Visualizing Relative Occurrences in Metabolic Transformations of Xenobiotics Using Structure-Activity Maps

Eric M. Gifford,[‡] Mark A. Johnson,*[,§] David G. Kaiser,[§] and Chun-che Tsai*[,‡]

Department of Chemistry, Kent State University, Kent, Ohio 44242, and Upjohn Laboratories, Kalamazoo, Michigan 49001

Received June 18, 1992

Structure maps are presented as an efficient means of indicating structure-reactivity relationships in metabolic pathway databases. The relative occurrence of N-demethylation and N-oxidation of *N*-methyl tertiary amines was examined using the structure map methodology. A new family of reaction site representations, the *n*-level representations, was developed to describe the *N*-methyl reaction sites of the compounds in the data set. It was possible to differentiate N-demethylation and N-oxidation reaction sites using a structure map constructed from a 3-level representation of the reaction sites.

I. INTRODUCTION

There is a growing interest in databases containing information on the biotransformation pathways of xenobiotics. These databases, whether machine-readable[1-3] or hard-copy documents,[4,5] provide information necessary to develop the rules used for predicting xenobiotic metabolism.

Many workers[2,3,6-9] are investigating computer-based approaches to metabolic prediction. The predictive ability of these computer-based methods will increase as further improvements in determining structure-reactivity relationships are made. In this paper we propose a method for gaining insight into the structure-reactivity relationships contained in metabolic pathway databases.

Metabolic pathways are presented as a series of pairwise connected molecules. An arrow between the two molecules indicates that a biotransformation from the reactant to the product has been observed. Reactivity, defined in a simple sense as the observed occurrence or the nonoccurrence of a given reaction, may be thought of as a molecular activity. By defining reactivity as a molecular activity, the data analysis techniques which are used in structure-activity relationship (SAR) studies become applicable.

In order to determine structure-activity relationships, one usually converts the chemical structural formula of a molecule to an appropriate SAR representation or descriptor. The selection of the representation is not always a simple matter, and even if a suitable representation is found, a method for effectively visualizing the relationships between these representations is needed. The representation developed for this study, called an *n*-level reaction site representation, includes information about the chemical structural environment of a reaction site on a molecule.

With a reaction site representation in hand, the next step is to visually inspect the data for structure-activity relationships. Although it is possible to simply list data in tables along with the associated chemical structures or representations, the tabular format does not necessarily indicate the relationships between the entries in the table. Of course the data table may be arranged in a defined order, for example, alphabetical lists of names or increasing values of a property, but even with the order defined it may still be difficult, if not impossible, to show any relationship between the chemical structures and the experimental data. A means of presenting data in a manner which facilitates the discovery of structure activity-relationships is needed.

There is increasing interest in the application of molecular similarity concepts to determine structure-activity relationships.[10] In this paper we apply molecular similarity concepts to construct a "map" of the structure-activity space of a data set. Structure-activity relationships, which may be difficult to establish from data tables, should become apparent once a suitable map of the structure-activity space has been constructed.

II. GENERAL CONCEPTS OF STRUCTURE-ACTIVITY MAPS

An intuitive idea of a structure-activity map is motivated by intercity distance maps. An example of a distance matrix underlying an intercity distance map is given in Figure 1, and the corresponding intercity distance map in Figure 2. Notice that one starts with a distance matrix giving the pairwise distances between the cities. From this, one constructs a graph whose vertices are cities and whose edges are labeled with the intercity distances. Here one is not entirely sure as to why some pairwise distances are represented by edges in the intercity distance map and others are not represented. However, one certainly suspects that the edge representing the distance between Syracuse and Boston was omitted because Albany lies in between these two cities.

A structure map bears many analogies with an intercity distance map. In both cases, the underlying graph is called a proximity graph. In a structure map, the vertices of the proximity graph are molecular structures rather than cities, and the edges denote pairwise molecular similarities, distances, or dissimilarities. The many molecular similarity indices that might be used to generate the underlying pairwise distance or proximity matrix are reviewed in Johnson;[11] additional ones are presented in other papers of this special issue. Moreover, the proximity graph comes in a variety of forms such as nearest-neighbor graphs, minimum spanning trees, relative neighborhood graphs,[12] Gabriel graphs,[13] and sphere of influence graphs.[14] In our development of a structure map, section IV, we use a similarity measure based on maximum common substructures[15] and a proximity graph closely related to the relative neighbor proximity graph.[14] However, the concept of a structure map is general. It can be based on any method of computing the underlying proximity graph and the pairwise

* To whom correspondence should be addressed.
‡ Kent State University.
§ Upjohn Laboratories.

0095-2338/92/1632-0591$03.00/0

	Albany	Bangor	Boston	Buffalo	New York	Philadelphia	Portland	Scranton	Syracuse	Watertown
Albany	0	405	165	283	157	249	272	179	132	204
Bangor	405	0	240	688	451	543	133	581	537	609
Boston	165	240	0	448	211	303	107	341	297	369
Buffalo	283	688	448	0	418	418	555	288	151	223
New York	157	451	211	418	0	92	318	130	267	339
Philadelphia	249	543	303	418	92	0	410	130	267	339
Portland	272	133	107	555	318	410	0	448	404	476
Scranton	179	581	341	288	130	130	448	0	137	209
Syracuse	132	537	297	151	267	267	404	137	0	72
Watertown	204	609	369	223	339	339	476	209	72	0

Figure 1. Intercity distance matrix.

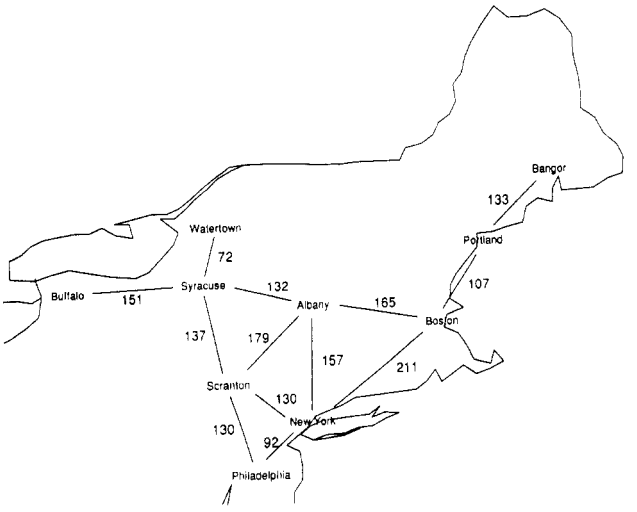

Figure 2. Intercity distance map.

matrix of molecular similarities from which that graph is derived.

The structure map seems to be an efficient way to visualize the increasing amount of information contained in chemical activity databases. The structure map of a set of compounds in conjunction with a given activity of these compounds constitutes a "map" of the chemical structure–activity space. The collection of these structure–activity maps for an entire database constitute an "atlas" for the chemist interested in chemical structure–activity relationships. The task of predicting properties of new compounds is facilitated with a suitable set of structure–activity maps. The activity may be predicted either directly from the structure–activity map, using local averaging methods such as nearest-neighbor predictors, or by using classical QSAR methods employing descriptors suggested by visual inspection of the structure–activity map.

