Diversity Measures for Enhancing ADME Admissibility of Combinatorial Libraries

Ferenc Darvas,* György Dormán, and Ákos Papp

ComGenex, Inc., Bem rkp. 33-34, H-1027, Budapest, Hungary

Received August 3, 1999

For general screening libraries, structural diversity descriptors and drug-likeness indicators still do not guarantee the in vivo bioavailability for the candidates, which is considered a major bottleneck in drug development. Early prediction of pharmacokinetics (log *P*, log *D*), metabolism, and toxicity makes it possible to deal with ADME (adsorption, distribution, metabolism, excretion) related diversity as an extension to the classical diversity concepts. It opens several new possibilities for optimization of a discovery library before doing any experimental screening. This new diversity concept is demonstrated on a subset of MeDiverse, which is a diverse collection of pharmacologically relevant compounds selected from our in-house library. From consideration of the ADME interface in living systems, virtual secondary libraries of metabolites and retrometabolites (prodrugs) can be generated. These additional libraries readily enhance both the structural and ADME related diversity. This new opportunity in library design can substantially improve the success rate for in vivo lead generation from in vitro hits.

INTRODUCTION

Combinatorial chemistry produces molecular libraries for high-throughput screening (HTS),[1] which rapidly evolved following the recent discovery of numerous, disease-related biological targets. Genomics, proteomics, automation, miniaturization, and data management contributed primarily to the birth of a new drug discovery paradigm.[2]

The contemporary drug discovery pipeline attempts to gather all the relevant information available into a single process in order to shorten the development span and to reduce the costs.[3] In order to find the expected lead molecules, the primary goal is to deliver libraries which represent maximum coverage of the property space with a minimum amount of compounds. In other words, if the compounds in a library are equally distributed in the chemical space, the probability of finding an "activity island" (hit and its neighborhood[4]) is much higher.

Maximizing the diversity of a virtual library is the central concept in combinatorial chemistry. It has two major elements: the concept of redundancy and the concept of total coverage not leaving gaps in the chemical space between regions, which is characterized as the "Swiss cheese" effect.[5]

One of the major approaches generating highly diverse libraries is to construct a small central core (scaffold or centroid) which is simultanously or consecutively "decorated" around with substituents.[6] The central core is frequently considered as a biologically inert skeleton which links together substituents oriented into different spatial regions. These substituents hold the recognition elements for protein binding.

The primary or regular diversity of a synthetic library is composed by the size, shape, and electronic characters of the scaffold and the preselected dissimilarity of the substituents.[3,7,8] Attempts to include medicinal chemistry knowledge (drug-likeness, biologically relevant structural patterns) and physicochemical parameters into the library design and selection process are currently under way.

Several efforts were made to give definitions to drug-likeness. Lipinski rules are widely recognized[9] as general filters, while most recently these requirements were supported by Ghose's analysis[10] on more than 6000 registered drugs collected in the Comprehensive Medicinal Chemistry (CMC) database.

EARLY ADME/TOX ASSESSMENT FOR IMPROVED ORAL BIOAVAILABILITY

The in vitro nature of HTS techniques generates "hit" compounds with often unfavorably low absorption and metabolic stability and short duration of action (fast clearence). Those factors are collectively responsible for the poor bioavailability of the in vitro hits in general.[11,12] Although there has been much effort in implementing new knowledge into the combinatorial drug discovery pipeline, several factors are neglected. Fecik recently characterized the present status: "the ADME considerations are rarely deliberately built into the library design".[13] A potential drug candidate has to cross several barriers, until it binds to the target and induces the desired response. These barriers are characterized as the "ADME interface" in general. Its major elements are as follows: absorption, distribution, metabolism, and excretion.

The present HT screening methods focus solely on pharmacological activity and produce many thousands of different molecules, which exhibit the required effects at relatively high concentration. On the other hand in drug development the major goal is oral bioavailability. Statistically about 80% of the initial in vitro hits fall out because of their poor pharmacokinetics.

These facts make the drug discovery process particularly difficult and lengthy since the original hits have to be structurally modified or filtered in order to obtain leads with optimal ADME properties and a low level of toxicity. This

* To whom correspondence should be addressed. Telephone: +36 1 214 2306. Fax: +36 1 214 2310. E-mail: df@comgenex.hu.

10.1021/ci990268d CCC: $19.00
Published on Web 02/08/2000

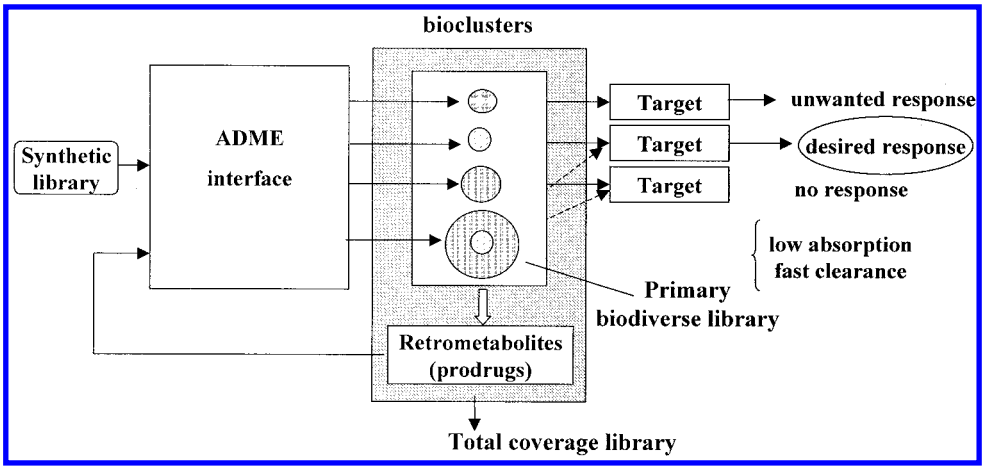

Figure 1. Major bioclusters deriving from the synthetic library passing through the ADME interface.

lead optimization as a separate action would be unnecessary if those parameters would be integrated either to the library design process with computer based methods ("in silico") or HT ADME/Tox assays[12,14] conducted parallel to the HT activity screening.

