Website Content Analysis

Are you spending hours reading websites for a client, a prospect, or while conducting market research?

Oftentimes analyzing a website requires manually scanning dozens of pages, then taking screenshots, downloading PDFs, eventually writing a report of the website’s commonly used phrases and concepts.

An expensive, often complicated solution is to hire or develop a web scrapping solution.

Working with us, give us the domain names you’re interested in, and you’ll receive results in 48 hours.

Our service allows you to analyze an entire website in minutes by scanning a single text document with the extracted text from all HTML pages, PDFs, and images available in the website. We also provide the top 500 most-used phrases of 3, 4, 5, and 6 words (called “N-grams”) in that website.

You’ll be able to quickly learn the language of the website’s owner. To help you craft a better story. And to better understand a particular industry. While saving time. You’ll be able to better serve more clients.

## Benefits

\* Fully automated, done for you

\* Fast turnaround: 48 hours

\* No coding. No IT involvement.

\* Flat fee pricing: US$200.00 per domain name. PDFs and image files are available for an extra US$100.00 per domain name. (Note: PDFs locked behind registration forms are not available).

## Who do we serve?

Our website content analysis service allows

\* Marketing agencies to:

 prepare better pitches for prospects by analyzing the prospect’s website

 develop effective branding campaigns for a client by analyzing the websites of client’s own clients

\* Market research agencies to:

 quickly analyze dozens, even hundreds of websites in a particular industry

## Deliverables

For each domain name that you define we’ll deliver to you:

\* a single text file with all text extracted from HTML, PDF, and image files

\* N-gram files (3, 4, 5, and 6 words)

\* PDFs and image files if desired

Please visit <https://DataSDR.com/website-content-analysis> to download free samples.

Contact:

 José C. Lacal, CTO

 Jose.Lacal@DataSDR.com

 {EU} +34 (674) 88 17 52

 {US} +1 (561) 777-2577

Description

This Word document contains the text extracted from the PDF files available at https://www.instem.com/

You can literally "navigate" this entire website in this single file.

Now you can analyze the entire contents of the website in your computer, even without an Internet connection.

Source PDF: /instem.com/www.instem.com/docs/Maintenance\_Package.pdf

 Maintenance Packages (3 Levels) Standard Select Basic
 Support Support Support
 (End of Life) (Academia)
 Software Update
 All new software releases of your licensed products (Genetic Tox products: updates only) ◼ ◼ ◼
 Information on new software releases ◼ ◼ ◼
 Product Support
 Unlimited Support Requests ◼ ◼
 Normal Business Hours Telephone Support for Designated Time Zones ◼ ◼ ◼
 Problem diagnosis & software error identification of:
 Software usage issues ◼ ◼ ◼
 Basic self-installation & upgrade issues ◼ ◼ ◼
 Basic instrument configuration issues ◼ ◼ ◼
 Basic 3rd party software configuration issues ◼ ◼ ◼
 Investigate faults ◼ \*
 Fix data by issuing Direct Database Updates ◼ ◼
 Provide known workarounds on customer’s version ◼ ◼ ◼
 Develop new workarounds (if possible) ◼ \*
 24 x 7 access to product Web Communities ◼ ◼ ◼
 Customer Service Notices ◼ ◼ ◼
 Invitations to Webcasts ◼ ◼ ◼
 Roadmap Information ◼ ◼ ◼
 On-line Incident tracking ◼ ◼ ◼
 Customer advocate meetings and status reports ◼ \*
 Bug Correction
 Create New fault fix releases ◼ \*
 Genetic Toxicology
 Repair/replacement of hardware supplied by Instem ◼
 Safety Pharmacology
 Data analysis optimization ◼
 Assistance in setting up & improving configurations for data acquisition & analysis ◼
 Assistance in optimizing simple Microsoft Excel® extraction models ◼
 Guidance on hardware utilization and optimization ◼
 SEND Product Support
 Basic FFDA configuration issues ◼ ◼
 SEND and Define Standards Support \*\* \*\*
 SEND Service Queries \*\* \*\*
 Audit Support
 Customer Audits ◼ \* \*\*
 On-site Product Support
 On-site assistance on specific requests, configuration check, answer to questions on \*\* \*\*
 software use
\* Additional charges apply for up to 2 additional years coverage after End of Life (Extended Support)
\*\* Additional charges apply
Notes:
Basic Support is for Academia only. Standard Support customers are automatically transferred to Select Support upon end of life.
Customers without a purchased plan will receive information on how to get back into good standing with support.
Software includes validation scripts only where an annual maintenance fee is paid specifically for script support.

