Biochemical Pharmacology, Vol. 44, No. 10, pp. 1935–1946, 1992.
Printed in Great Britain.

0006-2952/92 $5.00 + 0.00
Pergamon Press Ltd

QUANTITATIVE STRUCTURE–METABOLISM RELATIONSHIPS FOR SUBSTITUTED BENZOIC ACIDS IN THE RAT

COMPUTATIONAL CHEMISTRY, NMR SPECTROSCOPY AND PATTERN RECOGNITION STUDIES

FARIDA Y. GHAURI, CATHERINE A. BLACKLEDGE, ROBERT C. GLEN,* BRIAN C. SWEATMAN,* JOHN C. LINDON,* CHRISTOPHER R. BEDDELL,* IAN D. WILSON† and JEREMY K. NICHOLSON‡

Department of Chemistry, Birkbeck College, University of London, Gordon House, 29 Gordon Square, London WC1H 0PP; *Department of Physical Sciences, The Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS; and †Department of Safety of Medicines, ICI Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, U.K.

(*Received* 15 *June* 1992; *accepted* 3 *September* 1992)

Abstract—An extensive set of computed molecular properties, both steric and electronic, have been calculated using molecular orbital and empirical methods for benzoic acid (**1**) and a congeneric series of substituted benzoic acids, i.e. 2-, 3- and 4-fluorobenzoic acids (**2–4**), 2-, 3- and 4-trifluoromethyl benzoic acids (**5–7**), 2-, 3- and 4-methylbenzoic acids (**8–10**), 4-amino benzoic acid (**11**), 2-fluoro-4-trifluoromethyl benzoic acid (**12**), 4-fluoro-2-trifluoromethyl benzoic acid (**13**), 3-trifluoromethyl-4-fluorobenzoic acid (**14**). We have monitored the urinary excretion profiles and determined the metabolic fate of compounds **2–7**, **12–14** in the rat using high resolution ^{1}H and ^{19}F NMR spectroscopy. Corresponding data for compounds **1**, **8–11** are taken from the literature. In all cases phase II glucuronidation or glycine conjugation reactions dominated the metabolism of these compounds. Compounds **5**, **7**, **12**, **13** have ester glucuronides as their major metabolites; the rest primarily form glycine conjugates. Compounds (**1–12**) have been classified according to their calculated physicochemical properties using pattern recognition methods and principal components maps have been used as a novel type of structure–metabolism diagram. The maps of compounds in the physicochemical property space served to separate the compounds into the two major classes which related to their principal metabolic fate *in vivo*, namely glucuronidation versus glycine conjugation. Compounds **13** and **14** were used as further probes of the property space, and dominant metabolic fates of glucuronidation and glycine conjugation, respectively, were predicted from the previous "training set map". The metabolic fate of compounds **1–14** can thus be classified according to a simple set of physicochemical rules. Investigation of the physicochemical properties which are important in distinguishing the metabolic fate of the compounds may give insight into key features of the drug-metabolizing enzyme active sites and hence provide information on basic mechanisms of benzoate metabolism.

Knowledge of the metabolic fate and tissue half-life of drugs and other xenobiotics is of critical importance to the understanding of their mode of action and for the validation of toxicological studies. The recognition and exploitation of the relationships between chemical structure and biological activity have been crucial to many advances in pharmacology and experimental therapeutics and have given rise to the discipline known as quantitative structure–activity relationships (QSAR§). QSAR techniques have augmented the understanding of pharmacological activities based on structure and are an important part of modern drug discovery activities [1–3]. However, much less study has been devoted to quantitative structure–metabolism relationships (QSMR) and in the main this has been limited to the use of substituent parameters as pioneered by Hansch and co-workers. The present study is an attempt to classify the metabolic fate of a limited series of compounds using parameters which can be calculated or predicted and which therefore do not require the compound to be synthesized and experimental measurements made. We have attempted to investigate which molecular descriptors are important in the metabolic prediction using the well-known data reduction methods of non-linear mapping (NLM) [4] and principal components analysis (PCA) [5]. Other methods of course exist including for example comparative molecular field analysis (CoMFA) [6], and we intend to pursue such methods later. This complementary approach can be particularly valuable when the molecules being compared are chemically dissimilar. However, it does not produce a solution directly in terms of atom-centred or global molecular properties and our

‡ Corresponding author. Tel. (071) 380 7468; FAX (071) 380 7464.

§ Abbreviations: PR, pattern recognition; QSAR and QSMR, quantitative structure (metabolism)–activity relationships; FID, free induction decay; NLM, non-linear map; PCA, principal components analysis; PC, principal component.

intention in the present study is to determine whether such an approach is possible. It is clear that many of the theoretical properties which can be calculated for these molecules will be highly correlated and an important aspect of the present study is to investigate how to remove such correlations.

Modern drug discovery programmes particularly involve the synthesis and sequential functional group alteration of compounds aimed at producing specific pharmacological effects with appropriate efficacy and selectivity together with minimal toxicity to humans. An important factor in the generation of novel drugs is an understanding of how they are absorbed, distributed, metabolized and excreted (ADME studies). In contrast to the increasing component of design in drug discovery, there is still a high degree of empiricism used in optimizing half-life in the body by modifying the way potential metabolic routes affect the compounds. In addition, of course, metabolites of administered drugs can sometimes themselves be biologically active either beneficially or toxicologically. Furthermore, although significant effort is spent in the computational study of drug–receptor interactions, much less effort is expended in modelling the routes of metabolic modifications of drug substrates by endogenous enzyme systems. There is, therefore, a need to understand how combinations of molecular physicochemical properties quantitatively influence the metabolic fate of drugs.

Drug metabolism can be a complex process involving both the introduction of new functional groups into the compound itself (Phase I) and conjugation reactions of the compound, or a metabolic product from Phase I, with endogenous species (Phase II). However, the preferred route and extent of metabolism of a compound is essentially a product of absorption factors and subsequent binding to and reactivity towards one or more of these metabolizing enzyme systems. These factors may be related to particular physicochemical features of the drug molecule. Hitherto, many studies on structure–metabolism relationships have been limited to relatively simple models involving the calculation of conformational space-filling properties [7] or by applying standard substituent effects [8]. Although these factors are very important for some drug classes, given the spatial specificity of the active sites of many drug-metabolizing enzymes, there is at least as wide a range of properties ultimately influencing the metabolic fate of drugs. Many of these can be more complex than conformational properties alone and can be calculated and classified by theoretical chemistry methods [1–3]. For this preliminary investigation, using the computational chemistry and pattern recognition approach to study QSMR, we have chosen to investigate the metabolism of a simple series of substituted congeneric benzoic acids in the rat. In the main, these compounds are metabolized to glucuronide or glycine conjugates with the latter being much more widespread for metabolism of simple benzoic acids in the rat. This has resulted in a rather unbalanced set in which glucuronides predominate and, for this reason, we have chosen to classify the compounds into two groups, thus excluding unchanged compound as a group and in addition ignoring quantitative differences in metabolite production between the compounds within a class.