(1) Computer Based Methods ("in Silico"). Methods, which allow physicochemical predictions from molecular structure, are "badly needed in both early discovery and pharmaceutical development settings" (Lipinski[9]). This statement can be supplemented with the early prediction of the first pass effect (metabolic clearance) as well as toxicity assessment.[15] There are multiple advantages of computer-assissted prediction, models, and expert systems.[12] They are capable of screening vast numbers of molecules with relatively small investment in space and equipment and can also be used in a network environment or via the Internet. On the other hand, the users should be aware that computer models have accuracy with certain limitations and cannot reflect the true complexity of the biological systems. These systems are mainly rule and stucture based approaches, representaed by quantitative structure−pharmacokinetic (QPKR), structure−metabolism (QMR), and stucture−toxicity relationships (QTR).

Drug metabolism data[16] are accumulated over the years, and that knowledge is implemented into rule based expert systems, like MetabolExpert[17] and Meta.[17] As a result of the rapid development of molecular modeling, refined 3D enzyme structures can also be used in virtual metabolism screening of combinatorial libraries. The "metabolism sensitive" structures can be further evaluated. Those compounds are either quickly eliminated or metabolism simply alters their biological activity by in vivo production of various structural analogues. In summary, metabolism prediction can open new opportunities of evaluating synthetic combinatorial libraries, as discussed later.

Similarly, toxicity prediction is based on rules comprising the identified chemical substructure of the molecule responsible for the toxicity ("toxophores") (HazardExpert;[18] Derek[18]). The estimation is refined by calculating the effect of the neighboring structural environment.

(2) Biological Assays. The majority of the ADME/Tox HT methods utilize similar in vitro techniques, as developed for the functional assays.

The most commonly used in vitro model for determining the transport across the gastrointestinal membrane is the use of Caco-2 monolayers,[12] while brain microvessel endothelial cells are employed to predict the penetration through the blood−brain barrier (BBB). Three types of in vitro enzyme systems are utilized for studying metabolism: cellular (human liver slices), liver microsomal, and recombinant (Cytochrome P-450 isozymes) systems. To model the major biochemical functions, metalloporphyrins have been recently reported as synthetic livers or mimics of the in vivo metabolic processes.[19]

Finally in vitro toxicity screening assays are generally cellular based models, allowing estimation of cell toxicity, mutagenicity, and carcinogenicity. Many companies are turning to in vivo multiple cassettes or "*n*-in-one" dosing strategies, when multiple compounds are administered to a single animal.

For in vivo metabolism single dose administration of compounds results in metabolites, which can be examined by sensitive analytical methods (GC-MS or LC-MS). In vivo toxicology (whole body exposure) experiments can indicate which organ is the major target for unwanted side effects or toxicity, which also helps in providing the appropriate drug formulation and delivery system.

ADME/TOX IMPACT ON DIVERSITY: THE CONCEPT OF PRIMARY LIBRARY, SECONDARY LIBRARY, AND THE "TOTAL COVERAGE" LIBRARY

Oral drugs not absorbed through the gastrointestinal tract are eliminated unchanged, while drugs absorbed into the circulatory system may undergo metabolic transformations in the liver before reaching the target organ. A typical, unfiltered combinatorial library carries structures with completely different in vivo destiny. Additionally, some compounds or metabolites induce unwanted toxic responses at the target or at an entirely different site. Toxicity either derives directly from the original library or evolves during the interaction with the "ADME interface" as toxic metabolites.

The active, inactive, and toxic compounds among those that reach the target represent three different subsets (bioclusters), and an additional cluster collects all the other molecules, that are eliminated without interacting with the target (Figure 1).

Metabolic biotransformations can strongly affect the biovailability, potency, toxicity, distribution, and clearance of foreign compounds (xenobiotics), and, in fact, the

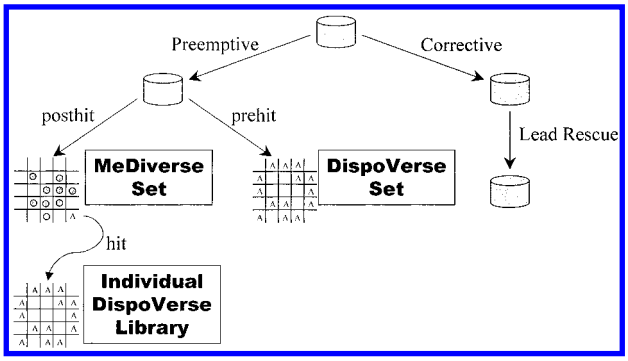

Figure 2. Preemptive and corrective measures of in vitro screening libraries.

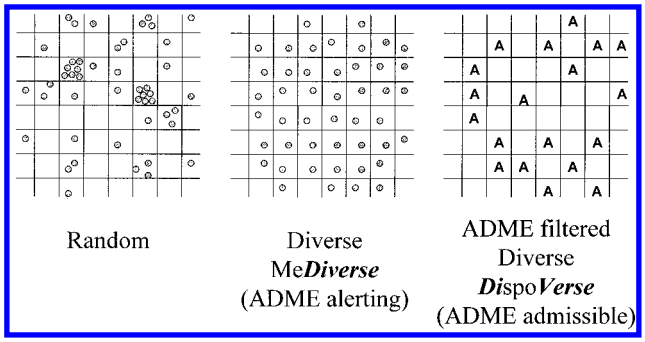

Figure 3. ADME alerting and ADME filtered structurally diverse libraries.

compounds we administer to the living systems exert their activity frequently through their metabolites. Thus, in a hypothetical case when we have reliable knowledge on the expected metabolic fate of our compounds, it is important to consider the structure of the metabolites together with the parent compounds listed in the library. In this way a more complete picture can be gained of the expected effects. While the library of synthesized compounds (termed "primary library") possesses a diversity, this diversity is likely modulated by the much wider library of the metabolites that are generated by the living system ("secondary library"). Diversity associated with those distinguished libraries will be referred to as "primary diversity" and "secondary diversity".

