Journal of Computer-Aided Molecular Design **17:** 119–127, 2003.

In silico prediction of drug toxicity

John C. Dearden
School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, England

Received 6 November 2002; Accepted for publication 10 December 2002

Summary

It is essential, in order to minimise expensive drug failures due to toxicity being found in late development or even in clinical trials, to determine potential toxicity problems as early as possible. In view of the large libraries of compounds now being handled by combinatorial chemistry and high-throughput screening, identification of putative toxicity is advisable even before synthesis. Thus the use of predictive toxicology is called for. A number of *in silico* approaches to toxicity prediction are discussed. Quantitative structure-activity relationships (QSARs), relating mostly to specific chemical classes, have long been used for this purpose, and exist for a wide range of toxicity endpoints. However, QSARs also exist for the prediction of toxicity of very diverse libraries, although often such QSARs are of the classification type; that is, they predict simply whether or not a compound is toxic, and do not give an indication of the level of toxicity. Examples are given of all of these. A number of expert systems are available for toxicity prediction, most of them covering a range of toxicity endpoints. Those discussed include TOPKAT, CASE, DEREK, HazardExpert, OncoLogic and COMPACT. Comparative tests of the ability of these systems to predict carcinogenicity show that improvement is still needed. The consensus approach is recommended, whereby the results from several prediction systems are pooled.

"*It is simply amazing that we can formulate any kind of QSAR. The (desired activity) is only the starting point. The truly formidable problem is that of toxicity, especially the difficult long-term toxicities resulting from chronic usage*". (Hansch & Leo [1])

Introduction

Traditionally, the search for new drugs has concentrated on the required activity, with considerations of bioavailability and toxicity being left until later in the development process. Numerous examples exist of drugs that have had to be withdrawn, because of unacceptable toxicity, in clinical trials and even after reaching the market-place. Such failures are very expensive, and have given rise to the saying; 'Fail early, fail fast, fail cheap(ly)'. Kennedy [2] reported that in 1997, of new drug entities that failed for reasons other than lack of efficacy, 16% failed in animal toxicity testing, and 14% failed because of adverse effects in man. Since a typical drug takes 10–12 years, and costs up to $500 million, to reach the market, it is clearly important to discover potential toxicity as soon as possible.

For that reason, in recent years pharmaceutical companies have brought toxicity testing, as well as ADME (absorption, distribution, metabolism, excretion) studies, earlier in the drug development process. The ultimate here would be to use computer-based (*in silico*) methods to predict toxicity even before a drug candidate was synthesised. There are, however, difficult problems to be overcome in this regard. Firstly, 'toxicity' covers a wide range of adverse effects; secondly, there is a paucity of data concerning, in particular, chronic toxicities, especially in humans; thirdly, the *in silico* methods currently available are class-specific and/or are of insufficient accuracy. There is little evidence yet that pharmaceutical companies are widely engaged in *in silico* toxicity prediction; for example, of the 208 presentations made at the 2002 European Symposium on QSAR, only 5 dealt with the prediction of drug toxicity.

* E-mail: j.c.dearden@livjm.ac.uk

This review paper examines the various approaches that have been and are being taken for *in silico* prediction of drug toxicity.

Quantitative structure-activity relationships (QSARs)

QSARs for congeneric series

A QSAR is a statistically derived rule that quantitatively describes a molecular property in terms of descriptors of chemical structure. The basis of QSAR was set out by the statement by Crum Brown and Fraser [3] in 1868–69 that 'there can be no reasonable doubt but that a relation exists between the physiological action of a substance and its chemical composition and constitution'. However, the modern science of QSAR can be said to date from the seminal 1962 publication of Hansch et al. [4], which established, *inter alia*, the use of multiple linear regression as an important tool for the correlation of biological activity with descriptors of chemical structure.

In order for a QSAR to model biological data well, and to be predictive, it is generally acknowledged that all the compounds involved should act by the same mechanism, since the physico-chemical and structural descriptors used in the QSAR are deemed to reflect mechanism of action. Since it is difficult to determine mechanism of action, the compounds involved in a QSAR correlation are usually restricted to a given chemical class in the hope that this will ensure a single mechanism of action.

Generally, the prime aim in developing a QSAR is so that it can be used for predictive purposes. It is therefore important that the statistics given with the QSAR give an indication of its predictivity. This is achieved by the use of either internal validation (cross-validation) or external validation (use of a test-set). The latter is to be preferred, but is not always possible because of, for example, the small size of the data-set. It has to be said that very few published QSARs, especially those published more than about five years ago, meet this criterion.

Another important point regarding the usefulness of a QSAR for toxicity is that the toxicity should be considered in relation to the desired activity. For example, a compound may be predicted to be toxic at a dose of 1 μmol kg^{-1}, but whether or not that is acceptable depends on the therapeutic dose; if the latter is 10 nmol kg^{-1}, the toxicity could be acceptable, but if the therapeutic dose is 10 μmol kg^{-1}, the toxicity is unacceptable. In other words, it is the therapeutic ratio, not toxicity alone, that should be considered. Sadly, most published toxicity QSARs do not address this point.

It is a curious fact that most of the published QSAR studies of drug toxicity were made in the period 1970–1990. Has the QSAR community, by and large, lost interest in drug toxicity prediction? It is true that, with the recent advent of combinatorial chemistry, there is a much greater interest in screening diverse libraries, which are not so amenable to classical QSAR analysis. Another reason may be the industry's preoccupation with the search for ever-increasing potency of drugs.

