Pharmaceutical Research, Vol. 5, No. 4, 1988

Research Article

Quantitative Structure–Activity Study on Human Pharmacokinetic Parameters of Benzodiazepines Using the Graph Theoretical Approach

Rodney S. Markin,[1] Wallace J. Murray,[2,3] and Harold Boxenbaum[4,5]

Received July 8, 1987; accepted October 19, 1987

The graph theoretical indices for a series of 13 benzodiazepines were calculated using a graph-path topological method. The total molecule, the ring fragments, and combinations of ring fragments were subjected to a quantitative structure–activity analysis using eight pharmacokinetic parameters. The metabolic clearance and the blood-to-plasma concentration ratios were most highly correlated with the graph theoretical indices, with R values of 0.975 and 0.938, respectively. These correlations were found when the diazepine + benzo fragment and phenyl fragment were used to calculate the graph-path indices. Terminal disposition half-life was correlated with the benzo + diazepine fragment, with $R = 0.969$. Truncating the graph-path codes by eliminating cycles in the total molecule markedly improved the correlation coefficients. When compared to the graph-path indices for the total molecule, the correlation coefficients for the terminal disposition half-life and metabolic clearance data rose from 0.721 to 0.935 and from 0.770 to 0.968, respectively, using the graph-path indices of the truncated molecule. Intrinsic clearance of unbound drug also was poorly correlated with the total molecule ($r < 0.7$) but rose significantly using the graph-path indices of the truncated moleucle ($r = 0.971$ and 0.975 for the well-stirred and parallel-tube models, respectively.)

KEY WORDS: quantitative structure–activity analysis (QSAR); topological approach; benzodiazepines; pharmacokinetics.

INTRODUCTION

Molecular topology has been defined as the mathematical description of molecular structure (1). Various methods have been devised that reduce molecular structure to a number—a topological index (2–4). In general, these methods apply concepts derived from graph theory in which constituent atoms of a molecule are represented as the vertices or nodes of a graph and bonds between atoms are the edges. Thus, the "graph" of a molecule represents the molecular skeleton of the compound. An adjacency matrix is constructed from the graph, and usually, the various topological indices are calculated from this matrix. The differences in indices arise from the mathematical formulations and assumptions of the particular method. However, what these methods have in common is that the calculated numbers result from knowledge of the molecular structure itself and are not dependent on laboratory measurement for their derivation.

When searching for a quantitative structure–activity relationship (QSAR), normally a physicochemical property such as the partition coefficient (P) is used as the structural descriptor. In the case of the partition coefficient, two parameters may be used: the log P value of the total molecule or the substituent constant, π. The latter parameter is derived with an assumption of additivity of functional groups to the total molecular property (5). The mathematical relationship between the property (log P or π) and the biological effect is empirically derived by searching for the best statistical "fit" using linear or nonlinear multiple regression analysis (6).

Topological indices offer us the same opportunity to search for a best fit in QSAR studies. The topological index for the total molecule can be calculated, and a correlation sought with biological activity. Analogous to the use of π, the molecule may also be divided into fragments and the topological indices for these fragments can be used in the correlations. If one compares the correlation coefficients obtained with the total molecule versus that found with each fragment, it may be determined whether only part of the molecule correlates with a property or activity.

The graph-path topological method developed by Randic and Wilkins (7) was used in this study. *Paths,* which are defined as sequences of adjacent bonds (edges) that connect atoms (vertices or nodes), with no atom appearing more

[1] Department of Pathology and Microbiology, College of Medicine, University of Nebraska Medical Center, Omaha, Nebraska 68105-1065.

[2] Department of Pharmaceutical Sciences, College of Pharmacy, University of Nebraska Medical Center, Omaha, Nebraska 68105-1065.

[3] To whom correspondence should be addressed.

[4] Pharmaceutics Section, University of Connecticut, Storrs, Connecticut 06268.

[5] Present address: Drug Metabolism Department, Merrell Dow Research Institute, 2110 East Galbraith Road, Cincinnati, Ohio 45215.

0724-8741/88/0400-0201$06.00/0

than once in a sequence, are counted. These enumerations are defined as *path numbers*. To compare differences in molecular structure quantitatively, Randic and Wilkins considered the path numbers of a molecule as coordinates in an *n*-dimensional, Euclidean space. Each molecule, therefore, assumes a position in that space, and similarity in structure can be estimated by calculating the Euclidean distance between molecules as defined by their path numbers. Molecules that are similar in structure will be "closer together" in that space, while dissimilar molecules will be further apart. Wilkins *et al.* (8) describe the use of this method in a QSAR study of aminotetralines.