The knowledge of the expected metabolic biotransformation pathways for a given combinatorial library allows generation of a virtual library of metabolites and retrometabolites as well.[20] The term retrometabolite comes from the analogy of synthetic chemistry when a collection of chemical transformations is used in a retrospective manner to generate all the possible precursors of a synthetic target. Another definition for retrometabolites is prodrugs. The difference between the two terminologies is by the fact that retrometabolites are also formed endogenously.[20]

If the primary rules for metabolism are used retrospectively, all precursors (prodrugs) constitute a new analogue library, which can be readily converted to the desired compounds after passing through the ADME interface. Retrometabolites of a synthesized primary library form another, but different, secondary library.

The secondary library differs not only structurally, but greatly extends the diversity of the primary library with respect to the pharmacokinetical (primarily absorption and distribution) properties. The secondary library has an important corrective or "rescue" function, in that it can replace in vitro hits from within a primary library which otherwise would fall out due to their poor absorption kinetics, and these gaps can be readily filled with retrometabolites. (Figures 2 and 3).

The primary library and the two types of secondary libraries compose a special collection of compounds which represents an authentic view of a combinatorial library in the pharma-related area. We call it a "total coverage library" possessing a "total coverage diversity".

METHODS

(1) Early ADME Prediction. Combinatorial libraries, which are designed to possess "regular diversity",[5] can be further characterized by additional information, which represents significant added value in the drug discovery process. The added ADME/Tox parameters allow several new ways of studying, filtering, and clustering a library, and new insights can also be obtained about the possible behavior of a library in the biological ADME interface. In our approach these opportunities are utilized in order to generate a library with a higher tendency to pass the ADME testing phase; thus, the library exhibits enhanced ADME admissibility. This strategy is composed of two major elements: First, a combinatorial library can be optimized (filtered, corrected) by using the predicted ADME parameters. Second, the predicted metabolites (retrometabolites) in the ADME interface form secondary libraries—as discussed earlier—which provide an extension of the classical diversity concepts to ADME diversity. This biologically generated, in vivo diversity has the role of filling out gaps in the chemical property space, providing better coverage in vivo.

In the present paper we provide evidence about that extension of regular diversity by these added new metabolic members. The tools, which allow early considering of ADME parameters for each member of a combinatorial library, are the integrated Pallas [Pallas is the software that contains ADME/Tox related prediction programs as modules; it got its name after the goddess of wisdom, Pallas Athene] modules and have been reported separately.[17,18,20,22−24]

We list here the parameters we predict in our approach, together with the software module that is responsible for the calculation.

(a) pK_a. Acidic and basic pK_a values (pK_a_ac1 and bas1 are the strongest) are calculated by pKalc.[22] [pKalc uses approximately 600 Hammett and Taft equations for the following prediction: $pK_a = pK_a^0 - \rho\sum\sigma$. Its accuracy is usually within an error of 0.25 pK_a units. It works for any organic compounds, includiing aromatic, mono- and polyheteroaromatics, and small peptides.]

(b) log *P* and log *D* at pH = 2...7.4...10. Octanol/water distribution coefficients are predicted by ProlgD[23,24] at various pH values (pH = 2, pH = 7.4, and pH = 10, representing the pH of the stomach, the blood, and the mouth, respectively). [ProlgD calculates accurate log *P* values (negative logarithm of the *n*-octanol/water coefficient) for organic molecules. The log *D* prediction is based on pK_a and log *P* prediction of the micro- and macrospecies of the compound. For the prediction of the log *P* values it uses three different methods. The CDR method is based on Rekker's fragmental method.[29] This method sums the fragmental contributions and the correction terms: log *P* =

$\sum a_i f_i - \sum b_j F_j$. The ATOMIC and ATOMIC5 methods are based on Broto[30] and Ghose[31] atomic methods, respectively. These methods use atomic contributions: $\log P = \sum n_i a_i$. log *P* and log *D* prediction accuarcy is, in most cases, within an error which is not significantly higher than the measurement error of the log *P*, log *D* value.]

The log *D* calculation is an essential tool where hydrophobicity should be estimated more correctly than just neglecting the ionization effect.

(c) Toxicity Assessment. Toxicity class is predicted by HazardExpert[18] for the structures of the parent compound and the first-level metabolites. We use this estimation only for alarming purposes, indicating the structures, which include highly-probable or probable toxic fragments. The prediction method is primarily based on searching for defined toxic fragments. [HazardExpert modules identify toxic fragments in the analyzed compounds. The toxicity estimation is carried for humans or animals in seven different classes: oncogenicity, mutagenicity, teratogenicity, irritation, sensitivity, immunitoxicity, and neurotoxicity. Bioavailability estimation for toxicants is based on pK_a and log *P* prediction. Toxicity prediction of the metabolites is also available. Accuracy results for the basic toxophore groups are comparable to literature data.]

(d) Metabolism Prediction (First-Pass Effect). The metabolism of the compound is predicted in two levels, and then the possibility of fast elimination is investigated by taking into account the most probable phase II reactions. [MetabolExpert is a rule-based system with an open architecture. The rules are based on theoritical metabolic pathways and completed with other general transformation rules extracted from experimental metabolic tree in different species. The core of the model is a biotransformation graph, modeling the composition of the living system together with transport processes and metabolic pathways. Transformations were formulated in the model as if...then rules. Every rule is composed of four elements: opening bonds during transformations (active substructure); bonds and atoms introduced by the transformation (replacing substructure); a list of substructures at least one of which must be present in the molecule for the fragment to match (positive condition); a list of substructures whose presence prevents the match of the fragment (negative condition). Available knowledge bases for humans and animals: the rules have been collected from transformations on different species of mammals, mainly human, rat, mouse, guinea pig, and dog. The total number of rules within the database is 179. For plant metabolism (for pesticides) and for phorodegradation additional rules are collected. There are two types of generation of metabolites: frameless generation and preliminary metabolic simulation, which uses the common biotransformation set automatically without user control. The output is an estimation of the structural formula of metabolites, which might be formed from a substance in human, animal, or plant. The module is also suitable for estimation of structures formed by photodegradation.

(e) Retrometabolite Prediction (Prodrugs). The retrometabolites are predicted by Retro-Metabolexpert[20] (RetroMex) [Retro-Metabolexpert is a rule based expert system for retrometabolism design. It contains 64 rules; the retrometabolites are generated by a simple structural input. The module also provides a retrometabolic tree]. Retrometabolism prediction was used to create a retrometabolic virtual library using the same rules retrospectively as in MetabolExpert.