Published drug toxicity QSARs cover over 30 different end-points, from carcinogenicity to neurotoxicity to gastric irritancy. It is not possible here to discuss all of these, but it is appropriate to look at a number of examples. In 1973 Hansch and Clayton [5] modelled the acute toxicity of barbiturates to the mouse using only the octanol-water partition coefficient (P), a measure of hydrophobicity:

$$\log 1/\mathrm{LD}_{50} = 1.02\log\mathrm{P} - -0.27(\log\mathrm{P})^2 + 1.86 \quad (1)$$

$$n = 13 \quad r^2 = 0.852 \quad s = 0.113$$

where LD_{50} = dose to kill 50% of mice, n = number of compounds used in developing the QSAR (the training set), r = correlation coefficient, and s = standard error of the estimate.

The use of the $(\log \mathrm{P})^2$ term, and its negative sign, indicate that toxicity passes through a maximum as hydrophobicity increases. This is commonly observed [5], and results from the differing ability of drugs with different hydrophobicities to enter and leave lipid membranes en route to the site of action.

It is rare to find human drug toxicity studies. One such was carried out by King and Moffat [6], who used as a toxicity index (TI) of barbiturates the number of U.K. deaths in a given year from a specified barbiturate divided by the number of prescriptions issued for that barbiturate:

$$\log\mathrm{TI} = 1.48\log\mathrm{P} - 1.28 \quad (2)$$

$$n = 5 \quad r^2 = 0.983 \quad s = 0.078$$

Although the data-set is very small, the statistics are remarkably high, reflecting probably the accuracy of the data. The limited log P range of the 5 barbiturates probably explains the lack of observed biphasic dependence on hydrophobicity.

Numerous QSAR studies have been published of the toxicity of anti-tumour drugs, and frequently it was found that toxicity was closely related to the anti-tumour activity, as the following example [7], relating to aromatic nitrogen mustards, shows:

$$\log 1/ED_{50} = -0.263\pi_R - -1.623\sigma_R + 3.660 \quad (3)$$

$$n = 9 \quad r^2 = 0.704 \quad s = 0.448$$

$$\log 1/LD_{50} = -0.269\pi_R - 1.577\sigma_R + 3.476 \quad (4)$$

$$n = 9 \quad r^2 = 0.859 \quad s = 0.273$$

where π_R = hydrophobic constant of substituent R, and σ_R = Hammett constant of substituent R.

Equations 3 and 4 are so similar that there is clearly no possibility of designing a compound with a good therapeutic ratio. However, that statement must be qualified by an acknowledgment of the very small size of the data-set.

Some antibiotics and antifungal agents are known to possess haemolytic activity. A study of amides of amphoteracin B [8] showed that selective toxicity (ST) was a parabolic function of hydrophobicity:

$$\log ST = 0.125\Sigma\pi - 0.079(\Sigma\pi)^2 + 1.995 \quad (5)$$

$$n = 16 \quad r^2 = 0.992 \quad s = 0.060 \quad F = 778.6$$

where $\Sigma\pi$ = sum of hydrophobic substituent constants, and F = Fisher statistic.

The cytotoxicity of a series of aminodiol HIV protease inhibitors has been correlated [9] with hydrophobicity, measured as the HPLC retention index k':

$$\log CC_{50} = -1.189\log k' + 2.747 \quad (6)$$

$$n = 24 \quad r^2 = 0.802 \quad \text{s not given}$$

The authors commented that very hydrophilic analogues, with low cytotoxicity, were also poorly active against HIV-infected cells, but that compounds with intermediate polarity generally demonstrated good antiviral activity along with acceptable cytotoxicity.

Despite the importance of cardiotoxicity, there have been few QSAR studies in this area. Schön et al. [10] reported for a series of bradycardic tetraalkylbispidines, a QSAR for selectivity between bradycardia and contractility:

$$\log(\text{selectivity}) = 0.37MR_1 - 0.01(MR_1)^2 + 0.17MR_{3,4} - 0.0043(MR_{3,4})^2 + 0.43 \quad (7)$$

$$n = 16 \quad r^2 = 0.950 \quad s = 0.194 \quad F = 38.3$$

where MR_i = molar refractivity of substituent at position i.

From a host of other toxicity end-points, two recent examples are selected. The gastric irritation caused by non-steroidal anti-inflammatory drugs is due to inhibition of cyclo-oxygenase-1 (COX-1), so there is great interest in COX-1/COX-2 selectivity. A study of a series of 1,2-diarylimidazoles [11] reported the following correlation:

$$\log(IC_{50}(\text{COX-1})/(IC_{50}(\text{COX-2})) = -1.09E_{HOMO} + 1.00\Sigma\sigma + 29.8q_8 + 5.60q_1 - 0.667\pi_4 + 1.35 \quad (8)$$

$$n = 40 \; r^2(\text{adj}) = 0.710 \; Q^2 = 0.557 \; s = 0.420 \; F = 20.0$$

where IC_{50} = concentration producing 50% inhibition, E_{HOMO} = energy of the highest occupied molecular orbital, $\Sigma\sigma$ = sum of Hammett substituent constants, q_i = charge on atom i, π_4 = hydrophobic constant of substituent at position 4, $r^2(\text{adj})$ = square of correlation coefficient adjusted for degrees of freedom, and Q^2 = square of cross-validated correlation coefficient (leave-one-out procedure).