A variety of pharmacokinetic parameters has been determined for the benzodiazepines (9), but as yet, no QSAR analysis has been conducted. Our approach in examining the QSAR of the pharmacokinetic parameters has been to ask the question, Is the topology of one portion of the molecule more useful than that of others in predicting a given parameter? In answering this question we dissected the benzodiazepine molecules into three areas: the diazepine nucleus, the benzo ring, and the phenyl ring (Fig. 1). This approach is similar to the approach used previously by Hall and Kier (10).

The graph theoretical indices were calculated for each fragment, and correlation was attempted using multiple regression analysis. Combinations of fragments were also used, namely, diazepine + benzo, diazepine + phenyl, and benzo + phenyl, as well as the total molecule. In addition, the diazepine ring was truncated to reduce the number of cycles (Fig. 2). Correlation of the graph paths of the total molecule and fragments with the pharmacokinetic properties was sought in an attempt to arrive at deductions regarding the role of the various fragments in characterizing pharmacokinetic behavior.

The pharmacokinetic parameters used in these QSAR studies were all obtained from human data (9):

(1) terminal disposition half-life, $t_{\frac{1}{2},z}$;
(2) unbound or free fraction in plasma, fu;
(3) unbound fraction in blood, fu_b;
(4) blood-to-plasma concentration ratio, λ;
(5) terminal exponential volume of distribution (referenced to blood), V_b;
(6) modified unbound volume of distribution, V_T/fu_T;
(7) metabolic clearance (referenced to blood), $CL_{m,b}$ (metabolism accounts for virtually all elimination); and
(8) intrinsic clearance of unbound drug, CLu_{int}, calculated (11–13) by assuming applicability of either the well-stirred ($CLu_{int,ws}$) or the parallel-tube ($CLu_{int,pt}$) model for hepatic disposition.

One major difficulty in analyzing most pharmacokinetic data is the hybridized nature of parameters. For example, the metabolic clearance ($CL_{m,b}$, referenced to blood) is dependent on the intrinsic clearance of unbound drug (CLu_{int}), unbound fraction of drug in plasma (fu), and blood-to-plasma concentration ratio (λ). The volume of distribution (V_b, referenced to blood) is dependent on the unbound drug fractions in blood (fu_b) and tissues. Since the terminal disposition half-life ($t_{\frac{1}{2},z}$) is dependent on both $CL_{m,b}$ and the volume of distribution, $t_{\frac{1}{2},z}$ will consequently be dependent on all the aforementioned parameters. On first consideration, therefore, we expect the regression for $t_{\frac{1}{2},z}$ to show dependence on all topological indices used as predictors of the more primary parameters. However, this is clearly not the case. Whereas $t_{\frac{1}{2},z}$ correlates best with the diazepine + benzo graph-path indices, fu correlates best with only the phenyl moiety path index (*vide infra*). Since $t_{\frac{1}{2},z}$ is biologically dependent on fu, why should $t_{\frac{1}{2},z}$ not also require the phenyl moiety as a predictor? The answer probably relates to the phenomenon of "collapsing." In the complex array of parameter interrelationships, the indices for the phenyl ring probably cancel each other out (in a mathematical sense). We know, for example, that enhanced values of fu increase both the volume of distribution and the metabolic clearance (approximately equal to the total clearance for these benzodiazepines). Since half-life is determined by the ratio of

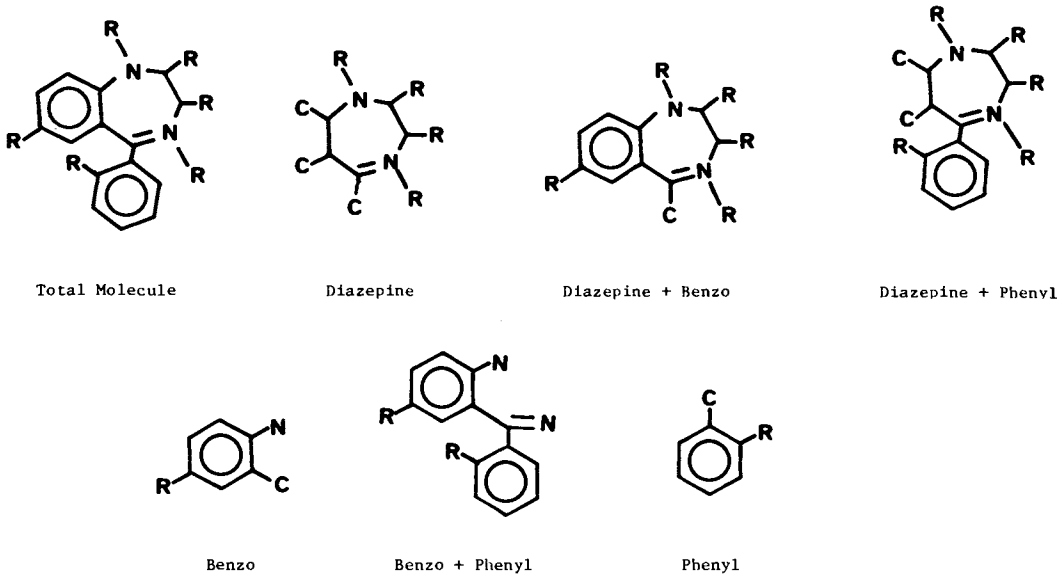

Fig. 1. Molecular graphs of benzodiazepines and fragments used to calculate the graph theoretical indices.

