Methods for Assessing the Effects of Mixtures of Chemicals
Edited by V. B. Vouk, G. C. Butler, A. C. Upton, D. V. Parke and S. C. Asher

Structure–Activity Relationships within and between Chemical Classes

Bernard D. Astill

1 INTRODUCTION

The idea has been current for many years that the conventional classes of chemicals, particularly organic chemicals, may provide a useful means of classifying biological and physiological activity due to chemicals, and possibly their toxic potential. Thus organic acids might be held to be irritants, halohydrocarbons to be potential hepatotoxins, aromatic amines to be carcinogens, acid chlorides to be corrosive to skin, eye, the lung, and so forth (Table 1). The idea is superficially attractive because of the tendency, common to chemists and biologists, to group scientific phenomena in terms of structure and behaviour. Chemists have traditionally thought of chemistry in terms of the broad divisions and categories of organic chemistry, such as aliphatic, aromatic, heterocyclic compounds, the subgroups of acids, alcohols, etc., and in terms of the impressive structure of the periodic table of the elements. Many of these classes and subgroups are, however, very large in terms of the number of chemicals which can be allocated to a classification. In addition, many chemicals, because of the variety of their substituents, may belong to several classes. Further, the range and types of biological activity within a class may be very considerable; substituents in a basic molecular type may enhance, change, diminish, or abolish biological activity. Activity may be exhibited by only one or two members of a class, and is frequently associated with lower molecular weight compounds rather than with higher molecular weight homologues; in a sense the activity may be idiosyncratic. Finally, metabolic changes to chemicals can frequently be responsible for biological activities which are characteristic of other groups of chemicals, rather than of the original classification of a chemical. For these reasons, therefore, it is clear that, by themselves, conventional chemical classes offer only a very approximate indication of the types of biological activity to be expected from their constituent members.

An alternative approach would be to consider chemical classifications which have arisen in a more or less empirical manner, and which refer to function or use. Such classes would include alkylating agents, organophosphate and

Table 1 Some chemical classes and associated health effects

Chemical class	Health effects associated with one or more substances in the class
Inorganic acids, alkalies, acid chlorides, anhydrides	Skin, eye, lung, and mucous membrane irritation and corrosion
Halogens	Skin, eye, lung, and mucous membrane irritation, sensitization, and corrosion
Metal and metal oxide dusts	Pulmonary sensitization, pneumoconiosis, fibrinogenesis, carcinogenesis, skin and eye irritation and skin sensitization
Aliphatic hydrocarbons	Skin and eye irritation, neuropathy, narcosis, asphyxiation
Aliphatic ketones	Skin and eye irritation, neuropathy
Aliphatic acids, acid chlorides, anhydrides	Primary irritation, sensitization, corrosion
Aliphatic esters, alcohol, ethers	Primary irritation, central nervous system (CNS) depression, narcosis
Aliphatic halohydrocarbons	Primary irritation, hepatotoxicity, nephrotoxicity, narcosis, mutagenesis, carcinogenesis, convulsions
Oxiranes, aziridines (alkylating agents)	Pulmonary irritation, mutagenesis, carcinogenesis
Aliphatic amines	Primary irritation, sensitization
Aromatic hydrocarbons	Primary irritation, bone marrow depression
Monocyclic	CNS depression
Polycyclic	Carcinogenesis
Halogenated aromatic hydrocarbons	Primary irritation, hepatotoxicity, neuropathy, CNS depression
Phenols	Primary irritation, sensitization, CNS depression, convulsions
Aromatic amines	Sensitization, methaemoglobinaemia, carcinogenesis
Aromatic nitro compounds	Methaemoglobinaemia, bone marrow depression, nephrotoxicity, sensitization
Organophosphate esters	Cholinesterase inhibition, neuropathy, respiratory paralysis

carbamate insecticides, dyestuff intermediates and precursors, antioxidants, synthetic monomers for polymers, heavy metals, and so forth, which usually arouse concern when encountered by health professionals. Such an approach, however, is difficult to systematize, because of the wide range of chemical structures encountered within any particular functional class.