There is world-wide concern about malaria, particularly with the emergence of drug-resistant strains of *Plasmodium falciparum*. A study of a series of anti-malarial quassinoids [12] yielded QSARs for both anti-malarial activity and cytotoxicity, the latter of which was:

$$\log 1/ED_{50} = 0.411f_2 - 1.96I_{1-OH} + 1.05 \quad (9)$$

$$n = 13 \quad r^2 = 0.869 \quad s = 0.470 \quad F = 33.2$$

Where f_2 = hydrophobic fragmental constant for substituent at position 2, and I_{1-OH} = indicator variable for the presence of −OH at position 1. Thus, for low cytotoxicity, substituent R_2 should be hydrophilic and substituent R_1 should be −OH.

In 1987 Hansch et al. [13] propounded their principle of minimal hydrophobicity, which states that 'without convincing evidence to the contrary, drugs should be made as hydrophilic as possible without loss of efficacy'; the reason is that hydrophilic compounds do not accumulate readily in the body, as they are rapidly excreted. On the other hand, it should be noted that rates of metabolism generally increase with

hydrophobicity, as the body attempts to eliminate xenobiotics. An example is the microsomal demethylation of a group of miscellaneous drugs [14]:

$$\log 1/K_m = 0.693 \log P + 2.900 \quad (10)$$

$n = 14 \quad r^2 = 0.846 \quad s = 0.330$

It will have been noticed that many of the QSARs reported above contain a hydrophobicity term. Indeed, some 70% of all published QSARs include such a term. Consequently, much effort has gone into the calculation of octanol-water log P values, and a number of commercial software packages are now available for that purpose. Mannhold and Dross [15] published in 1996 a comparative study of 14 such packages; Dearden et al. [16] have recently updated that, and have also reported a similar study on commercially available packages for the calculation of aqueous solubility [17].

QSARs for non-congeneric series

It was stated earlier that all compounds used in a QSAR study should preferably act by the same mechanism, otherwise good correlations are unlikely. The vast majority of published QSARs have been developed from congeneric series, but with the recent use of large, diverse chemical libraries, there is increasing interest in modelling activities and toxicities of heterogeneous collections of compounds. Benigni and Giuliani [18] have suggested that it is often possible to model data from such collections, provided that one is prepared to accept a lower accuracy of prediction (Figure 1).This is not unreasonable, since, at least for an initial indication of toxicity, a classification of high, moderate or low may well be adequate. The words of Aristotle (384-322 B.C.) are apposite here: 'It is the mark of an instructed mind to rest easy with the degree of precision which the nature of the subject permits, and not to seek an exactness where only an approximation of the truth is possible'. In fact, there are a number of published QSAR studies based on non-congeneric data-sets, and several of these are outlined here.

Dearden et al. [19] modelled human skin permeability coefficients (K_p) of 91 diverse compounds:

$$\begin{aligned} \log K_p = {} & -0.626\Sigma Ca - 23.8\Sigma(Q^+)/\alpha \\ & - 0.289SsssCH - 0.0357SsOH - 0.482BI \\ & + 0.405nRB + 0.834 \end{aligned} \quad (11)$$

$n = 91 \quad r^2 = 0.832 \quad s = 0.563 \quad F = 69.2$

where ΣCa = sum of hydrogen bond acceptor abilities, $\Sigma(Q^+)/\alpha$ = sum of positive charge per unit polarisability (effectively positive charge per unit volume, since polarisability correlates highly with volume), SsssCH = total electrotopological state index for CH groups, SsOH = total electrotopological state index for hydroxyl groups, BI = Bonchev index, and nRB = number of rotatable bonds. The last-named is especially interesting, since it suggests that flexible molecules penetrate skin more quickly.

It can be argued that since most compounds, other than those that have an active transport mechanism, are transported across skin by the same or similar mechanisms, this is not a good example of a non-congeneric QSAR. However, the 91-compound data-set contains some compounds whose skin penetration is affected by P-glycoprotein (Pgp) efflux, but which nevertheless are well modelled by equation 11. Of course, Pgp efflux itself can be modelled, and is shown to be a relatively non-specific effect, as would be expected. The reduction R in blood-brain barrier penetration of a very diverse group of drugs, determined using Pgp-knockout and wild-type mice, was modelled using 2-D descriptors [20]:

$$\begin{aligned} \log R(-/+) = {} & 0.113\,{}^3\chi_p^v + 22.6\alpha/V \\ & -0.104Ct_{SdssC} - 0.0435N_{circ} \\ & -0.317 \end{aligned} \quad (12)$$

$n = 22 \quad r^2 = 0.854 \quad Q^2 = 0.788 \quad s = 0.182$
$F = 24.9$

where ${}^3\chi_p^v$ = 3rd order valence path molecular connectivity, α/V = polarisability per unit volume, Ct_{SdssC} = no. of C atoms forming one double bond, and N_{circ} = no. of circuits (all possible rings).

