Collection and Evaluation of (Q)SAR Models for Mutagenicity and Carcinogenicity

Dr. Romualdo Benigni

General background and aims of the project

The concept of a relationship between the structure and the toxic activity of chemical mutagens and carcinogens has been widely investigated since the recognition that some chemicals can induce mutations and/or cancer. Structure-activity concepts have also been exploited to “domesticate” the chemicals, and make them less or no harmful (Ariens, 1984). During the last decades, all this activity has lead to the formalization of a large number of Structure-Activity Relationships (SAR) and Quantitative Structure-Activity Relationships (QSAR) models relative to mutagenicity and carcinogenicity (Benigni, 2005). These can be roughly divided into: a) models focusing on individual chemical classes (mostly QSARs); and b) “all-purposes” models for noncongeneric chemicals, hopefully able to cope with any and every chemical class. Most of the commercial systems which are popular today belong to the latter group.

Structure-activity models are generated empirically by analyzing a set of chemicals (training set) for which experimental biological data exist. Chemical descriptors of the molecules are calculated, and descriptors correlated with the biological activity are identified through statistical/mathematical methods. The maximum amount of information and the most reliable models are obtained when the training data set is designed ad hoc, i.e. by testing chemicals selected to represent the chemical and biological properties of the chemical class (e.g., adequate number of chemicals tested, low correlation coefficients between descriptors, equilibrated spanning of the chemical space, balance between the numbers of actives and inactives). Unfortunately, this ideal situation is not a common occurrence in toxicology, where the modelers usually find it necessary to base their analyses on the toxicological data that are available in the literature. Toxicological experimentation follows priorities (e.g., level of human exposure) different from those of the modelers. This limited and unbalanced representation of chemical types among those tested experimentally, affects at the same extent the QSARs on individual chemical classes and the commercial (Q)SAR models.

Whereas in mechanistic investigations even studies based on biased data sets may provide usable information, the extensive regulatory use of (Q)SARs foreseen by REACH requires that the models fulfill severe quality criteria, such as those exemplified by the OECD validation principles. In recent
years, a number of validation studies have been performed on the most popular commercial predictive models, whereas little attention has been given to the non-commercial (Q)SARs for mutagenicity and carcinogenicity (Benigni, 2005). However, they should reserve more attention since many of them are classical QSARs that provide both mechanistic information and predictive tools (Hansch & Leo, 1995).

This proposal is aimed at critically reviewing the non-commercial (Q)SARs for mutagenicity and carcinogenicity, mainly in the light of regulatory purposes and OECD principles (OECD, 2004), and propose a reasoned list of promising models.

**Work programme**

The first step will be reviewing the currently available, non-commercial (Q)SARs for predicting mutagenicity and carcinogenicity. In our laboratory we have constantly followed the evolution of this field. This activity has generated, together with original research articles (see CVs), a number of reviews (Benigni et al., 1989; Benigni & Giuliani, 1996; Benigni, 2003; Benigni, 2004; Benigni, 2005). Our latter review paper (Benigni, 2005) is quite extensive and recent, and will provide the basis for the present work. This review will be updated with the most recent papers, and a new check of previous models will be performed as well.

In the second step we will assess how the models follow the OECD validation principles (OECD, 2004). We will consider both formal criteria (e.g., statistics) and contribution to mechanistic information. It can be foreseen that the (Q)SAR models vary to a large extent in terms of accordance with OECD principles. For example, only a few models refer to sets of chemicals with a sufficient number of representatives, and very rarely authors have explicitly taken into consideration and defined the applicability domain of the models.

The above situation will lead to the third step of the programme, which will be the generation of a list of promising models which have satisfactory concordance with the OECD principles (OECD, 2004).

In the fourth step we will perform additional work on the short-listed models, to obtain missing information. Given the non-standardized formats of the published models, this work will be necessary and will consist of: a) a number of statistical re-analyses and checks of the data provided in the original papers. This will be aimed at reconfirming the results of the authors, and to complete the information on, for example, goodness-of-fit, validation and applicability domain; b) whenever possible, further QSAR analyses.
Among the QSAR analyses, major consideration will be given to the possibility of performing external validation. Together with internal validation procedures (e.g., cross-validation) and even more so, the assessment of the ability of a model to correctly predict the behavior of chemicals that have not been considered in the phase of the development of the model is a crucial test. By the way, it contributes to a better definition of the applicability domain of the model. Unfortunately, for practical reasons this is seldom possible in toxicology. Since we have recently established a database of chemical carcinogens and made it available freely on the internet (http://progetti.iss.it/ampp/hhhh/hhhh.php?id=233 or through the DSSTox site http://www.epa.gov/nheerl/dsstox/ExternalPublicDatabases.html#ISSCAN), we will use this resource to check if further chemicals belong to the chemical classes of the selected (Q)SAR models. Other literature sources will be used as well. If further (external to the model) chemicals are available, we will apply the models to theoretically estimate their activity: the agreement (or disagreement) with the actual experimental data will provide a crucial information to judge the reliability (and limits) of the model. Particular attention will be given to the classes most represented in the toxicity database and most important from an industrial and environmental point of view.