In this study a structure–activity map will be developed in the area of xenobiotic metabolism. The activity which will be examined is a chemical reactivity, the relative occurrence of N-demethylation and N-oxidation biotransformations. Such maps should prove useful in the design of new drugs. The predictive success depends on the underlying structure–activity similarity principle which, in this case, states that compounds with structurally similar reaction sites generally have similar reactivities. It should be apparent from the structure–activity map which relatively small structural changes are responsible for dramatic changes in reactivity.

III. DATA SET SELECTION

The metabolic predictive problem is difficult for many reasons. Suppose two compounds can undergo the same metabolic transformation under the same linear kinetics rate

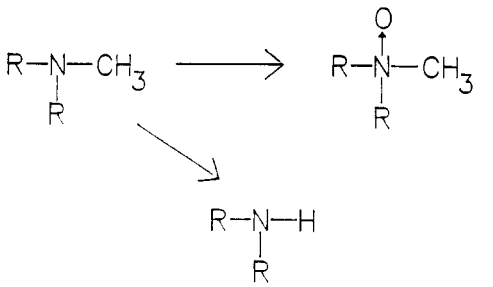

Figure 3. N-Demethylation and N-oxidation reactions.

constant. If these compounds have different elimination rate constants (again under linear kinetics), the amount of metabolite formed in the two cases can differ dramatically. Thus, any method of predicting absolute metabolic rates or the occurrence or nonoccurrence of a particular metabolic transformation is bound to be unreliable unless the elimination rate of the intact parent compound has been appropriately taken into account.

Under linear kinetics, the rate at which the intact parent compound has been eliminated can be adjusted out by studying the relative occurrence of two metabolic reactions rather than the absolute occurrence of one metabolic reaction. One can show that, under linear kinetics, the *ratio* of the amounts formed of two direct metabolites of a parent compound is independent of the elimination rate of the parent compound.

For this study, we have examined the relative occurrence of N-demethylation and N-oxidation of *N*-methyl tertiary amines (see Figure 3). We have chosen these two competitive reactions because they involve the simplifying feature of having identical reaction sites and because they occur frequently in drug design problems.

The compounds in the study were obtained from Volumes 1 and 2 of D. R. Hawkins' *Biotransformations*.[4,5] This is a survey of recently documented biotransformations in animals. The N-demethylation/N-oxidation data set is a subset of the biotransformation pathways contained in these volumes. The criteria for a parent compound's inclusion in the N-demethylation/N-oxidation data set are as follows:

1. The parent compound must be N-demethylated or N-oxidized at the nitrogen of an *N*-methyl tertiary amine.
2. The study referenced must have been conducted in one of the following four test systems: human in vivo, human microsomes, rat in vivo, or rat microsomes.

The first criterion was set to investigate the prediction of relative occurrences of these biotransformations. This criterion enables one to study the relative occurrence of a pair of reactions, but not the absolute occurrence of a single reaction. Primary and secondary amines are excluded from the analysis since N-oxide products are not observed for these substrates.

The second criterion was set to investigate the correlation between animal and human models from the microsomal level to the in vivo level. Both of the above criteria must be met for a compound to be included in our data set. The biotransformation pathways selected for inclusion to our data set are then further analyzed to determine the methods employed to isolate and/or characterize metabolic products. Each observation in our data set corresponds to an *N*-methyl reaction site in a documented metabolic transformation pathway study,[4,5] *not* necessarily a unique compound.

IV. CONSTRUCTION OF THE STRUCTURE–REACTIVITY MAP

A. Reaction Site Representations. A structure–reactivity map may be constructed once a representation for a chemical

structure or, as in this study, a chemical transformation site has been chosen. The choice of chemical structure representations is dependent on several factors including, but not limited to, the ability of the representation to be easily manipulated by a computer, the ease with which a chemical structure can be converted to a representation and then back to a chemical structure, and finally a chemical rationale for choosing the important structural features to include in the representation. Each *N*-methyl reaction site in the data set is converted to a chemical reaction site representation. Thus, a molecule may have multiple reaction site representations associated with it; one for each *N*-methyl site which is N-demethylated and/or N-oxidized. Both the N-demethylation and N-oxidation reactions may be represented as having the same reaction center, namely, the *N*-methyl nitrogen atom. This representation describes the overall chemical structural change, not necessarily those atoms involved in an actual reaction mechanism. Our N-dealkylation/N-oxidation reaction site representations include the reaction site (the methyl substituted nitrogen atom), surrounding atoms, and a steric bulk descriptor for the remainder of the molecule. This type of representation follows from our previous work in this area[16,17] and is similar to the "focus" concept in the DARC system.[18]

Specifically, the reaction site of a molecule is described by an *n*-level representation. This type of representation is defined by labeling the reaction site of a molecule. The atoms are numbered according to their bond distance away from the reaction center (see Figure 4). Atoms and bonds with a distance greater than *n* are then replaced with graph-theoretic steric bulk descriptors. Although the 1-level and 2-level reaction site representations were investigated in preliminary studies, the reaction site representation chosen for this analysis is a 3-level reaction site representation. It has been observed that a 1-level reaction site representation is useful for differentiating primary, secondary, and tertiary amines. But since our data set contains only tertiary amines, separation between reaction sites is achievable only by the differences in bulk descriptors. The 2-level reaction site representation contains, in addition to all of the information contained in the 1-level reaction site representation, information about the hybridization of the atoms α to the reaction center along with the atom type of the β atoms. Although the 2-level reaction site representation yields interesting results in the analysis of the N-demethylation/N-oxidation data set, more satisfying results were obtained using the 3-level reaction site representation. The 3-level representation contains the reaction center (the methyl substituted nitrogen atom) and all nonhydrogen atoms connected by a bond path distance of less than or equal to three bonds. This 3-level representation contains electronic information about the reaction site encoded in the atom labels and bond types.

These 3-level representations are augmented with a bulk descriptor to describe the size of that portion of the molecule which is not directly included in the 3-level representation. A series of "bullet" atoms (solid circles) was added to each of the terminal atoms of the representation. The number of bullet atoms indicates the number of atoms in the chemical structure which are not included in the 3-level representation. A range of atoms was defined as follows (see Figure 4): If between 1 and 3 atoms are deleted, a single bullet atom is added. If between 4 and 9 atoms are deleted, two bullet atoms are added. If between 10 and 27 atoms are deleted, three bullet atoms are added. And if over 27 atoms are deleted, four bullet atoms are added. The bullet atoms are a crude approximation of the position and steric bulk of that portion of the molecule not explicitly contained in the *n*-level representation. The reaction sites on noncyclic molecules can be effectively described in this straight forward manner. An example is provided in Figure 4.