At present, the largest obstacle to a systematic approach to the development of an intellectually compelling framework for structure–activity relationships is the general absence of knowledge of toxic mechanisms, and of the factors which

affect those mechanisms. In consequence, therefore, any systematic approach has been mostly confined to the development of structure–activity relationships in highly specific areas. These approaches are exemplified by the quantitative relationships between specific molecular features and particular types of activity, and the use of physicochemical parameters within a closely related series of compounds to predict activity. Such approaches (quantitative structure–activity relationships, QSARs) are highly sophisticated, imply the availability of, or the ability to acquire, experimental data, and are commonly of the greatest value in drug design. These approaches do little to alleviate the burden of the health professional confronted with the need to predict the likely toxic potential of a wide range of chemical substances. Many if not most of such substances may have received little testing beyond that for lethality and irritation.

Generally speaking, most toxicologists possess, or have recourse to, sufficient knowledge to examine the structure of a chemical and make a prediction as to its activity, based on experience, judgement, intuition and knowledge of its reactivity. There have been attempts to place the prediction process on a more systematic basis, and structure–activity relationships have been developed for selected effects such as carcinogenesis, mutagenesis, and teratogenesis. These relationships range from positive or negative predictions to ranges of activity but at present their general acceptance is slow. They require computerized analysis of structures and machine-readable data bases to operate. The development of such procedures also usually requires the existence of relatively large data bases. These are barely adequate for the existing procedures, and need to be developed for most other toxic phenomena.

Under these circumstances, therefore, the development of structure–activity relationships in predicting toxic potential must be regarded as an empirical process which depends on flexible classification schemes. Such schemes should promote judgement, permit modification, and allow for collaborative expert knowledge, and also allow for the interpolation of QSAR treatments or computerized predictions when available.

This report will list and discuss a chemical classification scheme which may be of use in toxicology predictions, review aspects of QSARs which are applicable to toxicity prediction, and discuss briefly some other approaches that have been used in this context.

2 STRUCTURE GROUPINGS AND ACTIVITY CLASSES

A number of generalized approaches to structure class and activity have been developed. These depend on the identification of small subgroups, on the possession or absence of substructures in a molecule, or on the possession of certain types of reactivity. It is useful to consider a multiphasic approach which will allow for the use of selected chemical classes or substituents, multiplicity of fused rings, the effect of chain branching, and so forth.

2.1 Role of Substructures

Substructures need to be chosen on the basis of the frequency of occurrence in compounds which exhibit a certain type of effect, and also need to be related to that effect negatively, i.e. in their absence the effect does not occur. It is clearly necessary to achieve some sort of compromise in the selection of classes and substructures; thus certain substructures, such as hydroxyl, amino, nitro, carbonyl, carboxyl, and sulphydryl, possess chemical reactivity which may be expressed biologically if their molecular milieu promotes their interaction with biological molecules.

On the other hand, much larger assemblages of atoms, in themselves independent molecular entities, possess characteristic activities (e.g. polycyclic hydrocarbons, chlorinated *p*-benzodioxins, aromatic amines, hexocarboketones, and alcohols). While a classification based on the empirical selection of classes and subgroups may yield results lacking in precision, the results may, however, be somewhat better than guesses. The alternative is to ignore the presumption that chemicals which possess structural features associated with adverse responses may themselves exhibit those responses to some extent.

