Regulatory Toxicology and Pharmacology 72 (2015) 335–349

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

A practical application of two in silico systems for identification of potentially mutagenic impurities

Nigel Greene [a,*], Krista L. Dobo [b], Michelle O. Kenyon [b], Jennifer Cheung [b], Jennifer Munzner [b], Zhanna Sobol [b], Gregory Sluggett [c], Todd Zelesky [c], Andreas Sutter [d], Joerg Wichard [d]

[a] Computational Sciences CoE, Worldwide Medicinal Chemistry, Pfizer Worldwide Research and Development, Groton, CT 06340, USA
[b] Genetic Toxicology, Drug Safety Research and Development, Pfizer Worldwide Research and Development, Groton, CT 06340, USA
[c] Analytical R&D, Pharmaceutical Sciences, Pfizer Worldwide Research and Development, Groton, CT 06340, USA
[d] Bayer Health Care, Investigational Toxicology, Müllerstr. 178, 13353 Berlin, Germany

ARTICLE INFO

Article history:
Received 9 April 2015
Available online 15 May 2015

Keywords:
ICH M7
Computational toxicology
Genotoxic impurities
Structure–activity relationships

ABSTRACT

The International Conference on Harmonization (ICH) M7 guidance for the assessment and control of DNA reactive impurities in pharmaceutical products includes the use of *in silico* prediction systems as part of the hazard identification and risk assessment strategy. This is the first internationally agreed guidance document to include the use of these types of approaches. The guideline requires the use of two complementary approaches, an expert rule-based method and a statistical algorithm. In addition, the guidance states that the output from these computer-based assessments can be reviewed using expert knowledge to provide additional support or resolve conflicting predictions. This approach is designed to maximize the sensitivity for correctly identifying DNA reactive compounds while providing a framework to reduce the number of compounds that need to be synthesized, purified and subsequently tested in an Ames assay. Using a data set of 801 chemicals and pharmaceutical intermediates, we have examined the relative predictive performances of some popular commercial *in silico* systems that are in common use across the pharmaceutical industry. The overall accuracy of each of these systems was fairly comparable ranging from 68% to 73%; however, the sensitivity of each system (i.e. how many Ames positive compounds are correctly identified) varied much more dramatically from 48% to 68%. We have explored how these systems can be combined under the ICH M7 guidance to enhance the detection of DNA reactive molecules. Finally, using four smaller sets of molecules, we have explored the value of expert knowledge in the review process, especially in cases where the two systems disagreed on their predictions, and the need for care when evaluating the predictions for large data sets.

1. Introduction

The recently adopted ICH M7 guidance provides a framework for the assessment and control of DNA reactive impurities in pharmaceuticals (ICH M7, 2014). The intent of the guideline is to focus efforts on the identification and control of substances that can cause DNA damage at low levels, and therefore cause mutations and increase cancer risk. These types of compounds are usually detected in the Ames bacterial reverse mutation assay (McCann et al., 1975; Zeiger, 1987, 1998). Analyses have shown a good correlation of Ames test data with carcinogenicity data for genotoxic carcinogens and non-carcinogens (Contrera et al., 2005; Kirkland et al., 2005; McCann and Ames, 1976). However, genotoxic agents with non-linear dose–response relationships and a negative outcome of the Ames test typically do not pose an increased cancer risk at levels present as impurities (Dobo et al., 2012). The concept of using structural alerts to predict the mutagenic activity of chemicals was introduced more than 30 years ago (Ashby and Tennant, 1991). Computational, or *in silico* methods that make use of known relationships between chemical structure and mutagenicity have been developed based on the findings of Ashby and Tennant, and have been continuously improved over time. Similarly, computational algorithms that use a variety of statistical correlation methods to make associations between structural features and a chemical's mutagenic activity have been developed since the first publication of large data sets of compounds and their respective results in the Ames assay. The ICH M7 guideline calls for both of these approaches to be used as part of the hazard identification process along with an expert review process. Recommendations for the implementation of these methods have been previously published (Sutter et al., 2013) but a systematic evaluation of systems using a common data set was not performed nor were specific examples used to illustrate the expert review process.

* Corresponding author.

http://dx.doi.org/10.1016/j.yrtph.2015.05.008
0273-2300/

The implications of implementing these guidelines are now being realized within the pharmaceutical industry and as such there remain some important considerations that need to be better understood. For example, the question of which two systems work best together and how to deal with conflicts in predictions from the different methodologies. The present paper systematically looks at some of the *in silico* (quantitative) structure–activity relationship ((Q)SAR) systems most commonly used in the pharmaceutical industry and compares their respective performances for a set of 801 compounds assembled as part of a data sharing consortium, that represent common intermediates and starting materials used in the synthesis of pharmaceuticals. A large proportion of the data set used has not been previously published and so is unlikely to form part of the training sets for the statistical algorithms. However, these data may have been used to enhance the structural alerts contained within the Derek software. In addition, we have taken 4 small subsets of structures where the computational algorithms did not always agree with one another to look at how expert review may help differentiate mutagenic and non-mutagenic compounds.