The high cost in both time and resources required for conventional studies on xenobiotic metabolism has imposed severe practical limitations on the systematic investigation of structure–metabolism relationships. However, it has been amply demonstrated that comprehensive metabolic data on the fate of xenobiotics can be generated using multinuclear and multipulse high resolution NMR techniques. Furthermore, NMR methods [9–17] require only a fraction of the time and resources needed for conventional radiochemical/chromatographic techniques. In addition to drug metabolic information, ^{1}H NMR spectroscopy of biological fluids can also show biochemical or toxicological effects of the compounds under investigation [18].

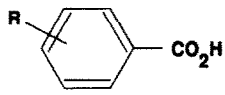
However, even when operating at a ^{1}H NMR observation frequency of 600 MHz (the highest commercially available) there may be circumstances where signals from drug metabolites are partially or wholly obscured by resonances from endogenous compounds. As a result we have devised a number of simple NMR-monitored chromatographic and enzymatic procedures to aid metabolite identification [19, 20]. We have also availed ourselves of the opportunities presented by the availability of a wide range of fluorinated compounds to use ^{19}F NMR where possible. Where fluorinated compounds are being studied, ^{19}F NMR is an obvious probe for drug metabolism [9, 11, 17, 21]. ^{19}F NMR has 83% of ^{1}H NMR sensitivity, but there is negligible interference from endogenous fluorinated compounds. The organic compound ^{19}F chemical shift range is >200 ppm and ^{19}F shifts may be sensitive to structural changes in molecules occurring up to eight bonds from the ^{19}F atom [9, 17]. Concentrations as low as 2–3 μM can be detected in a variety of intact biological materials [21]. A useful feature of ^{19}F NMR spectroscopy in metabolic studies is that the ^{19}F nucleus can act as a "handle" with which to probe the "metabolism-directing" effects of other functional groups substituted on the aromatic ring. Also, it is unusual for fluorine to be removed to any significant extent during metabolism and hence it is possible to explore the metabolism of a wide range of fluoroaromatic compounds with sequentially modified substituents. As described below we have used ^{1}H and ^{19}F NMR to generate metabolic data on a series of congeneric substituted benzoic acids and related this metabolic data to semi-empirical molecular orbital and structural physicochemical parameters analysed using pattern recognition (PR) methods which we have already successfully applied in the analysis of NMR spectra of biofluids in relation to toxicity effects [22–24].

MATERIALS AND METHODS

Compounds. All compounds were of analytical grade and were obtained commercially (Fluorochem, Glossop, U.K.) and used without further purification. The structures and numbering system are given in Table 1.

Animals and treatments. Male Sprague–Dawley

Table 1. Compounds used in the metabolism prediction study



Compound No.	R
1	H
2	2-F
3	3-F
4	4-F
5	2-CF_3
6	3-CF_3
7	4-CF_3
8	2-CH_3
9	3-CH_3
10	4-CH_3
11	4-NH_2
12	2-F, 4-CF_3
13	2-CF_3, 4-F
14	3-CF_3, 4-F

rats (200–250 g) were divided into a series of groups of rats each dosed with one from a range of substituted benzoic acids (Tables 1 and 2) at 100 mg/kg. Animals were housed individually in plastic metabolism cages, in a well-ventilated room with regular light cycles (12 hr). Food and tap water were provided *ad lib*. Following an acclimatization period of 2 days each rat received a single i.p. injection of the appropriate substituted benzoic acid. Urine was collected over ice for 24 hr prior to dosing and at 8, 24 and 48 hr post dosing. Urinary volume and pH were measured and the urine was centrifuged at 3000 rpm for 10 min at 4°; the samples were then frozen and maintained at −20° prior to NMR measurements.

^{1}H *NMR analysis of urine.* Measurements were made at ambient probe temperature (15 ± 1°) either on a Bruker WH400 spectrometer operating at a proton resonance frequency of 400 MHz or a JEOL GSX500 spectrometer operating at a proton resonance frequency of 500 MHz. For each sample 64 free induction decays (FIDs) were collected into 16,384 computer points using a 45° pulse and a spectral width of 5000 Hz centred on the water resonance frequency. This gave an acquisition time of 1.7 sec and a further 3 sec delay was added between pulses to allow T_1 relaxation to occur. Water suppression was achieved by a gated (off during acquisition) secondary irradiation field applied at the water resonance frequency. Exponential line weighting functions of 0.2–0.5 Hz were supplied to the FIDs prior to Fourier transformation. Chemical shifts were referenced to internal sodium 3-trimethylsilyl-[2,2,3,3-$^2H_4$]1-propionate, sodium salt (δ0.0). Resonance assignments were confirmed by a combination of chemical shift, spin–spin coupling patterns, coupling constants, pH dependence of chemical shifts and ultimately by standard addition.

^{19}F *NMR analysis of urine.* Measurements were made at 15° either on Bruker AM400 wide-bore or on Varian VX400 standard-bore spectrometers operating at a ^{19}F frequency of 376 MHz. Samples were placed in 5 mm o.d. precision NMR tubes for measurement. For each sample 256 FIDs were collected into 32,768 computer points using a 45° pulse, and a spectral width of 10,000 Hz. This gave an acquisition time of 0.5 sec and a further 6 sec delay was added between pulses to allow full T_1 relaxation to occur. An exponential line broadening function of 0.5 Hz was applied to the FIDs prior to Fourier transformation. Chemical shifts were referenced to the external secondary reference trifluoroethanol (δ-77 ppm from $CFCl_3$). Proton-decoupled ^{19}F NMR spectra were collected for monofluorinated aromatic compounds. Broadband proton decoupling was not necessary for trifluoromethylated compounds. Resonance assignments were confirmed by a combination of standard additions, proton–fluorine coupling constants, enzy-

Table 2. Mean 24 hr excretion profiles of nine substituted benzoic acids in the rat

Compound	N	Parent compound	Total glucuronides*	Glycine conjugate	Unknown
2	5	2.5 ± 0.6	0.1 ± 0.1	*72.3 ± 7.6*	ND
3	4	11.5 ± 4.1	ND	*26.7 ± 4.9*	ND
4	5	0.1 ± 0.1	ND	*43.7 ± 13.0*	ND
5	4	ND	*39.0 ± 9.2*	ND	ND
6	4	2.1 ± 0.2	1.3 ± 0.4	*37.8 ± 9.7*	0.3 ± 0.2
7	5	5.8 ± 0.9	*27.7 ± 2.1*	ND	ND
12	3	*16.7 ± 1.4*†	*15.3 ± 2.4*†	ND	5.2 ± 1.4
13	3	*47.0 ± 1.5*†	*6.7 ± 1.6*†	ND	ND
14	3	2.8 ± 0.2	10.7 ± 0.9	*26.9 ± 1.5*	0.5 ± 0.0

Dominant urinary species in italics.

Values are means ± SEM, % of dose excreted in 24 hr, determined from quantitative ^{19}F spectroscopy of urine.

N, No. of animals.

ND, Not detected.

* Includes all glucuronide transacylation products.

† Glucuronide conjugate hydrolyses extensively to parent compound in urine at pH > 7.

matic hydrolysis and cross-referenced proton NMR spectra [17].