All calculations are semiautomated, because Pallas can communicate with any structural database handling software (like ISIS) or other calculation software (like Cerius2−MSI) throughout its chemical structure import/export capability (with MOL-files, SD-files or RD-files). The export of the results *delimited* text files is also possible.

ADME/Tox Characterization in Practice.[21] In combinatorial drug discovery each member of the library can be classified after prediction. The main classes are as follows: (1) *ADME dangerous* library members (extreme hydro/lipophilicity, extreme pK_a values, significant risk for first pass effect, and increased probability for toxic symptoms (including metabolites); (2) *ADME favorable* library members (established similarity to known agents with favorable ADME profile, benign metabolism profile, and benign toxicity profile). Naturally, there is an intermediate class, but the definition of each group is highly dependent on the requirements for the targeted drug class and the organ.

Any library can be easily classified, and the prediction gives an ADME alerting feature to the library.

Early consideration of the ADME parameters at screening can be at two levels:

(1) The in vitro level represents preemptive measures. In the *posthit* strategy first the synthetic library is screened with no ADME parameters considered, followed by a second focused library where the initial hits are filtered for ADME admissibility. In another approach, (*prehit* strategy) ADME prediction precedees the screening, thus only ADME optimized library members are screened. In Figure 3, a structurally diverse ADME optimized library is shown. The holes correspond to the ADME dangerous members that have been removed. The structurally diverse selection of compounds ("regular library"), which is characterized by the ADME/Tox parameters for each member, is an "ADME alerting" library (called *MeDiverse* [MeDiverse is an ADME alerting combinatorial library (10 000 compounds synthesized in-house at ComGenex and selected from about 100 000 compounds); the name of MeDiverse refers *medicinal chemistry* relevance]), which allows prescreening or postscreening filtration. On the other hand the ADME filtered library, which contains only ADME admissible elements with optimized disposition behavior, is labeled as a DIspoVERSE collection.

(2) The in vivo level ADME corrective measures can be taken and the retrometabolites can be rescreened. Using that process, the in vitro hits with poor bioavailibility can be "rescued" by generating retrometabolites with possibly better pharmacokinetics. In summary one might be either preemptive or corrective if wanting to consider ADME admissibility of a library at the beginning of a discovery project.

(2) Establishing Second Generation Libraries. The early characterization of the metabolic stability can be profitable in two different senses: (1) helping in identifying better candidates through metabolism information or (2) contributing to removal of unwanted candidates that are unsuitable for in vivo testings.

In general, a good knowledge of the expected metabolism profile can contribute positively to further screening efforts by providing an abundance of interesting metabolites which might possess different or enhanced biological activity. As

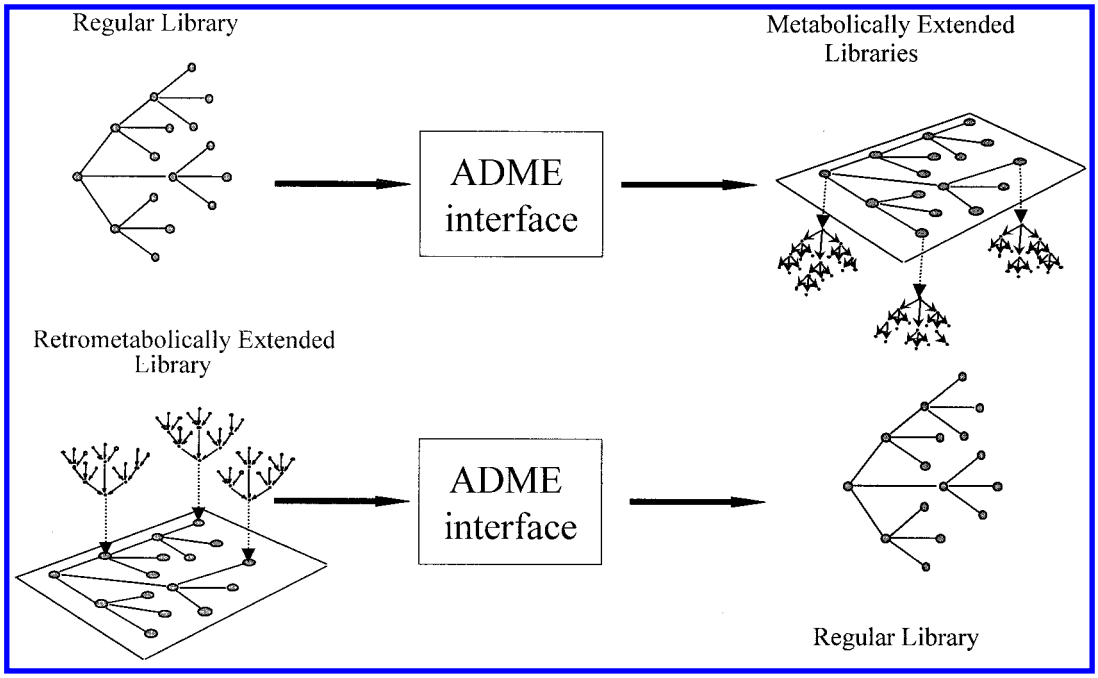

Figure 4. Extension of synthetic ("primary") libraries by secondary sets that tend to be generated by in vivo biotransformations in the ADME interface.

shown in Figure 2, the ADME filtered library, which had lost elements in covering the chemical space leaving gaps behind, would gain new members which might fill out the discontinuity. The library members of MeDiverse are connected to *metabolic* and *retrometabolic* diversity in addition to that of the compound collection. In other words any members that fall into the ADME dangerous category can be substituted either by structurally isosteric metabolites derived from other members or by generation of retrometabolites (prodrugs) with favorable pharmacokinetic properties. The whole concept is presented in Figure 4, where the upper portion indicates the metabolic tree, which displays several new members added to the library generated, while passing through the ADME interface. The lower portion represents the inverse process. Thus, all the retrometabolites generated from the original library reproduce the starting collection in the ADME interface being subjected to the same metabolic transformations as above. Retrometabolites represent a significant ADME diversity in a sense of various absorption rates and transporting properties, while resulting practically in the same product from different precursors (Figure 4).