2.2 Activity Classification

Accompanying the process of identifying key structures or substructures is the need to define in a reasonably systematic manner the associated toxic effects. It seems particularly important to do this, to avoid having merely a collection of molecular entities to which random suspicion is attached. It is proposed to focus attention on human health effects for which there have been attributable chemical causes. Animal testing in general is directed toward providing models by which to judge the effect of chemicals on humans. While much animal testing is directed to exploring mechanisms of toxicity, there is a considerable amount of testing which is carried out to define suspicion aroused by structural consideration. A wide range of chemically induced health effects has been described in humans, and generally speaking these have been associated with one or more target organs (Table 2). These effects may in turn be classified in a number of ways depending on the type of impairment, the time of onset of the effect, and its reversibility or irreversibility. Overall, we may consider the highest level of concern to be directed towards those effects which are life threatening. With these considerations in mind, it is proposed that the types of activity most usefully related to chemical classes or substructures would be the following: carcinogenesis, mutagenesis, adult reproductive failure, foetotoxicity, teratogenesis, nervous system disorders, pulmonary disorders, disorders of the immune system, hepatic, gastrointestinal and renal disorders, blood and bone marrow disorders, skin, eye and mucous membrane disorders. Clearly, there are other chemically induced disease classifications, but it is assumed that they are limited in number,

Table 2 Target organs and chemically induced effects in humans

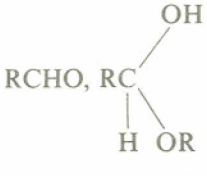
Target organ or tissue	Effects	Target organ or tissue	Effects
Skin	Altered appearance Irritation Sensitization Corrosion	Musculoskeletal system	Osteoporosis
		Liver	Mixed function oxidase induction Cholestasis Neoplasia Adenoma, carcinoma Cirrhosis Necrosis
Eye	Irritation Corneal opacity Retinal damage Corrosion		
Mucous membrane	Irritation Corrosion	Kidney	Aminoaciduria, proteinuria Uraemia Lithiasis
Lung	Irritation Sensitization Fume fever Pneumoconiosis Fibrosis Neoplasia Adenoma, carcinoma Asphyxiation	Reproductive system	Germ cell mutation Embryotoxicity Teratogenesis Infertility
		Blood and haematopoietic system	Methaemoglobinaemia Bone marrow depression Aplastic anaemia Leukaemia
Nervous system	Behavioural changes Peripheral neuropathy Central nervous system depression Cholinesterase inhibition Locomotor ataxia Narcosis Convulsions Respiratory paralysis	Immune system	Immune suppression

or more associated with specific chemicals rather than with classes. In considering effects associated with structures, both human and animal data are used.

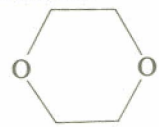
2.3 Substructures and Health Effects

A listing of substructures with which one or more of the above impairments have been associated can be constructed. It is not proposed that the effects are general features of all members of a class or substructure, but rather that some members have exhibited the activity or activities with them in a number of instances. Factors which tend to diminish or alter an effect are sometimes given. In general, increase in molecular weight is an ameliorating factor in toxicity, since the movement of a chemical across biological membranes will usually tend to diminish, or at least follow a parabolic relationship. Metabolic considerations may also be involved, in that strongly ionized low molecular weight materials may be rapidly eliminated unchanged, higher molecular weight lipophilic materials tend to undergo biotransformation, higher molecular weight absorbed materials may be eliminated by the enterohepatic system, and still higher molecular weight materials may not be absorbed at all.

It is proposed that the absence of a class or substructure from such a list would indicate that it is of a lower level of concern than those on the list with respect to the impairments involved. However, individual chemicals may be implicated in these and other impairments without it necessarily being felt that a general structural relationship exists. The list is not complete; it may be regarded as exemplary rather than comprehensive. Because much toxicological work has been directed towards a few key areas such as carcinogenicity, a much larger number of structures associated with carcinogenic potential is seen than for any other effect. It would be incorrect to conclude that the distribution of carcinogenic substructures amongst the active substructures presented here corresponds to the distribution of carcinogenic subgroups amongst all substructures.

