Disseminated with permission from Bentham Science Publishers:

*Current Drug Metabolism*, volume 4, edition 5, pages 411-422, copyright 2003;

The article entitled: A *Human Drug Metabolism Database: Potential Roles in the Quantitative Predictions of Drug Metabolism and Metabolism-Related Drug-Drug Interactions*;

Authored by Paul W. Erhardt
A Human Drug Metabolism Database: Potential Roles in the Quantitative Predictions of Drug Metabolism and Metabolism-Related Drug-Drug Interactions

Paul W. Erhardt*

Center for Drug Design and Development, The University of Toledo College of Pharmacy, USA

Abstract: Previous attempts to predict drug metabolism and drug-drug interaction possibilities by deploying databases that house established drug metabolism results have been only marginally successful. Consideration of some of the key issues and concerns derived from these efforts suggests that three major hurdles loom in front of using xenobiotic metabolism databases more effectively in the future. These hurdles include: the need for an improved treatment of chemical structure in three-dimensions (3D); a better quantitative accounting of competitive and complementary biotransformation pathways; and, the critical need for a comprehensive, human drug metabolism database (hDMdb) that can serve as a bio/chemoinformatic resource pertaining to drug metabolism and drug metabolism-related drug-drug interactions in general. Approaches that might be taken to traverse each of these hurdles are discussed herein where the first involves maturation of chemical structures from simple 2D entries into more sophisticated 3D displays that can also account for interactions with relevant biological surfaces; the second involves a systematic, pair-wise comparison of various metabolic options in a statistically driven manner relative to chemical structure descriptors; and, the third involves mounting a hDMdb on the Internet in a user-friendly manner that it is available via a non-profit format.

Key Words: Human Drug Metabolism, Database, Bioinformatics, 3D Chemical Structure, Chemoinformatics, Drug-Drug Interactions.

INTRODUCTION

The gradual accumulation of data from decades of drug metabolism studies has afforded a vast field of information that offers the potential to be used for predicting the metabolic profiles for new drug candidates along with their likelihood to become involved in metabolism-related drug-drug interactions. Several expeditions have ventured into this field to provide maps of varying detail, some of which are available commercially as databases with chemical structure searching capability. Emphasizing chemical structure, two expert systems emerged as early leaders in this area. MetabolExpert [1-3] and META [4-7] were constructed by compiling literature intentionally limited to well-established sources such as respected textbooks and authoritative reviews so as to assure the quality of the information contained within the model dataset. By using computer driven queries, these systems can identify sites on a new molecule where metabolic biotransformations are likely to occur. However, because these early systems indiscriminately combined metabolism data available from studies that employed a variety of mammalian species, their programs tend to predict all of the metabolic possibilities for a given xenogenous material when the latter is placed in a "theoretical 'average' mammal" [5]. Recognizing the need to prioritize the often numerous metabolic possibilities that are suggested, a priority number based on a scale of 1 (fast biotransformation) to 9 (slow biotransformation) also accompanies each prediction from the META program.

Subsequent to the development of these early expert systems, two major database initiatives that represent significant collections of drug metabolism data have also become available. Metabolite [8, 9] is a broad compilation of metabolism data that is being accumulated without bias as to literature source. Thus, the strength of this database may eventually lie in its extensive quantity of data rather than in the quality of each of its data entries. Alternatively, the Accelrys Metabolism Database [10, 11] represents an electronic version of data taken from the well-respected, hard-copy series entitled Biotransformations [12]. Substructure queries across these databases can be used to predict biotransformations for new compounds by finding data for actual compounds within the database that are closest in structure to each query. Finally, METEOR [13, 14] is a recently released expert system that, like META, has sought to provide a ranking for the predicted metabolic possibilities arising from a new structure query. These five commercial products are summarized in Table 1.

A survey on the use of drug metabolism databases within the pharmaceutical industry has previously been assembled as part of a text devoted to this topic [15]. This survey's consensus points become informative toward appreciating the pros and cons of today's attempts to develop and deploy such data during the early stages of drug discovery. The survey's consensus points are highlighted in Table 2 along

*Address correspondence to this author at the Center for Drug Design and Development, The University of Toledo College of Pharmacy, USA; Tel: (419) 530-2167; Fax: (419) 530-1994; Email: paul.erhardt@utoledo.edu

© 2003 Bentham Science Publishers Ltd.
Table 1. Commercially Available Drug Metabolism Databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Vendor</th>
<th>Approx. No. (K)</th>
<th>Search Paradigms</th>
<th>Recommended Platform</th>
<th>Compatible Software</th>
<th>Approx. Cost (SK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MetaboExpert</td>
<td>CompuDrug, Inc.</td>
<td>-</td>
<td>Chemical structure; Transformation; Test system</td>
<td>PC</td>
<td>PALLAS</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>South San Francisco, CA 415-271-8800 <a href="http://www.compubdrug.com">www.compubdrug.com</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>META</td>
<td>Multicase, Inc.</td>
<td>-</td>
<td>Chemical structure</td>
<td>VAX; Open VMS; Windows based PC</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Beachwood, OH 216-831-3740 <a href="http://www.multicase.com">www.multicase.com</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolite</td>
<td>MDI Information Systems Inc. San Leandro, CA 510-895-1315 <a href="http://www.mdil.com">www.mdil.com</a></td>
<td>50</td>
<td>Chemical structure (similar and sub-); Transformation; Data searches</td>
<td>PC; MAC</td>
<td>ISIS; REACCS</td>
<td>15-70</td>
</tr>
<tr>
<td>Metabolism Database</td>
<td>Accelys, Ltd. Leeds, U.K. 44 113 224 9788 <a href="mailto:name@accelys.com">name@accelys.com</a></td>
<td>3</td>
<td>Chemical structure; Transformation; Test system; Author; Journal name; Keywords</td>
<td>SGI; Windows NT; Sun Solaris; Open VMS</td>
<td>ISIS; REACCS; ACCORD Database Explorer</td>
<td>15-70</td>
</tr>
<tr>
<td>METER</td>
<td>LHASA, Ltd. Univ. Leeds Leeds, UK 44 113 233 6531 <a href="http://www.chem.leeds.ac.uk/link">www.chem.leeds.ac.uk/link</a></td>
<td>-</td>
<td>Chemical structure; clust; Transformation; Author; Journal name</td>
<td>PC; Windows NT or 2000</td>
<td>Metabolite; ISIS eLogP</td>
<td>50/26</td>
</tr>
</tbody>
</table>