Intrinsic Retrometabolites. In a regular library, a retrometabolic relationship can be identified between a certain portion of the library members. From these retrometabolic relatives, those which show the most favorable disposition parameters for the desired target can be selected. To demonstrate this idea, we targeted the identification of so-called *intrinsic retrometabolites* (cryptolites) of a synthetic library, which are also contained within the original library. Thus, combinatorial libraries can be characterized according to how many retrometabolite and regular library member pairs are found, which represents primarily retrometabolic redundancy. Carefully filtered diverse libraries rarely contain retrometabolites; on the other hand, from the ADME point of view retrometabolites represent a pharmacokinetic diversity and regular libraries should be supplemented in order to increase their ADME admissibility. Finally there are a few reversible metabolic transformations which can be converted into each other, depending on the enzymatic site and environment. Those molecules can be found in the metabolic and retrometabolic clusters as well. For example a novel core structure (spirooxindoles) with diverse substituents around showed about 10% overlap between the metabolic and retrometabolic clusters.

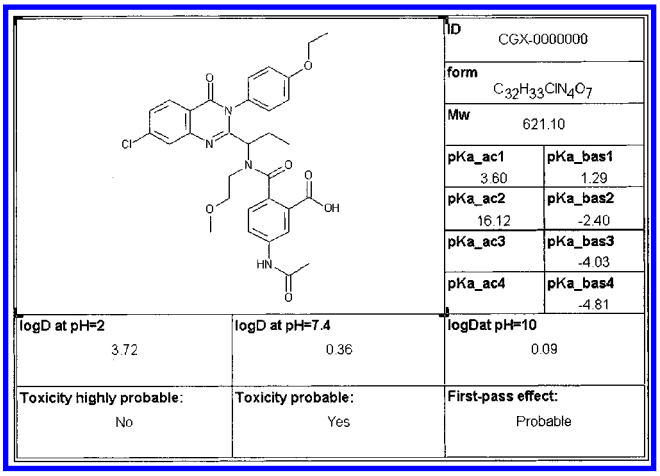

Figure 5. Computer snapshot of a typical member of the MeDiverse library with the ADME parameters predicted.

RESULTS AND DISCUSSION

(1) MeDiverse: A Preselected Library with Extended Diversity and Total Coverage. In the previous section we introduced the MeDiverse library as an ADME alerting collection, where each library member is characterized with disposition data. In Figure 5, a general computer screen snapshot illustrates how all the data are presented. Now we extend the Mediverse ADME focused concept with diversity measures and provide evidence that metabolic transformation, other than those leading to first pass effects, could contribute

Table 1. Average Values of the Eight Major "Drug-likeness" Factors in the Reg, Met, and Retro Subsets

			log *D*						
		log *P*	pH = 2	pH = 7.4	pH = 10	mol wt	rotatable bonds	H-bond acceptor	H-bond donor
Reg	av	5.07	3.67	4.98	5.06	438.17	8.85	4.21	0.73
Met	av	4.44	3.09	4.27	4.04	447.15	9.58	4.93	1.36
Retro	av	4.89	2.34	4.43	4.83	446.36	8.96	4.49	0.95

to the lead hunting efforts by broadening in vivo the choice for the biological target. Metabolic transformations are primarily induced by oxidative enzyme families in the liver, which is the major "natural chemical reactor" of the body, and provide several bioclusters: toxic entities, fast eliminating products, and finally bioactive derivatives, which are difficult to selectively synthesize. Some of these are much more potent in certain tests than the parent compound, like 19-hydroxyprostaglandin E_1 (19-OH−PGE_1) compared with PGE_1 (Dormán, G. Unpublished results). Directed metabolism as a diversity generating tool is rarely studied. Biomimetic systems, e.g. metalloporphyrines as synthetic livers, can provide a practical tool for generating such libraries.[19]

The MeDiverse set of compounds is a selected, diverse library of 10 000 heterocyclic organic compounds. In that collection all possible metabolites are predicted for each member; therefore, it should be considered as an in vivo metabolically extended library. In our initial diversity comparison study we did not exclude the members that are predicted to be eliminated quickly; thus, starting from a 1000 structure subset of MeDiverse, about 6500 metabolites were predicted using MetabolExpert. The simultanously performed retrometabolic transformations resulted in more than 8000 precursors from the same cluster. The most interesting questions are as follows:

(1) How are the major physicochemical parameters and structural properties altered?

(2) What diversity shift can be detected in the new subsets?

The libraries of the 1000 regular members of MeDiverse, its metabolites, and retrometabolites are labeled Reg, Met, and Retro, respectively. The most relevant properties were calculated by a customized version of MSI Cerius^2 software. AlogP [Alog P is a calculated log P value] and the 3D parameters have been omitted from the default set of descriptors, which consist of electronic, spatial, structural, thermodynamic, and topological properties,[25−27] while values calculated by PrologD at three different pH values and by PrologP have been inserted.

Table 1 shows eight major factors and the mean values among the three subsets.

It is easy to observe that metabolism resulted in about a 13−15% decrease in lipophilicity at all pH values. This phenomenon contributes to the fast clearance and increased water solubility. On the other hand, it is rather unexpected that the retrometabolic precursors possess similarly reduced lipophilicity, particularly with respect to the major decrease at pH = 2. Interestingly, the regular library shows also low lipophilicity at that pH value, but it is designed to be suitable for efficient absorption through oral administration. Metabolism resulted in a significant increase in rotatable bonds and the number of H-bond donors, while both clusters showed an increase in molecular weights. It can also be concluded that the libraries basically meet the drug-likeness requirements summarized in the Lipinski's rules.[9] A previous study on our libraries in comparison with Comprehensive Medicinal Chemistry Databases showed similar results.[28]

The individual distribution curves for the different properties give a dynamic picture about the clusters (Figure 6). In conclusion the decrease in lipophilicity is a general trend among the MeDiverse related library members. That is largely dependent on the original chemical structure, which determines the mean absorption behavior, and metabolites or retrometabolites can complement the original library character in either direction.