Group or structure	Suspected impairment or organ affected
Acetals and aldehydes 	Carcinogenesis, mutagenesis, foetotoxicity, skin and mucous membrane
(predominantly lower MW aliphatics)	
Unsaturated aldehydes	Carcinogenesis (promotion)
Aromatic aldehydes	Skin (sensitization)

Group or structure (*Cont'd*)	Suspected impairment or organ affected (*Cont'd*)
Acid halides RCOX	Skin and mucous membrane
Carbamoyl halides NCOX	Carcinogenesis, mutagenesis, foetotoxicity; skin and mucous membrane
Aliphatic alcohols: saturated ROH (predominantly lower MW)	Carcinogenesis (promotion), foetotoxicity, reproductive failure, neurotoxicity, hepatic impairment
Aliphatic alcohols: unsaturated $R \cdot CH{=}CH \cdot CH_2OH$ Conjugation usually decreases epoxidation.	Carcinogenesis (after metabolism to epoxide)
Alkylating agents— see epoxides, aziridines, phosphamides, alkyl sulphonates, alkyl sulphates, lactones	
Amines: aliphatic RNH_2	Nervous system, skin and mucous membrane
Amines: aromatic RNH_2 (Electron-donating substituents may enhance carcinogenicity, electron-withdrawing may diminish. Toxicity may be enhanced by additional amino or nitro groups, and is usually retained with the addition of fused aromatic rings.)	Carcinogenesis or mutagenesis (after metabolism), nervous system, blood and bone marrow, skin and mucous membrane, hepatic and renal damage
Anhydrides: carboxylic $(RCO)_2O$ (usually low MW aliphatic)	Skin and mucous membrane
Azo and azoxy compounds N=N, N=N ↓ O (Toxic effects may be related to metabolism; electron-withdrawing substituents usually diminish toxicity, as do water-solubilizing groups.)	Carcinogenesis or mutagenesis (after metabolism), foetotoxicity (?), blood and bone marrow
Aziridines N $H_2C{-}{-}{-}CH_2$	Carcinogenesis, mutagenesis, foetotoxicity, adult reproductive performance impairment

Group or structure (*Cont'd*)	Suspected impairment or organ affected (*Cont'd*)
Boron, borates, boranes B, B_nH_{n+2}, BO_3^{3+}	Nervous system, skin and mucous membrane
Carbamates O ‖ RNC–O–R H	Carcinogenesis, foetotoxicity, nervous system, immune system, renal damage
Carboxylic acids	Skin and mucous membrane
Dioxanes 	Nervous system, hepatic and renal damage, skin and mucous membrane
Ethers: aliphatic ROR (lower MW)	Nervous system
Glycols, glycol ethers $-OCH_2 \cdot CH_2 \cdot OR$	Teratogenicity, nervous system, blood, hepatic and renal, reproductive performance
Haloaliphatic amines $X-(CH_2)_R N-$	Carcinogenesis, mutagenesis, foetotoxicity, skin and mucous membrane
Haloformic esters $XCO \cdot OR$	Skin and mucous membrane
Halogenated aliphatic hydrocarbons	
Saturated $-C \cdot X$	Carcinogenesis, mutagenesis, nervous system, hepatic, skin and mucous membrane, adult reproductive failure
Unsaturated $-C=C-X$	Carcinogenesis, mutagenesis, immune system, hepatic and renal damage
Dihalo aliphatics (α–ω)	Carcinogenesis, mutagenesis
(Increasing molecular weight usually diminishes reactivity and toxicity.)	
Halogenated aromatic hydrocarbons $R \cdot X$	Carcinogenesis, nervous system, immune system, hepatic damage

