DRUG DISCOVERY TODAY BIOSILICO

Computational prediction of genotoxicity: room for improvement

Ronald D. Snyder and Mark D. Smith

Decades of mutagenesis and clastogenesis studies have yielded enough structure–activity-relationship (SAR) information to make feasible the construction of computational models for prediction of endpoints based on molecular structure and reactivity. Although there is cause for optimism that these approaches might someday reduce or eliminate the need for actual genotoxicity testing, we are in fact a long way from this. We provide an overview of the state of the art of such approaches, dissecting out how these models are suboptimal. It is clear that current programs still have limited predictive capabilities. We propose that one of the major contributing factors for the inherent lack of sensitivity (typically 50–60%) is inadequate coverage of non-covalent DNA interactions. Suboptimal specificity can be partly attributed to chemical space considerations with associated non-causal activity correlations.

There is a general appreciation that a complete understanding of all toxicities associated with a new drug candidate is crucial to its successful development and marketing. Fortunately, genotoxicity can be measured directly by long-standing and universally accepted assays, such as the Ames test for bacterial mutagenicity, chromosome aberration assays in human lymphocytes or other mammalian cells in culture, *in vivo* cytogenetics studies, and a host of 'second tier' assays which, although not always uniformly concordant, are applied in a weight-of-evidence context. Of necessity, these regulatory agency-mandated studies [1,2] have typically been conducted rather late in development after preclinical efficacy has been established and in the same time frame as the general toxicology studies. But every pharmaceutical company has stories of how otherwise safe and effective molecules have been forced out of further development owing to unexpected genotoxicity seen during these regulatory studies.

There is also a need to characterize the genotoxic potential of metabolites, degradants, impurities and, in the occupational health arena, process intermediates. Today nearly all large Pharma companies have early gene-tox screening programs usually employing a scaled down 'mini'-Ames and an *in vitro* assessment of chromosome damage in cultured mammalian cells. Genotoxicity is thus revealed early on and structure–activity-relationship (SAR) techniques can usually guide subsequent chemical syntheses to avoid genotoxicity. Most large Pharma companies also use computational programs to aid in the prediction of genotoxicity and a combination of *in vitro* screening and *in silico* analysis is widely used.

Genotoxicity should be easier to predict than other types of toxicity because genotoxicity typically arises from direct chemical/DNA interaction dependent to a large extent on electrophilicity. Specific organ toxicities, on the other hand can arise by any of several pharmacological or chemical mechanisms not necessarily related to or obvious from chemical structure analysis. In fact, the advent of microarray technologies has made it possible to establish specific organ toxicity gene expression signatures which

Ronald D. Snyder
Genetic and Molecular Toxicology,
Schering-Plough Research Institute,
Lafayette,
NJ 07848,
USA
e-mail:
ronald.snyder@spcorp.com
Mark D. Smith
Safety Assessment,
GlaxoSmithKline,
King of Prussia,
PA 19406,
USA

1359-6446/05/$ – see front matter PII: S1359-6446(05)03505-1

may someday allow prediction of organ toxicities without the need for longer-term preclinical animal studies.

Unfortunately, even genotoxicity has proven to be substantially refractory to prediction based on two dimensional structure analysis despite the existence of computational programs whose 'intelligence' is based on very large numbers of compounds and attendant genotoxicity data.

This review will briefly describe the principal computational programs and their performance characteristics in predicting genotoxicity. It is not our intention to describe in detail the history, the evolution, or the chemico-biological/statistical basis of such systems, as many reviews have already covered this. Instead, we will discuss the inherent strengths and weaknesses of these programs and prospects for improvement.

The principal players and some newcomers

Two excellent reviews have been published [3,4] examining the most commonly used computational mutagenicity programs and the interested reader is encouraged to consult these and the respective program websites (see text) for greater detail. The following general descriptions are provided.

DEREK (Deductive Estimation of Risk from Existing Knowledge)

Created by Lhasa Ltd (http://www.chem.leeds.ac.uk/luk/derek/index.html), DEREK is a knowledge- and rule-based expert system that makes semi-quantitative estimations as to whether or not a DNA reactive (subdivided as to general genotoxic, mutagenic, or chromosome damaging) moiety is present on the input chemical structure. An experienced user is able to determine if a flagged alert is in the proper chemical context to be genotoxic relative to the compound(s) upon which the DEREK rule was based. The learning set for DEREK was created using both bacterial mutagenicity and all other available genotoxicity data. Query outputs define the structural alert recognized, the type of genotoxicity (bacterial mutagenicity, *in vitro* cytogenetics, etc.) associated with the alert, specific examples of genotoxic compounds sharing the alerting moiety, detailed mechanistic comments relevant to the alert, and literature references. Derek can be customized by the user.