(2) Diversity Comparison of Regular, Metabolic, and Retrometabolic Libraries. The diversity characterization of the three libraries could provide a special insight into the diversity generation of the living systems.

As previously mentioned, the modified default set of individual descriptors (46 pieces) has been applied to characterize all three libraries in the property space. This operation was critical since identical treatment for each of the library members allowed occupation of a common "virtual space" in a comparable way.

For reduction of the number of the descriptors and removal of the effect of correlation, we used principal component analysis (PCA).

In our study we selected a five-dimensional property space based on the attached results in Table 2. Five principal components (PC1-5, including the largest eigenvalues) typically cover 80−90% or more of the variance in most data sets.

This representation also enables simple diversity calculation by using the approximate distances between the molecules in the property space.

To speed up the calculations, we selected 1000 random members of the MeDiverse library and named it "regular library" (LibReg). The metabolites (metabolic library, LibMet) and retrometabolites (retrometabolic library, LibRetro) of LibReg have also been predicted and used for adding additional members to LibReg, which fills the diversity holes. For this purpose the PCA was done only for LibReg, and the resulting five principal components have been used in unchanged form for Met and Retro. The hole identification method located the largest discontinuous areas (holes) in the principal property space not covered by the regular (LibReg) library. For hole filling we used the LibSelect module of the MSI Cerius^2 software. For the selection of the additional 1000 members either from LibMet or LibRetro we used distance-based method including 3000 Monte Carlo steps. As a result we prepared two new libraries, LibReg+LibMet and LibReg+LibRetro, containing 2000 members of each. The diversity of the libraries has been compared to each other and visualized in distance histograms, as shown in Figure 7. This way two libraries can be compared by qualitative visual inspection of the compounds in the principal component space or the library comparison functionalities (LibCompare module) to provide a quantitative measurement

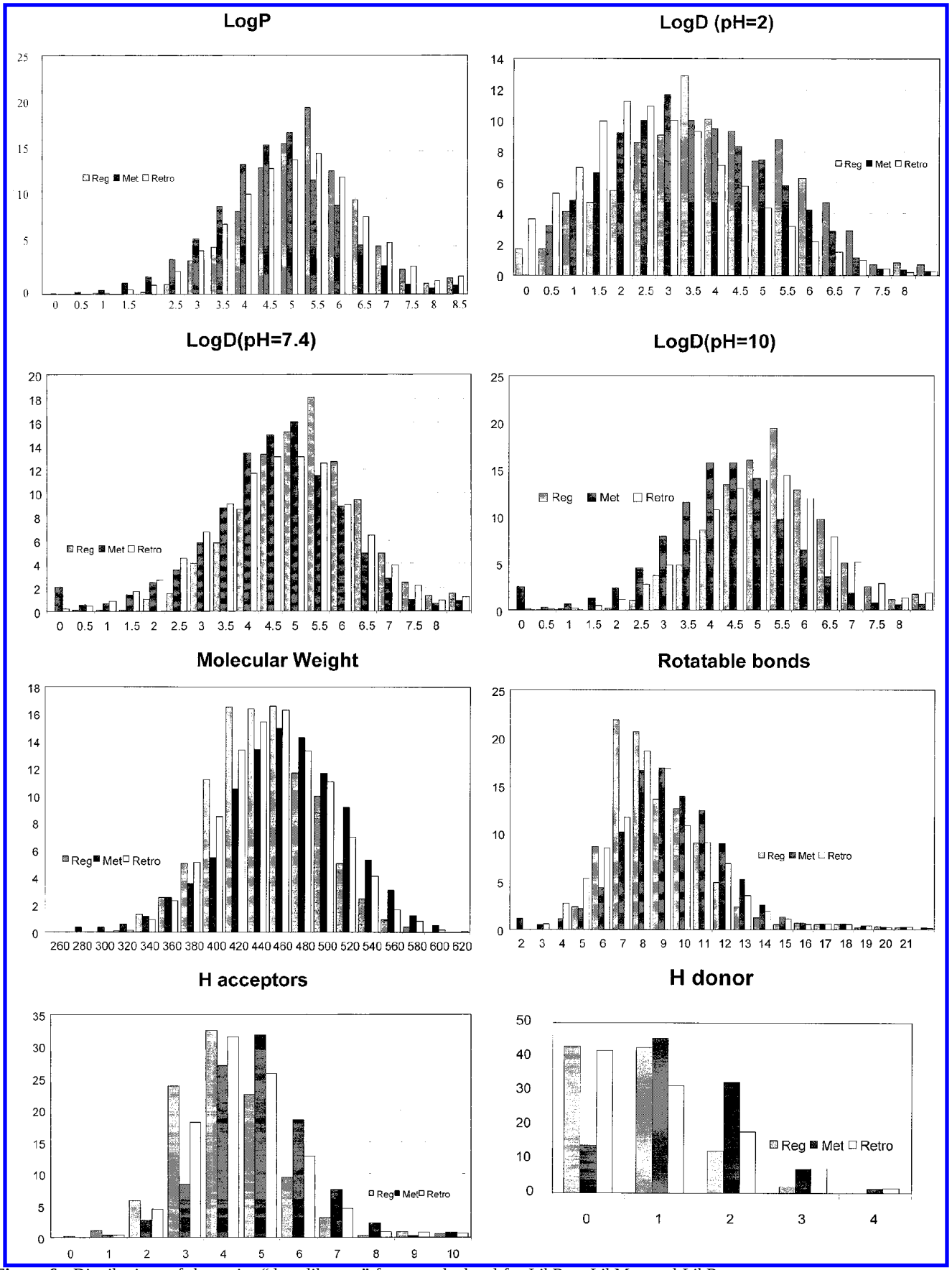

Figure 6. Distributions of the major "drug-likeness" factors calculated for LibReg, LibMet, and LibRetro.

which should confirm the visual assessments. Thus, the libraries can be characterized by the accumulated minimum distance value. These values are as follows: for LibReg, 4.204; for LibReg+LibMet, 3.752; for LibReg+LibRetro, 4.085. These values indicate that adding randomly selected 1000 metabolites to the regular library caused an 11% of increase in covering the virtual property space and 3% of diversity enlargement by adding 1000 retrometabolites.