Group or structure (*Cont'd*)	Suspected impairment or organ affected (*Cont'd*)
Halogenated alicyclic hydrocarbons	Carcinogenesis, mutagenesis, nervous system, immune system, reproductive failure
Halo ethers $ROCH_2X$	Carcinogenesis, mutagenesis
Heterocyclic compounds: –N-containing	
Pyrroles and pyrrolidines	Skin and mucous membrane
Pyridines	Nervous system, skin and mucous membrane
Morpholines	Skin and mucous membrane
Quinolines	Carcinogenesis, mutagenesis, skin and mucous membrane
Phthalimides	Foetotoxicity, mutagenicity
s-Triazines	Renal
Acridines	Carcinogenicity
Phenanthridines	Carcinogenicity
Hydrocarbons	
Aliphatic	Skin and mucous membrane
C_6 straight chain	Nervous system
Aromatic: monocyclic	Carcinogenesis, blood and bone marrow, skin and mucous membrane
polycyclic	Carcinogenesis
Hydroxylamines –NHOH	Carcinogenesis, blood and bone marrow
Hydrazines $NHNH_2$	Carcinogenesis, blood and bone marrow, hepatic, skin and mucous membrane
Isocyanates, aryl –N=C=O	Pulmonary, skin and mucous membrane
Ketones RCOR	Skin and mucous membrane
C_6 ketones and precursors of 2,5-hexanedione	Nervous system
Lactones (strained ring, unsaturated)	Carcinogenesis, mutagenesis

Group or structure (*Cont'd*)	Suspected impairment or organ affected (*Cont'd*)
Metals: Target organs, effects on reproduction and carcinogenesis are reviewed by Beliles (1975).	
Nitriles (cyanides) –CN	Nervous system, blood and bone marrow
Nitrites, alkyl O · NO	Nervous system, blood and bone marrow
N-Nitrosamines RRN · NO (Usually restricted to dialkyl or mixed nitrosamines; aryl amines are less active or inactive.)	Carcinogenesis, mutagenesis, hepatic damage
Nitro compounds, aromatic $-NO_2$	Carcinogenesis, mutagenesis, foetotoxicity, blood and bone marrow depression, methaemoglobinaemia
Organophosphates	Nervous system, skin and mucous membrane
Phenols R · OH Chlorinated phenols (Fused ring homologues have decreased activity; electron-withdrawing substituents usually decrease activity.)	Nervous system, pulmonary, skin and mucous membrane Reproductive failure, nervous system
Quinones (more specifically anthraquinone)	Carcinogenesis, mutagenesis, skin and mucous membrane
Steroids	Foetotoxicity, nervous system, reproductive failure, immune suppression
Sulphates, alkyl $(RO)_2SO_2$	Carcinogenesis, mutagenesis, skin and mucous membrane, sensitization
Sulphonates, alkyl RSO_2OR	Carcinogenesis, mutagenesis
Sulphides and disulphides SH, –SSH Chloroalkyl sulphides: XCH_2CH_2S-	Reproductive failure, nervous system, skin and mucous membrane Carcinogenesis, mutagenesis, pulmonary, skin and mucous membrane
Urethanes—see carbamates	

3 QUANTITATIVE STRUCTURE–ACTIVITY CONSIDERATIONS

Extensive attempts have been made to establish correlations between structural and physicochemical features and biological activities in the area of drug design, a process which has been facilitated by the use of computers. Structural fragments and physicochemical properties can be rapidly correlated in a manner not possible in the precomputer epoch (Olson and Christofferson, 1979). Two general approaches have emerged: the Hansch approach (van Valkenburg, 1972), useful principally in dealing with congeneric series in which a single substituent is changed systematically; and the Free–Wilson approach (Free and Wilson, 1964), which treats a wide range of substituents in an additive manner. The Hansch approach usually requires extensive physicochemical data, and the Free–Wilson approach requires a wide range of chemical structures to avoid singularity.