MCASE (Multiple Computer Automated Structure Evaluation)

MCASE (http://www.multicase.com) dissociates each input molecule into 2–10 atom fragments and statistically evaluates the strength of association of those fragments (biophores), and similar fragments from its database, with an associated mutagenicity score (a value based on the observed mutagenic potency). It generates a quantitative prediction of mutagenicity which is then further refined through taking into consideration physico-chemical properties as well as the existence of potential 'deactivating fragments' or biophobes. The original MCASE model was based solely on bacterial mutagenicity data derived from 2032 compounds from the National Toxicology Program (NTP), the U.S. Environmental Protection Agency (USEPA) Genetox programs, and 204 pharmaceuticals (the latter of which were all negative in the Ames test). A more recent version is based on a set of 3000 compounds and includes Drosophila mutation data. About to be released is yet another version created by the FDA in collaboration with MCASE in which 16 separate modules allow predictions of mutagenicity in individual Salmonella strains in the presence and absence of either rat or hamster S9 activating systems. MCASE can be readily customized by the user.

TOPKAT (Toxicity Prediction by Komputer Assisted Technology)

TOPKAT (http://www.accelrys.com/products/topkat/index.html) uses 'electro-topological' descriptors rather than chemical structures to predict mutagenic reactivity with DNA and, as such, is an extension of classical quantitative structure–activity relationship (QSAR) analysis. The intelligence of TOPKAT was derived solely from bacterial mutagenicity data. TOPKAT was initially designed by Health Systems Inc. and is now marketed by Accelrys, San Diego, CA, USA. The Ames prediction module consists of 1866 compounds divided into individual models based on chemical class analogy. Unlike DEREK, TOPKAT provides a measure of the similarity between a test molecule and the chemical space covered by the program excluding from further analysis any molecules deemed to have insufficient coverage. TOPKAT cannot be readily customized by the user.

QSAR models

In addition to the above programs, numerous QSAR models have been designed and evaluated [5–12]. QSAR models use algorithms based on various types of chemical descriptors such as chemical substructure, logP, electronics, geometrical attributes, and surface area to yield a predictive value. Most QSAR genotoxicity models predict and are based on bacterial mutagenicity data, an exception being that developed by Serra *et al.* [9] which predicts and is based solely on chromosome aberration data. Remarkably, this chromosome aberration QSAR model required only three topological descriptors for prediction. At the present time, only one QSAR model, CSGenotox (www.ChemSilico.com), has been evaluated in side by side trials with other computational programs against a common tester set of molecules to establish comparative performance characteristics [7]. That study compared the predictivity of three QSAR models to that of MCASE and DEREK for 217 non-drugs and 30 drugs. Of the descriptors found to be predictive, 40% were related to well-known structural genotoxicity alerts. The results of that study were interpreted as indicating that the QSAR approach had better specificity, but used unsupervised 'out of the box' calls for MCASE and DEREK for comparison which biases the results. Nevertheless, QSAR approaches offer a

TABLE 1

Comparison of sensitivities of computational programs

	MCASE	DEREK	TOPKAT
Ames			
Overall	13/27 (48%)	14/27 (52%)	10/23 (43%)
Structurally alerting (12)[a]	92%	67%	82%
Non-alerting (15)	20%	27%	18%
***In vitro* cytogenetics**			
Overall	10/47 (21%)	15/47 (32%)	11/48 (23%)
Structurally alerting (20)	45%	60%	35%
Non-alerting (28)	4%	11%	14%
MLA			
Overall	7/23 (30%)	8/24 (33%)	3/24 (13%)
Structurally alerting (8)	75%	75%	25%
Non-alerting (16)	6%	13%	6%
***In vivo* cytogenetics**			
Overall	6/25 (24%)	10/30 (33%)	8/28 (29%)
Structurally alerting (11)	60%	73%	50%
Non-alerting (19)	0%	11%	17%

[a]Number of drugs in parentheses. Data adapted with permission from [22].

different way of categorizing new chemical entities. Further evaluation may prove QSAR modeling to be superior or at least complementary to the non-QSAR models.