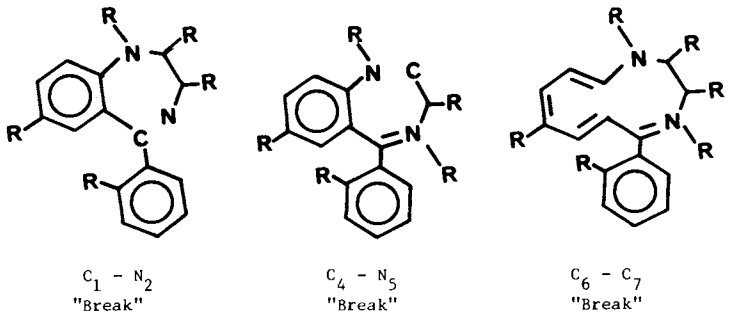

Fig. 2. Truncated graphs of benzodiazepines showing various bond breaks used in the calculation of the graph theoretical indices.

volume of distribution to clearance, the influence of the phenyl ring may be nullified. This same generalized phenomenon could be occurring with all curve fits.

METHODS

The details of applying the graph-path theoretical approach to arrive at a numerical representation of molecular structure have been previously described (7,8,14–17). The method is based on a graph representation of molecular structures from which ordered sequences of numbers are derived. These values are derived from the graph paths. The graph-path numbers are calculated using the BASIC program of Randic *et al.* (7,17). Application of the method is illustrated in Table I with the isomers of hexane. Each carbon atom is represented as a node; bonds are represented as the lines connecting nodes. As with other topological approaches (4), a hydrogen suppressed graph is employed, i.e., hydrogen atoms are not included to simplify the calculations. *Paths* are sequences of adjacent bonds (lines) connecting the carbon atoms (nodes), with no atom appearing more than once in any given sequence. The sum of paths of length zero is simply the number of carbon atoms in the molecule; the sum of paths of length one corresponds to the number of C–C bonds. In both cases these are identical for all isomers. For paths of length one or greater, the sum of the path lengths must be divided by two, since each bond is counted twice (once from both ends of the graph). Moreover, the number of paths of different lengths may not be independent. For example, for noncyclic structures the total number of paths is $[1/2N(N + 1)]$, where N is the number of atoms.

When this method is applied to the series of hexane isomers, the result is the sum of the number of each path of a particular length (Table I). These sequences of sums, called the *path codes*, are the numerical representations of the graphs of the four hexane isomers; the differences in the codes can be related to differences in their structures. This is more evident with larger molecular structures. Figure 3 shows the molecular graphs and codes for larger, somewhat similar molecules. The molecular codes for the isomeric forms of dimethylcycloheptanes and trimethylcyclohexanes reflect the similarities and dissimilarities of these molecules. Analyzed qualitatively, the codes of cycloheptyl derivatives show a high frequency of code values 11 and 12, while the cyclohexyl series has the value 13 or higher appearing frequently. Since these molecules are isomers, they all have codes which begin identically with the number 9, 9. However, this observation of similarity must be converted to a *quantitative* and systematic algorithm if the method is to be of any use in QSAR.

A suitable *quantitative measure of similarity* is derived by considering the path codes as coordinates in an n-dimensional, Euclidean space, described earlier by Randic *et al.* (17). Each molecule occupies a point in that space, and the similarity between molecules can be calculated as a mathematical difference. Molecules which are similar in structure and which have similar molecular path codes will be very close in space; those that are dissimilar will be far apart. This simple approach can be modified to include various normalization procedures, e.g., division of path codes by the lowest common denominator, or truncations of codes (exclusion of graph paths of specified lengths, e.g., use only graph paths of distances one through six).

Modifications do not appear to disable the use of Euclidean distances to characterize similarity in molecular structure quantitatively (17). Equation 1 summarizes the relationship between the two hypothetical sequences a and b:

$$S^{-1} = D_{ab} = \sum_{i=0}^{N} [(a_i - b_i)^2]^{1/2} \qquad (1)$$

Singularity, i.e., $S^{-1} = \infty$ or $D_{ab} = 0$, would denote identical sequences. A further extension of this method, which was used in this study, incorporates the type of bond between two atoms, i.e., single, double, or fractional as in aromatic systems (17).