*Entries reflect status at end of 2000.
*CompuDrug, Inc.
*Budapest, Hungary
*Royalty fee/Annual update fee.

Table 2. Consensus Points (†) from Industry Survey About Using Drug Metabolism Databases [15] and More Recent Concerns (§) that Additionally Pertain to Predicting Drug-Drug Interactions

- Predicting all metabolic possibilities within a theoretical "average" mammal creates a dense forest of information from which it becomes difficult to discern the specific pathways that might be associated with the human response.
- In most cases, many of the suggestions were already suspected from a simple visual inspection of the query molecules whereas in other cases, biotransformations were missed despite the existence of specific literature precedent.
- The need for conveying statistically derived metabolic probabilities rather than a list of possibilities is critical.
- Priority rankings represent a step toward assigning relative probabilities but to date the derivations of such values have not adequately accounted for the potentially competing and/or complementary nature of various metabolic pathway options available to a given xenobiotic.
- Absolute quantitation of a given metabolic prediction is not possible by using any of these approaches.

with some additional concerns that particularly pertain to predicting drug-drug interactions.

The specific points listed in Table 2 can be grouped into two general themes associated with today's bio/chemoinformatic databases. These two themes involve initial biological data entry/sorting, followed by chemical structural entry/inquiry. In terms of biological-related issues, concerns about indiscriminately combining metabolic data from different species have already been alluded to. In most cases where systematic studies have been conducted across different species, qualitative as well as quantitative differences have been observed between the metabolic profiles obtained for a given set of selected xenobiotics, even when the species have been restricted to various mammals [16]. Likewise, the potential for obtaining significantly divergent results within the same species when comparing in vitro versus in vivo data, should also be emphasized. For example, the metabolic profile for a drug that is distributed away from the liver, e.g. processed in the lungs, after intravenous administration could be quite different from that which would be suggested.
by its in vitro study using hepatocytes or liver microsomal fractions wherein there is a constant, static exposure of the drug to these enzymes. Again, in most cases where systematic studies have been conducted across different tissues within the same species, differences have been observed between the metabolic profiles obtained for any given series of selected xenobiotics [17]. Table 3 conveys the differing metabolic and excretion capabilities displayed by several different tissues.

Interestingly, the factors pertaining to distribution and pharmacokinetic half-life may actually be greater concerns when attempting to cross-relate data obtained from in vitro versus in vivo studies in the same species than when cross-relating data from in vivo studies obtained between two different species. For example, some analyses suggest that there is a statistically significant correlation between the half-lives of a wide range of drugs when determined in vivo in two very different species, namely rat data versus human data [18, 19]. The final concern in the biological area has become encompassed by the newly coined field of "pharmacogenetics." Within the present context, this involves the growing appreciation that there can be significant differences in drug metabolism due to just subtle, individual differences in enzymatic genotype or phenotype. Thus, the metabolism of the same drug studied in the same species using the same method (in vitro or in vivo) can still vary from individual to individual based upon classic differences in population phenotypes [20], upon disease-related differences in phenotype [21, 22], or due to the impact upon phenotype resulting from an individual's total metabolic history of previous and ongoing exposures to xenobiotics [23]. Table 4 lists some of the classical, human drug metabolizing phenotypes.

**Table 3. Differing Metabolic and Excretion Capabilities Displayed by Selected Tissues**

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Metabolic/Excretion Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI Mucosa</td>
<td>Rich in CYP 3A but essentially all CYPs are represented, gradually peaking in the duodenum region and then falling off towards the ileum; Significant glucuronide and sulfide conjugation pathways; Significant monooxygenase activity.</td>
</tr>
<tr>
<td>Blood</td>
<td>Rich in various esterase enzymes.</td>
</tr>
<tr>
<td>Liver</td>
<td>Rich in essentially all CYPs and conjugation pathways. Excrete compounds into bile having MW ≥ 500 gm, particularly when highly polar, conjugated molecules.</td>
</tr>
<tr>
<td>Lungs</td>
<td>Represented by most of the CYPs and conjugation pathways with certain levels being comparable or even higher than levels found in the liver, e.g. CYP 2A13 as recently implicated in the carcinogenicity of nicotine.</td>
</tr>
<tr>
<td>Vasculature</td>
<td>Neither the BBB, the placenta nor the mammary gland 'barriers' exhibit significantly enhanced overall xenobiotic metabolizing capabilities.</td>
</tr>
<tr>
<td>Kidneys</td>
<td>Rich in the various protease enzymes. Excrete compounds having MW ≤ 500 gm; highly polar, water soluble compounds are not reabsorbed from the urine.</td>
</tr>
</tbody>
</table>

BBB = Blood brain barrier. CYP = Cytochrome P450 enzyme. MW = Molecular weight.