Source PDF: /instem.com/www.instem.com/docs/sot-2021/Acute-QSAR-final.pdf

 Contents lists available at ScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepag e: www.el sevier.com/loc ate/yrtph
A cross-industry collaboration to assess if acute oral toxicity (Q)SAR models
are fit-for-purpose for GHS classification and labelling
Joel Bercu a, Melisa J. Masuda-Herrera a, Alejandra Trejo-Martina, Catrin Hasselgren b,
Jean Lord n, Jessica Graham c, Matthew Schmitzd, Lawrence Milchak e, Colin Owens e,
Surya Hari Lal f, Richard Marchese Robinson f, Sarah Whalley f, Phillip Bellion g,
Anna Vuorineng, Kamila Gromek h, William A. Hawkins i, Iris van de Gevel j, Kathleen Vriens j,
Raymond Kemperk, Russell Naven k, Pierre Ferrer l, Glenn J. Myatt m,\*
aGilead Sciences, 333 Lakeside Drive, Foster City, CA, USA
bGenentech, Inc., 1 DNA Way, South San Francisco, CA, 94080, USA
cBristol Myers Squibb, 1 Squibb Dr, New Brunswick, NJ, 08903, USA
dAbbVie Inc., North Chicago, IL, USA
e3M Company, 3M Center, St. Paul, MN, 55144-1000, USA
fSyngenta Crop Protection, Product Safety Department, Jealott’ s Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, UK1
gDSM Nutritional Products, Kaiseraugst, Switzerland
hGalapagos SASU, 102 Avenue Gaston Roussel, 93230, Romainville, France
iGlaxoSmithKline, Park Road, Ware, Hertfordshire, SG12 0DP, United Kingdom
jJanssen Pharmaceutical Companies of Johnson & Johnson, 2340, Beerse, Belgium
kVertex Pharmaceuticals Inc., Discovery and Investigative Toxicology, 50 Northern Ave, Boston, MA, USA
lDepartment of Veterinary Physiology and Pharmacology, Interdisciplinary Faculty of Toxicology Program, Texas A& M University, 4466 TAMU, College Station, TX,
77843-4466, USA
mLeadscope (an Instem company), 1393 Dublin Rd, Columbus, OH, 43215, USA
nUltragenyx, 60 Leveroni Court, Novato, CA, 94949, USA
ARTICLE INFO ABSTRACT
Keywords: This study assesses whether currently available acute oral toxicity (AOT) in silico models, provided by the widely
Acute oral Toxicity (Q)SAR employed Leadscope software, are fit-for-purpose for categorization and labelling of chemicals. As part of this
In silico 3Rs Expert review study, a large data set of proprietary and marketed compounds from multiple companies (pharmaceutical, plant
Expert rule-based Statistical-based model protection products, and other chemical industries) was assembled to assess the models’ performance. The ab-
Classification and labelling solute percentage of correct or more conservative predictions, based on a comparison of experimental and
CLP/GHS GHS predicted GHS categories, was approximately 95%, after excluding a small percentage of inconclusive (inde-
 terminate or out of domain) predictions. Since the frequency distribution across the experimental categories is
 skewed towards low toxicity chemicals, a balanced assessment was also performed. Across all compounds which
 could be assigned to a well-defined experimental category, the average percentage of correct or more conser-
 vative predictions was around 80%. These results indicate the potential for reliable and broad application of
 these models across different industrial sectors. This manuscript describes the evaluation of these models,
 highlights the importance of an expert review, and provides guidance on the use of AOT models to fulfill testing
 requirements, GHS classification/labelling, and transportation needs.
1. Introduction general degrees of toxicity and understand the potential for a compound
 to cause life-threating effects from an acute exposure. Regulatory au-
 The purpose of the acute oral toxicity (AOT) study is to characterize thorities often require the AOT testing of substances in order to
 \*Corresponding author. Present address: Leadscope, Inc., 1393 Dublin Road, Columbus, OH, 43,215, USA.
 E-mail address: gmyatt@leadscope.com (G.J. Myatt).
 1 Any example workflow or guidance outlined in this paper is not currently endorsed or approved by Syngenta.
https://doi.org/10.1016/j.yrtph.2020.104843
Received 24 September 2020; Received in revised form 19 November 2020; Accepted 14 December 2020