DNA docking model

Another approach under development is a 3-dimensional DNA docking model for the identification of molecules capable of non-covalent DNA interaction, ie. groove-binding, DNA intercalation or both [13–15]. This model evaluates the ability of a molecule to fill the space between two adjacent DNA base pairs and calculates the strength of that fit based on H-bonding and van der Waals contacts. Initial studies have demonstrated the utility of this model in predicting genotoxicity of non-covalent DNA binding molecules [16,17]. Subsequent work has shown that many non-structurally alerting clastogenic drugs, not predicted by standard computational programs, may act via DNA intercalation (manuscripts in preparation).

Consensus modeling

Basing genotoxicity predictions on the combined outputs of multiple models – consensus modeling – generally improves the accuracy of prediction but this is at the expense of decreased chemical coverage. White *et al.* [18], for example, reported that CASETOX, TOPKAT and DEREK had specificities of 78 to 82% when used as single programs. In combination with one or both additional programs, the specificity increased to 92 to 100% (with no substantive change in sensitivity) but the number of molecules that could be assessed by multiple programs (based on the criterion that both programs had to have made the same call on a particular molecule) dropped by 60 to 80%. Consensus modeling from multi-QSAR modules has also been reported to provide additional predictive power when applied to generalized polycyclic aromatics [19] or to thiophenes [20].

Correlating predictions

Accurate prediction of Ames mutagenicity is essential for drug development because, except for drugs being developed for life-threatening diseases or conditions, a positive Ames is usually the death knell for a molecule. But, there is an equal need for early genotoxicity prediction of *in vitro* and *in vivo* cytogenetics, the other two components of the regulatory testing battery. There is no reason to believe that a positive computational prediction based on Ames' SAR, would not also pertain to other genetic toxicology endpoints because all true genotoxicity is based on the same capability for covalent DNA addition, be it in bacteria, mammalian cells, in culture or *in vivo*. Attesting to this is the considerable overlap between Ames-positivity and positivity in other genotoxicity assays among marketed pharmaceuticals: 20 of 23 (87%) Ames-positives are also positive in other genotoxicity assays [21]. Considering only the major computational programs, structure/genotoxicity data of anything other than Ames results has been built only into DEREK, although new modified versions of MCASE, at least, are being developed as specific models for prediction of mouse lymphoma and chromosomal aberration assays and the QSAR model of Serra *et al.* [9] is based only on chromosome aberration data. It will be interesting to see if these more broadly-based models will be more predictive for either global or specific genotoxicity endpoints.

Sensitivity

Lack of sensitivity (the measure of a program's ability to correctly identify true positives) is a problem if one relies heavily on computational analysis to advance a compound to the candidate stage or in making worker safety decisions. However, even in the absence of a computational prediction of genotoxicity, most companies would most probably evaluate the genotoxicity of any molecule of interest, particularly if the molecule possessed a structural component of concern. Lack of sensitivity might therefore be more of an academic curiosity than a drug development hurdle. Nevertheless, a more complete understanding of the nature of false negatives is required to improve computational models. As mentioned earlier, direct comparison of models is difficult because for the most part, models have not been tested side-by-side with the same tester sets.

Table 1 summarizes the sensitivity of the three most commonly used programs when challenged with the drug set from the Physician's Desk Reference (PDR) [21,22]. Excluded from the test set were molecules with known mechanism-based genotoxicity (e.g. nucleoside analogs,

TABLE 2

Summary of performance characteristics of computational programs in the assessment of bacterial mutagenicity of marketed drugs

	Sensitivity	Specificity	Concordance
MCASE	13/27 (48%)	307/330 (93%)	320/357 (90%)
DEREK	14/27 (52%)	260/346 (75%)	274/373 (74%)
TOPKAT	10/23 (43%)	267/316 (85%)	277/339 (82%)

Ratios are correct calls/total calls. Data adapted with permission from [22].

fluoroquinolones, etc.), as the genotoxicity of these molecules is not attributable to a specific molecular fragment or moiety and is, therefore, not predictable by standard computational analyses (see below). Inclusion of these molecules would have the consequence of further reducing sensitivity.

It is clear that all three programs have approximately the same (poor) overall sensitivity for detecting Ames positives (43 to 52% of positives correctly identified). This was also reported in a study of proprietary pharmaceuticals in which 49/90 (54%) true Ames positives were missed by DEREK and in which 44/63 (70%) of Ames positives were missed by MCASE [3]. Similar poor sensitivity was reported by White *et al.* [18] for all three programs challenged with a tester set of over 500 proprietary pharmaceuticals.