Table 2. Five Principal Components, Which Cover 80% of the Property Space with the Eigen and SSVAR Component Values

	eigenvalue	SSVAR component, %	SSVAR accumulated
PC1	25.27	54.93	54.93
PC2	5.92	12.86	67.78
PC3	4.29	9.32	77.10
PC4	2.89	6.28	83.39
PC5	2.56	5.57	88.96

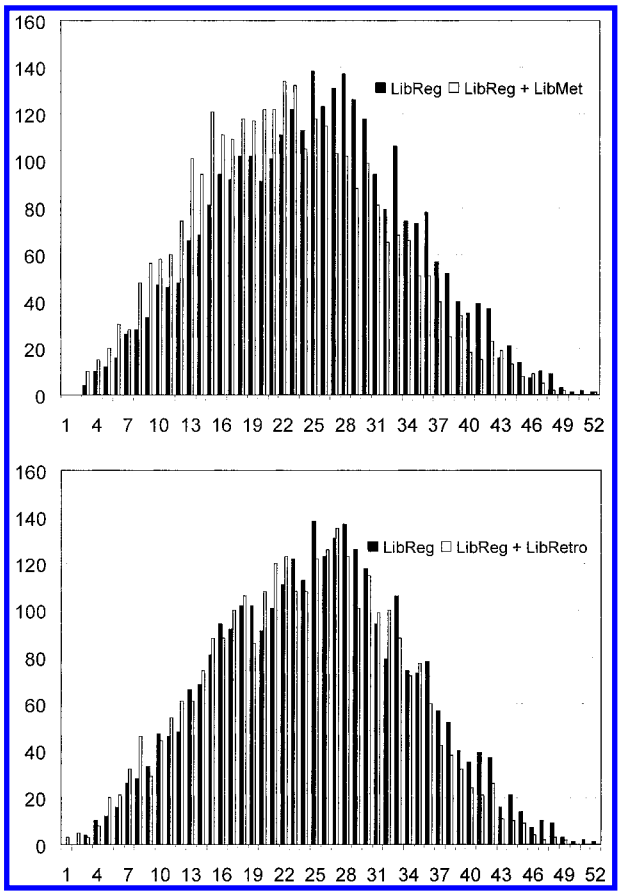

Figure 7. Diversity enhancement of LibReg+LibMet and LibReg+LibRetro virtual libraries demonstrated with minimum distances.

The same results can be read from the comparative distance histograms. Comparing candidate libraries against a reference library, the frequency distribution at shorter distances of a candidate library should shift to larger frequency values to represent higher coverage. This increase can be clearly observed when comparing LibReg with LibReg+LibMet. Simlarly those compounds that correspond to the left-hand side of the histogram are those for which there exists at least one close ("nearest neighbor"[4]) compound in the reference library. The compounds that correspond to the right-hand side of the histogram are those for which there are no close (structurally similar) compounds in the reference library.

In summary, as originally hypothesized, the addition of both metabolites and retrometabolites cause an increase in diversity, although in a slightly different extent.

CONCLUSION

ADME Integrated Combinatorial Drug Discovery. Earlier the ADME related bioavailibility topics were restricted to the preclinical development. The early assessment of ADME parameters has an alerting, correcting function and allows clustering the library into sublibraries based on the data.

For instance, metabolism and toxicological prediction or early evaluation gives an alerting signal about which site of the molecules represents a metabolically labile or toxophoric group that should be replaced during library design or optimization.

There is a strong preference to incorporate more and more knowledge into the design phase of a library. Advanced library design should implement as much knowledge about the ADME/Tox interface as possible; therefore any virtual or in vitro hits would be considered as in vivo hits. On the other hand HTS, which is originally devoted to potency testing of a certain target, will integrate ADME related aspects, leading the way in the future for whole body prediction. In the initial phase iterative links are created between HTS and ADME/Tox features in three aspects:

(1) If ADME prediction is preceding HTS, only *ADME filtered library* members will be tested.

(2) If HTS is preceding ADME evaluation, the in vitro hits can be *optimized* with metabolically stable groups or rescued by transforming them into bioavailable analogues or prodrugs by (retrometabolites).

(3) Secondary libraries enhance structural and ADME diversity, which represent new opportunities in library design.

In conclusion the ADME/Tox assessment would move from the preclinical development to the drug discovery research, and lead discovery and optimization would also become a single integrated effort. This integration would also result in a new concept of diversity, taking into account the contribution of the living systems. The incorporation of these aspects would remarkably reduce the time and improve the cost-effectiveness of the whole drug development pipeline.

ACKNOWLEDGMENT

We are grateful to Mr. Sándor Marokházi for the definition of the retrometabolic rules. We would like to thank Tamás Nagy for his participation in the selection of the MeDiverse library. We are also grateful to Paul W. Erhardt (University of Toledo) and Nicholas Bodor (University of Florida) for helpful discussions.

REFERENCES AND NOTES

(1) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions. *J. Med. Chem.* **1994**, *37*, 1385−1401.

(2) Warmus, J. S. Integrating Combinatorial Chemistry into the Drug Discovery Pipeline. *Drug Discovery Today* **1999**, *4*, 7−9.

(3) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening−An overview. *Drug Discovery Today* **1998**, *3*, 160−178.

(4) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity" Descriptor. *J. Med. Chem.* **1996**, *39*, 3049−3059.

(5) Kauvar, L. M.; Laborde, E. The Diversity Challenge in Combinatorial Chemistry. *Curr. Opin. Drug Discovery Dev.* **1998**, *1*, 66−70.

(6) Hogan, J. C. Directed Combinatorial Chemistry. *Nature* **1996**, *384*, 17−21.

(7) Martin, E. J.; Critchlow, R. E. Beyond Mere Diversity: Tailoring Combinatorial Libraries for Drug Discovery. *J. Comb. Chem.* **1999**, *1*, 32−45.