For the class of the aromatic amines, which constitute almost one fifth of the entire carcinogenicity database, we will build a database including structures, mutagenicity and carcinogenicity data. The database will have the same format of the DSSTox files and will be freely available on the website of our Institute (http://progetti.iss.it/ampp/hhhh/hhhh.php?id=233). The structure of the database will be Chemical-Relational, and will consist of: a) an Excell file with all relevant information; b) a .sdf file, with the same information of the Excel file, plus a machine readable code of the chemical structures; c) a .pdf file with the graphical representation of the chemical structures. The database will permit both biological and chemical interrogation (and their crossing), and the chemical structures will be directly usable for computations. This database will be also useful in the future to other investigators and regulators that may want, for example, check the similarity of new aromatic amines with those in the database, and assess in this way their activity.

Since many of the non-commercial (Q)SAR models are in the form of the classical Hansch (or extra-thermodynamic) approach, and thus provide mechanistic information, we will attempt to exploit also this information by translating it into a qualitative language. For example, the QSARs on aromatic amines (Debnath et al., 1992; Benigni & Passerini, 2002) point to a remarkable influence of the substituent bulkiness in certain positions of the molecules. When translated into qualitative language, this information becomes knowledge about Structural Alerts. The identification of Structural Alerts is
very important, because they are at the basis of the prediction by human experts and they can be incorporated into computerized Expert Systems.

In the final, fifth step we will formulate a detailed judgment on each of the short-listed models, from a scientific point of view and regarding their suitability for the different purposes of the regulatory process.

References


Deliverables and Milestones

The deliverables will include all those explicitly requested by the European Commission (Points 1 to 6, Section 7 of Technical Annex A). In addition, we will make freely available on the website of our Institute a new Chemical-Relational database on the aromatic amines, including structures, mutagenicity and carcinogenicity data (together with additional information such as CAS, SMILES, names, synonyms, etc...). The mutagenicity data will include potency in the main Salmonella strains. The carcinogenicity data will include yes / no response, and potency in the different species (mouse and rat). Another additional deliverable will be information on Structural Alerts, as derived from the analysis of the individual (Q)SAR models.

Public dissemination of the results (e.g., papers in peer-reviewed journals, presentations at scientific meetings) will be discussed with JRC.

In particular:

Within 4 weeks: kick-off meeting with ECB at JRC, to discuss in detail the work to be carried out;
Within 6 weeks: delivery of the minutes of the kick-off meeting;
Within 4 months: identification of non-commercial (Q)SAR models which have satisfactory concordance with the OECD principles;
Performance of further statistical and QSAR analyses (when necessary);
Building of the database on aromatic amines (mutagenicity and carcinogenicity), and location on the web site of Istituto Superiore di Sanita’;
Within 5 months: organization of an Expert Consultation meeting on the results of the project;
Within 6 months: delivery of the final report on the “Evaluation of (Q)SAR models for mutagenicity and carcinogenicity”, including the minutes of the Expert Consultation meeting;
delivery of a CD containing the data relative to the short-listed models (in Excel-readable format);
delivery of a copy of all background materials used in the preparation of the final report. The final report (In English) will be delivered in 3 copies, together with an electronic version (Microsoft Word format), the CD and the background material.
## Team composition and experience of the proponents

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Position</th>
<th>Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romualdo Benigni</td>
<td>Senior Scientist</td>
<td>Mutagenicity and carcinogenicity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QSARs of mutagens and carcinogens</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QSAR methodology</td>
</tr>
<tr>
<td>Alessandro Giuliani</td>
<td>Senior Scientist</td>
<td>QSARs of mutagens and carcinogens</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QSAR methodology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multivariate statistical analyses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protein modeling</td>
</tr>
<tr>
<td>Cecilia Bossa</td>
<td>Post-Doc</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QSAR</td>
</tr>
</tbody>
</table>

Romualdo Benigni (see CV) has experience in the assessment of regulatory dossiers on chemicals submitted in compliance with EU legislation, and has been Project Leader of a number of projects.

Romualdo Benigni and Alessandro Giuliani (see CVs) have experience in the organization of Scientific Meetings.