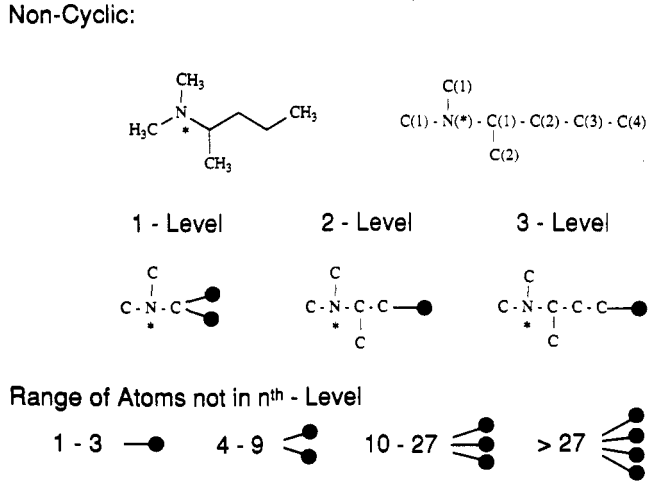

Figure 4. Reaction site is identified and labeled "*". Each atom is then labeled based on its bond distance from this reaction center. Each *n*-level is then constructed by deleting atoms with labels greater than *n*. The deleted atoms are then represented by bullet atoms (solid circles). The number of bullet atoms added is based on the number of deleted atoms and the above scale. If 1–3 atoms are deleted from the structure to obtain the *n*-level representation, then one bullet atom is added to the *n*-level atom of the representation. If 4–9 atoms are deleted from the structure to obtain the *n*-level representation, then two bullet atoms are added. If 10–27 atoms are deleted, then three bullet atoms are added, and if more than 27 atoms are deleted, then four bullet atoms are added. The bullet atoms are a crude approximation of the position and steric bulk of that portion of the molecule not explicitly contained in the *n*-level representation.

Reaction sites on cyclic molecules require a method for representing the positioning about the ring system of the low-level atoms and the steric bulk of the high-level atoms. The approach used is illustrated in Figure 5. The reaction site is determined, the *N*-methyl nitrogen atom, and the *n*-level representation is constructed with the following modifications. The bullet atoms are not added until the ring closure has been indicated. If the terminal atoms of the *n*-level representation are part of a ring system, then they are connected via bonds to a bullet atom; this maintains the cyclic nature of the graph. The ring closure is indicated by dashed bonds from the terminal atoms of the *n*-level representation to the ring closure bullet atom. The second step accounts for those atoms not included in the newly formed representation. The size and location of the remaining portion of the molecule is represented by bullet atoms. The remaining atoms are assigned to either a terminal atom or the ring closure bullet based on the shortest bond path distance. The range of atoms not in the *n*-level representation is shown with the same bullet atom scale previously described. If a terminal atom, at the *n*-level ring closure step, has two atoms as its nearest neighbors, then that terminal atom is bonded to a single bullet atom. Nearest-neighbor atoms are those atoms for which the bond path distance between the terminal atom and the neighbor atom(s) is a minimum. Each atom not explicitly contained in the *n*-level representation is assigned to the nearest terminal or ring closure bullet atom based on bond path distance. It is possible to describe where on the ring system the remaining portion of the molecule is located by assigning bullet atoms to terminal or ring closure bullet atoms.

An example of the construction of the *n*-level representations of *N*-methylpiperidine, assuming the nitrogen atom is the reaction site, is shown in Figure 5. Atoms which are equidistant from both terminal atoms are assigned to the ring closure

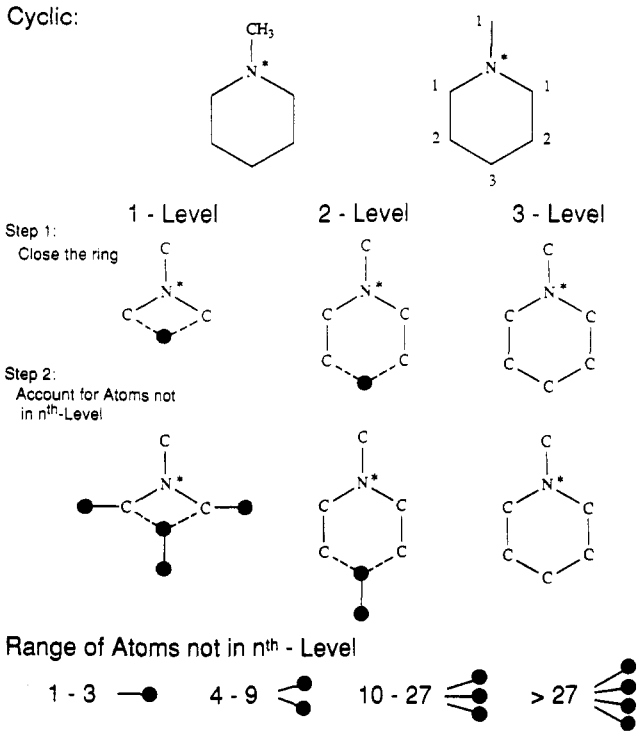

Figure 5. Step 1: Reaction center is identified and labeled "*". Each atom is then labeled based on its bond distance from the reaction center. Each *n*-level is then constructed by deleting atoms and bonds between atoms with labels greater than *n*. If the *n*-level atoms were part of a ring, they are connected by bonds to a ring closure bullet atom. Step 2: The deleted atoms are then represented by bullet atoms based on the previously defined scale. The arrangement of the bullet atoms depends on the location of the deleted atoms. Each deleted atom is assigned to its nearest-neighbor *n*-level atom. The sum of the deleted atoms assigned to each *n*-level determines the number of bullet atoms assigned. Atoms which are equidistant from each *n*-level atom are assigned to the ring closure bullet atom.

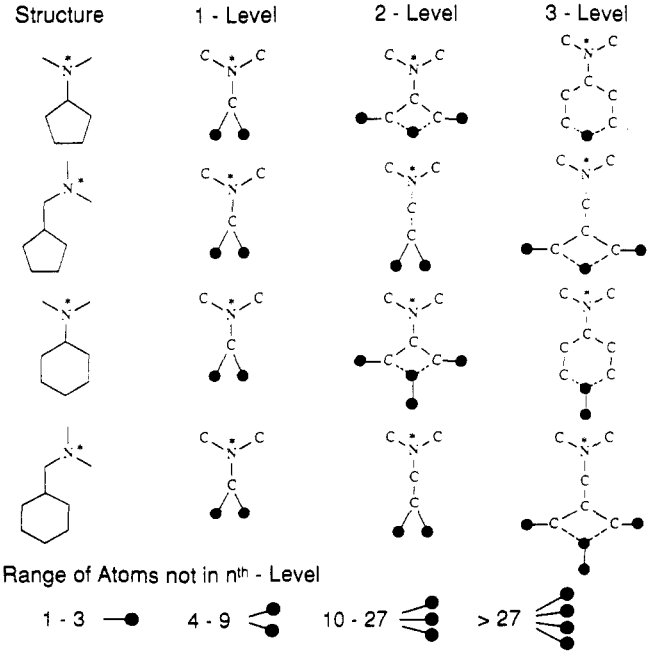

Figure 6. The *n*-level representations are constructed for these simple five- and six-member ring systems.

bullet atom. Figure 6 illustrates the possible *n*-level degeneracies obtained from different structures.