In order to use these data in a QSAR study, the compounds must be ordered and assigned a relative value. This is accomplished by using the method of partial ordering described by Randic and Wilkins (14). This results in an ordered series, with the first compound in the series designated the "reference" compound. The reference compound

Table I. Path Enumerations for Hexane Isomers

	Path length					
	0	1	2	3	4	5
N-Hexane	6	5	4	3	2	1
2-Methylhexane	6	5	5	4	1	
2,3-Dimethylhexane	6	5	6	4		
2,2-Dimethylhexane	6	5	7	3		

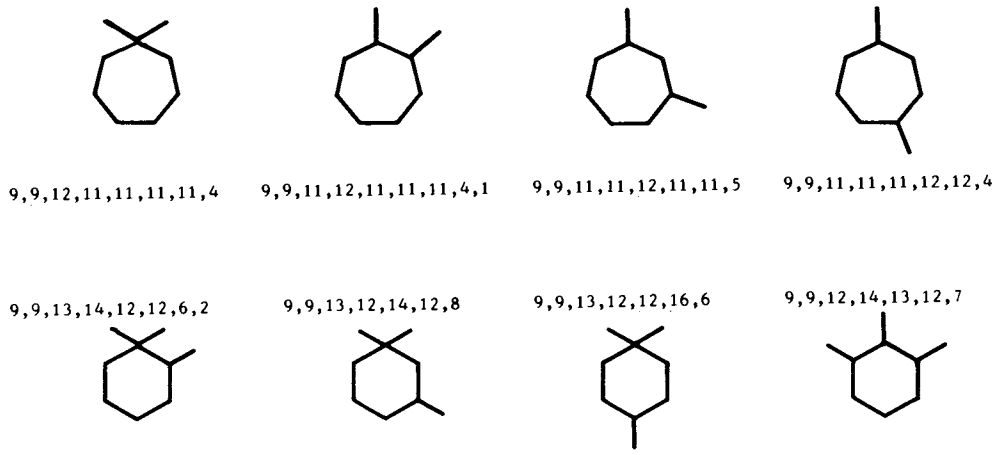

Fig. 3. Molecular graphs and path codes used to compare and differentiate isomeric cyclohexanes and cycloheptanes.

has (by definition) a value of zero for each of the coordinate axes, while the remaining compounds have values assigned by Eq. (1). From this rank order, the series of compounds is compared to the experimental data (which may be physical property values or biological activity data). In this study, a standard, nonweighted least-squares method was used.

The sources of the data and derivations of the pharmacokinetic parameter values have been described previously (9). CLu_{int} was selected as an especially appropriate pharmacokinetic parameter for these QSAR analyses because of its fundamental importance in determining unbound, pharmacologically active, steady-state blood levels for hepatically metabolized drugs orally administered. All the benzodiazepines investigated are virtually completely metabolized, presumedly by the liver. Additionally, CLu_{int} is independent of plasma protein binding, tissue distribution, and hepatic blood flow. The pharmacokinetic models, assumptions, and derivations have been discussed previously (9,11–13). CLu_{int} values were calculated assuming applicability of either the well-stirred or the parallel-tube model (9,11–13). Calculations based on each of these models produce CLu_{int} values that have a mean absolute difference of 3.36% (a difference considerably less than experimental error). The largest difference was for triazolam, 13.3%. Therefore, QSAR results based on either model would be virtually identical.

Individual values for CLu_{int} are actually equal to $\Sigma_{i=1}^{N} CLu_{int_i}$, where N is the number of metabolic pathways for intact drug. Consequently, CLu_{int} values quantitate overall drug metabolism rates and do not necessarily reflect the formation rate of any metabolite in particular.

The modified, unbound volume of distribution, V_T/fu_T, is equal to the actual physical volume of tissue into which drug distributes (V_T) divided by the weighted, mean unbound fraction of drug in those tissues (fu_T) (9). V_T usually equals the total-body water less the blood volume. Therefore, the quotient V_T/fu_T is a measure of tissue uptake and binding and increases as tissue binding increases. More importantly, V_T/fu_T is independent of binding of drug to blood components.

The unbound fraction of drug in plasma, fu, is dependent on the number of binding proteins, their concentrations, the number of binding sites per protein molecule, and the equilibrium association (affinity) constants. Taking protein concentrations to be constant in all binding experiments (only healthy subjects were used) and assuming binding of all benzodiazepines to be qualitatively similar, a decreased value for fu reflects a greater affinity and/or more protein binding sites.