Mumtaz et al. [21] used 44 group contributions to model the lowest observable adverse effect level (LOAEL) of a diverse set of 234 chemicals; they obtained a good model, with $r^2 = 0.839$, $s = 0.411$ and $F = 28.6$. Barratt [22] used a 2-dimensional principal components plot to distinguish fully between 27 diverse organic acids that either did or did not give rise to skin corrosivity. Gao et al. [23] carried out a binary QSAR analysis of 410 diverse œstrogen receptor ligands using 13 descriptors; they were able to model actives with 87% accuracy and inactives with 95% accuracy.

The last example is perhaps of greatest significance for the modelling and prediction of toxicity of heterogeneous libraries of drugs. Even if, by virtue of the disparate nature of the compounds in a library, it were not possible to develop a QSAR model for the quantitative prediction of toxicity, then the classification approach exemplified above could be a realistic alternative, using as many classes (e.g. low, medium and high toxicity) as appropriate. This is already widely used in the modelling and prediction of, for example, biodegradability [24].

Expert systems

An expert system has been defined [25] as 'any formal system, not necessarily computer-based, which enables a user to obtain rational predictions about the toxicity of chemicals. All expert systems for the prediction of toxicity are built upon experimental data representing one or more toxic manifestations of chemicals in biological systems (the database), and/or rules derived from such data (the rulebase)'. Expert systems for the prediction of toxicity are broadly of four types: automated QSAR, knowledge-based systems, automated rule derivation and decision trees. A number of reviews have been published on the use of expert systems for toxicity prediction [25–29].

The automated QSAR system is exemplified by TOPKAT (www.accelrys.com). This software uses both continuous and dichotomous (binary) measures to predict a number of toxicity end-points, including carcinogenicity, mutagenicity, teratogenicity, irritation, allergic contact dermatitis (ACD), acute toxicity and Ah receptor binding; it also predicts skin permeability. Continuous end-points, such as LD_{50}, are modelled using multiple linear regression QSARs, whilst dichotomous measures, such as carcinogenicity, are modelled using linear discriminant regression. TOPKAT utilises large, heterogeneous data-bases with carefully selected data, and its QSARs employ mainly topological, sub-structural and electronic descriptors. It also provides an estimate of confidence in the prediction. TOPKAT can be operated in batch mode from SD file input.

Enslein et al. [30] reported on the use of TOPKAT in the prediction of carcinogenicity and mutagenicity, whilst Enslein et al. [31] discussed the development of a dermal sensitisation module in TOPKAT. The latter, based on a training set of 315 chemicals, yielded a cross-validated specificity (correct prediction of non-sensitisers) of between 81 and 91%, and a cross-validated sensitivity (correct prediction of sensitisers) of between 85 and 95%.

The CASE technology (www.multicase.com) uses a very different approach, through the creation of its own structural alerts. Each molecule is broken down into all possible fragments from two to ten heavy (non-hydrogen) atoms. These are then classified statistically as biophores (associated with toxicity) or biophobes (not associated with toxicity). The results are then combined into an equation of the form:

$$\begin{aligned}\text{CASE units} \;=\; &\text{constant} + \text{a(fragment 1)} + \\ &\text{b(fragment 2)} + \ldots . \qquad (13)\end{aligned}$$

It is able to take account of interaction between fragments, since large fragments automatically encompass smaller ones (for example, C-C(=O)-O-C-C contains C-C(=O)-O and O-C-C). The CASE suite, which can be used in batch mode, covers a similar range of endpoints to TOPKAT, including carcinogenicity, mutagenicity, teratogenicity, irritation, ACD, acute toxicity, CYP450 2D inhibition and cellular toxicity.

The developers of CASE, G. Klopman and H.S. Rosenkranz, have published very widely. A detailed exposition of CASE has been given by Rosenkranz et al. [32], and Cunningham et al. [33] have discussed its application to the prediction of carcinogenicity; they reported a cross-validated concordance of 71% between experimental and predicted rat carcinogenicity results for a 745-compound data-set. Gomez et al. [34] found a predictive performance for CASE of 74% correct predictions of developmental toxicity in hamsters for a diverse set of chemicals.

It may be noted that the CASE technology has been adapted by the U.S. Food and Drug Administration (FDA). The new software, known as MCASE QSAR-ES, was designed to improve the prediction of carcinogenic potential of pharmaceuticals. In a test of the system using a 934-compound data-set, Matthews and Contrera [35] found an overall concordance of 75%, with carcinogens being better predicted than non-carcinogens.

DEREK is a knowledge-based expert system originally devised at Schering, and now developed by LHASA Limited at the University of Leeds, UK (www.chem.leeds.ac.uk/luk). The development of DEREK is overseen by a collaborative group comprising representatives from commercial and educational organisations. It offers a similar range of endpoints to those of TOPKAT and CASE, including carcino-

genicity, mutagenicity, teratogenicity, irritation, skin sensitisation, acute toxicity and neurotoxicity, and, like them, can be operated in batch mode; it also offers an estimate of skin permeability. Its main strengths lie in the prediction of carcinogenicity, mutagenicity and skin sensitisation [29]. Greene et al. [36] have discussed the DEREK system, and have given an example of its performance regarding mutagenicity; for a diverse data-set of 266 chemicals, DEREK correctly predicted 84% of mutagens. Steger-Hartmann et al. [37] have reported on rule improvement in DEREK, whilst Barratt et al. [38] have highlighted the development of the DEREK rulebase for the prospective identification of photoallergens.

