechnol.* **1988**, *6*, 148.
(229) European Center for Ecotoxicology and Toxicology of Chemicals. The Role of Bioaccumulation in Environmental Risk Assessment: The Aquatic Environment and Related Food Webs. October 1995.
(230) Newsome, L. D.; Nabholz, J. V.; Kim, A. In Designing Safer Chemicals, DeVito, S.; Garret, R, Eds. American Chemical Society: 1996, Vol. 640, 172.

CR9003105