(8) Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. S. Virtual Compound Libraries: A New Approach to Decision

Making in Molecular Discovery Research. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1010−1023.
(9) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(10) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1*, 55−68.
(11) Berman, J.; Howard, R. J. Combinatorial Drug Discovery: Concepts. In *Combinatorial Chemistry and Molecular Diversity in Drug Discovery;* Gordon, E. M., Kerwin, J. F., Eds.; Wiley-Liss: New York, 1998; pp 447−452.
(12) Leahy, D. E.; Duncan, R.; Ahr, H. J.; Bayliss, M. K.; Boer, A. G.; Darvas, F.; Fentem, J. H.; Fry, J. R.; Hopkins, R.; Houston, J. B.; Houston, J. B.; Karlsson, J.; Kedderis, G. L.; Pratten, M. K.; Prieto, P.; Smith, D. A.; Straughan, D. W. Pharmacokinetics in Early Drug Research. In *The Report and Recommendations of ECVAM Workshop* 22, ATLA 25; Fentem, J. H., Ed.; Russell & Burch House: Nottingham, U.K., 1997; pp 17−30.
(13) Fecik, R. A.; Frank, K. E.; Gentry, E. J.; Menon, S. R.; Mitscher, L. A.; Telikepalli, H. The Search for Orally Active Medications through Combinatorial Chemistry. *Med. Res. Rev.* **1998**, *18*, 149−185.
(14) Gibbons, J. A.; Taylor E. W.; Braeckman, R. A. ADME/PK Assays in Screening for Orally Active Drug Candidates. In *Combinatorial Chemistry and Molecular Diversity in Drug Discovery;* Gordon, E. M., Kerwin, J. F., Eds.; Wiley-Liss: New York, 1998; pp 453−469.
(15) Darvas, F.; Dormán, G. Early Integration of ADME/Tox Parameters into the Design Process of Combinatorial Libraries. *Chim. Oggi* **1999**, *17* (Jul/Aug), 10−13.
(16) Erhardt, P. W. Drug Metabolism Data: Past, Present and Future Considerations. In *Drug Metabolism, Databases and High-Throughput Testing During Drug Design and Development*; Erhardt, P. W., Ed.; Blackwell Science: Cambridge, MA, 1999; pp 2−15.
(17) Darvas, F.; Marokházi, S.; Kormos, P.; Kulkarni, G.; Kalász, H.; Papp, Á. MetabolExpert: Its Use in Metabolism Research and in Combinatorial Chemistry. In *Drug Metabolism, Databases and High-Throughput Testing During Drug Design and Development*; Erhardt, P. W., Ed.; Blackwell Science: Cambridge, MA, 1999; pp 237−270. Klopman, G.; Tu, M. META: A Program for the Prediction of the Products of Mammal Metabolism of Xenobiotics. In *Drug Metabolism, Databases and High-Throughput Testing During Drug Design and Development*; Erhardt, P. W., Ed.; Blackwell Science: Cambridge, MA, 1999; pp 271−276.
(18) Darvas, F.; Papp, Á.; Allardyce, A.; Benfenati, E.; Gini, G.; Tichy, M.; Sobb, N.; Citti, A.; Overview of Different Artificial Intelligence Approaches Combined with a Deductive Logic-Based Expert System for Predicting Chemical Toxicity. In *Conference Book: Predictive Toxicology of Chemicals, American Association for Artificial Intelligence, Spring Symposium Series;* Stanford University: Stanford, CA, 1999; pp 94−99. Greene, N. Knowledge Based Expert Systems for Toxicity and Metabolism Prediction. In *Drug Metabolism, Databases and High-Throughput Testing During Drug Design and Development*; Erhardt, P. W., Ed.; Blackwell Science: Cambridge, MA, 1999; pp 289−296.
(19) Keserü, G.; Balogh, G.; Karancsi, T.; Bertók, B. Chemical Models of Cytochrome P450 Catalyzed Insecticide Metabolism. Application to the Oxidative Metabolism of Carbamate. *J. Agric. Food Chem.* **1999**, *47*, 762−769.
(20) Darvas, F. *Retrometabolism Design Concepts and Realization for Combinatorial Libraries, 2nd Conference on Retro-Metabolism Based Drug Design and Targeting*, May 11−14, 1999, Amelia Island, FL.
(21) Darvas, F. In Vivo Leads from Diverse Libraries: New Diversity Approaches. *CHI Conference: Strategies and Techniques for Identification of Novel Bioactive Compounds*, Zurich, Switzerland, Oct 7−9, 1998.
(22) Csizmadia, F.; Szegezdi, J.; Darvas, F. Expert System Approach for Predicting pK_a, Trends. In *QSAR Molecular Modelling 92*; Wermuth, C. G., Ed., ESCOM: Leiden, The Netherlands, 1993; pp 507−510.
(23) Csizmadia, F.; Panderi, I.; Tsantili-Kakoulidou, A.; Darvas, F. Prediction of Distribution Coefficient from Structure. 1. Estimation Method. *J. Pharm. Sci*. **1997**, *86*, 865−871.
(24) Csizmadia, F.; Panderi, I.; Tsantili-Kakoulidou, A.; Darvas, F. Prediction of Distribution Coefficient from Structure. 2. Validation of Prolog D, an Expert System. *J. Pharm. Sci*. **1997**, *86*, 1173−1179.
(25) Warr, W. A. Commercial Software Systems for Diversity Analysis. In *Perspectives in Drug Discovery and Design;* Willett, P., Ed.; Kluwer: Dordrecht, The Netherlands, 1997; pp 115−130.
(26) Brown, R. D. Descriptors for Diversity Analysis. In *Perspectives in Drug Discovery and Design;* Willett, P., Ed.; Kluwer: Dordrecht, The Netherlands, 1997; pp 31−49.
(27) Martin, Y. C.; Brown, R. D.; Bures, M. G. Quantifying Diversity In *Combinatorial Chemistry and Molecular Diversity in Drug Discovery;* Gordon, E. M., Kerwin, J. F., Eds.; Wiley-Liss: New York, 1998; pp 369−383.
(28) McGregor, M. J.; Pallai, P. V. Clustering of Large Databases of Compounds: Using the MDL "Keys" as Structural Descriptors. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 443−448.
(29) Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity*; VCH Verlag: Weinheim, Germany, 1992.
(30) Ghose, A. K.; Crippen, G. M. *J. Comput. Chem*. **1986**, *7*, 565.
(31) Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med.Chem.−Chim. Ther*. **1984**, *19*, 71.

CI990268D