HazardExpert is another knowledge-based expert system, which is part of the Pallas suite of programs developed and marketed by CompuDrug Limited (www.compudrug.com). It too offers a range of end-points, including carcinogenicity, mutagenicity, teratogenicity, irritation, skin sensitisation, immunotoxicity and neurotoxicity. The user defines the species, route of administration, dose level and duration of exposure. Smithing and Darvas [39] have described the rationale behind HazardExpert, and its potential applications. Dearden et al. [25] reported a test of HazardExpert's ability to predict human and animal carcinogenicity; for 192 chemicals evaluated in IARC Monographs, 55% of the chemicals were predicted within $\pm$ 1 of the IARC classification when the 'high' exposure condition was chosen, rising to 75% when the 'low' exposure condition was chosen. Brown et al. [40] examined HazardExpert's ability to predict the carcinogenicity of 80 chemicals tested by the National Toxicology Program (NTP) of the US National Institutes of Health; it was found to be good at identifying non-carcinogens (81% correct predictions), but poor at identifying carcinogens (36% correct predictions).

OncoLogic is, as its name implies, designed to predict only carcinogenicity. It is a knowledge-based system developed and marketed by LogiChem (www.logichem.com), and uses an hierarchical, decision-tree structure for each of four separate sub-systems for estimating the carcinogenicity of fibres, metals and metal-containing compounds, polymers, and organics respectively [41]. The organics sub-system is the largest of these, with over 40,000 program rules based on over 10,000 organic chemicals. Within this sub-system there are distinct modules for about 50 chemical classes. A mechanism-based justification is provided for each carcinogenicity evaluation.

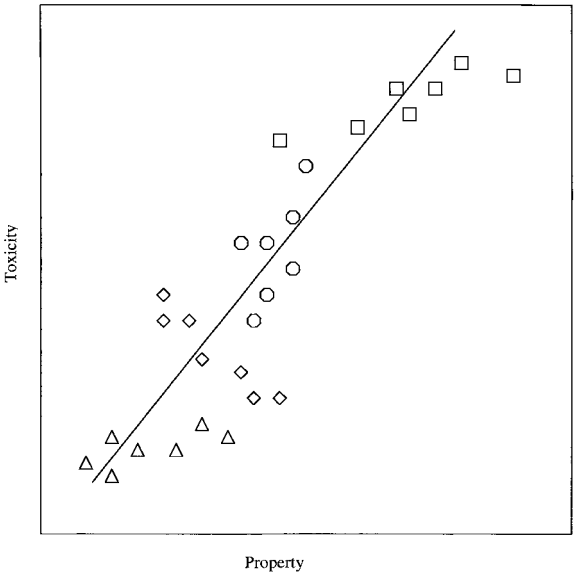

Figure 1. Showing how toxicity data from non-congeneric libraries of compounds may be modelled using QSAR (after Benigni and Giuliani [18]).

Another decision-tree approach to carcinogenicity prediction is that of Purdy [42]. It is not commercially available, but the information given by Purdy [42] is sufficient for it to be used by others. It is based on known or hypothetical chemical reactivity considerations and on the mechanisms of chemical interactions in biological systems. It uses a set of eleven sequentially applied major classification rules based primarily on feature identification. Each rule feeds either directly into an activity assignment, another feature identification query or a QSAR based on calculated properties. The method performed well when applied to a test set of 301 organic chemicals, correctly classifying over 90%.

Another interesting approach to toxicity prediction is COMPACT, developed by Lewis and co-workers at the University of Surrey, UK; it is not commercially available, but the information needed to use the approach is available in Lewis's many publications (see, for example, [43, 44, 45]). COMPACT essentially predicts the potential of a chemical to act as a substrate for one of the cytochromes P450, and is based on the ability of a chemical to fit onto, and interact with, the relevant binding site on the enzyme. It relies largely on two descriptors, molecular planarity and electronic activation energy (the difference between the energies of the highest occupied and lowest unoccupied molecular orbitals, ΔE). Molecular planarity is defined as a/d^2,

where a = molecular cross-sectional area and d = molecular depth. A COMPACT ratio is then defined as $(a/d^2) \times \Delta E$, and values above 0.25 are indicative of P4501 specificity (carcinogenicity), whilst values below 0.15 are indicative of other P450 specificities. In a study of the use of COMPACT to predict the mutagenicity of 100 NTP chemicals, Lewis et al. [46] found an overall concordance of 64%.

It is clear from the above that the predictive ability of expert systems still needs considerable improvement. Richard [47], writing in 1998, stated that 'all of the current commercial methods for toxicity prediction are limited in very real ways by available data and knowledge, and we must be careful not to place unrealistic expectations on their predictive capabilities'. She also went on to say that 'to best address the challenge of structure-based toxicity prediction, we should seek a convergence of modelling approaches toward a more unified approach for the future'. It is interesting, in this respect, to note the work of Lewis et al. [48] in combining the COMPACT and HazardExpert approaches for the prediction of carcinogenicity. Using a small data-set of 14 human carcinogens, they found that COMPACT alone predicted 71% correctly, whilst HazardExpert alone predicted 57% correctly. However, the two systems used in conjunction predicted 100% correctly. For the foreseeable future, the consensus approach would appear to hold out the best promise of being able to make acceptable predictions of chemical toxicity.