RESULTS AND DISCUSSION

The pharmacokinetic parameters for 13 benzodiazepines are listed in Table II. These data were compared to the numerical values calculated for the benzodiazepines using the modified graph theoretical approach of Randic *et al.* (17). Figures 1 and 2 illustrate the structural characteristics of the benzodiazepine nucleus and the fragments which were used in the correlation studies (including the three "broken" structures). Several different equations were used to evaluate the relationship between the path numbers and the measured pharmacokinetic parameter. Table III summarizes the correlation coefficients, the reference compound, and the mathematical model which best fit the data.

The regressions for the graph theoretical indices for the total molecules result in correlation coefficients, R, of greater than 0.7 for only two parameters: (1) the metabolic clearance ($CL_{m,b}$) of the benzodiazepines resulted in an R value of 0.770 when fit to the curve $y = mx^3 + b$ using flunitrazepam as the reference compound; and (2) the terminal disposition half-life, $t_{1/2,z}$, is correlated with an R value of 0.721 [for the curve $y = m\,(1/x) + b$]. All other correlation coefficients using the graph path indices of the total molecule were less than 0.770.

Since the total molecule was not highly correlated to any of the selected pharmacokinetic parameters, the benzodiazepine structure was selectively partitioned into the structural subgroups shown in Fig. 1. When regressions of the pharmacokinetic parameters with the graph-path indices of the individual fragments were performed, it was found that certain fragments of the benzodiazepine molecule had better correlation coefficients than the total molecule. Using $CL_{m,b}$ as the dependent parameter, a correlation coefficient of 0.975 was obtained using the diazepine + benzo fragment; a correlation coefficient of 0.969 was found using the diazepine fragment alone. In both cases, the best mathematical model was $y = mx^3 + b$ (using as reference those com-

Table II. Benzodiazepine Pharmacokinetic Data (Mean Values) from Human Studies (9)[a]

ID No.	$t_{\frac{1}{2},z}$ (hr)	fu	fu_b	λ	V_b (liters/kg)	V_T/fu_T (liters/kg)	$CL_{m,b}$ (ml/min/kg)	$CLu_{int,pt}$ (ml/min/kg)	$CLu_{int,ws}$ (ml/min/kg)
1. Chlornordiazepam	79.5	0.0454	0.0811	0.560	2.75	33.0	0.434	5.40	5.45
2. Lorazepam	12.4	0.105	0.164	0.642	1.67	9.76	1.56	9.82	10.1
3. Triazolam	2.25	0.110	0.196	0.560	1.26	6.07	6.24	36.5	42.1
4. Chlordiazepoxide	9.38	0.0605	0.0913	0.663	0.466	4.34	0.564	6.25	6.32
5. Demoxypam	30.0	0.217	0.329	0.660	0.703	1.92	0.283	0.865	0.870
6. Bromazepam	12.0	0.477	0.714	0.668	0.900	1.16	0.770	1.10	1.11
7. Oxazepam	7.33	0.0427	0.0411	1.040	0.673	14.7	1.06	26.3	26.9
8. Nordiazepam	45.5	0.034	0.0607	0.560	0.807	12.1	0.205	3.39	3.40
9. Flunitrazepam	13.5	0.220	0.293	0.750	4.58	15.4	3.92	14.5	15.8
10. Diazepam	36.0	0.0134	0.0231	0.580	1.91	79.7	0.672	29.5	29.9
11. Medazepam	5.13	0.0070	0.0125	0.560	1.17	88.0	2.64	223.0	236.0
12. Clorazepam	35.8	0.134	0.239	0.560	5.48	22.6	1.64	7.09	7.33
13. Nitrazepam	21.7	0.137	0.221	0.620	2.92	12.9	1.64	7.67	7.93

[a] $t_{\frac{1}{2},z}$ is terminal exponential half-life, fu is unbound fraction in plasma, fu_b is unbound fraction in blood, λ is blood-to-plasma concentration ratio, V_b is volume of distribution during the terminal exponential phase (referenced to blood), (V_T/fu_T) is volume of tissue into which drug distributes divided by weighted, unbound fraction of drug in these tissues, $CL_{m,b}$ is metabolic clearance (referenced to blood), $CLu_{int,pt}$ is intrinsic clearance of unbound drug (parallel-tube model), and $CLu_{int,ws}$ is intrinsic clearance of unbound drug (well-stirred model).

pounds designated in Table III). The metabolic clearance also had a high correlation (R = 0.943) with the benzo + phenyl fragment.

The correlation coefficients are generally highest for metabolic clearance, especially when correlations are made with indices representing the diazepine fragment. Benzodiazepines are metabolized predominantly on the diazepine portion. This includes ring hydroxylation in the 3 position, N^1 dealkylation, and in the case of triazolam, hydroxylation of the methyl group attached in the triazolo ring. In this study, the triazolo ring was treated as being part of the diazepine fragment. Surprisingly, CL_{int}, which is a less hybridized index of drug metabolism, correlates best with the benzo + phenyl indices. As these moieties are generally not appreciably involved in drug metabolism in humans, this points to a lack of mechanistic correspondence between the QSAR model and benzodiazepine metabolism.