The Hansch procedure is exemplified by the use of octanol–water partition coefficients. The ability to cross a biological membrane and thus interact with receptor sites is clearly a key factor in the biological activity of a chemical, and is related to the distribution of a chemical in a water–octanol system. If this is expressed on a numerical scale as log P, a relationship emerges between log P and log B (some biological activity) which is roughly parabolic. That is, at lower log P values in a homologous series, a direct relationship exists between log P and log B, but at higher log P values activity will decrease. This expresses a physical reality: increasingly negative log P values imply increasing water solubility, so that a compound will localize in aqueous compartments, and increasingly positive log P values imply greater lipid solubility, so that a compound localizes in the first lipid compartment it encounters. Thus aromatic hydrocarbons have high log P values, but show decreasing activity as log P increases (Hansch and Fujita, 1964). A direct relationship between log P and some toxicity value is thus unlikely, and cannot be assumed without knowledge of the parabolic region. In addition, the parabolic region is also likely to be the area where phase I metabolic reactions will be encountered, and thus produce activities not encountered in regions of lower or greater lipophilicity.

In general, therefore, the non-additive parabolic approach is unlikely, except in special circumstances, to be useful in a general structure–activity prediction process for toxicity. The range of chemicals studied is at present insufficient for most of the parameters employed, even though a range of such parameters (Table 3) has been proposed (Hopfinger *et al.*, 1981). To be of value ultimately, knowledge of mechanisms and extensive measurements on congeners will be needed.

The additivity approach has been exemplified by the approach of Free and Wilson (1964), in which *de novo* values are assigned to various substituents using regression analysis. It is postulated that:

$$\text{Biological activity} = a\text{A} + b\text{B} + c\text{C} \ldots + n\text{N}$$

Table 3 Molecular descriptors used in toxic QSAR studies analogous to popular medicinal structure–activity studies

Descriptor	Explanation
Log P (π, f)	Water–1-octanol partition coefficient
MR	Molecular refractivity
σ	Hammett's electronic factor
Q	Del Re charge density
E_s, V	Taft's steric parameter and volume parameters
R_v	Van der Waal radius
X_μ^v	Molecular connectivity indices
F_{H_2O}	Aqueous free energy
F_{oct}	1-Octanol free energy
B, C, D	Principal component physicochemical descriptors
I_x	Indicator variable for property x
S_x	Structure fragments

From Hopfinger *et al.* (1981).

where A, B, C are substituents (or substructures) and a, b, c are coefficients selected to weight the contributions of A, B, C to a particular type of activity. The equation works best where there is a commonality of mechanism for all compounds in which A, B, C, etc., are substituents, even if the mechanism is unknown. Even if mechanisms differ, some level of correlation may be obtained which is usable.

For a given type of activity, as data are accumulated, weighting factors are constantly under review, and additional substituents may be introduced. A peculiarity of the system is that substructure analysis may result in structural features (e.g. numbers of fused rings, tertiary C atoms, types of linkage) which have no clear biological or chemical role in the activity under consideration. Additionally, weighting factors may be arbitrary, depending on the data base, and thus may be formally opposite to conventionally expected values. In this sense, therefore, substituents, structural features and weighting factors may be unreal in terms of direct structure–activity experience.

Examples of the additivity approach are the LD_{50}, mutagenesis, carcinogenesis and teratogenesis predictions due to Enslein and Craig (1982a) and described in their various publications. These employ regression analyses of structural features based on available data bases. The mutagenesis data base is derived from the Environmental Mutagenesis Information Center file, using results in the Ames test with five strains and applying criteria from the EPA Gene-Tox Program. The carcinogenesis data base is derived from the Reviews of the International Agency for Research on Cancer (IARC), using the categories of

definite (definite animal, definite human) and indefinite (negative and indefinite). The IARC categories of suspected animal and suspected human are unused in the Enslein and Craig model, Teratogenesis Centre file, and the information collections of Shepard (1980) and Schardein (1976). It is to be noted that data in these compendia are carefully evaluated by a variety of acceptable criteria. Other data bases susceptible to this treatment would be fish toxicity, bioconcentration (partition coefficient), biodegradation (BOD), the developing data bases in short-term tests, and, possibly, irritation and sensitization.