B. Computing Distances between Reaction Site Representations. The distances between each pair of reaction site representations of the data set were determined as follows. Each *N*-methyl reaction site was converted to it's 3-level representation. For example, for the nitrogen reaction site, the nitrogen is labeled N*, the α-carbons are labeled C_1, the β-carbons are labeled C_2, and the γ-carbons are labeled C_3, and so on. The representations of the reaction site retain these labels throughout the analysis. The number of nonsuperimposible bonds between each pair of representations is computed. This number is called the edge deletion distance as it is the smallest number of bonds that can be deleted so as to obtain a common substructure of the two representations. During this superpositioning, atom and bond labels are preserved. A C_1 will only be matched with another C_1. A double bond will only be matched with another double bond.

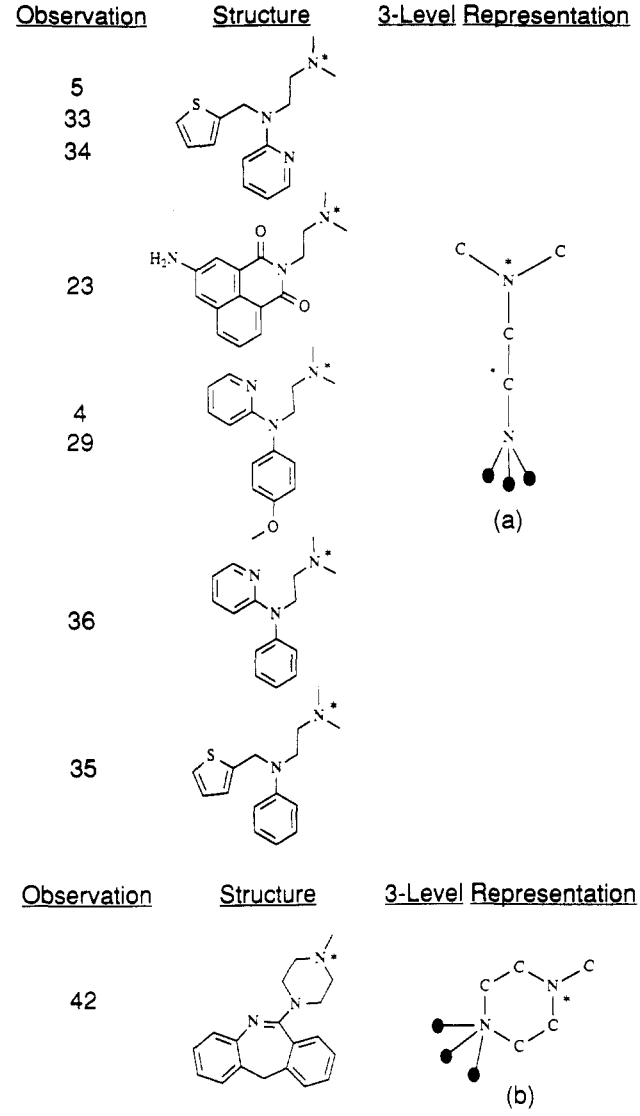

Figure 7. *N*-Methyl reaction sites for reaction sites 4, 5, 23, 29, 33, 34, 35, and 36 are all represented by the noncyclic 3-level representation (a) shown in the figure. The reaction site number indicates a documented N-demethylation/N-oxidation biotransformation site, *not* necessarily a unique chemical structure. The *N*-methyl reaction site for each of these observations is represented by the same noncyclic 3-level descriptor. *N*-methyl reaction site for reaction site 42 is represented by the cyclic 3-level representation (b) shown in the figure.

Graphs (molecular structures or 3-level representations) with an edge deletion distance of zero are identical, while an increase in edge deletion distance indicates a decrease in similarity between the graphs. For example, reaction sites 4, 5, 23, 29, 33, 34, 35, and 36 are all represented by the same 3-level representation (a) as shown in Figure 7a. Their edge deletion distance from each other is zero. The representation for reaction site 42 (b) is shown in Figure 7b. The distance between these two representations is determined by counting

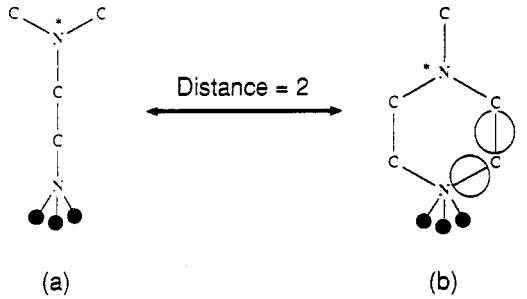

Figure 8. Edge deletion distance (equal to 2) between the 3-level representations (a) and (b) from Figure 7 is determined from the number of non-superimposible bonds between the two representations. The circled bonds in (b) of this figure indicate these non-superimposible bonds.

the number of non-superimposible bonds between them (the circled bonds in Figure 8).

C. Mapping the Space of Reaction Sites. The relationships between 3-level representations can be represented by a metagraph in which the vertices are the 3-level representations and the labeled edges indicate the edge deletion distances between representations. The complete graph quickly becomes very complicated. The number of edges of a complete graph is $m(m-1)/2$, where m is the number of vertices. We are primarily interested in edges which represent the more interesting structural changes as defined by the nearest-neighbor proximity graph. The proximity graph is obtained using the triangle inequality. Each triangle, a complete graph with three vertices, is selected from the complete proximity graph. Each edge is labeled with the edge deletion distance between the two adjacent vertices. An edge is marked if the other two edge labels are of lower value. After all triangles have been examined, all edges receiving a mark are deleted (see Figure 9). The remaining edges largely represent those smaller structural differences into which the larger structural changes can be decomposed. It is this decomposition of complex structural changes into simpler structural changes that makes the edges of the proximity graph interesting.

The resulting proximity graph may be drawn in any arrangement as long as the topology (connections between vertices) remains the same. Since the proximity graph is a topological representation, it is important that the edges of