The half-lives of the benzodiazepines were also well correlated using the molecular fragment approach rather than the total molecule (Table III). The best correlation coefficient (R = 0.969) was observed when $t_{\frac{1}{2},z}$ was correlated with the diazepine + benzo moieties, using compound 5 as the reference and using the mathematical model, $y = m(1/x) + b$. In addition, the benzo + phenyl fragments gave a correlation coefficient of 0.881 using compound 2 as the reference with the relationship $1/y = mx + b$.

The blood-to-plasma concentration ratios correlate well with the phenyl moiety (R = 0.938) using compound 1 as the reference [with the mathematical relationship, $y = m(1/x) + b$]. The diazepine fragment has a lower correlation (R = 0.707). The fragments with the most lipophilic character (the benzo fragment, benzo + phenyl, and diazepine + benzo) resulted in R values of 0.819, 0.811, and 0.818, respectively. These data suggest that the diazepine ring is not the major benzodiazepine structural feature that correlates with blood-to-plasma partitioning.

Regression analyses were also performed with fu and fu_b parameter estimates. The best correlation was found between fu and the indices for the phenyl fragment (R = 0.777) and between fu_b and the diazepine + benzo fragments (R = 0.719). Surprisingly, the fu and fu_b individually have a correlation coefficient of <0.700 when correlated with the benzo + phenyl fragments together.

Data of Lucek and Coutinho (18) indicate that the extent of benzodiazepine plasma protein binding is related to the substitution pattern on the phenyl and benzo rings. The determining factor is attributed to *ortho* substitution on the phenyl ring, which apparently restricts the rotation of the phenyl moiety. Substitution on the phenyl ring at this position is hypothesized to cause the phenyl ring to lie perpendicular to the diazepine ring, and this rotational isomer appears to have a greater binding affinity. If Lucek and Coutinho are correct, then the findings that fu has a higher correlation with one fragment, the phenyl, rather than the combined benzo + phenyl fragments, points to a weakness mentioned earlier, i.e., statistical correlations do not necessarily imply corresponding mechanistic dependence.

The modified volume of distribution (V_T/fu_T) shows a significant correlation (R = 0.898) with the benzo fragment when compound 9 is the reference and the mathematical model is $y = mx^3 + b$. Noteworthy is the observation that plasma protein binding (characterized by fu) and tissue binding (characterized by V_T/fu_T) correlate best with different fragments.

Inherent in the calculation of graph-path indices of large polycyclic molecules are paths of extreme length. These long path lengths in large polycyclic molecules increase the dissimilarities, or Euclidean distances, between their path numbers. Since some of the endogenous compounds which bind to benzodiazepine receptors in the brain are di- and tripeptides (19), the similarity in structure between the benzodiazepines and selected di- tripeptides was examined in a preliminary study. It was found that elimination of one of the C–N or C–C bonds in the benzodiazepine ring produced structures with a marked similarity to the native peptides (20).

We wished to examine the effect on correlations if the molecules were mathematically truncated, reducing the

Table III. Correlation Coefficients: Pharmacokinetics Parameters vs Graph-Path Indices

	Total molecule	Diazepine moiety	Diazepine + benzo	Diazepine + phenyl	Benzo + phenyl	Benzo moiety	Phenyl moiety	Total molecule C_1–N_2 break[j]	Total molecule C_4–N_5 break[j]	Total molecule C_6–C_7 break[j]
$t_{V2,z}$	<0.721[a] (9)	0.832[a] (6)	0.969[a] (7)	0.842[a] (9)	0.881[b] (2)	<0.7[k]	<0.7[k]	0.932[b] (4)	0.935[b] (11)	0.928[b] (9)
fu	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	0.777[i] (8)	<0.7[k]	<0.7[k]	<0.7[k]
fu_b	<0.7[k]	<0.7[k]	0.719[e] (6)	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]
λ	<0.7[k]	0.707[d] (7)	0.818[e] (3)	0.715[i] (3)	0.811[i] (6)	0.819[i] (6)	0.938[a] (1)	0.732[i] (6)	<0.7[k]	<0.7[k]
V_b	<0.7[k]	<0.7[k]	0.767[c] (11)	<0.7[k]	0.793[h] (3)	<0.7[k]	<0.7[k]	0.805[a] (9)	0.867[h] (3)	0.730[h] (3)
V_T/fu_T	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]	0.898[c] (9)	<0.7[k]	<0.7[k]	<0.7[k]	<0.7[k]
$CL_{m,b}$	0.770[c] (9)	0.969[c] (9)	0.975[c] (1)	0.866[c] (9)	0.943[d] (10)	<0.7[k]	<0.7[k]	0.960[d] (2)	0.968[d] (9)	0.967[d] (9)
$CLu_{int,pt}$	<0.7[k]	0.706[a] (3)	0.828[a] (3)	0.850[d] (1)	0.894[f] (2)	<0.7[k]	<0.7[k]	0.975[g] (2)	<0.7[k]	0.720[d] (1)
$CLu_{int,ws}$	<0.7[k]	0.705[a] (3)	0.829[a] (3)	0.892[d] (11)	0.894[f] (2)	<0.7[k]	<0.7[k]	0.971[g] (2)	<0.7[k]	0.806[d] (11)