**Table 4. Differing Metabolic Capabilities Displayed by Selected Human Phenotypes**

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Characteristic Metabolic Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debrisoquine²</td>
<td>Extensive versus poor metabolizers. Inherited defect in CYP 2D6 expression; 5-10% Caucasian, 2% Oriental and 1% Arabic populations.</td>
</tr>
<tr>
<td>Phenyltrimine²</td>
<td>Normal versus poor metabolizers. Inherited defect in CYP 2C9; 15-20% Oriental and 2-6% Caucasian populations.</td>
</tr>
<tr>
<td>Aldehyde Oxidase²</td>
<td>Lack of indicated capability. Significant polymorphism among the Oriental population, ca. 50%</td>
</tr>
<tr>
<td>Acetylation²</td>
<td>Fast and slow acetylators. Individual variations in Nat 2; 30-40% Caucasian, 80-90% Oriental and 100% Eskimo populations are fast.</td>
</tr>
<tr>
<td>Neonatal</td>
<td>Overall, CYP isoform activities typically lower than children to adult range except for the 3A subfamily. Blood esterase activity appears to be about 50% at birth [62]. Immature UGT pathways fairly common leading to decreased clearance of bilirubin or gray baby syndrome.</td>
</tr>
<tr>
<td>Elderly</td>
<td>Most studies directed toward enzymatic activity per se suggest that levels remain comparable to younger populations. Alternatively, decreased hepatic blood flow can significantly lead to a decrease in the overall metabolism and excretion of xenobiotics.</td>
</tr>
</tbody>
</table>

CYP = Cytochrome P450 enzymes, NAT2 = N-Acetyl Transferase isofrom 2. UGT = Uridine diphosphoglucuronate Transferase.

²This phenotype extends to numerous other drugs.

²This phenotype is largely characterized by an altered pathway for ethanol metabolism.
Clearly, many of the issues that have complicated earlier systematic investigations within the field of drug metabolism, still remain and serve to exacerbate the additional challenges associated with today's construction and deployment of much larger databases. Nevertheless, at least some of the biological issues might be addressed as a series of initial sorting steps when data is being entered. While this strategy would be cumbersome initially, it would result in a series of biologically intelligent and, perhaps, more manageable databases. Certainly the current move toward constructing "smarter" databases [24] would favor this approach. Alternatively, if enough searchable terms pertaining to the concern areas are entered into a single, relational database along with the actual metabolism data, then appropriately factored searching paradigms, coupled with rational, physiologically-based pharmacokinetic (PBPK) considerations throughout, could be devised to surmount these issues on the query end.

In terms of chemical-related issues, there are two areas that come to immediate attention. The first involves a fundamental question: what is the proper role to expect the entire structure to be playing versus the more discrete roles of its displayed functional groups? After all, it is localized functional groups that actually have metabolic conversions within an overall molecule and, just like a "metabophore" defines the pattern of localized structural elements requisite for a drug's interaction with a biologic receptor or enzyme active site, it will be the particular array of selected functional groups and their immediate molecular environments that dictate the metabolic events that will occur. Numerous issues stem from this question, all leading to a common concern about how much detail should be included for a given chemical structure data entry/query. While the level of detail that can be applied toward molecular description/searching spans a considerable range, a second chemical issue additionally presents itself at this juncture, namely the accuracy of the initial chemical depiction. For example, one extreme would be a thorough, three-dimensional (3D) electronic surface map across the entire structure as obtained from x-ray analysis and/or rigorous computational treatments, additionally coupled with experimental physicochemical information or descriptors such as logP to also address the distribution issues. The other extreme would be a simple two-dimensional (2D) figure, perhaps energy minimized with only the aid of an automated drawing program. Historically, practicality has dictated that the most simple structural data entry possible be used due to the sheer amount of valuable information that is already available. However, even in this case it may be expecting too much from a given database to be able to predict the entire metabolic outcome for a new compound using a single structural query. Rather, it may be more appropriate to think in terms of a multi-step approach that might, first include analyses across a parent database in terms of various metabolic functional groups so as to initially produce a series of specific SMR maps which characterize key 'metabophore' [25, 26] elements in terms of the probabilities that they might undergo a particular metabolic reaction based upon such occurrences relative to other possibilities. The metabophores may then need to be better refined by inputting considerably more precise structural detail. As for handling the biological data and issues, such a strategy would certainly be in line with today's "trend toward smarter libraries" as well as with the trends being utilized to "integrate various bioinformatic and chemoinformatic systems" [24]. Employing sets of species specific atlasses, each having its own list of refined metabophore/relative metabolic probability maps, in conjunction with a new drug structure query would then become a third step in an overall searching paradigm that will probably also need to be continuously guided in a rational manner by appropriately considering distribution and pharmacokinetic issues throughout.

From the foregoing discussions, three major hurdles loom in front of using xenobiotic metabolism databases to more effectively predict human drug metabolism and metabolism-related drug-drug interactions. The first involves the need for an improved treatment of 3D structure within chemical-related databases in general, the second involves the need to better quantify the competing and/or complementary nature of drug metabolism that results from multiple biotransformation possibilities, and the third involves the critical need to ultimately be able to correlate the various types of metabolism-related data to the human clinical experience. Each of these hurdles is further discussed below.