NMR-monitored solid phase extraction chromatography and enzymatic treatment of urine samples. In order to partially purify and identify xenobiotic metabolites from some samples, solid phase extraction followed by stepwise elution procedures were employed with NMR monitoring. Acidified urine samples were applied to C18 solid phase extraction columns (Bond Elut®, Analytichem) as previously described [19, 20] and after washing the column with acidified water were eluted with a series of methanol–water mixtures (5 mL washes) starting with 20:80 and progressing through 40:60, 60:40 and 80:20 steps to a final wash with 100% methanol. The individual fractions were collected and methanol removed with a stream of dry nitrogen and the residual water was removed by freeze-drying, followed by dissolution in 100% ^2H_2O prior to NMR analysis.

Calculation of physicochemical properties. Molecular structures for each compound studied were built using the SYBYL [25] molecular modelling software on a DEC-VAX 8550 system. The molecular geometries were optimized using MOPAC (AM1 PRECISE [26]) on a Cray X-MP and overlaid by a least squares fit of the benzene ring and the principal functional group (benzoic acid carbonyl group). In all, 39 theoretical molecular properties were calculated using PROFILES [1]. These were atom-centred partial charges on ring carbons, ATCH; energy of the highest occupied molecular orbital, EHOMO; energy of the lowest unoccupied molecular orbital, ELUMO; total dipole moment and its x, y and z components, DIP, DIPX, DIPY, DIPZ; nucleophilic and electrophilic superdelocalizabilities, NSDL, ESDL at each ring carbon [27, 28]; polarizability, POL; all derived from the semi-empirical AM1 method in MOPAC [26]. In this version of the calculation of superdelocalizability, the algorithm as originally defined by Fukui *et al.* [27] is modified as the Huckel parameter (λ) is replaced by the orbital energy E. E may tend to zero but in general is not zero. As such this parameter may be useful and correlate well with molecular properties [28]. In addition, a number of steric parameters were calculated. These were dead space volume, DSV [29]; van der Waals volume, VDWV [29]; collision diameter, CD [29]; closest approach diameter, CAD [29]; surface area, SA [30]; moments of inertia in x, y and z directions, MIX, MIY, MIZ [31]; and principal ellipsoid axes in x, y and z directions, PEAX, PEAY, PEAZ [31]. In addition, the octanol–water partition coefficient, CLOGP, and the molar refractivity, CMR, were calculated from the MEDCHEM program [32]. Molecular mass was also included as a descriptor. In principle, experimental molecular descriptors could be substituted in some cases, for example it would be possible to use logP instead of CLOGP, but in this series of compounds CLOGP is a good approximation to logP for those compounds where a measured value exists, and one of the goals of this exercise was to be able to predict the metabolic class without the need to synthesize the compound.

Computer data analysis. The calculated physicochemical data for the compound set were analysed by two unsupervized PR methods, i.e. NLM [4] and PCA [5] using ARTHUR [33]. NLM and PCA reduce the multi-dimensional parameter space where each dimension is a calculated physicochemical property. The technique of NLM accepts the data values as coordinates in n-dimensional space and then compresses these into a NLM which is a 2-dimensional or 3-dimensional approximation to the true multi-dimensional interpoint distances. Two points which are close on the map should be more similar in terms of input variables than two distant points. PCA is a well established multivariate technique for dimension reduction in which principal components (PCs) are new variables created from linear combinations of the starting variables with appropriate weighting coefficients. The properties of these PCs are such that: (i) each PC is orthogonal with all other PCs, (ii) the first PC describes the largest part of the variance of the data set with subsequent PCs containing correspondingly smaller amounts of variance. PCs may be plotted one against the other to reveal the degree of classification contained in the PCs. Analysis of the factors contributing to the PCs in a plot which gives good classification of the data may reveal underlying trends which can be used to explain the classification mechanism. For example, in this case it may be possible to gain insight into the physicochemical factors involved in the recognition, enzyme-binding and metabolism of the compounds.

RESULTS

The major urinary metabolites of the selected benzoates under study were determined using ^1H and ^{19}F NMR methods or were identified from the literature. A combination of directly applied NMR techniques or solid phase extracted metabolites as described previously for other compounds was used to study urine samples [16]. ^{19}F NMR has proved particularly useful in these metabolic studies, providing an inexpensive and relatively accessible alternative to radiolabelling, enabling excretion balance and metabolic profile data to be collected simultaneously. Typical results for compound **4**, 4-fluorobenzoic acid, are shown in Fig. 1. This compound is excreted in urine in the rat predominantly as the glycine conjugate (Structure 1) and resonances from this molecule are seen in the ^1H spectrum as shown and assigned on Fig. 1. Most of the peaks arise from endogenous metabolites which have been assigned previously [18] but resonances relating to the xenobiotic can also be seen. The resonance from the CH_2 group of the glycine conjugate appears as a doublet because of spin coupling to the NH group, confirmed by making the pH alkaline whereupon the coupling collapses because of fast proton exchange. In addition the aromatic ring protons of the drug metabolite can also be assigned.

Similarly for compound **7**, 4-trifluoromethylbenzoic acid, ^1H and ^{19}F NMR (Fig. 2) allowed the identification of the principal metabolism route as glucuronic acid conjugation (Structure 2). However, this compound exhibits multiple acyl

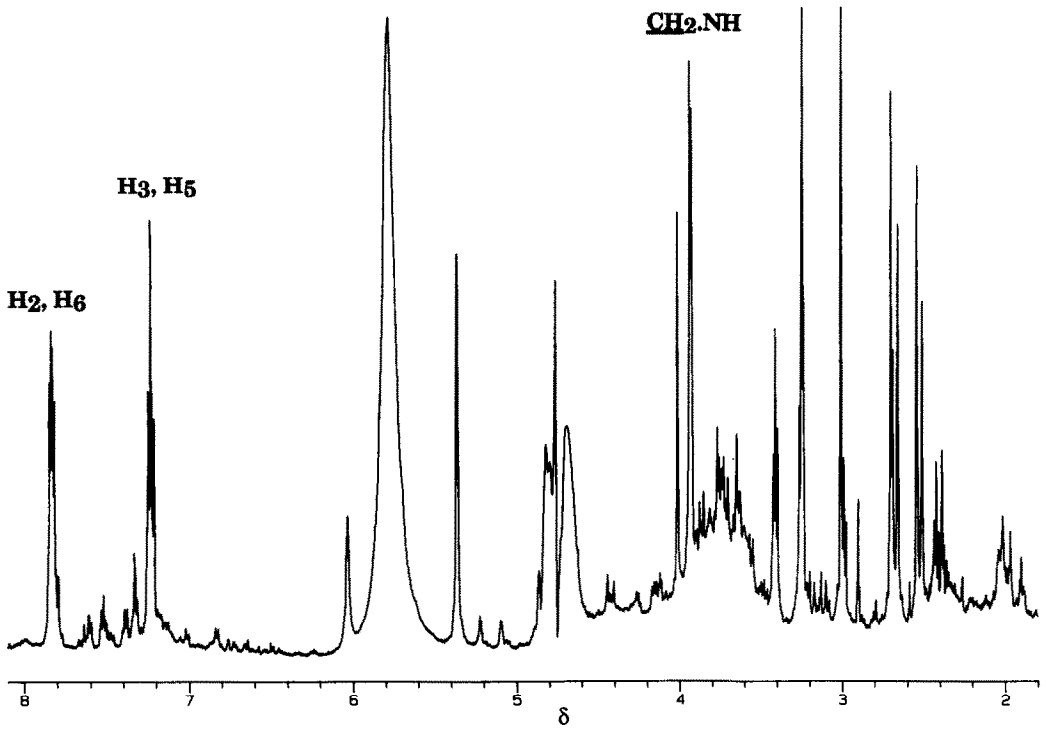
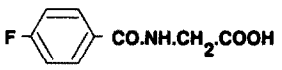
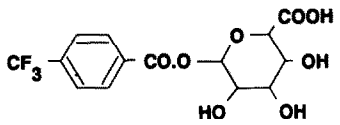