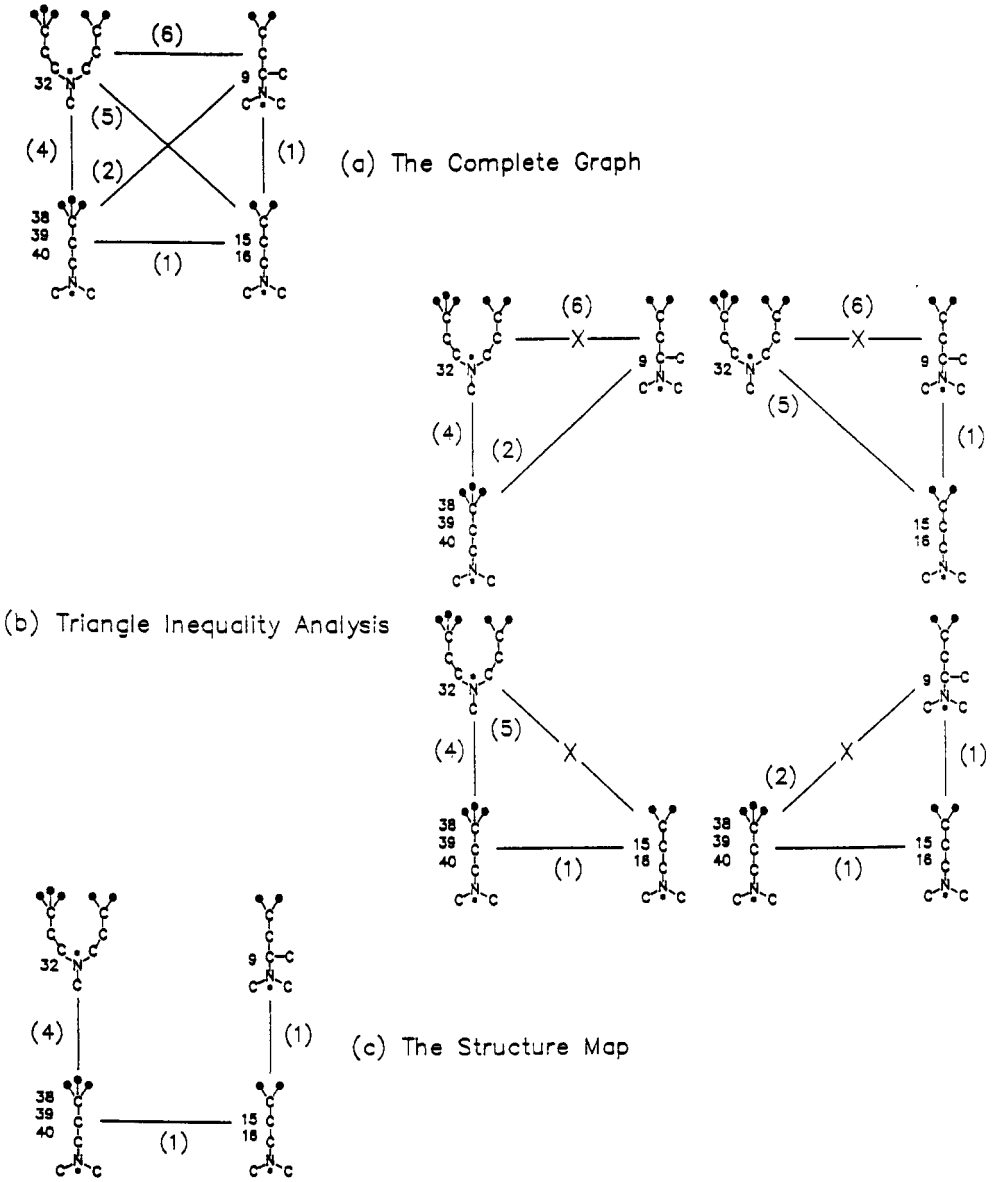

Figure 9. This is an illustration of the construction of a structure map for a portion of the data set based on the method described in the text. The complete graph is obtained for the data set (a). Each vertex is a 3-level representation, and each edge is labeled with the edge deletion distance between vertices. The complete graph is then analyzed using the triangle inequality (b). Each triangle of the complete graph is examined, and the longest edge of the triangle is deleted. The deleted edges are indicated by an "X". Once an edge is deleted, that edge is removed from the complete graph. The remaining edges are then redrawn in the form of the structure map (c). The resulting structure map links the 3-level representations (vertices) by the minimum edge deletion distances of the data set.

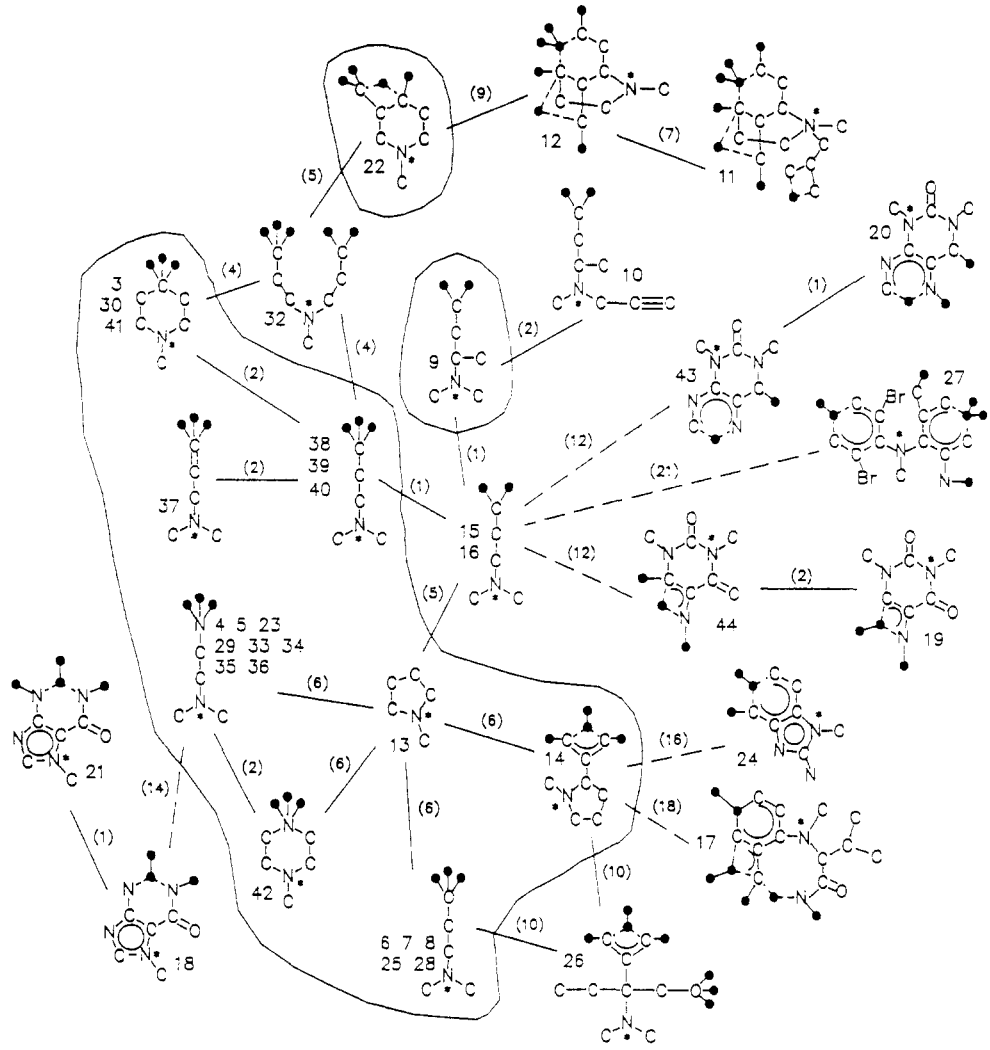

Figure 10. Structure map is constructed based on the 3-level representations of the N-demethylation/N-oxidation data set. Solid edges indicate relationships between representations linked by a edge deletion distance less than or equal to 10. Dashed edges indicate the link between representations differing by a edge deletion distance greater than 10. The reaction center of each 3-level representation is labeled with an "*". The edge deletion distance between each vertex is shown as the edge label.

the graph be easily recognized. In order to clearly see the edges of the graph, edge crossings should be kept to a minimum. An aesthetically pleasing, e.g., a minimum of edge crossings, arrangement of the proximity graph in this study (Figure 10) was obtained using an interactive computer graphics program. One can use multidimensional scaling methods for arranging a proximity graph on a plane,[19] or one can order the vertices according to property of their corresponding graphs. However, the purposes of those methods often conflict with the aesthetic goal of minimizing the number of edge crossings.