Verification of additivity type estimates is proceeding; most recently, of 908 compounds tested in the LD_{50} model, 80% were ± 5 times the measured value. In the mutagenesis model, 50 compounds were tested, and 82% were correctly identified, with 14% false negatives and 5% false positives. Of 35 compounds tested in the carcinogenesis model, 71% were correctly classified (Enslein and Craig, 1982b). It should be noted that the additivity estimate in these applications is generally only semiquantitative in manner; thus results tend to be positive or negative for a given end-point, e.g. carcinogenicity, mutagenicity and teratogenicity.

A variant of the additivity principle is that developed by Hodes *et al.* (1977) and applied to screening for antitumour activity. The algorithm depends on the selection of arbitrarily defined fragments in a computer-analysable data base of chemical structures, with a regression analysis to relate the contribution of such fragments to a particular type of activity. As employed by Hodes, a data base of 1000 compounds was successfully assembled, to give good predictability for the controlled end-point used (antitumour activity). This process has been extended by Tinker (1981) to bacterial mutagenesis prediction. The data base employed Ames test data for 1000 compounds selected from the scientific literature, and structures given in Chemical Abstracts Connection tables or Mechanical Codes Ciphers. The structural feature selected was any assemblage of three connected non-hydrogen atoms with terminal bonds; clearly such a substructure could possess only a coincidental biological activity. Unit(s) would be weighted for occurrence in active or inactive molecules. Results for 34 compounds tested in the programme and independently in the Ames test gave an 88% agreement within one category of assigned activity. It should be noted that the activity predictions can be obtained in categories or levels of activity, an output which is somewhat more quantitative than the Craig–Enslein approach.

The Hodes–Tinker process would be amenable to applications similar to the Enslein and Craig additive procedure (i.e. carcinogenesis, teratogenesis, LD_{50}, fish toxicity, partition coefficient, etc.), and evaluation of such procedures is currently under way. In general, data bases with uniform or accepted procedures to define an activity, combined with adequate distributions of positives and negatives, are amenable to these semiquantitative approaches to toxicity prediction. Both extant procedures, however, need more extensive validation, and the absence of data bases beyond those available at present limits the application of these techniques.

4 OTHER APPROACHES

Extensive use of structural features has been made in a decision tree approach worked out by Cramer *et al.* (1978). The tree makes a decision as to whether a chemical belongs in one of three categories (I, low concern; II, moderate; III, serious concern), and employs exposure and oral toxicity data in making a decision about the final category for a chemical. However, carefully selected chemical structural features are employed to make an initial decision. Such features include those conferring suspicion of toxicity, such as organophosphate, secondary amines, azo linkages, unsaturated lactones, heavy metals, strongly ionized groupings, etc. Most of the functions included in part 2.3 above play some role in the decision tree, and commonly lead to a category II or category III designation. Features which confer little or no suspicion of toxicity are also employed, such as aliphatic hydrocarbons and carbohydrates. The ranking is intended to lead to the appropriate level of toxicity testing. It would seem that the decision tree could be used in a classification system, although the advantage to be derived thereby over a less rigorous inspection approach is unclear. In addition, the scheme gives no guidance as to the type of toxicity to be expected in categories II or III.

A further descriptive approach has been recently evolved by the US Food and Drug Administration (USFDA, 1982) in setting up structure classes of food additives to establish the proper level of concern for their safety. Similarities to known toxicants are used as criteria, and the assignment to a category includes possible or known metabolites, components of mixtures and contaminants. A decision tree approach is also employed after functional groups have been evaluated. Three categories of concern are used: A, low probable toxicity; B, intermediate or unknown probable toxicity; and C, high probable toxicity. The categorization process will require an expert committee. Three classes of substructures are set up (as Tables A, B, and C) which correspond to categories A, B, and C. Nine groups of substructures of apparently low toxicity entities are set up, including aliphatic hydrocarbons, fatty acids, polysaccharides, monofunctional aldehydes, ketones, etc. A second set of eight intermediate level functions include non-conjugated olefins, inorganic salts of certain metals, amino acids, proteins and polypeptides. The third group includes a wide range of active substructures and classes, including aziridines, unsaturated lactones, organohalides, aromatic amines, a range of heterocycles, carbamates, etc. However, allocation of structures to a type of activity is not included.