**HANDLING CHEMICAL STRUCTURE**

Handling chemical structures and chemical information within the setting of large databases represents a specialized exercise complicated enough to merit its own designation as a new field, namely that of "chemoinformatics" [27-29]. As mentioned, there is a significant need for improvement in the handling of chemical structures beyond what appears to be occurring within today's database assemblies. For example, that "better correlations are sometimes obtained by using 3D displays of a molecule's chemical structures than by using 2D displays," only testifies to the fact that we are still not doing a very good job at developing the latter [30]. In general, the handling of small molecules and of highly flexible molecular systems [31] is controversial with the only clear consensus being that treatments of small molecules for use within database collections have, to date, been extremely inadequate. Furthermore, one common reason why the data fields within a chemoinformatics database typically do not fit together is because "each was designed to optimize one [particular] process and the data relationships and structures are [often] not consistent" [32]. Certainly a variety of automated, 3D chemical structure drawing programs are available that can start from simple 2D representations by using Dreiding molecular mechanics or other user-friendly, automated molecular mechanics-based algorithms, as well as by using data expressed by a connection table or linear string [33]. Some programs such as Chem Pen 3D, CS Chem Draw 3D, Corina and JChem, are able to derive 3D structure from several different types of import formats [34]. Furthermore, several of these programs can be directly integrated with the latest versions of more sophisticated quantum mechanics packages such as Gaussian 98 MOPAC (with MNDO/d) and
extended Huckel [29, 33]. Thus, electronic handling of chemical structures, and to a certain extent comparing them in 3D formats, has already become reasonably well worked-out [27–29, 33–36]. Table 5 provides a listing of some of the 3D molecular modeling products that have become available during the 1990’s [35].

The fundamental problem that remains, however, is how the 3D structure is initially derived in terms of its chemical correctness, the latter being dependent upon what assumptions might have been made during the process of energy minimization. This situation is further complicated by the additional need to understand how a given drug molecule’s conformational family behaves during its interactions with each of the biological environments of interest, namely those associated with all of the compartments that must be traversed, as well as with the compartments associated with each of the specific metabolizing enzymes that the drug will eventually encounter. In order to track these conformational behaviors in a comprehensive manner, it becomes necessary to consider a drug’s multiple conformational possibilities. The common approaches that can be used to assess chemical conformation include: (i) X-ray which becomes subject to molecular interactions unique to the crystalline state; (ii) Solution spectroscopic methods such as NMR which can often be done in both polar and non-polar media; and, (iii) Computational approaches which can be done with various levels of solvent and heightened energy content, but are subject to the assumptions and approximations that need to be taken in order to simplify the mathematical rigor so as to allow computational solutions to be derived in practical time periods. Analogous to the simple, drawing program starting points, programs are also available for conversion of x-ray and NMR data into 3D structures [37]. Detailed descriptions of computational approaches that can be taken to discern chemical structure relative to drug metabolism possibilities can be found elsewhere [38].

As mentioned above, structures need to be considered by ascertaining what their relevant conformations might be during interactions within various biological milieu. It can

Table 5. 3D Molecular Modeling Packages that Became Available During the 1990s [36]

<table>
<thead>
<tr>
<th>Package</th>
<th>Company</th>
<th>Platform</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nano Vision</td>
<td>ACS Software</td>
<td>MAC</td>
<td>Simple, effective tool for viewing and rotating structures, especially large molecules and proteins</td>
</tr>
<tr>
<td>Ball &amp; Stick</td>
<td>Cherwell Scientific</td>
<td>MAC</td>
<td>Model building and visualization; analysis of bond distances, angles</td>
</tr>
<tr>
<td>MOBY</td>
<td>Springer Verlag</td>
<td>IBM (DOS)</td>
<td>Model building and visualization; classical and quantum mech. computations; large molecules and proteins; PDB files</td>
</tr>
<tr>
<td>Nemesis</td>
<td>Oxford Molecular</td>
<td>IBM (Wind) MAC</td>
<td>Quick model building and high-quality visualization; geometry optimization (energy minimization)</td>
</tr>
<tr>
<td>CSC Chem. 3D/Chem 3D Plus</td>
<td>Cambridge Scientific</td>
<td>MAC</td>
<td>Easy-to-use building and visualization; geom. opt.; integrated 2D program and word processing</td>
</tr>
<tr>
<td>Alchemy III</td>
<td>Tripos Assoc.</td>
<td>IBM (DOS, Wind) MAC</td>
<td>Quick model building; energy minimization; basic calc.; easy integration to high-end systems</td>
</tr>
<tr>
<td>PC Model</td>
<td>Serena Software</td>
<td>IBM (DOS) MAC</td>
<td>Low cost with sophist. calc.; platform flexibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cache</td>
<td>Tektronics</td>
<td>MAC</td>
<td>Sophist. comp. tools; distributed processing</td>
</tr>
<tr>
<td>HyperChem</td>
<td>Auto Desk</td>
<td>IBM (Wind) Silicone Graph</td>
<td>Easy-to-use array of comp. tools (classic &amp; semi-empirical quant. mech.)</td>
</tr>
<tr>
<td>Lab Vision</td>
<td>Tripos Assoc.</td>
<td>IBM (RISC-6000) Silicone Graph, DEC, VAX</td>
<td>Sophist. but practical modeling for research</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYBL</td>
<td>Tripos assoc.</td>
<td>IBM (RISC-6000) Silicone Graph, DEC VAX, Sun 4, Convex</td>
<td>Integrated comp. tools for sophist. struct. determ. &amp; anal., database management</td>
</tr>
<tr>
<td>CERIUS</td>
<td>Molecular Simulations</td>
<td>Silicone Graphics IBM (RISC-6000) Student Titan</td>
<td>Suite of high-performance tools for building and simulating properties</td>
</tr>
</tbody>
</table>
be imagined that for predicting metabolism and drug-drug interaction possibilities, a useful range of such environments to be considered should ideally include: aqueous solutions of acidic and neutral pH, namely at ca. 2 (stomach), 5.3 (estimated microscopic ‘virtual’ pH relevant to absorptive surfaces) [39], and 7.4 (macroscopic physiological), respectively; one or more lipophilic settings, such as might be encountered during passive transport through membranes; and finally, specific biological receptors and/or enzyme active site settings that are of particular interest from a metabolism and drug-drug interaction standpoint. With time this initial list can then be expanded to also include; several distinct environmental models deemed to be representative for interaction with various transportophore relationships; several distinct environmental models deemed to be relevant for interaction with specific metabophore relationships such as within the active site of a specific cytochrome P450 metabolizing enzyme (see the article by Dr. Lewis in this issue); and finally, several distinct environmental models deemed to be relevant for interaction with specific toxicophore relationships. If x-ray, NMR and computational approaches can be further applied to assess any one or combination of these types of interactions, then a composite approach that invokes as many as possible of these techniques will certainly represent the most ideal way to approach future conformational considerations within the variously biased settings. Advances toward experimentally studying the nature of complexes where compounds are ‘Docked’ into real and model biological environments are proceeding rapidly in all of these areas. In addition to the experimental approaches, computational schemes such as those described by Lewis in this issue will likely be of great value, because they can provide the relative energies associated with all of the different species. Furthermore, computational methods can be used to derive energy paths to get from a first set of unbiased structures to a second set of environmentally accommodated conformations in both aqueous media and at biological surfaces. Eventually, these paths and their energy differences will then be able to be compared within database settings, along with the direct comparison of the structures themselves, while attempting to uncover and define correlations between chemical structure and some other informational field.