The edge deletion distance between each vertex is shown as the edge label. The dashed lines indicate connections in which the edge deletion distance between representations is greater than 10. A small edge deletion distance indicates similar representations. A large edge deletion distance arises from a lack of intervening representations between two representations.

A usefully arranged proximity graph with its vertices replaced by the corresponding structural representations is called a structure map or, more specifically for this case, a 3-level site map. A structure map for the N-demethylation/N-oxidation data set is shown in Figure 10. The closed curves represent an aspect of relative reactivity that will be discussed later. Each vertex of the proximity graph in this example is a 3-level representation.

Since reaction sites on different molecules may have the same 3-level representation, each vertex may contain information about many compounds. Figure 11 is obtained by replacing the 3-level site representations in Figure 10 with the set of reaction sites having those respective representations. It will be shown in the next section that the structure map, constructed using the 3-level reaction site representations, effectively clusters the N-oxidation reaction sites. These N-oxidation reaction sites are found within the closed curves in Figure 10.

D. Coloring a Reaction-Site Map. Just as the region surrounding a city in an intercity distance map might be colored so as to indicate the altitude at that city, we can color a structure map so as to indicate the value of some property associated with the structures. In general, there are many properties that might be associated with the structures giving rise to many different colorings of a structure map. We can combine these different colorings using one of the many methods available for representing multivariate data.[20]

We modify the usual presentation of "snowflake" representations of multivariate data[21] in two regards. We replace the snowflake by a regular polygon in which each sector represents one property or unit of the multivariate data. The shading of the sector represents the value of the corresponding property. For example, in Figure 12, the shading of the upper-right sector of a compound's hexagon indicates whether a value was available for that compound in an oral rat study, and if so, whether N-demethylation, N-oxidation, or both were observed. Figure 10, together with Figure 12, defines

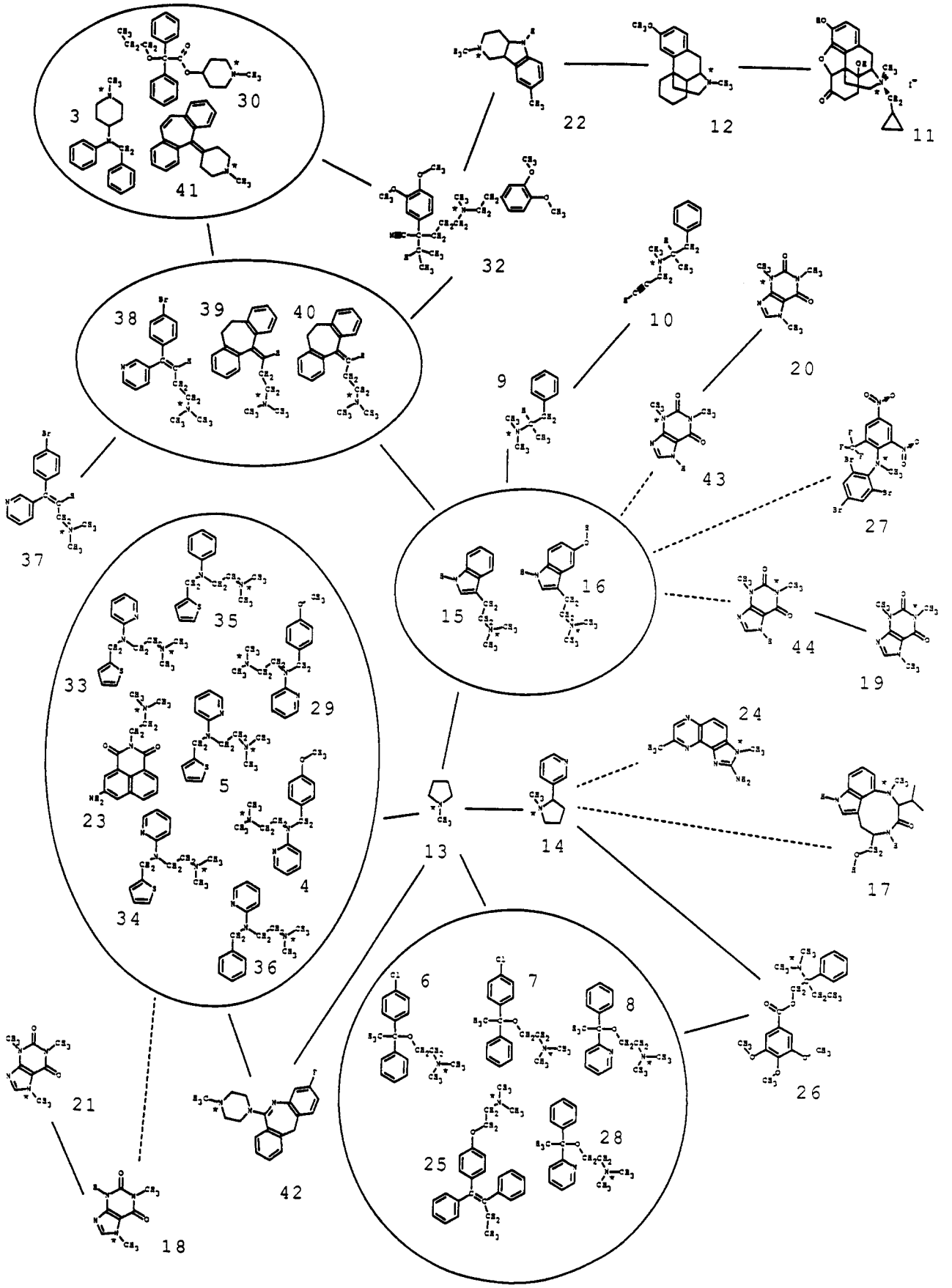

Figure 11. Structure map arranged as in Figure 10, but now displaying the actual structures included in the data set. The reaction sites are labeled with a "*".

an equivalence structure–activity map or, more specifically, an equivalence 3-level site-reactivity map.

Figure 12 is colored to indicate the observed reaction for each observation. It has been observed that there is a high proportion of N-oxidation reaction sites in some equivalence classes and not in others. This observation is exemplified by the cluster of reaction sites 4, 5, 23, 29, 33, 34, 35, and 36 (see Figure 10), all of which are N-oxidized. Moreover, visual inspection strongly suggests that similar reaction site representations have similar reaction products.