It is also important to note that, in general, expert systems are better at predicting toxicity than at predicting lack of toxicity [49, 50]. This is because a rule relating to a positive prediction represents specific information pertaining to a restricted region of the chemical universe; however, characterisation of negative toxicity space, lacking any organising mechanism-based principles, is an inherently more difficult problem [50]. The expert systems were able to point to the presence of chemical or structural alerts, but were not able, in general, to modulate the toxicity classification [51].

Carcinogenicity

It is not possible here to review in detail the *in silico* prediction of the many toxic end-points of interest to the drug designer and developer. Since the toxicity of greatest concern in the development of new drugs is undoubtedly carcinogenicity, it is pertinent briefly to consider its prediction. There is a vast literature dealing with the prediction of carcinogenicity, although, surprisingly, little if any of that deals with drugs. The following examples illustrate the various approaches taken.

A number of publications have dealt with carcinogenicity prediction within congeneric series of chemicals. Leo et al. [52] developed the following QSAR for the tumour-promoting ability of aniline mustards:

$$\log 1/C = -1.17\sigma + 3.30\sigma^2 - 1.70I_4 + 5.03 \qquad (14)$$

$$n = 11 \quad r^2 = 0.819 \quad s = 0.482 \quad F = 10.6$$

where σ = Hammett substituent constant, and I_4 = indicator variable for substitution in position 4 of the aromatic ring. It should be noted that this QSAR is statistically unacceptable, since it fails the Topliss and Costello rule [53] that for every descriptor in a QSAR there should be at least five compounds in the training set, in order to minimise the risk of chance correlations.

Dunn and Wold [54] used their SIMCA approach to model the carcinogenicity of a series of 4-nitroquinoline 1-oxides. With 43 descriptors reduced to four principal components they obtained 89% correct predictions for the 18 carcinogens and 70% correct predictions for the 10 non-carcinogens.

Rippmann [55] observed a biphasic dependence of the tumour-promoting activity of phorbol esters on hydrophobicity alone:

$$\log(\text{nrtpa}) = 0.441\log P - 0.738\log(\beta P + 1) - 2.571 \qquad (15)$$

$$n = 42 \quad r^2 = 0.728 \quad s = 0.326 \quad F = 33.8$$
$$\log\beta = -5.026$$

where nrtpa = numerical relative tumour promoting activity.

Benigni et al. [56] modelled the carcinogenic potency of aromatic amines in the mouse with a 5-term QSAR:

$$\text{Log}1/CP_{50} = 0.56\log P + 1.03E_{HOMO} - 1.19E_{LUMO} - 0.79\Sigma MR_{2,6} - 0.93MR_3 - 0.22E_S(R) + 8.51 \qquad (16)$$

$$n = 37 \quad r^2 = 0.714 \quad s = 0.485 \quad F = 12.5$$

where CP_{50} = millimolar dose to produce tumours in 50% of mice, E_{LUMO} = energy of lowest unoccupied molecular orbital, and $E_S(R)$ = Charton steric constant for substituent on amino nitrogen.

Carcinogenicity is, of course, not one end-point but many, and it would therefore seem highly unlikely that any QSAR modelling could be carried out of the carcinogenic potency of diverse groups of chemicals. Surprisingly, a number of reasonably successful attempts have been made to do this.

Benigni et al. [57] used an electrophilic reactivity descriptor *inter alia* to model the carcinogenicity of 142 diverse chemicals, and obtained 97% correct predictions; of a test set of 14 chemicals, 13 were correctly predicted. Jurs et al. [58], in a pattern recognition study of a diverse set of 130 carcinogens and 79 non-carcinogens, used their ADAPT descriptors to obtain 90% correct predictions for carcinogens and 78% correct predictions for non-carcinogens. Gini et al. [59] used a back-propagation neural network to model the numerical carcinogenic potency of 104 aromatic nitrogen-containing compounds; they obtained a cross-validated r^2 of 0.69, which increased to 0.82 after the removal of 12 outliers. Zhang et al. [60] modelled the carcinogenicity of 239 aromatic hydrocarbons and heterocyclics using hydrophobicity and E_{HOMO}.

For 112 carcinogens, Blake et al. [61] used 13 descriptors to discriminate between genotoxic and non-genotoxic carcinogens with an accuracy of 94.5%; for 93 non-genotoxic chemicals, they used 24 descriptors to discriminate between carcinogens and non-carcinogens with an accuracy of 95.2%.

Moriguchi et al. [62] used their fuzzy adaptive least squares method to predict the carcinogenicity of a diverse set of 586 chemicals. They divided the chemicals into 8 chemical classes, and obtained cross-validated (leave-one-out) correct predictions ranging from 78.3 to 92.3%.