Fig. 1. ^{1}H NMR spectrum of urine from a rat dosed with 4-fluorobenzoic acid (**4**) as described in the text. The urine was collected for 8 hr after dosing and the spectrum was measured with presaturation of the H_2O resonance.



Structure 1.



Structure 2.

migration [10, 11] resulting in transfer of the benzoic acid moiety from C1 of the glucuronic acid to sequentially C2, C3 and C4, with a consequent ability of these latter isomers to exist as α and β anomers. Thus seven species should be observed. Again the ^{1}H NMR spectrum (Fig. 2a) is dominated by endogenous metabolites but resonances from the various glucuronide species can be seen in the δ5–6 region and the benzoic acid aromatic proton resonances can clearly be seen between δ7.8 and 8.2.

Assignment of the ^{19}F NMR signals from the various glucuronide species is tentative and is based on the time-course of equilibration following hydrolysis at slightly alkaline pH. Thus initially the β-1 glucuronide intensity decreases with a concurrent increase in the level of the parent compound and the 2-glucuronide, followed later by a reduction in the 2-glucuronide level and an increase in the concentration of the 3-glucuronide, and similarly later for the 4-glucuronide.

The resolution of separate signals from the α and β anomers provides a dramatic illustration of the sensitivity of the changes in ^{19}F chemical shifts of ^{19}F atoms to metabolite structure even when they are remote from the site of metabolism [11].

The percentage of total excretion and the proportions of the major urinary metabolites of nine substituted benzoic acids not previously studied in the rat (**2–7**, **12–14**) are given in Table 2. Four of these compounds are metabolized and excreted in urine as ester glucuronides and show different degrees of transacylation (Ghauri *et al.*, unpublished data). Compounds **1**, **8–11** have been studied previously by chromatographic methods and are know to form glycine conjugates (>80%) and this has been confirmed for **8–11** in recent studies by us (single animals dosed only) [34–38].

The 14 compounds can therefore be classified into two groups, namely glycine and glucuronic acid conjugates. A series of theoretical chemistry properties has been calculated for all 14 compounds and PR methods have been applied to these data in an attempt to evaluate those properties which are relevant to the metabolic route. Compounds **1–12** were treated as a training set for the generation of maps of compounds in the property space, and

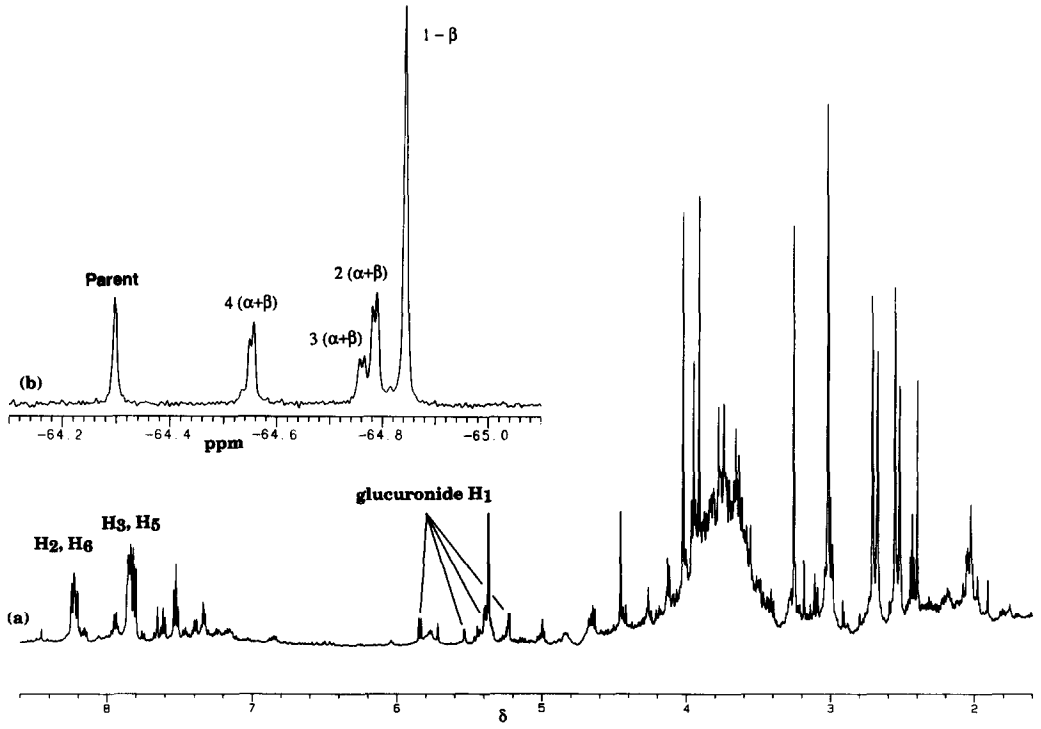


Fig. 2. ^{1}H and ^{19}F NMR spectra of urine from a rat dosed with 4-trifluoromethylbenzoic acid (**7**) as described in the text. (a) ^{1}H NMR spectrum from a sample lyophilized and reconstituted into D_2O; (b) ^{19}F NMR spectrum.

compounds **13** and **14** were used to test whether the observed groupings were useful for predicting the metabolic route.

Fig. 3a thus shows the NLM of the compounds **1–12** treated as a training set based upon the full 39 parameter descriptor set. The three glucuronide metabolite species appear at the left side of the figure and this trend is also reproduced in Fig. 3b, the corresponding map of PC1 versus PC2. Adding in compounds **13** and **14** as a test of the classification power of the maps produced Fig. 3c and d, the corresponding NLM and PC maps, respectively. It can be seen that compound **13**, which forms a glucuronide conjugate, maps close to the other glucuronide forming benzoic acids. Compound **14**, which forms a glycine conjugate, also maps close to the rest of the glycine conjugate class especially in the NLM (Fig. 3c). However, compound **6**, a glycine conjugate is also well separated from other members of its class and is close to the glucuronide class, especially in the PC (Fig. 3d).

The fact that the compounds tend to cluster according to their metabolic propensities using physicochemical descriptors indicates that there is an underlying set of physicochemical characteristics which determine the selection of the metabolic route.