It is also seen from Table 1 that the sensitivities in the non-bacterial genotoxicity assays is even worse than for Ames, for all three programs. An overwhelming majority of correct calls for all assay types by all three programs was made on drugs with obvious 'structural alerts' as originally defined by Ashby [23,24]. This is to be expected as each computational system has these structural motifs contained within their respective training sets. On the other hand, the highest sensitivity value observed for 'non-Ashby' alerting PDR drugs was 27%. This indicates that, for whatever reason, these programs were not adequately trained to detect the non-alerting drugs, the ones of greatest interest! From within the PDR test set at least 84 non-alerting genotoxic drugs were identified [22]. They collectively constituted a troubling 56, 58, 67, and 63% of all positive Ames, *in vitro* cytogenetics, mouse lymphoma and *in vivo* cytogenetics results, respectively. Twelve of these molecules might have been missed owing to their known requirement for metabolic activation; seventeen appear to have been missed because they were classical intercalating agents, information on which is greatly underrepresented in these models. Why the remaining 55 drugs were missed is unknown but recent studies suggest that many structurally-diverse molecules are capable of functional DNA intercalation even though they do not have classical (fused ring, planar) intercalating structures [16,17]. Because only a small number of classical and non-classical intercalating structures are included in the databases of these programs, it would seem likely that further characterization of non-covalent DNA interaction and inclusion of this information into computational models might substantially increase the sensitivity. The need to generate more information relative to intercalating potential is important for better predictivity of bacterial frameshift mutagenesis. But non-covalent DNA binding could be even more important in the context of positive chromosome aberration results because intercalating agents are well-known topoisomerase inhibitors and, as such, generate largely irreparable DNA double strand breaks [25].

The sensitivity of these computational models is somewhat better if one tests chemicals rather than drugs [26]. This is most probably related to the fact that because drugs are underrepresented in these models, the chemical 'space' – or extent of chemical diversity represented by the molecules used to construct each program – is enriched with non-pharmaceutical-type structures. The difficulties of using programs based on a global non-congeneric set of molecules have also been discussed in the QSAR context [11]. The same is true of databases constructed from a set of structurally-related proprietary molecules, which is then challenged with additional structures within that same chemical space. Thus, these programs are inherently more accurate when testing structurally restricted rather than global tester sets.

Specificity

Table 2 compares the three primary programs with respect to their total performance characteristics when challenged with PDR drugs. The MCASE and DEREK values used in this table are derived using expert evaluation of the program outputs, not 'out of the box' calls. Without user input, performance characteristics are considerably poorer (unpublished observations). Concordance, the overall percentage of correct positive and negative calls, is often misleading and is usually biased by the larger number of true negatives in most tester sets. As shown in Table 2, high concordance belies the very poor sensitivity of all of these programs.

Aside from the poor sensitivity, it is clear from Table 2 that specificities could also use some improvement. Specificity is defined as the ratio of true negatives to the sum of true negatives and false positives. Thus if there were 50 true negative predictions but 25 false positive predictions, the specificity would be 67%, unacceptably low. The generation of false positives can result in the 'throwing out' of useful drugs. This would probably not be an issue in a combinatorial chemistry program where a certain proportion of molecules must be filtered out anyway to limit the number of molecules that are taken forward to the next Discovery step. However, if the computational analysis is done later on when, say, 20 compounds are being rank ordered, discarding false positives can be more problematic. Also, a false positive genotoxicity prediction, even if tested directly by bioassay and found to be 'clean', might require additional testing or discussion with regulatory agencies. This would be particularly true for molecules carrying structural alerts.

What causes this high number of false positive predictions? Published proclamations as to the extremely poor

specificities of computational programs generally refer to 'out of the box' calls and rarely take expert input into consideration. As stated above, false positives could be markedly reduced by including an expert opinion as to whether the identified biophore is in a chemical context consistent with it being DNA reactive. This determination can usually be made by a chemist or toxicologist. But there are other sources of false positive calls. An ongoing exercise by the creators and developers of MCASE has concluded that predictive value can be dependent on such things as the number of molecules, the ratio of positive to negative compounds, [27,28] and the nature of the molecules, e.g. drugs versus, for example, pesticides, in the learning set [29]. Furthermore, poor specificity could also be due in part to the incorrect assignment of a fragment as being responsible for the mutagenicity of a molecule that, in fact, has no true biophore. For example, a mutagenic nucleoside would not carry an actual structural alert, its mutagenicity instead being due to inhibition of DNA polymerization. Correlative approaches such as MCASE would nevertheless seek to define a fragment(s) to ascribe to mutagenicity and such assignments would be necessarily incorrect. Subsequent test molecules carrying this same fragment may be incorrectly identified as positives. We removed non-alerting positive gene-tox structures from the MCASE learning set and rechallenged with the PDR database. Some improvement of specificity (ie less false positives) did occur but it is clear that such incorrect biophores account for only a small fraction of the false positives seen with MCASE (Braunstein *et al.*, presented at Annual Society of toxicology meeting, Baltimore, MD, 2004).