Regardless of how they are exactly evolved, what this section ultimately points to is that, the more sophisticated chemical structural databases of the future will probably have several ‘tiers’ [40] of organized chemical and conformational information available which can be distinctly mined according to the specified needs of a directed searching scheme, while still being able to be completely mixed within an overall relational architecture such that undirected, ‘knowledge-generating mining paradigms’ can also be undertaken [41-46]. Certainly, simple physicochemical data will need to be included among the parameters for chemical structure storage. Likewise, searching engines will need to allow for discrete sub-structure queries as well as for assessing overall patterns of ‘similarity’ and ‘dissimilarity’ [47-54] across entire electronic surfaces.

It can be noted that it is probably already feasible to place most of the clinically used drugs into a structural database that could at least begin to approach low to mid-tier levels of 3D sophistication because considerable portions of such data and detail are likely already available within the literature, even if spread across a variety of technical journals for each drug. On the other hand, it should also be clear that an alternate strategy will be needed to handle the mountains of research compounds associated with just a single HTS survey. Taken together, the present discussions suggest that we still have a long way to go toward achieving the aforementioned tiers of conformational treatments when dealing with large databases and applying them to the process of drug discovery. Nevertheless, because of the importance of cheminformatics toward understanding, fully appreciating and ultimately implementing bioinformatics along the practical avenues of new drug discovery, it can be imagined that future structural fields within databases, including those associated with drug metabolism and drug-drug interactions, may be handled according to a variety of more sophisticated scenarios, such as that described in the prior discussion and summarized within Fig. (1).

It follows that chemical structure search engines of the future will probably also be set up so that they can be undertaken at several tiers of sophistication, the more sophisticated requiring more expert-based inquiries and longer search times for the attempted correlations to be assessed. A reasonable hierarchy for search capability relative to the structural portion of any database query might become: (i) Simple 2D structure with and without physicochemical properties; (ii) 3D structure at incremented levels of refinement; (iii) 2D and 3D substructures; (iv) Molecular similarity/dissimilarity indices; (v) ‘Fuzzy coordinate’ matrices [55, 56]; (vi) Docked systems from either the drug’s or the receptor/active site’s view at various levels of specifiable precision; and finally, in the more distant future, (vii) Energy paths for a drug’s movement across various biological milieu including the trajectories and molecular motions associated with drug-receptor/active site docking scenarios and mutual molding paradigms. Emphasizing informatics flexibility, this type of approach where data entry can occur rapidly for starting structure displays and then be gradually matured to more sophisticated displays as conformational details are accurately accrued, coupled with the ability to query at different levels of chemical complexity and ‘visual displays’ [57] at any point during database maturation, should allow for chemically creative database mining strategies to be effected in the near term, as well as into the more distant future.

QUANTIFYING COMPETITIVE AND COMPLEMENTARY METABOLIC PATHWAYS

The second major hurdle toward more effectively deploying drug metabolism databases involves the critical need to prioritize one biotransformation pathway’s contributions versus that of another pathway’s contributions while accounting for a given compound’s in vivo distribution pattern. This situation is difficult to address by using a series of distinctly separate in vitro experiments, mixtures of e.g. CYP enzymes, or even when using intact hepatocytes or liver slices, thus prompting the need for intact human data. However, a database replete with intact human data (see next section) will still represent only the starting point for the type of informatic studies that will be needed to confidently
traverse this particular hurdle in order to better predict drug metabolism and drug-drug interaction possibilities. An example of an informatic study that attempts to surmount this hurdle is described below.