In order to explore the structure–metabolism relationships further a reduced data set was produced by removing those parameters having an inter-parameter correlation coefficient of $r > 0.85$. The redundant parameters were DSV, CD, NSDL2, NSDL4, NSDL5, NSDL6, CMR, EDSL4, ESDL5, MIY, POL, MW, ELUMO, EHOMO and SA leaving a total of 24 partially decorrelated parameters. The resulting PC1 versus PC2 map for the training set plus test set of all 14 compounds shows that compound **13** classifies close to the other glucuronides but so does compound **6** which forms a glycine conjugate. Compound **14** remains in an ambiguous position well separated from the other glycine conjugates. One advantage of PCA is that the key descriptors which are responsible for classification of data can be identified by consideration of their individual contributions to the overall variance in the data set. From inspection of the parameters contributing to the first two PCs, those which contributed >2% of the variance and had an inter-parameter correlation coefficient of less than 0.85 were selected for further investigation. This procedure extracted 10 parameters, i.e. CLOGP, ATCH2, ATCH3, ATCH6, MIX, MIZ, PEAX, PEAY, ESDL1 and ESDL6. A new plot was constructed using only these 10 parameters and again the same physicochemical clustering behaviour is maintained, i.e. compounds **6** and **14** are anomalous.

An alternative method of determining which parameters are important in distinguishing the two classes of metabolic outcome is to examine the descriptors one-by-one to identify those which are significant in separating the two predefined classes. We have used the unpaired Student's *t*-test for this purpose. The two classes were predefined as

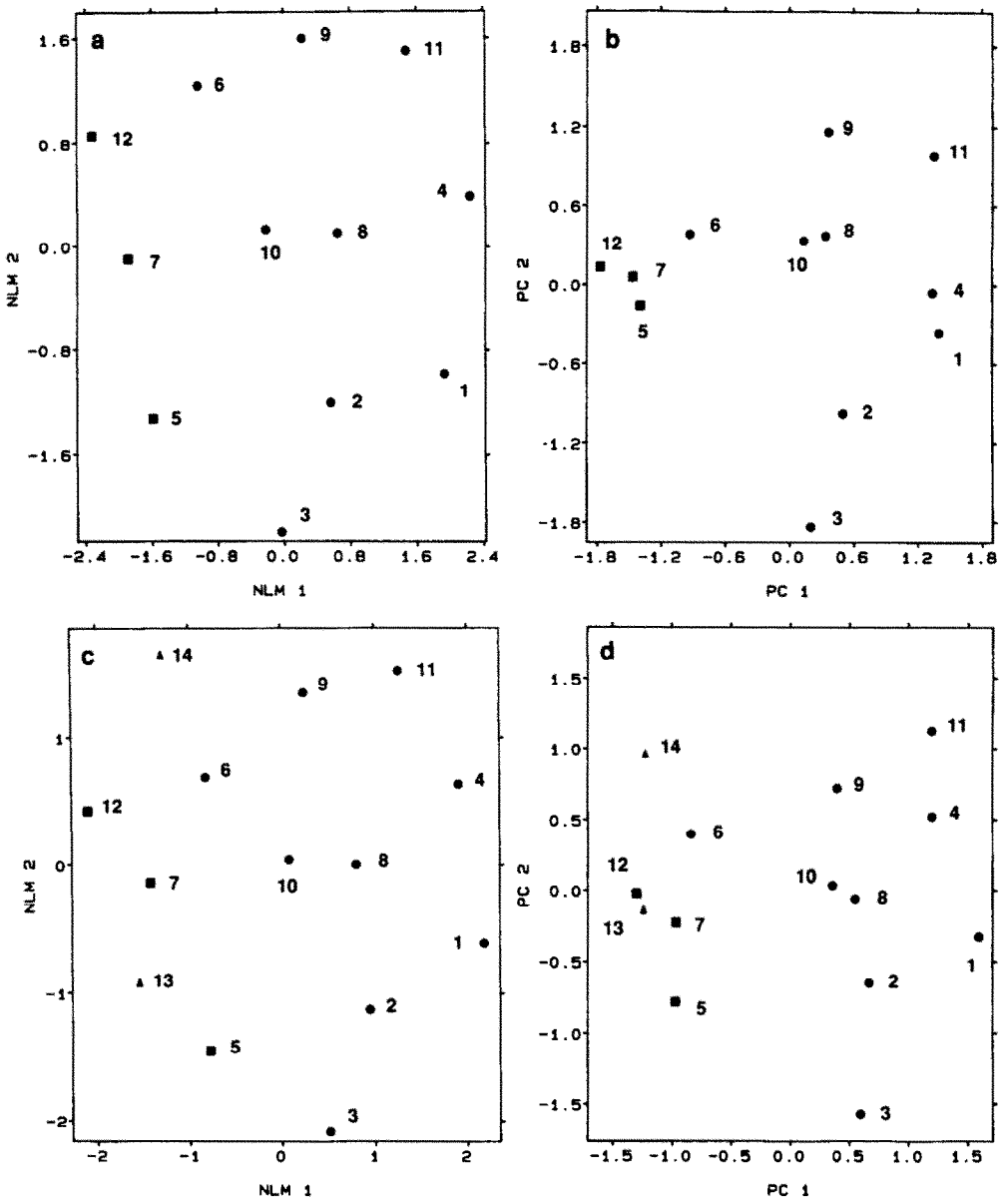


Fig. 3. (a) NLM and (b) PC maps of the 12 compound training set based on 39 parameters. (c) NLM and (d) PC maps of 14 compounds based on the same 39 parameters. Squares represent compounds which form glucuronide conjugates, circles represent glycine conjugates and triangles denote the two test compounds.

glucuronides (compounds **5**, **7**, **12** and **13**) and glycines (compounds **1–4**, **6**, **8–11** and **14**). The training set of data (i.e. excluding compounds **13** and **14**) was used to elucidate those descriptors which were significant in separating the classes at a significance level of $P < 0.05$. The 15 resulting parameters were CLOGP, CMR, POL, ATCH4, ATCH5, DIPX, EHOMO, VDWV, SA, MW, DSV, CD, ESDL1, ESDL3 and ESDL5. A PC map produced from these 15 parameters shows good separation of both classes except that compound **6** still maps close to the glucuronide group. Subclasses are also apparent. Adding in the two test compounds and remapping using the *same* parameters gave a PC map with compound **13** correctly positioned with the glucuronides and compound **14** very close to compound **6**, a glycine conjugate, but both still close to the glucuronide group. The method was refined further by removing highly correlated parameters ($r > 0.85$) from the 15 determined above. This resulted in the map shown in Fig. 4a for the 14 compound combined set, and was produced from seven parameters derived from only the training set, namely CLOGP, POL, ATCH4, ATCH5, DIPX,

