MUT/MIN/2020/1.5

Minutes: Lhasa presentation on the use of *in silico* for predictions of genotoxicity

Dr Robert Foster presented an update on the Lhasa Limited *in silico* prediction models for genotoxicity. Lhasa, was established in 1983, and has its Head Quarters in Leeds, United Kingdom. It is a Not-for-profit & Educational Charity. Licence holders become members of Lhasa, all members are encouraged to share and publish their data, working collaboratively with Lhasa scientists to improve their products.

Dr Foster introduced (Q)SAR systems, using Derek & Sarah Nexus as examples, before discussing the performance of (Q)SAR systems and model development with respect to genotoxicity.

Derek was one of the first pieces of software to include reactive groups in the form of structural alerts that flagged potential genotoxicity to the user. In the early years, alert refinement focused on alert specificity as the 'Ashby & Tennant' alerts were too general. Over subsequent years, improvements were made to the interface and efforts were made by the scientists at Lhasa to identify new toxicophores and structure activity relationships from the general literature. These were then encoded as new structural alerts in the system. For mutagenicity endpoints, the number of alerts increased from 58 in 1996, to 91 in 2010, to 132 alerts in the most recent release.

Derek Nexus is an expert rule-based model with a knowledge base incorporating 132 alerts for mutagenicity in 2018. It has been built using public and confidential data in collaboration with regulators and industry members. Thirty five percent of alerts for mutagenicity are based upon proprietary data. A new release to be published soon will comprise of 220 alerts for genotoxicity: 148 mutagenicity alerts; 99 chromosome damage alerts; and 5 non-specific genotoxicity alerts. Rules written by experts, incorporating chemical reactivity, metabolism, toxicology expertise make the SAR more relevant, this is an advantage of a rule-based system.

Sarah Nexus is a statistical system. Statistical systems are built on a large training set and use a computer algorithm to look for associations. In Sarah the compounds in the training set are fragmented, as DNA reactivity is associated with the chemical reactive group, Ames activity is then associated with each fragment.

In Sarah the training set contains 9882 individual structures from the public domain and member donations. This is made up of 4716 (48%) mutagens and 5166 (52%) non-mutagens These molecules are fragmented. Sarah then generates a *hypothesis* for a fragment. For a query structure the program carries out the same process, takes the fragments and assigns the activity as in the training set.

Although, Derek and Sarah are different types of models, both demonstrate a large amount of information to the user about the prediction for them to review. A key attribute of the Lhasa software is transparency of results to provide the reasoning behind the alert implementation (Derek) or prediction (Sarah) so users have the necessary information to review the prediction for their query. Lhasa systems are designed to be fully transparent to present the user with the information to show why the alert writer implemented the alert in Derek Nexus and the training set is shown in Sarah Nexus. These models perform better on public data as the compounds in the training sets are mainly public chemicals.

For mutagenicity it is accepted that these models perform very well, they are accepted for regulatory purposes. The ICH M7 guidelines state that one expert rule-based and a statistical-based model can be reviewed with expert knowledge to support the final conclusions for the mutagenic potential of impurities.

Improving the algorithm or display of results may be important for newer models but established systems such as Derek and Sarah may focus on improvements through increasing the chemical space coverage by addition of data from public and/or proprietary sources. Donation of proprietary data encourages collaboration to benefit scientific community. In Derek 35% of alerts for mutagenicity are constructed or refined using proprietary data.

This also benefits members as this improves the models in their chemical space. The chemical structure is generalised and benefits other members. The data is used in Sarah will have to be published and be in the public space.

Dr Foster then went on to give examples of how public and private data have been used to develop Lhasa products.

The first examples looked at public data and genotoxicity for alylbenzenes. These compounds are relevant for food as well as pharmaceutical domains. For this group of chemicals three alerts were written: the Ames alert, was the most restrictive as S9 may not be metabolically capable, an alert for chromosomal aberrations and the UDS as a non-specific genotoxicity. An advantage of rules-based system they can be very specific and use expert knowledge.

The second example focused on the use of proprietary data. In a collaboration with Japanese Pharmaceutical Manufacturers Association which is a consortium of 11 companies. A large data set was shared which predicted with low sensitivity. The data were then curated and clustered, and alerts were refined, or new alerts were implemented. This led to a large increase in the predictive performance against the dataset.