Genotoxic potency

The PDR provides a unique opportunity to evaluate genotoxicity results derived from harmonized study design and subjected to the same interpretative criteria. The database is also inherently weakened, however, by the lack of information relating to the strength of the response and the dose at which that genotoxic response was observed. This is a feature unfortunately shared by many publicly available toxicology database collections (Register of Toxic Effects of Chemical Substances [RTECS], NIH collection, etc). Recently, at least two independent initiatives have tried to address this shortfall in suitable SAR data. The first of these is VITIC, a database concept initially formed from a Health and Environmental Science Institute (HESI) initiative now under development with Lhasa in which publicly available toxicology data are presented in a database format with searchable chemistry and links to the original biological assay data. The second, DSSTox, is a database concept championed by Ann Richard of the USEPA, in which data are shared as SDF files in a regular ASCII format [30].

It is hoped that this potency data will aid in developing risk-estimation paradigms. Most of these genotoxic responses are likely to be very weak and to occur at concentrations far beyond those targeted in the clinic. In truth, these positive responses, although almost certainly genuine, probably carry minimal risk to the patient population and a better means of assessing the actual risk associated with these drug exposures would be highly desirable. It is instructive in this regard that in the PDR marketed drug dataset, of 50 genotoxic drugs with two-year bioassay results, only 26 (52%) were carcinogenic in at least one species. Of 151 non-genotoxic drugs, 51 (34%) tested positive for carcinogenicity [21]. This latter observation is explained by the fact that much carcinogenicity arises via other than genotoxic mechanisms. The former observation suggests that 'weak' genotoxicity does not necessarily translate into carcinogenicity. This is most probably because of exposure considerations rather than an 'incorrect' genotoxicity result but this has not been rigorously evaluated. In the only such study reported, a clear linear relationship was observed between *in vivo* genotoxicity potency and carcinogenic dose defined in two-year bioassays of 50 molecules (only 5 of which were drugs) [31]. It might be further enlightening to compare 'weak' versus 'strong' *in vitro* genotoxins with respect to drug plasma levels and carcinogenicity outcome in the two-year rodent bioassay. Such studies would be helpful in determining if rational risk assessment could be based on potency (DNA reactivity) and clinical plasma drug concentration.

Conclusions

It is without question that the currently available *in silico* approaches to genotoxicity prediction add value to the drug development process. However, it is equally apparent that the available tools fall short of the desired degrees of both sensitivity and specificity. Although marked improvements in concordance can be made via expert interpretation of the 'out of the box' calls, it is clear that each computational model still requires extensive modification before the ultimate goal of replacing biological assays can be attained. It is proposed that one area in which significant improvements can be made is in the coverage of non-bacterial genotoxic responses. In particular, inclusion of specific genotoxicity data pertinent to non-covalent DNA interactions should lead to a marked increase in predictive value.

References

1 ICH S2A: Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals, CPMP/ICH/141/95.

2 ICH S2B: Genotoxicity: a standard battery for genotoxicity testing for pharmaceuticals, CPMP/ICH/174/95.

3 Green, N. (2002) Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Res.* 54, 417–431

4 Cariello, N.F. *et al.* (2002) Comparison of computer programs DEREK and TOPKAT to predict bacterial mutagenicity. *Mutagenesis* 17, 321–329