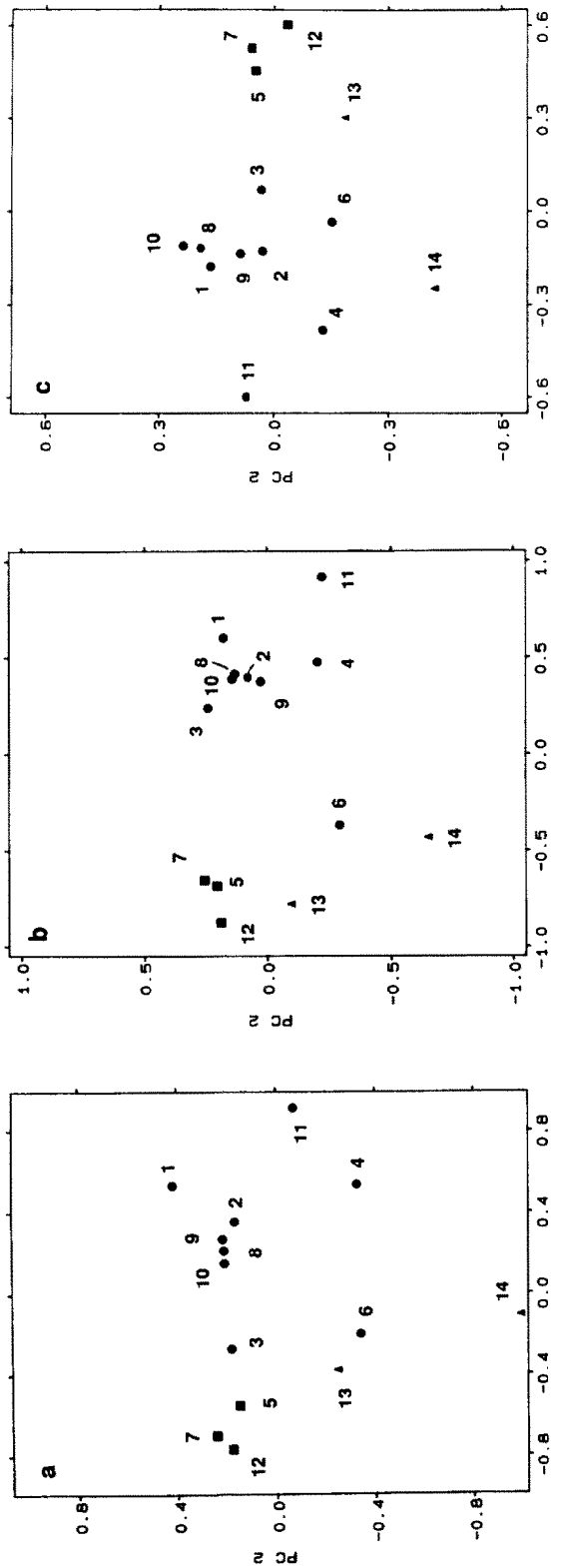


Fig. 4. (a) PC maps produced for the 14 compound set based on seven parameters significant by *t*-test at the $P < 0.05$ level in the 12 compound training set and with inter-parameter correlation coefficient of $r < 0.85$. (b) PC map based on six parameters for the 14 compound sets as in (a) but using a significance level of $P < 0.001$. (c) PC map for the 14 compound set using only the two most significant parameters, ATCH5 and ESDL5. Symbols are as defined in Fig. 3.

ESDL3 and ESDL5. This shows compound **6** close to the main glycine group. This also shows compound **13** mapping some distance away from the glucuronide cluster and actually still close to compound **6**, a glycine conjugate. Compound **14** remains distant from all compounds.

Repeating the procedure but at the more stringent significance level of $P < 0.001$ on the 12 compound training set resulted in a six parameter descriptor set, i.e. ATCH5, EHOMO, SA, MW, ESDL1 and ESDL5. Mapping all 14 compounds using this descriptor set resulted in Fig. 4b and unambiguous clustering is observed in that compound **13** lies close to the glucuronide class and compound **14** is well separated from the glucuronides, although it remains close to compound **6** and both are some distance from the main glycine class.

Finally, Fig. 4c shows a PC map of all 14 compounds using just two parameters, the most significant ones from the above analysis, i.e. ATCH5 and ESDL5. This provides an excellent classification; as good as that in Fig. 4b.

DISCUSSION

Biotransformations of xenobiotic carboxylic acids

Glucuronidation and glycine conjugation are the most important routes of metabolism of xenobiotic carboxylic acids involving a very broad range of substrates and with widespread species occurrence [34, 39]. The xenobiotics can react with UDP-glucuronic acid under the catalytic influence of UDP-glucuronyl transferase to form an ester which links the C-1 hydroxy group of β-D glucuronic acid and the carboxyl function [39]. At least two UDP-glucuronyl transferase isoenzymes exist with different substrate binding preferences; one being concerned with small planar molecules of relatively low lipid solubility and the other binding larger non-planar lipophilic substrates. The other major route of metabolism for xenobiotic carboxylic acids involves amino acid conjugation [34, 35] involving the activation of the acid to a high energy CoA thioester (via adenylate) with subsequent linkage to an amino acid at the amino functional group. Small planar carboxylates (e.g. benzoates) normally conjugate with glycine in primates and sub-primates, but arylacetic and aryloxyacetic acids may also be conjugated with glutamine and taurine in primates, and other amino acid conjugation reactions are also known [7, 35]. In the presence of *ortho* or *para* (to the carboxyl) trifluoromethyl groups substituted benzoic acids undergo glucuronidation reactions whereas the *meta* position of the CF_3 results in glycine conjugation *in vivo* [11]. There is a set of structural and electronic physicochemical rules which underlie the route of metabolism and biotransformation of xenobiotic carboxylates in different species but these are poorly understood in terms of quantified molecular physicochemical properties. In the case of arylacetic acids the degree of amino acid or glucuronide conjugation is strongly influenced by the substitution pattern on the carbon atom adjacent to the carboxylate function (the spacer group) and the *ortho*-substitution pattern on the aromatic ring [1]. The interpretation of glucuronide and amino acid conjugate excretion data in terms of "selection" of the conjugation and excretion pathway is complicated by the propensity for certain xenobiotic glucuronides to hydrolyse back to the parent compound, which may be greatly enhanced in the presence of high levels of tissue or serum esterases; but this can be corrected for using esterase inhibitors [40].

An important reason for studying the metabolic fate of xenobiotic carboxylic acids is because of the potential for metabolism-related hypersensitivity reactions that are thought to occur with certain non-steroidal anti-inflammatory drugs. Ester glucuronides of such drugs also undergo varying degrees of acyl migration so that up to seven structural isomers may be present in a biological sample, i.e. acyl conjugation at the β-1 position of the glucuronic acid followed by transfer of the acyl group to the 2, 3 and 4 positions of the glucuronic acid, these isomers then being capable of mutarotating to their α-forms. A number of non-steroidal anti-inflammatory drugs have been withdrawn from clinical use because of hypersensitivity reactions [40]. Such compounds have been shown to undergo glucuronidation followed by extensive acyl migration of the glucuronide ring to produce novel positional isomers which can react with the plasma proteins to form covalently bound adducts [38] (probably by Schiff-base formation). In our studies the acyl migration reactions can be readily followed by fluorine NMR spectroscopy, and this is well illustrated by 4-trifluoromethyl benzoic acid glucuronide (Fig. 2).