Improving predictivity in Sarah: Sarah uses published literature. In 2019 Lhasa focused on increasing coverage of nitrosamines due to issues in pharmaceuticals. Lhasa's data team rapidly increased the coverage of nitrosamines form published data and added 95 added data sets to Vitic. All of these have gone into the Sarah training set.

Finally, Dr Foster discussed *in silico* predictions of genotoxicity in particular. He referred to the recent EFSA Supporting publication 2019:EN-1598. (Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across. doi:10.2903/sp.efsa.2019.EN-1598.)

This report concluded that QSAR predictions work well for mutagenicity but does not work so well for other endpoints. The reasons for this are given as the Ames test is very well accepted the protocols are standardised, for other genotoxicity tests the models are not as well established. There is also far greater Ames data available for model building compared to other tests for genotoxicity such as chromosome aberration and micronucleus tests.

A validation of Derek against chromosome aberration data showed that it performed well on chemicals which are expected to be DNA reactive. But Derek has low sensitivity for prediction of a set of compounds known to interact with either topoisomerase or tubulin. In Derek chromosomal damage (CD) alerts primarily cover NDA/protein reactive compounds. This is an issue with rule-based systems where creating a valid SAR is incredibly difficult for complex, poly(hetero)aromatic ring systems.

Dr Foster demonstrated how a statistical system may be able to complement the rule-based system by creating a Sarah model for the prediction CD. Data were taken predominantly from Vitic Nexus. Each time a compound is positive in both *in vitro* CA or *in vitro* MN data sets it is counted as positive in CD. This model is significantly more sensitive for prediction of chromosome damage compared to Derek. However, it is important to note that Sarah was designed for the prediction of mutagenicity *in vitro* and, in line with the report by EFSA, additional refinement would be required to the model should it be considered for use for prediction of chromosome damage *in vitro*.

Dr Foster Concluded his presentation with a summary:

In silico systems can provide predictions of genotoxicity. Very strong performers in mutagenicity. Performance accepted by regulators under ICH M7.

The push towards other genotoxicity endpoints - this still requires more research. This is discussed in the EFSA report.

Lhasa is addressing the issues highlighted in the EFSA report. One way would be to increase the amount of data used to develop the models. QSARs don't necessarily need to predict the endpoint but could be used, for example, in adverse outcome pathways.

A member of COM asked if there are examples of requests for further tests such as CA, for Lhasa to fill data gaps. This does happen where data gaps have been identified and Lhasa members will contribute further data. Members can also challenge the predictions. This dialogue helps to improve the Lhasa products.

In terms of the robustness of mutagenicity data there is a large volume of mutagenicity data but much less for chromosomal aberrations, the quality of the data can also be a factor with a number of false positives. The data is quality controlled before it is incorporated into the Sarah chromosome damage (CD) model, a reliability tag is added. These tags are used to quality control the data before it goes into Sarah. But there has to be balance between removing data which may be flagged and having sufficient data sets to build a model - ensuring the user is then aware of the quality of the data which has been used and therefore the reliability of the predictions is essential for review

There has been a lot of debate as to whether *in vivo* data should be used instead of *in vitro* data if available However, there is a lot of *in vivo* MN but few *in vivo* CA data. So how can these data sets be combined? The CD model was built on *in vitro* MN and CA data. There are over 1500 compounds with *in vivo* data, this data could still be made available to a user for review but not used in a prediction. The predictivity of the models is only as good as the predictivity of the data upon which they are based. QSAR predictions can be used in a weight of evidence assessment and to guide testing.

QSARs which predict AMES positive results are basically showing compounds which can produce an electrophilic DNA reactive species. QSARs can be used to predict initiating events and not the endpoint this is a way to look at mechanisms in an adverse outcome pathway assessment. But this is outside the scope of Derek or Sarah.

There was discussion with regard to the availability of negative data. Pharmaceutical companies will submit both negative and positive data, that is proprietary data. It is often herder to find negative data in the open literature. This can cause some bias in the models.

The Chair thanked Dr Foster for a very informative presentation and discussion.