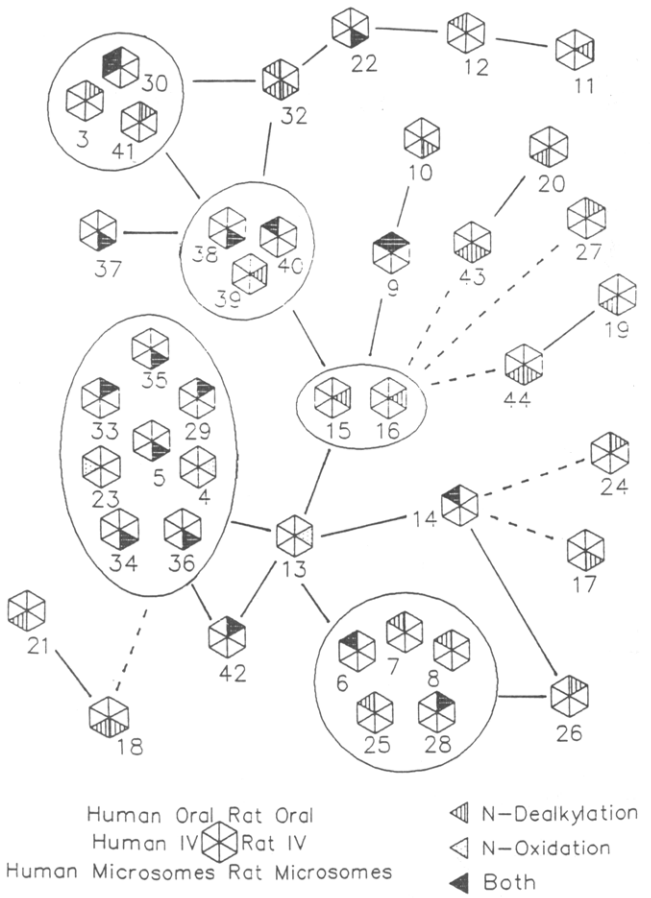

Figure 12. Test system structure–reactivity map for the N-demethylation/N-oxidation data set. Each sector of the hexagon vertex of the structure–reactivity map is shaded to indicate which test system was employed for each reaction site in the data set.

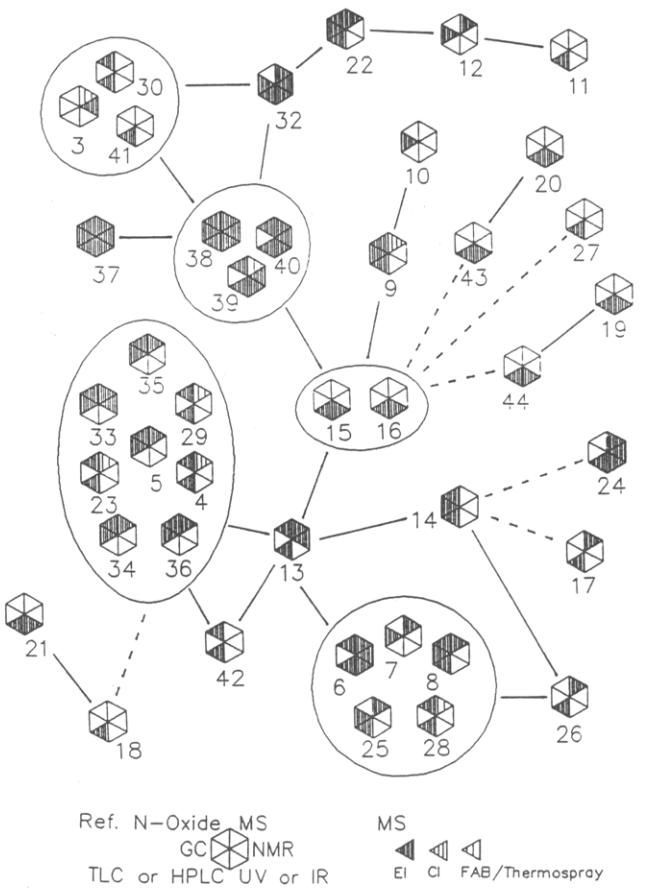

Figure 13. Experimental method structure–reactivity map for the N-demethylation/N-oxidation data set. Each section of the hexagon vertex of the proximity graph is shaded according to the experimental techniques used to determine the metabolic products for each observation in the data set.

The 3-level site representations associated with the relative occurrence of N-oxidation are circled in Figure 10. Edges which cross these curves, for example, the edge between site 22 and site 12 in the top center portion of the structure map, suggest that N-oxidation is less likely to occur relative to N-demethylation in sterically hindered sites. This suggestion is supported by a large majority of other edges that cross the closed curves. This information may be incorporated into more quantitative models of metabolism.

Figure 12 also provides information about which animal test system was used in the documented study. For example, it was noted that both N-oxidation and N-demethylation are observed in a rat and a human in vivo study for reaction site 9 (roughly the center of Figure 12). Comparisons between test systems and the observed products are possible using such colorings. In this particular instance, the structure–reactivity map can only show us that there are too few data points to support any definitive conclusions regarding the relationships between animal models and human metabolism.

The mode of administration of a xenobiotic may also influence the metabolic fate of that compound. This attribute of the documented metabolic pathway study is also easily included in the reactivity map. It has been observed that only *N*-oxide products (or no N-demethylation products) are formed in those studies in which the compound was administered intravenously (see reaction sites 4, 13, and 23).

The structure map is also very useful for summarizing and displaying the multiplicity of analytical methods used in each study. The structure map in Figure 13 has been colored to indicate the analytical methods employed in determining the N-demethylation/N-oxidation products in our data set.

Five techniques of metabolic product determination were commonly employed: mass spectrometry (MS); nuclear magnetic resonance (NMR); Ultraviolet (UV) or infrared spectroscopy (IR); thin-layer (TLC) or high-performance liquid chromatography (HPLC); and gas chromatography (GC). It was also of interest as to whether the *N*-oxide reference compound was synthesized. Since these six parameters were relevant to the analysis, a hexagon was chosen as the label for each observation.

Each sector of the polygon corresponds to one of the parameters. If a technique was employed in the documented observation, then the corresponding sector of the polygon is shaded. The shading may be a simple yes or no, shaded or not shaded, as in the NMR/ UV/IR, TLC/HPLC, GC, or reference *N*-oxide sectors. The sector may also contain more detailed information in the form of a gradient of shading, indicating variations of the same technique as illustrated in the MS sector of the vertex labels of Figure 13. The shading in this block is used to distinguish between electron impact (EI), chemical ionization (CI), and fast ion bombardment (FAB) mass spectroscopy. In Figure 13, a completely filled hexagon indicates a study that included all of the indicated techniques (e.g., observation 38), while a hexagon such as observation 18 indicates a study in which only one technique (in this case, liquid chromatography) was employed for metabolite characterization.