PR using computational chemistry properties

As can be seen from Fig. 4b, compounds **5**, **7** and **12** cluster close together and are metabolized to glucuronides. Compound **13** is also metabolized primarily to a glucuronide and maps in an intermediate position between the glucuronide and glycine conjugation groups. It would in principle be possible to treat the cases which apparently show a preponderance of unchanged compound as a separate metabolic class. However, with such a small data set and with the imbalance within it, it would be of questionable merit. Moreover, the high level of parent compound for **12** and **13** reflects the fact that the glucuronide, which is the real metabolic product, is unstable in urine and spontaneously hydrolyses to the parent compound. This can be observed if the NMR spectra of the urine are measured immediately after urine collection when the glucuronide can be observed. This decreases with time as the hydrolysis proceeds with the parent compound increasing. The true level of glucuronidation can then be obtained by back extrapolation and shown to be the dominant route of metabolism. A common feature of compounds **5**, **17**, **12** and **13** is that they possess a $-CF_3$ group *ortho* or *para* to the carboxyl group and this appears to have a powerful electronic influence on the route of metabolism. It is unlikely to be a steric effect because *ortho* and *para* substituted toluic acids, with a methyl group of similar size to the trifluoromethyl group, are metabolized to glycine conjugates.

On the other hand, the compounds with *meta* CF_3 groups, **6** and **14**, map distantly from the other CF_3

group-containing compounds and are metabolized to glycine conjugates, but **14** has a significant level of the glucuronide metabolite detected. It is interesting that **6** and **14** map close together and that **14** with some glucuronidation metabolism is further from the main glucuronide group (**5**, **7**, **12**, **13**) than compound **6**, implying that the descriptors used in Fig. 4b may not be optimal.

The descriptors which provide the best separation and which were used for Fig. 4b were ATCH5, EHOMO, SA, MW, ESDL1 and ESDL5. Four of these are electronic descriptors and only the surface area and molecular mass can be regarded as representing the size of the molecule. Figure 4c shows that reasonable classification can be obtained with only two parameters, namely the partial atomic charge at C5, i.e. an unsubstituted carbon *meta* to the carboxyl group, and electrophilic superdelocalizability at the same ring position. Both these parameters describe the electronic properties of that position and suggest that for small benzoates, positional electronic effects dominate the route of metabolism. The actual significance of these two parameters cannot be rationalized easily because, amongst other reasons, they are highly correlated with other electronic parameters of the molecules and they were used for Fig. 4c only by a process of decorrelation and elimination. In addition they are themselves reasonably correlated ($r = 0.61$).

Interestingly, Table 2 shows metabolic differences between positional isomers and this provides a justification for the use of atom-centred properties which can reflect this. In addition, prediction of such isomeric differences in metabolism would not be likely if only global parameters such as CLOGP and CMR which are calculated on a fragment basis were used.

The PR approach appears to represent an efficient way of reducing a plethora of computer chemistry-derived parameters into interpretable subsets. Initially, only "unsupervized" methods were used. These do not use the category of the sample, in this case the preferred metabolic outcome, in an attempt to maximize the separation of the two classes. The unsupervized approach as shown in Fig. 3 was not successful in that, at best, compound **6** which formed a glycine adduct mapped close to others which are predominantly glucuronide conjugates. Although most of the variance in the data is explained by the first two PCs, it may well be that the information necessary to classify the metabolism is present in some of the other PCs.

In an attempt to assess which descriptors were important in describing the metabolism, an alternative "supervized" method was used. This was based on the significance of each descriptor individually to classify the sample in terms of metabolic route. This proved successful giving rise to a limited number of descriptors which, when high inter-parameter correlations were taken into account, resulted in only six parameters being required and yielded a plot of the first two PCs (Fig. 4) which gave a good and unambiguous separation of the classes. The results indicate that only a small number of physicochemical properties are required to classify which conformationally fixed and otherwise metabolically inert benzoates may be recognized by the drug-metabolizing enzymes responsible for glucuronidation and glycine conjugation and, therefore, knowledge of these properties for a particular compound should be of value in predicting the extent of glycine versus glucuronidation routes *in vivo*.

The metabolism of benzoates may require the recognition of the planar phenyl ring by a metabolizing enzyme system. As the structural information of the compound set used here is largely confined to two dimensions and none of the molecules are chiral any effects of molecular shape and problems associated with recognition of chiralty will be minimized. Functionalization *ortho* to the carboxylate group may impose steric constraints on benzoate metabolism but in the case of all the compounds studied here, the molecules were calculated to exist with the COOH group in the plane of the benzene ring. Electronic distributions around both the benzene nucleus and functional groups, and how these dictate the character of the enzyme–substrate transition states, will play a critical role in determining enzyme–ligand binding and hence the metabolic fate of these compounds. Knowledge of the physicochemical parameters which are associated with the classification of metabolic fate is potentially important because of the clues it gives with regard to the molecular recognition properties of the active site of the UDP-glucuronyl transferase isoenzyme responsible for benzoate glucuronidation [8].

Although this work is at an early level of refinement with regard to selection and calculation of parameters it shows that useful classification of metabolic fate data can be achieved in this way and establishes that the PR approach to the understanding of QSMR may be highly revealing. The next stage of development is currently being undertaken and is an exploration of the physicochemical property spaces for a wider range of benzoic acids with more diverse substituent groups in order to refine the prediction of metabolic fate based on PR principles (as illustrated with respect to the three disubstituted benzoic acids shown). This will include the specific synthesis of compounds chosen to fit into particular regions of the maps. Once this has been achieved and a robust set of descriptors found, then rule-induction methods [41] could be used to derive a chemically-based "expert system" for predicting metabolism. These arguments can also be extended to the study of other compound classes, different animal models and relating *in vitro* to *in vivo* results. Currently available "expert systems" for drug metabolism prediction have found limited favour. This is partly because the systems are purely knowledge-based and take little account of the chemical properties of the drugs themselves and hence may give poor predictions. Furthermore, these systems give little or no mechanistic information in structure–metabolism relationships. In contrast, the present study represents a radically different molecular approach to studying structure–metabolism relationships and, in principle, the information generated should be widely applicable to other types of molecule.

Acknowledgements—We thank Professor C. D. Flint and Drs D. B. Davies and A. F. Drake for their helpful comments, and SERC and MRC for supporting this and related work; and the National Institute for Medical Research (MRC) for providing 400 MHz ^{19}F NMR facilities.