Note. The correlation coefficients were calculated using the number in parentheses as the reference compound as described under Methods. The superscript refers to the equation listed in the footnote below for which the data had the largest correlation coefficient. For the x value being the graph theory indices (independent variable) and y being the pharmacokinetics parameter (dependent variable), the notation x vs $1/y$ would yield the equation $1/y = mx + b$.

[a] $y = m(1/x) + b$.

[b] $1/y = mx + b$.

[c] $y = mx^3 + b$.

[d] $y^2 = mx + b$.

[e] $y = m\log x + b$.

[f] $\log y = mx + b$.

[g] $\log y = m(1/x) + b$.

[h] $y = m\log(1/x) + b$.

[i] $\log(1/y) = mx + b$.

[j] The notation "break" refers to the modification of the structure formed by the denoted bond as shown in Fig. 2.

[k] Correlation coefficients of less than 0.700 are not reported.

number of paths. In order to truncate the graph-path numbers in a uniform manner, three bonds in the diazepine molecule were selected to be structurally broken (Fig. 2). These bonds were chosen so that in the case of the C_1-N_2 and the C_4-N_5 breaks, the diazepine cycle would essentially become a linear chain with the benzo and phenyl functional groups on the chain. The C_6-C_7 break was utilized to eliminate the two cycles of the diazepine + benzo fragment and make it just one cycle. Since the maximal path length is usually less than or equal to the number of atoms in the molecule, truncation of the path numbers would not uniformly affect the graph theoretical indices. If these molecules were all isomers, i.e., had the same molecular formula as the hexanes given in the example under Methods, then truncation of the path codes would be a reasonable method for this approach.

As shown in Table III, breaking the diazepine ring of the total molecular graph causes the correlation coefficients to increase dramatically as compared to the intact graph of the total molecule. When comparing the $t_{V2,z}$ data using the C_1-N_2 break of the total molecular graph, there is a notable rise in the correlation coefficient, from 0.721 for the intact graph to 0.932 for the truncated diazepine nucleus. This result is somewhat surprising, and the immediate suggestion would be that the diazepine moiety does not actually contribute to the molecular representation or that the calculation is in error. However, when the graph theoretical indices for the diazepine moiety with the C_1-N_2 break are calculated and compared with $t_{V2,z}$, the correlation coefficient is not statistically significant ($R = 0.106$; not listed in Table III). This suggests that the increase in the correlation coefficient is a result of the path code truncation by means of reducing the graph-path cycles.

When comparing the previously described correlation coefficients for the intact total molecule with those of the total molecule with specific breaks, there is a significant increase in the correlation coefficients for every pharmacokinetic parameter listed in Table III. The values of R for the molecules with the C_4-N_5 break are all larger than those for the molecules with the C_6-N_2 breaks. However, the differences are slight with the exception of the terminal exponential (blood) volume of distribution (V_b). These variations in the total graph yield similar graph-path numbers. This is expected since the maximal path length is the same in both cases. The differences are probably due to the relative ring placement of the functional groups.

OVERVIEW

The method of fragmenting the benzodiazepine molecule into molecular subsets and using the graph theoretical indices as structural descriptors of these subsets to examine the QSAR of their pharmacokinetic properties (parameters) leads to useful information. One could, for example, readily generate (predict) an entire array of pharmacokinetic profiles for a multitude of benzodiazepine molecules. The same could be done for anticipated metabolites. Based on preliminary calculations, pharmacologists could select for biological testing (once synthesized) those compounds most likely to have the desired pharmacokinetic profile. Thus, for clinical situations requiring a short duration of activity (induction of anesthesia, anesthesia for tooth extraction, etc.), those compounds with predicted large CLu_{int} values and short elimination half-lives could be given priority in biological screening. The predictions could also be used to help focus chemists on the synthesis of analogues with desired pharmacokinetic properties.

Another intellectually gratifying feature of this study was the demonstration that graph theoretical indices may serve as good descriptors of molecular structure upon which to base correlations of pharmacokinetic properties. One does not need to measure a physical property such as a partition coefficient to perform a property–activity relationship.