Mention has already been made of expert systems for the prediction of carcinogenicity [25, 30, 33, 40, 41, 42, 44, 48] and other toxicity end-points. In recent years the NTP invited the developers of expert systems, and others with an interest in predicting carcinogenicity, to participate in a trial. The NTP prepared a list of 40 substances that had not been tested for carcinogenicity, and asked the trial participants to predict whether or not each substance was carcinogenic; they then carried out rodent carcinogenicity testing on the substances. Of the expert systems discussed here, none performed well. The concordances found [49, 51] were: DEREK 59%, TOPKAT 58%, COMPACT 54% and CASE 49%. In a second similar NTP exercise, involving 26 substances, the following concordances were found [50]: OncoLogic 67%, COMPACT 44%, DEREK 38%, Purdy 35%, CASE 18%. These figures are disappointing, since even the toss of a coin should give 50% concordance. However, it has been pointed out [50] that the chemicals used in the NTP prediction exercises were not a random sample from the chemical universe, but included a majority of suspect chemicals. The real goal was not to separate carcinogens from non-carcinogens, but rather to separate actual carcinogens from potential carcinogens; this is a difficult task, because of the sometimes subtle interplay of molecular features that can modulate the carcinogenic potential of a primary structural feature, such as an aromatic amino group.

Conclusions

In view of the enormous costs of drug failure due to toxicity found late in the development process, toxicity should be determined as early as possible, so as to guide synthesis. This indicates the use of computational prediction of toxicity, and a number of approaches to such prediction have been discussed.

Published QSARs exist for a wide range of toxicity endpoints of drugs. There is great potential for further QSAR modelling of toxicity endpoints, provided that adequate experimental data are available for the development of the QSARs. In this respect, it would be immensely helpful if pharmaceutical companies' in-house toxicity data could be released, perhaps to an 'honest broker' who could develop QSARs whilst guarding the confidentiality of the data.

In view of the increasing use of libraries of diverse chemicals, it is useful to know that QSARs for toxicity prediction have been developed for very heterogeneous groups of chemicals, although few if any of those relate specifically to drugs.

A number of expert systems are available for the prediction of a wide range of toxicity endpoints. Currently such systems do not have particularly high accuracy, and it is recommended that a consensus approach be taken with their use.

References

1. Hansch, C. and Leo, A., *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, Amer. Chem. Soc., Washington, DC, 1995.
2. Kennedy, T., Drug Disc. Today, 2 (1997) 436.
3. Crum Brown, A. and Fraser, T.R., Trans. Roy. Soc. Edinburgh, 25 (1868–69) 151.