REFERENCES

1. Glen RC and Rose VS, A computer programme suite for the calculation, storage and manipulation of molecular property and activity descriptors. *J Mol Graph* **5**: 79–86, 1987.
2. Hyde RM and Livingstone DJ, Perspectives in QSAR: computer chemistry and pattern recognition. *J Comput Aid Mol Design* **2**: 145–155, 1988.
3. Lewis DFV, MO-QSARs: a review of molecular orbital-generated quantitative structure–activity relationships. *Prog Drug Metab* **12**: 205–255, 1990.
4. Chatfield C and Collins AJ, *Introduction to Multivariate Analysis*, 3rd Edn. Chapman and Hall, London, 1986.
5. Kowalski BR and Bender CF, Pattern recognition: a powerful approach to interpreting chemical data. *J Am Chem Soc* **94**: 5632–5639, 1972.
6. Cramer RD III, Patterson DE and Bunce JD, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* **110**: 5959–5967, 1988.
7. Waterbeemd H, Testa B and Caldwell J, The influence of conformational factors on the metabolic conjugation of aryloxyacetates. *J Pharm Pharmacol* **38**: 14–18, 1986.
8. Kim KH, Quantitative structure–activity relationships of the metabolism of drugs by uridine diphosphate glucuronosyltransferase. *J Pharm Sci* **80**: 966–970, 1991.
9. Wade K, Troke J, MacDonald CM, Wilson ID and Nicholson JK, ^{19}F NMR studies of the metabolism of trifluoromethylaniline. *Methodol Surveys Biochem Anal* **18**: 383–388, 1988.
10. Ghauri FYK, Wilson ID and Nicholson JK, ^{19}F and ^{1}H NMR studies of the metabolism of 4-trifluoromethylbenzoic acid in the rat. *Methodol Surveys Biochem Anal* **20**: 321–324, 1990.
11. Ghauri FYK, Blackledge CA, Wilson ID and Nicholson JK, Studies on the metabolism of fluorinated xenobiotics in the rat using ^{19}F-NMR and ^{1}H-NMR spectroscopy. *J Pharm Biomed Anal* **8**: 939–944, 1990.
12. Bales JR, Sadler PJ, Nicholson JK and Timbrell JA, Urinary excretion of acetaminophen and its metabolites as studied by proton NMR spectroscopy. *Clin Chem* **30**: 1631–1636, 1984.
13. Everett JR, Jennings K, Woodnut G and Buckingham MJ, Spin echo 1H-NMR spectroscopy: a new method for studying penicillin metabolism. *J Chem Soc Chem Commun* 894–895, 1984.
14. Bales JR, Nicholson JK and Sadler PJ, Two dimensional proton NMR "maps" of acetaminophen metabolites in human urine. *Clin Chem* **31**: 757–762, 1985.
15. Tulip K, Timbrell J, Wilson ID, Troke J and Nicholson JK, A proton magnetic resonance study of the metabolism of *N*-methyl formamide in the rat. *Drug Metab Dispos* **14**: 746–749, 1986.
16. Nicholson JK and Wilson ID, High resolution nuclear magnetic resonance spectroscopy of biological samples as an aid to drug development. *Prod Drug Res* **31**: 427–479, 1987.
17. Wade KE, Wilson ID, Troke JE and Nicholson JK, Magnetic resonance strategies for metabolic studies on fluorinated xenobiotics—application to flurbiprofen 2-(2-fluoro-4-biphenyl)propionic acid. *J Pharm Biomed Anal* **8**: 401–410, 1990.
18. Nicholson JK and Wilson ID, High resolution proton NMR spectroscopy of biofluids. *Prog NMR Spectrosc* **21**: 449–501, 1989.
19. Wilson ID and Nicholson JK, Solid phase extraction chromatography and nuclear magnetic resonance spectrometry for the identification and isolation of drug metabolites in urine. *Anal Chem* **59**: 2830–2832, 1987.
20. Wilson ID and Nicholson JK, Solid phase extraction chromatography and NMR spectroscopy (SPEC-NMR) for the rapid identification of drug metabolites in urine. *J Pharm Biomed Anal* **6**: 151–165, 1988.
21. Hull WE, Port RE, Herrmann R, Britsch B and Kunz W, Metabolism of 5-fluorouracil in plasma and urine as monitored by ^{19}F nuclear magnetic resonance spectroscopy for patients receiving chemotherapy with or without methotrexate treatment. *Cancer Res* **48**: 1680–1688, 1988.
22. Gartland KPR, Sanins SM, Lindon JC, Beddell CR, Sweatman BC and Nicholson JK, Pattern recognition analysis of high resolution ^{1}H NMR spectra of urine. A non-linear mapping approach to the classification of toxicological data. *NMR Biomed* **3**: 166–172, 1990.
23. Gartland KPR, Beddell CR, Lindon JC and Nicholson JK, A pattern recognition approach to the comparison of PMR and clinical chemical data for classification of nephrotoxicity. *J Pharm Biomed Anal* **8**: 963–968, 1990.
24. Gartland KPR, Lindon JC, Beddell CR and Nicholson JK, Application of pattern recognition methods to the analysis and classification of toxicological data derived from proton nuclear magnetic resonance spectroscopy of urine. *Mol Pharmacol* **39**: 629–642, 1991.
25. SYBYL, Tripos Associates, St Louis, MO, U.S.A., Molecular Modelling Software **V5.3**, 1989.
26. MOPAC Version 5, Quantum Chemical Program Exchange, Department of Chemistry, University of Indiana, Bloomington, IN, 1989.
27. Fukui K, Yonezawa T and Nagata C, Theory of substitution in conjugated molecules. *Bull Chem Soc Jpn* **27**: 423–427, 1954.
28. Gomez-Jeria JS, Calculation of the nucleophilic superdelocalisability by the CNDO/2 method. *J Pharm Sci* **71**: 1423–1424, 1982.
29. Edward JT, Molecular volumes and the Stokes-Einstein equation. *J Chem Ed* **47**: 261–270, 1970.
30. SAREA, Pearlman RS, *QCPE Bull* **1**: 16, 1981.
31. Margenaux H and Murphy GM, *The Mathematics of Physics and Chemistry*, 2nd Edn. van Nostrand and Co., New York, 1956.
32. CLOGP Version 3.54, Pomona College Medicinal Chemistry Project, Medchem software, Daylight Chemical Information Systems, Irvine, CA, U.S.A., 1984.
33. ARTHUR 81 Version 3, B & B Associates, P.O. Box 85505, Seattle, WA, U.S.A., 1981.
34. Bridges JW, French MR, Smith RL and Williams RT, The fate of benzoic acid in various species. *Biochem J* **118**: 47–51, 1970.
35. Caldwell J, Conjunction of xenobiotic carboxylic acids. In: *Metabolic Basis of Detoxification* (Eds. Jakoby WB, Bend JR and Caldwell J), pp. 271–290. Academic Press, London, 1980.
36. Mikulski P, Wiglusz R, Galuszko E and Delag G, Reciprocal metabolic effects of benzene and its methyl derivatives in rats. *Bull Inst Marine Tropical Med* **30**: 77–78, 1979.
37. Tardif R, Brodeur J and Plaa GL, Simultaneous high performance liquid chromatographic analysis of hippuric acid and *ortho*, *meta* and *para* methyl hippuric acids in urine. *J Anal Toxicol* **13**: 313–316, 1989.
38. Katoaka K, Mabobe S, Nalcase S and Makita M, Determination of hippuric acids and *o*-, *m*- and *p*-methyl hippuric acids in urine by capillary gas

chromatography. *J Pharm Biomed Anal* **9**: 699–704, 1991.

39. Bray E, Clowes HG, Thorpe RC, White K and Wood PB, The fate of certain organic acids and amides in the rabbit. *Biochem J* **50**: 583–587, 1950.
40. Spahn-Langguth H and Benet LZ, Acyl glucuronides revisited: is the glucuronidation process a toxification as well as a detoxification process? *Drug Metab Rev* **24**: 5–48, 1992.
41. Quinlan JR, Induction of decision trees. *Machine Learning* **1**: 81–106, 1986.