The choice of analytical methods used to characterize metabolites depends on the metabolite(s) which interest an investigator. If several techniques were employed in a study, one would expect that the chances of observing even minor metabolites should increase. If a study indicates that a

reference compound of a specific metabolite has been included in the study, one has greater confidence that the investigators were actually looking for that compound. Figure 13 indicates that in all but two cases (sites 8 and 39) an *N*-oxide product was observed when the *N*-oxide reference compound was available in the study.

V. CONCLUSIONS

A new family of reaction site descriptors, the *n*-level reaction site representations, have been presented. The reaction site representations contain both electronic (atom labels and bond types) and steric (the bullet atoms) information. In this study, a 3-level reaction site representation effectively described the local environment of the N-demethylation/N-oxidation reaction site of tertiary *N*-methyl amines.

The structure maps constructed using a 3-level reaction site representation have shown very promising results in indicating structural features which may play a role in differentiating between the relative occurrence of the N-demethylation and N-oxidation of tertiary amines. In addition, the structure maps facilitate the analysis of documented metabolic pathway information by indicating in one drawing the values of several experimental parameters (i.e., test system, route of administration, characterization techniques, etc.). The structure map with its associated "colorings" has the ability to display complex data sets in a concise manner and is much easier to interpret than data presented in a tabular format.

The structure map's effectiveness in indicating clear structure–activity relationships is useful for suggesting what information to include in a database. In addition to the observed biotransformations, it may be useful to include information about the test system (animal model or human metabolism), the level of the system studied (i.e., microsomes, hepatocytes, in vivo), and the experimental conditions employed for the analysis of metabolites in a metabolic pathway database. Whether or not reference compounds were used in the documented study is of interest for two reasons. First, it suggests the metabolic focus of the investigator, and second, there is an apparent correlation between the relative occurrence of N-oxidation and the use of a reference *N*-oxide in the experimental metabolism studies of our data.

An atlas of structure–reactivity maps constructed from a metabolic pathway database would provide a useful tool for researchers interested in predicting drug metabolism. By examining these structure maps it may be possible to determine which structural modifications are necessary to influence the metabolism of a new compound.

Although the structure maps presented in this paper have been used to provide a visualizable method for the analysis of the relative occurrence of N-demethylation and N-oxide formation in xenobiotic metabolism, the methodology presented clearly applies to structure–activity relationships in general.

ACKNOWLEDGMENT

We would acknowledge the support, encouragement, and helpful suggestions of Gerald Maggiora, Director of Computational Chemistry at Upjohn Laboratories. Jim Caputo, Animal Health Research/Performance Enhancement Research at Upjohn Laboratories, helped in clarifying the conditions under which a consideration of relative rates adjusts out the effect of the elimination rate on the amount of formed metabolites. Support for Eric Gifford as a summer employee at Upjohn Laboratories is also gratefully acknowledged. Finally, we would like to thank the reviewers for many stylistic improvements and thoughtful suggestions.

REFERENCES AND NOTES

(1) Barcza, S.; Kelly, L. A.; Lenz, C. D. Computerized Retrieval of Information on Biosynthesis and Metabolic Pathways. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 243–251.

(2) Darvas, F. Metabolexpert: An Expert System for Predicting the Metabolism of Substances. In *QSAR in Enwironmental Toxicology II*; Kaiser, K. L. E., Ed.; D. Reidel Co.: Dordrecht, 1987; pp 71–81.

(3) Darvas, F. Predicting Metabolic Pathways by Logic Programming. *J. Mol. Graphics* **1988**, *6*, 80–86.

(4) Hawkins, D. R. *Biotransformations: A survey of the biotransformations of drugs and chemicals in animals*; The Royal Society of Chemistry: Letchworth, Herts, U.K., 1989; Vol. 1.

(5) Hawkins, D. R. *Biotransformations: A survey of the biotransformations of drugs and chemicals in animals*; The Royal Society of Chemistry: Letchworth, Herts, U.K., 1990; Vol. 2.

(6) Spann, M. L.; Chou, K. C.; Wipke, W. T.; Ouchi, G. J. Use of Computerized Methods To Predict Metabolic Pathways and Metabolites. *J. Enviro. Pathol. Toxicol.* **1978**, *2*, 123–131.

(7) Wipke, W. T.; Ouchi, G. I.; Chou, J. T. Computer-Assisted Prediction of Metabolism. In *Structure–Activity Correlations as Predictive Tools in Toxicology*; Golberg, L., Ed.; Hemisphere: Washington DC, 1983; pp 151–166.

(8) Varkony, T. H.; Smith, D. H.; Djerassi, C. Computer-Assisted Structure Manipulation: Studies in the Biosynthesis of Natural Products. *Tetrahedron* **1978**, *34*, 841–852.

(9) Tinker, J. F.; Gelernter, H. Computer Simulation of Metabolic Transformations. *J. Comput. Chem.* **1986**, *7*, 657–665.

(10) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley Interscience: New York, 1990.

(11) Johnson, M. A. A Review and Examination of the Mathematical Spaces Underlying Molecular Similarity Analysis. *J. Math. Chem.* **1989**, *3*, 117–145.

(12) Lefcovitch, L. P. A Nonparametric Method for Comparing Dissimilarity Matrices, A General Measure of Biogeographical Distance and Their Application. *Am. Nat.* **1984**, *123*, 484–499.

(13) Matula, D. W.; Sokal, R. R. Properties of Gabriel Graphs Relevant to Geographical Variation Research and the Clustering of Points in the Plane. *Geogr. Anal.* **1980**, *12*, 205–222.

(14) Toussaint, G. T. The Relative Neighborhood Graph of a Finite Planar Set. *Pattern Recog.* **1980**, *12*, 261–268.

(15) Johnson, M. A. Relating Metrics, Lines and Variables Defined on Graphs to Problems in Medicinal Chemistry. In *Graph Theory and Its Applications to Algorithms and Computer Science*; Alavi, Y., Ed.; John Wiley: New York, 1985; pp 457–469.

(16) Johnson, M. A.; Gifford, E.; Tsai, C.-c. Similarity Concepts in Modeling Chemical Transformation Pathways. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley Interscience: New York, 1990; pp 289–319.

(17) Gifford, E.; Johnson, M. A.; Tsai, C.-c. A Graph-Theoretic Approach to Modeling Metabolic Pathways. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 303–322.

(18) Dubois, J. E.; Mercier, C.; Panaye, G. DARC Topological System and Computer Aided Design. *Acta Pharm. Jugosl.* **1986**, *36*, 135–169.

(19) Kruskal, J. The Relationship Between Multidimensional Scaling and Clustering. In *Classification and Clustering*; Van Ryzin, J., Ed.; Academic Press: New York, 1977; pp 17–44.

(20) McDonald, G. C.; Ayers, J. A. In *Graphical Representation of Multivariate Data*; Wang, P. C. C., Ed.; Academic: New York, 1978; pp 183–197.

(21) *Encyclopedia of Statistical Sciences*; Kotz, S., Johnson, N. L., Eds.; Wiley Interscience: New York, 1981; Vol. 8, pp 533–535.