5 Votano, J.R. *et al.* (2004) Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis* 19, 365–377

6 Basak, S.C. *et al.* (2001) Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: A hierarchical approach.

J. Chem. Inf. Comput. Sci. 41, 671–678
7 Votano, J.R. (2005) Recent use of topological indices in the development of *in silico* ADMET models. *Curr. Opin. Drug Discov. Devel.* 8, 32–37
8 Mattioni, B.E. *et al.* (2003) Predicting the genotoxicity of secondary and aromatic amines using data subsetting to generate a model ensemble. *J. Chem. Inf. Comput. Sci.* 43, 949–963
9 Serra, J.R. *et al.* (2003) Development of binary classification of structural chromosome aberrations for a diverse set of organic compounds from molecular structure. *Chem. Res. Toxicol.* 16, 153–163
10 Benigni, R. and Giuliani, A. (1996) Quantitative structure activity relationship (QSAR) studies of mutagens and carcinogens. *Med. Res. Rev.* 16, 267–284
11 Benigni, R. and Richard, A. (1996) QSARs of mutagens and carcinogens: Two case studies illustrating problems in the construction of models for noncongeneric chemicals. *Mutat. Res.* 371, 29–46
12 Livingstone, D.J. *et al.* (2002) Modeling mutagenicity using properties calculated by computational chemistry. *SAR QSAR Environ. Res.* 13, 2–33
13 Hendry, L.B. *et al.* (1994) Design of novel antiestrogens. *J. Steroid Biochem. Mol. Biol.* 49, 269–280
14 Hendry, L.B. *et al.* (1998) The ligand insertion hypothesis in the genomic action of steroid hormones. *J. Steroid Biochem. Mol. Biol.* 65, 75–89
15 Hendry, L.B. *et al.* (1999) Multidimensional screening and design of pharmaceuticals by using endocrine pharmacophores. *Steroids* 64, 570–575
16 Snyder, R.D. *et al.* (2004) Evaluation of DNA intercalation potential of pharmaceuticals and other chemicals by cell-based and three-dimensional computational approaches. *Environ. Mol. Mutagen.* 44, 163–173
17 Snyder, R.D. *et al.* The influence of N-dialkyl and other cationic substituents on DNA intercalation and genotoxicity. *Mutat. Res.* (in press)
18 White, A.C. *et al.* (2003) A multiple *in silico* program approach for the prediction of mutagenicity from chemical structure. *Mutat. Res.* 539, 77–89
19 He, L. *et al.* (2003) Predicting the genotoxicity of polycyclic aromatic compounds from molecular structure with different classifiers. *Chem. Res. Toxicol.* 16, 1567–1580
20 Mosier, P.D. *et al.* (2003) Predicting the genotoxicity of thiophene derivatives from molecular structure. *Chem. Res. Toxicol.* 16, 721–732
21 Snyder, R.D. and Green, J.W. (2001) A review of the genotoxicity of marketed pharmaceuticals. *Mutat. Res.* 488, 151–169
22 Snyder, R.D. *et al.* (2004) Assessment of the sensitivity of the computational programs DEREK, TOPKAT and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. *Environ. Mol. Mutagen.* 43, 143–158
23 Ashby, J. *et al.* (1989) Classification according to chemical structure, mutagenicity to Salmonella, and level of carcinogenicity of a further 42 chemicals tested for carcinogenicity by the U.S. *National Toxicology Program. Mutat. Res.* 223, 73–103
24 Ashby, J. and Tennant, R.W. (1991) Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S.N.T.P. *Mutat. Res.* 257, 229–306
25 Ferguson, L.R. (1998) Inhibition of Topo II enzymes: A unique of environmental mutagens and carcinogens. *Mutat. Res.* 400, 271–278
26 Snyder, R.D. *et al.* (2003) A cross-platform comparison of *in silico* models for predicting genotoxicity using marketed pharmaceuticals. *Environ. Mol. Mutagen.* 41, 207
27 Rosenkranz, H.S. and Cunningham, A.R. (2001) SAR modeling of unbalanced data sets. *SAR QSAR Environ. Res.* 12, 267–274
28 Rosenkranz, H.S. (2004) SAR modeling of genotoxic phenomena: the consequence on predictive performance of deviation from a unity ratio of genotoxicants/non-genotoxicants. *Mutat. Res.* 559, 67–71
29 Rosenkranz, H.S. and Cunningham, A.R. (2001) SAR modeling of genotoxic phenomena: the effect of supplementation with physiological chemicals. *Mutat. Res.* 476, 133–137
30 Richard, A.M. and Williams, C.R. (2002) Distributed structure-searchable toxicity (DSSTox) public database network: A proposal. *Mutat. Res.* 499, 27–52
31 Sanner, T. and Dybing, E. (2005) Comparison of carcinogens and *in vivo* genotoxicity potency estimates. *Basic Clin. Pharmacol. Toxicol* 96, 131–139