4. Hansch, C., Maloney, P.P., Fujita, T. and Muir, R.M., Nature, 194 (1962) 178.
5. Hansch, C. and Clayton, J.M., J. Pharm. Sci., 62 (1973) 1.
6. King, L.A. and Moffat, A.C., Med. Sci. Law, 23 (1983) 193.
7. Lien, E.J. and Tong, G.L., Cancer Chemother. Rep., 57 1973) 251.
8. Jarzebski, A., Falkowski, L. and Borowski, E., J. Antibiot. 35 (1982) 220.
9. Chen, P., Cheng, P.T.W., Alam, M., Beyer, B.D., Bisacchi, G.S., Dejneka, T., Evans, A.J., Greytok, J.A., Hermsmeier, M.A., Humphreys, W.G., Jacobs, G.A., Kocy, O., Lin, P.-F., Lis, K.A., Marella, M.A., Ryono, D.E., Sheaffer, A.K., Spergel, S.H., Sun, C., Tino, J.A., Vite, G., Colonno, R.J., Zahler, R. and Barrish, J.C., J. Med. Chem., 39 (1996) 1991.
10. Schön, U., Antel, J., Brückner, R. and Messinger, J., Med. Chem., 41 (1998) 318.
11. Dearden, J.C., Cronin, M.T.D. and Thomas, G.R., J. Pharm. Pharmacol., 51 (1999) Suppl., 267.
12. Dearden, J.C., Cronin, M.T.D. and Shah, S. J. Pharm. Pharmacol., 51 (1999) Suppl., 266.
13. Hansch, C., Bjorkroth, J.P. and Leo, A., J. Pharm. Sci., 76 (1987) 663.
14. Martin, Y.C. and Hansch, C., J. Med. Chem., 14 (1971) 777.
15. Mannhold, R. and Dross, K., Quant. Struct.-Act. Relat., 15 (1996) 403.
16. Dearden, J.C., Netzeva, T.I. and Bibby, R., J. Pharm. Pharmacol. 54 (2002) suppl., S-65.
17. Dearden, J.C., Netzeva, T.I. and Bibby, R., J. Pharm. Pharmacol. 54 (2002) suppl., S-66.
18. Benigni, R. and Giuliani, A., Med. Res. Rev., 16 (1996) 267.
19. Dearden, J.C., Cronin, M.T.D., Patel, H. and Raevsky, O.A., J. Pharm. Pharmacol., 52 (2000) Suppl., 221.
20. Dearden, J.C. and Scott, A.C., J. Pharm. Pharmacol., 54 (2002) Suppl., S-66.
21. Mumtaz, M.M., Knauf, L.A., Reisman, D.J., Peirano, W.B., Derosa, C.T., Gombar, V.K., Enslein, K., Carter, J.R., Blake, B.W., Huque, K.I. and Ramanujam, V.M.S., Toxicol. Lett., 79 (1995) 131.
22. Barratt, M.D., ATLA, 23 (1995) 111.
23. Gao, H., Williams, C.M., Labute, P. and Bajorath, J., J. Chem. Inf. Comput. Sci., 39 (1999) 164.
24. Dearden, J.C. and Cronin, M.T.D., In Peijnenburg, W.J.G.M. and Damborský, J. (Eds.) *Biodegradability Prediction*, Kluwer Academic Publishers, Dordrecht, 1996, pp. 93–104.
25. Dearden, J.C., Barratt, M.D., Benigni, R., Bristol, D.W., Combes, R.D., Cronin, M.T.D., Judson, P.N., Payne, M.P., Richard, A.M., Tichý, M., Worth, A.P. and Yourick, J.J., ATLA, 25 (1997) 223.
26. Benfenati, E. and Gini, G., Toxicol., 119 (1997) 213.
27. Richard, A.M., Knowledge Eng. Rev., 14 (1999) 307.
28. Helma, C., Gottmann, E. and Kramer, S., Stat. Methods Med. Res., 9 (2000) 329.
29. Greene, N., Adv. Drug Deliv. Rev., 54 (2002) 417.
30. Enslein, K., Gombar, V.K. and Blake, B.W., Mut. Res., 305 (1994) 47.
31. Enslein, K., Gombar, V.K., Blake, B.W., Maibach, H.I., Hostynek, J.J., Sigman, C.C. and Bagheri, D., Food Chem. Toxicol., 35 (1997) 1091.
32. Rosenkranz, H.S., Cunningham, A.R., Zhang, Y.P., Claycamp, H.G., Macina, O.T., Sussman, N.B., Grant, S.G. and Klopman, G., SAR QSAR Environ. Res., 10 (1999) 277.
33. Cunningham, A.R., Klopman, G. and Rosenkranz, H.S., Mut. Res. – Fund. Mol. Mech. Mut., 405 (1998) 9.
34. Gomez, J., Macina, O.T., Mattison, D.R., Zhang, Y.P., Klopman, G. and Rosenkranz, H.S., Teratol., 60 (1999) 190.
35. Matthews, E.J. and Contrera, J.F., Regul. Toxicol. Pharmacol., 28 (1998) 242.
36. Greene, N., Judson, P.N., Langowski, J.J. and Marchant, C.A., SAR QSAR Environ. Res., 10 (1999) 299.
37. Steger-Hartmann, T., Yih, T.D., Marchant, C.A. and Langowski, J.J., Naunyn-Schmiedeberg's Arch. Pharmacol., 363 (2001) Suppl., S-363.
38. Barratt, M.D., Castell, J.V., Miranda, M.A. and Langowski, J.J., J. Photochem. Photobiol. B – Biol., 58 (2000) 54.
39. Smithing, M.P. and Darvas, F., In Finlay, J.W., Robinson, S.F. and Armstrong, D.J., (Eds.) *Food Safety Assessment* (American Chemical Society Symposium Series No. 484), American Chemical Society, Washington, DC, 1992, pp. 191–200.
40. Brown, S.J., Raja, A.A. and Lewis, D.F.V., ATLA, 22 (1994) 482.
41. Woo, Y.-T., Lai, D., Argus, M. and Arcos, J., Toxicol. Lett., 79 (1995) 219.
42. Purdy, R., Environ. Health Perspect., 104 (1996) 1085.
43. Parke, D.V., Ioannides, C. and Lewis, D.F.V., ATLA, 18 (1990) 91.
44. Lewis, D.F.V., Ioannides, C. and Parke, D.V., Drug Metab. Rev., 30 (1998) 709.
45. Lewis, D.F.V., J. Chem. Tech. Biotech., 76 (2001) 237.
46. Lewis, D.F.V., Ioannides, C. and Parke, D.V., Mut. Res., 291 (1993) 61.
47. Richard, A.M., Mut. Res. – Fund. Mol. Mech. Mutagen., 400 (1998) 493.
48. Lewis, D.F.V., Bird, M.G. and Jacobs, M.N., Human Exper. Toxicol., 21 (2002) 115.
49. Benigni, R., Mut. Res. – Environ. Muta., 334 (1995) 103.
50. Richard, A.M. and Benigni, R., SAR QSAR Environ. Res., 13 (2002) 1.
51. Benigni, R., Mut. Res. – Rev. Mut. Res., 387 (1997) 35.
52. Leo, A., Panthananickal, A. and Hansch, C., J. Med. Chem., 24 (1981) 859.
53. Topliss, J.G. and Costello, R.J., J. Med. Chem., 15 (1972) 1066.
54. Dunn, W.J. and Wold, S., J. Med. Chem., 21 (1978) 1001.
55. Rippmann, F., Quant. Struct.-Act. Relat., 9 (1990) 1.
56. Benigni, R., Giuliani, A., Franke, R. and Gruska, A., Chem. Rev., 100 (2000) 3697.
57. Benigni, R., Andreoli, C. and Giuliani, A., Carcinogenesis, 10 (1989) 55.
58. Jurs, P.C., Chou, J.T. and Yuan, M, J. Med. Chem., 22 (1979) 476.
59. Gini, G., Lorenzini, M., Benfenati, E., Grasso, P. and Bruschi, M., J. Chem. Inf. Comput. Sci., 39 (1999) 1076.
60. Zhang, L.T., Sannes, K., Shusterman, A.J. and Hansch, C., Chem.-Biol. Interact., 81 (1992) 149.
61. Blake, B.W., Enslein, K., Gombar, V.K. and Borgstedt, H.H., Mut. Res., 241 (1990) 261.
62. Moriguchi, I., Hirano, H. and Hirono, S., Environ. Health Perspect., 104 (1996) 1051.