The use of molecular fragments and their graph theoretical indices also allows us to search for those portions of the molecule which best predict (correlate with) a given biological property. It is not necessary to use a total physical property to attempt to find a relationship where none may hypothetically exist.

Finally, by using the fragmentation scheme we can, in principle, begin the process of trying to isolate those structural features which best contribute (or correlate) to the biological action. We are not locked into using a nonspecific physical property such as "lipophilicity" as a means of rationalizing the biological behavior of molecules. Rather, the emphasis is on the molecular structure itself. Using molecular structure as a basis, one may hope to gain clues regarding specific effects of the molecule, be it interaction of drug functional groups with receptors or alterations in conformation caused by substitution patterns.

In another sense, however, the results of this study are not encouraging. The best correlations did not occur with indices from fragments selected a priori for their structural relevance. If QSAR techniques are to evolve, our constructs, images, and hypotheses of the ways in which drugs work must be more consonant with the system elements whence the data derive.

This brings us to an even more fundamental issue that unfortunately has not received much consideration, viz., the primitive state of the QSAR field. Lipophilicity, bond strength, bond location, steric effects, etc., may well correlate with a multitude of biological response variables, but in most cases, we simply do not know why. To maintain that lipophilicity is an important physical property because biological membranes and cellular structures have lipoidal characteristics represents a tautological legerdemain. The existence of parabolic QSAR relationships (where compounds with dramatically different lipophilic properties elicit quantitatively identical biological responses) may be explained summarily by introducing even more independent variables (correlates), but these often lack palpable credibility also. You feed a set of numbers into a computer and turn the crank, and out comes a second set of numbers. When it works, you accept the result and ask no further questions. When it fails, you try new variables or equations. This is why most QSAR data are consistent with a number of alternative explanations and hypotheses. Fortunately, the history of pharmaceutical science suggests that this is an intermediate (albeit prolonged) desideratum, eventually to be replaced by more heuristically fertile constructs of how chemicals interact with important homeostatic and other regulatory biological entities. Our black-box approach to QSAR reflects simply the early stage of our theorizing. Although

this present work may help focus the problems at hand, it has very little more to offer mechanistically than has previous work.

ACKNOWLEDGMENT

The authors wish to thank Gary A. Thompson for his critical review of this work.

REFERENCES

1. H. E. Simmons and R. E. Merrifield. *Proc. Natl. Acad. Sci.* 74:2616–2623 (1977).
2. D. H. Rouvray. *Am. Sci.* 61:729–735 (1973).
3. D. H. Rouvray. *J. Chem. Ed.* 52:768–773 (1975).
4. L. B. Kier and L. H. Hall. *Molecular Connectivity in Chemistry and Drug Research,* Academic Press, New York, 1976.
5. C. Hansch and A. Leo. *Substituent Constants for Correlation Analysis in Chemistry and Biology,* Wiley, New York, 1980.
6. Y. C. Martin. In G. L. Grunewald (ed.), *Medicinal Research Series, Vol. 8,* Marcel Dekker, New York, 1978, pp. 139–165.
7. M. Randic and C. L. Wilkins. *J. Chem. Inf. Comp. Sci.* 19:31–37 (1979).
8. C. L. Wilkins, M. Randic, S. M. Shuster, R. S. Markin, S. Steiner, and L. Dorgan. *Anal. Chim. Acta* 133:637–645 (1981).
9. H. Boxenbaum. *J. Pharmacokin. Biopharm.* 10:411–426 (1982).
10. L. H. Hall and L. B. Kier. *J. Pharm. Sci.* 67:1743–1747 (1978).
11. K. S. Pang and M. Rowland. *J. Pharmacokin. Biopharm.* 5:625–654 (1977).
12. K. S. Pang. *Trends Pharmacol. Sci.* 1980:247–251 (1980).
13. H. Boxenbaum. *J. Pharmacokin. Biopharm.* 8:165–176 (1980).
14. M. Randic and C. L. Wilkins. *J. Phys. Chem.* 83:1525–1531 (1979).
15. M. Randic and C. L. Wilkins. *Int. Symp. Quantum Chem. Quantum Biol.* 6:55–61 (1979).
16. C. L. Wilkins and M. Randic. *Theor. Chem. Acta* 58:45–51 (1980).
17. M. Randic, G. M. Brissey, R. B. Spencer, and C. L. Wilkins. *Comput. Chem.* 4:27–39 (1979).
18. R. W. Lucek and C. B. Coutinho. *Mol. Pharmacol.* 12:612–619 (1976).
19. J. H. Woolf and J. C. Nixon. *Biochemistry* 20:4263–4268 (1981).
20. J. L. Marx. *Science* 227:934–935 (1985).
21. R. S. Markin and W. J. Murray. *Pharm. Res.*, in press (1988).