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