Prioritization of biotransformation pathways will need to be conducted by performing computationally based, statistical analysis of data in a well defined, systematic manner across extremely large data sets so that conclusions and predictions are generated and strengthened by the sheer number of experimentally derived data entries that are available for such considerations. Specifically, with a large enough database it should become possible to examine the relative occurrence of competing metabolic reactions by taking a reactivity schemes approach. One of the tools that has already been developed in a preliminary fashion for the study of xenobiotic metabolism data in this manner is the structure-reactivity map [47-49]. Structure-reactivity maps represent a graph-theory-based map of the structure-reactivity space of a data set. It has been demonstrated that by constructing a suitable map of the structure-reactivity space, one may visualize structure-reactivity relationships which are not readily apparent from a conventional examination of data tables. The structure-reactivity map, in turn, can display the practical similarity, or conversely the dissimilarity, between structures or chemical descriptors in a database. In preliminary investigations using structure-reactivity maps, statistically significant relationships were established between several molecular structural descriptors associated with the relative occurrences of N-demethylation versus N-oxidation [58], as well as statistically significant relationships between the relative occurrences of N-demethylation versus non-methyl N-dealkylation [50]. In each of these studies, predictive indicators of the relative occurrence of each of the competing biotransformations were identified from an analysis of the structure-reactivity map. Likewise, preliminary studies using the Metabolite database were able to rank the likelihood that certain metabolic transformations will occur in a given system [8]. These rankings were obtained by tabulating the proportional occurrences of competing metabolic transformations contained in this particular database. Because these rankings were determined from a very large database having numerous, representative metabolic transformations, it became possible to associate statistical confidence intervals with the reliability of the rankings for a given new chemical entity. In addition to these studies, work has been directed toward refining structure-reactivity relationships of metabolic transformations using molecular mechanics and semiempirically calculated molecular descriptors. For example, initial results suggest that there are linear relationships between certain of
these descriptors and the relative rates of amide N-dealkylation [51].

Given the complexity of the considerations needed to most effectively utilize an even more abundant set of specific human data in a predictive manner, it is probably constructive to review three milestones that can be envisioned for progress in this area [59]. Taking a systematic approach, the first milestone could involve an examination of all major classes of metabolic transformations across a database using a 2 x 2 or paired comparison scheme. The endpoint would be a rank ordering of metabolic reactions that could be employed immediately by an expert user to predict the relative occurrence of competing biotransformations among discrete structural themes. Based upon an experimentally derived mechanistic understanding of the relative pathways, an expert would be able to use the data to rank the metabolic possibilities in a manner similar to some of the presently available commercial expert system products that include such ranking features. However, the present example would involve queries into the original data for each case such that the rankings of the metabolic options for a given structural query would be inherently tailored to reflect just that structure relative to the database’s similar structures rather than to the database as a whole.

A second milestone would be to re-examine the metabolic transformations in a relational manner across the entire database in order to strengthen conclusions about their co-relationships as multiply competing pathways for any given structure. Special focus could be placed upon paired comparisons having low numbers of data entry points and on competing pathways having similar statistical occurrences. Structure-reactivity maps would be used to develop and convey an improved structural/mechanistic understanding about all pathways. The finished product would then be a set of specific atlases of structure-reactivity maps indexed by pairs of competing reactions. Predictive rules could be associated with each of the maps in the atlases. An expert user would then be able to use the atlases and associated rules, by applying them to specific structural and mechanistic considerations, to better predict the relative likelihood of biotransformations which may occur over an entire query molecule. While still relying upon the original data, the rules contained in this atlas would also provide an excellent source of information for the further elaboration of an expert system that could be used to predict the probability of various biotransformations possible within a query molecule in a simplified manner for a non-expert user.

A final milestone could thus involve an ongoing refinement of specific structural descriptors associated with the reactivity maps and the relative rate tables, and further molecular parameterization of mechanistic details. This phase would also involve development of statistical method/software to assess/rank possible outcomes for a given structure based upon identifiable structure-reactivity elements, mechanistic parameters and relational table information as initially defined during milestones 1 and 2, and is further refined by the above. Presumably, these activities would closely parallel the developments suggested in the section on handling chemical structures. Table 6 lists some of the experimental details that would be associated with pursuing this approach.

Table 6. Experimental Details Associated with Assessing Competing Metabolic Reactions Within Large Data-Base Settings [59]

- Identify the 10 most commonly reported metabolic reaction classes which are also well represented by multiple entries within the database.
- Construct a 10 x 10 table of each of the potential reaction environments so as to set all of the possibilities up as pair-wise competing reactions.6
- Construct structure-reactivity maps for each of the 45 pair-wise competing reactions. (Symmetry of the 10 x 10 table reduces the 100 interactions to \( \frac{(10 \times 10)-1}{2} \) where \( n \) is the number of interactions to be studied.
- Convert reaction-sites on structures to reaction-site representations, i.e. 3-dimensional fragments or other suitable structure plus physicochemical representations.
- Compute similarity distance between each reaction-site.
- Map structure space by connecting most closely related reaction-site representations. Label structure-reactivity map according to reactivity, modes of administration, etc.
- Develop reactivity structural descriptors and statistically evaluate the proposed descriptors.6
- Superimpose a composite, relational survey upon the 2 x 2 box analyses within the 10 x 10 table.6
- During all phases, selected drug examples having at least two quantified metabolic results, but which are deliberately absent from the database under statistical examination, should be used to test the identified relationships.6

6The diagonal will contain the number of compounds containing the individual reaction site. Each block of the table will contain a 2 x 2 box indicating the four possible pair-wise interactions, i.e. both reactions are observed, either reaction is observed, or one reaction is observed to the absence of the other.