5 CONCLUSIONS

A variety of schemes for relating structure to toxic activity has been reviewed, ranging from highly specific QSAR treatments to decision tree approaches to ranking for testing. All possess deficiencies which are inherent in our present

incomplete knowledge of toxic mechanisms. It is noteworthy that most success has been achieved where data bases deal with relatively few closely related congeners, or a large number of disparate structures, related to a well-defined end-point. However, a very wide range of information on chemicals is now available, as seen in collections such as *Patty's Industrial Hygiene and Toxicology* (Clayton and Clayton, 1981). A systematic attempt should be made on a large scale to classify this information in terms of (1) selected key health effects, (2) active substructures, (3) substructures which diminish or alter activities, and (4) modifiers of metabolism. Whenever possible, computerization should be employed, with structural analyses undertaken, by either the additive (weighting) or pattern recognition procedures.

6 REFERENCES

Beliles, R. P. (1975). Metals. In Casavett, L. J., and Doull, J. (Eds.) *Toxicology*, pp. 454–502. Macmillan, New York.

Clayton, G. D., and Clayton, F. E. (1981). *Patty's Industrial Hygiene and Toxicology*, Vols. 2a, b, c. John Wiley & Sons, New York.

Cramer, G. M., Ford, R. A., and Hall, R. L. (1978). Estimation of toxic hazard: a decision tree approach. *Food Cosmet. Toxicol.*, **16,** 255–276.

Enslein, K., and Craig, P. N. (1982a). Carcinogenesis: a predictive structure activity model. *J. Toxicol. Environ. Health*, **10,** 521–530.

Enslein, K., and Craig, P. N. (1982b). Validation of a set of predictive structure activity models of toxicity endpoints. In *Proceedings of the American Chemical Society Meeting, Kansas City, Missouri.*

Free, S. M., and Wilson, J. W. (1964). A mathematical contribution to structure activity studies. *J. Med. Chem.*, **7,** 395–399.

Hansch, C., and Fujita, T. (1964). Analyses. A method for correlating biological activity and chemical structure. *J. Am. Chem. Soc.*, **86,** 1616–1626.

Hodes, L., Hazard, G. F., Geran, R. I., and Richman, S. (1977). A statistical–heuristic method for automated selection of drugs for screening. *J. Med. Chem.*, **20,** 469–475.

Hopfinger, A. J., Pearlstein, R., Potenzone, R., Kibuchi, O., and Klopman, G. (1981). Devising descriptors for toxic structure activity studies. In McKinney, J. D. (Ed.) *Environmental Health Chemistry*. Ann Arbor Science, Ann Arbor, Michigan.

Olson, E. C., and Christofferson, R. E. (1979). *Computer-assisted Drug Design*. ACS Symposium Series 112. American Chemical Society, Washington, DC.

Schardein, J. L. (1976). *Drugs as Teratogens*. CRC Press, Cleveland, Ohio.

Shepard, T. H. (1980). *Catalog of Teratogenic Agents*, 3rd edition. Johns Hopkins University Press, Baltimore, Maryland.

Tinker, J. F. (1981). Relating mutagenicity to chemical structure. *J. Chem. Inf. Comput. Sci.*, **21,** 3–7.

USFDA (1982). *Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food.* Bureau of Foods, US Food and Drug Administration.

van Valkenburg, J. W. (1972). Biological correlations—the Hansch approach. *Adv. Chem. Ser.*, **114,** R7–R8.