2. Methods

The data set used in this analysis consisted of 801 compounds taken from the Intermediates Data Sharing Group organized and maintained by Lhasa Limited. Only compounds that had been tested in a standard Ames assay using 5 strains or demonstrated to be mutagenic in a screening assay format were included in the evaluation set.

Computational algorithms evaluated at the time of analysis were as follows:

1) Derek Nexus (Derek) v3.01 using the 2012.1.0 version of the knowledge base available from Lhasa Limited (http://www.lhasalimited.org/). This is classed as a rule-based expert system.
2) Sarah Nexus (Sarah) v1.0 Pre-release prototype also available from Lhasa Limited. This is classed as a statistical algorithm.
3) Leadscope Model Applier v1.6.0-17 using the Genetic Toxicity Suite Microbial in vitro – Salmonella v3 model available from Leadscope Inc (http://www.leadscope.com/). This is classed as a statistical algorithm.
4) Case Ultra using Modules AZ2/AZ3 for extended Salmonella Ames Mutagenicity available from Multicase Inc (http://www.multicase.com/). This is classed as a statistical algorithm.

It should be noted that development of all of these algorithms is a continuous process and so newer versions of these applications may now be available. Specifically, the final production version of Sarah saw some significant improvements in the predictive performance with a sensitivity of 75.8% and overall accuracy of 72.3%. However, this analysis represents a snapshot in time intended to demonstrate the impact of using two systems and illustrate the importance of the expert review process with specific examples rather than to compare and contrast the merits and limitations of each system.

2.1. Examples of using expert review

Four subsets of compounds were selected for expert review, each one based on a common chemical substructure. These sets of compounds were selected due to the high number of discordant results between the *in silico* systems and so would provide typical examples of where expert review would be needed.

2.2. The expert review process

In general, Derek mutagenicity alerts and Leadscope Salmonella Mut Microbial Gene Mutation predictions were reviewed for each structure being analyzed. Structural features contributing to a Leadscope prediction were reviewed by looking at the training structures and the bacterial mutagenicity result to ascertain whether features predicted positive were likely due to that feature or another mutagenic substructure. When examining the supporting structures in the training set, if a significant number of Ames positive structures also contained other likely mutagenic substructures (e.g. nitroaromatics) that were not present in the query structure or the training set Ames negative structures, then relevance of the positive prediction was considered questionable. Further evaluation of each queried structure was generally conducted using structure-searchable databases to identify mutagenicity data for exact structures or structurally similar compounds with the same alerting structural features. This included the Chemical Carcinogenesis Research Information System (CCRIS) and GENETOX databases within the United States National Library of Medicine TOXNET Toxicology Data Network, Vitic Nexus (Vitic) databases (Lhasa Limited), excluding the Vitic Intermediates database, as well as a Pfizer database summarizing both proprietary mutagenicity data and mutagenicity data contained in the Benchmark data set for *in silico* prediction of Ames mutagenicity (Hansen et al., 2009). Additionally, mutagenicity data for structural analogs identified in the Leadscope Genetox Level 2 Database were reviewed. In some instances, a chemistry-based assessment of how a substituent affects the electrophilicity of the structural feature predicted to be mutagenic was conducted. In the few cases where mutagenicity data was found for the exact structure being queried, the compound was defined as mutagenic or not based on the data and no additional structure searching was performed. Otherwise, reviewers classified structures as potentially mutagenic or non-mutagenic based on their review of the data. At least two other individuals, one a genetic toxicologist and the other a chemist, reviewed all of the structural classifications and the final classification was agreed upon by consensus where possible.

Set 1: Fourteen chemicals (see Table 3 in the results and discussion), 10 Ames positive and 4 Ames negative, were selected based on a common 2-aminopyridine substructure (see Fig. 1).

Set 2: Twenty-two chemicals (see Table 4 in the Section 3), 9 Ames positive and 13 Ames negative, were selected based on a common 2-aminobenzoic acid substructure (see Fig. 2).

Set 3: 12 chemicals (see Table 5 in the Section 3), 6 Ames positive and 6 Ames negative, were selected based on a common 4-aminopyridine substructure (see Fig. 3).

Set 4: Thirteen chemicals (see Table 6 in the Section 3), 4 Ames positive and 9 Ames negative, were selected based on a common sulphonylchloride substructure (see Fig. 4).

3. Results and discussion

The predictive performance of each computational program for the entire 801 compound data set is summarized in Table 1. When

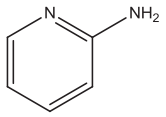

Fig. 1. 2-Aminopyridine substructure.