6These analyses would take the form of 2 x 2 contingency boxes. The association of the classification schemes is then examined via the chi-square test. Each labeled structure-reactivity map can be examined to discover important structural features affecting the relative occurrence of each reaction. Each hypothesis can be statistically evaluated.

6For example, relationships such as the relative occurrences defined by data for the two systems having pathways A & B and B & C, respectively, will be analyzed for their reliability in allowing extrapolation of the relative occurrence for pathway D versus A or B based upon a third system containing only C & D relative occurrence data (and likewise restricted for larger and more distant connections). Ultimately, the presence of such multiple reaction scenarios within the database can be accounted for as statistically discerned primarily over each 10 x 10 table with the apex representing the possibility in a single 10 x 10 box where all ten reactions have occurred within the same molecule. This approach can be used periodically as an internal check of the overall assessment, e.g. the aforementioned extrapolated relative occurrence data for D should be similar to actual D & A or D & B data when it also happens to be present and derived separately from within the database or from a compound being deliberately withheld from the database for use as a test case.

6In some cases, it may become advantageous to have such information generated via laboratory experiments undertaken by the research team itself.
The aforementioned milestones can be thought of as part of an integrated program designed to facilitate the development of predictive methods which may be useful in drug metabolism and drug-drug interactions research, as well as being of practical utility during new drug development. Armed with statistically supported predicative schemes obtained from published experimental studies, investigators should be able to confirm and sharpen what have otherwise been only intuitive notions about relative biotransformation rates, drug-drug interaction possibilities, and mechanistic toxicological possibilities.

**HUMAN DRUG METABOLISM DATABASE**

The third major hurdle toward more effectively deploying drug metabolism databases involves the critical need to better define the correlation of preclinical data and in silico predictive paradigms with what can ultimately be expected to occur in humans. Simply stated, there remains a desperate need for a ready mechanism or protocol that can be used to validate a given HTS or in silico model relative to the model’s ability to predict the corresponding metabolic-related event in humans. The model may reflect a directly measurable endpoint in itself, such as an observable human metabolite, or it may reflect a component within a set of multiple parameters that together provide for quantitation of the PK profile or competitive metabolic pattern observed in humans. In either case, a significant composite of human data becomes requisite in order to pursue such correlations in a statistically relevant manner. In this regard, however, it must be emphasized that the amount of metabolism data produced at the clinical level is quite small compared to the rather large amount of metabolism data available from preclinical studies. This situation has prompted an effort by our laboratories to establish a broad composite of specific human drug metabolism data [60, 61]. This relational database is currently under development. Coupled with substructure searching capability, it is being assimilated solely from human clinical results that can be continuously updated and expanded by all interested researchers collecting metabolic data in clinical trials. In turn, it is envisioned that the database will be made available on the Internet via a non-profit mechanism. Thus, the operation and utility of this metabolism database can be likened to that of the Cambridge X-ray collection, the protein databank, or to some of the newer gene-related informational resources that have been made available over the Internet on a non-profit basis.

Importantly, the sheer size of such a common database can overcome the anecdotal nature of the numerous smaller collections presently being held individually by the big pharmaceutical companies. As a result, the database’s growing size will eventually allow it to be utilized to develop more accurate and meaningful human SMRs, and to better assess the competing nature of metabolic possibilities (prior section) within intact humans. It is also imagined that selected aspects of the overall SMRs, in turn, will still be applicable to individuals in a proprietary fashion to better predict the metabolic fate of their own, specific structural motifs. Likewise, it is imagined that a specific human metabolism database would support, rather than compete with, the ongoing activities of the existing metabolism database vendors. The latter have already collected data from numerous species and various testing paradigms, all of which will still be very much required as useful road maps during new drug development for quite some time. Finally and perhaps most important, the assembly of this type of database may be the only way to assess and validate the actual utility of the ongoing explosion of biochemical and in vitro HTS metabolism data and in silico techniques currently being directed toward resolving metabolism issues at the earliest possible stages of drug discovery. The anticipated benefits of such a database are listed in Table 7.

The distinct informational fields presently being constructed within this database are shown in Fig. (2), which also depicts the database’s overall architecture. The treatment of 3D chemical structures has already been outlined in Fig. (1) which depicts the progression for

<table>
<thead>
<tr>
<th>Table 7. Utility of a Human Xenobiotic Metabolism Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>• To be readily available on the Internet, it will warehouse drug metabolism and drug metabolism-related drug-drug interaction data that can be queried by any visitor via a non-profit format.</td>
</tr>
<tr>
<td>• Will allow explicit structure searching to readily provide information about specific compounds and to allow for the selection of the most relevant standards across a variety of structural themes that can then be used to validate proprietary drug metabolism screens and screens intended to assess drug-drug interactions.</td>
</tr>
<tr>
<td>• Will allow substructure searching to identify analogous metabolic occurrences and drug-drug interactions within humans relative to proprietary compounds undergoing drug development.</td>
</tr>
<tr>
<td>• Will have a large number of biotransformation and drug-drug interaction entries so that statistically derived probability assessments can be made about all metabolic possibilities including that of competing and complimentary biotransformation pathways as well as drug-drug interaction possibilities.</td>
</tr>
<tr>
<td>• Will be relational in nature so that any type of searching paradigm can be deployed to identify novel relationships across all or selected segments of the data, e.g., potential correlations between a certain set of structural parameters with potential metabolic occurrences or with likelihoods for a particular type of drug-drug interaction.</td>
</tr>
<tr>
<td>• Will be updated on a periodic basis, including that of filling presently unoccupied data fields being reserved so as to also be able to include enhanced pharmacogenetic data as the latter is expected to unfold reasonably soon relative to human drug metabolism data.</td>
</tr>
</tbody>
</table>
DRUG-DRUG INTERACTIONS

Although the mechanistic basis for a given drug-drug interaction is ultimately defined at the level of biochemistry or molecular pharmacology, historically the manifestation of such events has been first observed at the clinical level. For drug metabolism-related events in particular, there is already a wealth of information about drug-drug interactions residing within the available clinical data. Indeed, the importance of including drug-drug interaction data within the human drug metabolism database was recognized at the outset of our program. Once assembled into a composite database that also contains an informational field to rigorously addresses chemical structure features, this information, in itself, is likely to yield useful structure drug-drug interaction relationships when the database is subjected to unbiased knowledge-generating mining paradigms, as well as when queried in a biased manner directed toward a specific metabolism-based drug-drug interaction. These structure-derived trends, in turn, can then be used to predict the probability for a new compound to engage in a given drug-drug interaction according to the same way that has been described for eventually being able to predict the probability.
for a new compound to undergo a given metabolic event. Likewise, returning to a mechanistic approach, the biochemical and molecular pharmacological models that may be derived to ascertain the potential for drug-drug interaction during the early phases of drug HTS, will also be able to utilize the human drug metabolism database as a standard set of data to validate the predictability of such models along various structural themes. In this manner, drug-drug interaction-related chemical structure themes become the key bridges that can be used to connect mechanism based in vitro or in silico models with clinical observations, regardless of which direction an investigator is attempting to traverse these types of data fields.

SUMMARY

The gradual accumulation of drug metabolism studies has afforded a vast field of data which offers the potential to be used for predicting the metabolic outcomes of new drug candidates and for predicting potential metabolism-related drug-drug interactions. Five major expeditions have ventured into this field to provide maps of varying detail which are available commercially as expert systems or databases with chemical structure/sub-structure searching capabilities. Previous analyses of several case studies indicate that attempts to predict metabolite outcomes for new structures by using these types of databases have been only marginally successful. The reasons for this shortcoming include both biological and chemical factors. Some of the biological issues include species and phenotypic variation, as well as how to properly inter-relate PB/PK-types of parameterization and HTS screening results so as to better reflect the intact human. The major chemical issues include a fundamental question about how much of a structure should be included within a metabophore during SMR assessments, and the long-standing issue of how to accurately derive 3D structure for small, highly flexible molecules. The discussions within this paper have considered these issues and have suggested some new approaches that could enhance the utility of future chemical structure-biological information databases in general. In particular, the suggestions may be useful toward the specific assembly of drug metabolism databases that might partner in a synergistic manner with HTS metabolism and drug-drug interaction data acquisition methodologies.

Over-riding all activities in this area, it is imperative that a large collection of human drug metabolism data be made available as a standard database so that all types of in silico or theoretical models, and all types of pre-clinical sets of data from whatever type of parameterized HTS screen, in vitro or in vivo experimental model, can then be monitored for their predictabilities in statistical terms, if not actually validated for their direct utilities. Likewise, in terms of chemical issues, it has been emphasized that while 2D representations may constitute a practical starting-point for the input of structures, it is imperative that methods to be evolved to mature these displays into accurate representation of the relevant 3D conformational families. Specific approaches toward the construction of a commonly held, human drug metabolism database and for the handling of 3D chemical structure have been elaborated herein. Both initiatives are now being further developed within our laboratories.

ACKNOWLEDGEMENTS

The author is gratefully appreciative of the seed-funding and support jointly provided for the human drug metabolism database project by the International Council of Scientific Unions (ICUSU), the International Union of Pharmacology (IUPHAR) and the International Union of Pure and Applied Chemistry (IUPAC).

REFERENCES

[25] This phrase is being considered for adoption by the International Union of Pure and Applied Chemistry's nomenclature working party on Drug Metabolism Terms being chaired by P.W. Erhardt.

[30] Personal communications while attending numerous workshops directed toward expediting drug development.


[37] QUANTA (system for determining 3D protein structure from x-ray data) and FELIX (system for determining 3D protein structure from NMR data) available from MSI-Pharmasset (www.msi.com).


[46] Bio Reason at www.bioreason.com; Columbus Molecular Software (Lead Scope) at www.columbus-molecular.com; Daylight Chemical Information Systems at www.daylight.com; IBM Intelligent Miner at www.ibm.com; Iaclyte (Life Tools and Life Prod) at www.iaclyte.com; Molecular Applications (Gene Mine 3.5.1) at www.mmg.com; Molecular Simulations at www.msi.com; Oxford Molecular (DIVA 1.1) at www.oxmol.com; Pangea Systems (Gene World 3.5) at www.pangsys.com; Pansight at www.pansight.com; SAS Institute (Enterprise Miner) at www.sas.com; SCI (Meta Set 3.0) at www.spi.com; Spotfire (Spotfire Pro 4.0 and Leads Discover) at www.spotfire.com; SPSS (Clementine) at www.spss.com; Tripes at www.tripos.com.