J. Bercu et al.
characterize their toxicity and assign hazard categories, which informs toxicity, are designated as “ not classified” according to the CLP
the labelling of products to indicate appropriate restrictions or pre- regulation.
cautions to be taken during their handling, transportation, or use There is a balance in toxicology research for understanding the
(Hamm et al., 2017). While the exact requirements for the content and hazards of chemicals versus the need for animal testing (OECD 2001).
formatting of labelling may vary by the product type, regulatory agency, The “ 3Rs” is a global initiative geared toward reducing animal use in
and use context, there have been numerous international efforts to research and stands for (1) Replacing animal-dependent study methods
harmonize hazard identification, and classification and labelling over with reliable/comparable alternative methods, (2) Reducing the number
the last several decades (Strickland et al., 2018). Examples of frame- of animals in a study, and/or (3) Refining studies to improve animal
works include the United Nations (UN) Recommendations on the welfare (Russel and Burch, 1959). Industry implements the 3Rs to
Transport of Dangerous Goods and the Globally Harmonized System accelerate scientific discovery, support innovation and technological
(GHS) of Classification and Labelling of Chemicals (UN 2019a; UN developments, and address societal concerns about animal research.
2019b). Each framework is regularly revised and updated to reflect There are ongoing national and international efforts to employ the 3Rs
national, regional and international experiences in implementing their across toxicology testing and gain regulatory endorsement (NC3Rs
requirements into laws, as well as the experiences of users who perform (2020); EFPIA (2019); AnimalResearch.Info (2018); Lautenberg Chem-
the classification and labelling (UN (2019a)). ical Safety Act (2016); Tox21 (2008)). Additionally, the EU Directive
 AOT studies are required for the majority of compounds as part of the 2010/63/EU mandated the application of reduction, refinement and
European Union’ s (EU’ s) legislation on the registration, evaluation, replacement across the EU (EU Directive, 2010/63/EU).
authorization and restriction of chemicals (REACH) produced at ≥ 1 There have been efforts to reduce the number of laboratory animals
tons per year and manufactured or imported in the EU or European needed for the existing in vivo methodologies utilized for determining
Economic Area (EEA) (EU 2006; ECHA 2015) as well as other interna- the AOT of compounds. The new OECD guidelines for AOT studies
tional compound registrations. AOT information is also utilized to define reduced the number of animals needed to define a point estimate while
labeling information for safety data sheets (SDS) and containers as also enabling a more harmonized approach to classifying compounds
defined by the UN’ s GHS for classification and labelling of chemicals (i. based on their AOT hazard (UN GHS 2005). Introduction of a limit dose
e., the purple book, EU’ s Classification, Labelling and Packaging (CLP)) (2000 mg/kg) and a maximum tested dose (5000 mg/kg) to define “ not
(UN GHS 2005; EU 2017). Finally, AOT information guides how a acutely toxic” , also reduced the number of animals required for com-
chemical should be packaged, labeled and, or transported (49 CFR, Part pounds of low toxicity as there was no need for excessive dosing (OECD
178; 16 CFR 1500.3; IATA 2020). The well-established practice and 2002a; OECD 2002b; OECD 2008; UN GHS 2005). The approval of the
widespread use of AOT studies for these intended purposes, as well as an Fixed Dose Procedure (OECD TG 420), Acute Toxic Class (OECD TG 423)
overall lack of non-animal alternatives, results in the mandated neces- and Up and down procedure (OECD TG 425) were also considerable
sity to continue to conduct these tests. advances as historical studies utilized ~100 animals per study and these
 The median lethal dose, LD50, is a general indicator of a chemical newer test guidelines utilize 2– 15 animals per study (Erhirhie et al.,
substance’ s acute systemic toxicity. The LD50 values from acute toxicity 2017). In addition, the fixed dose procedure relies on clear signs of
tests in rodents serve as the basis for the toxicological classification. The toxicity at fixed dose levels versus lethality, which reduces animals and
most commonly performed tests for acute toxicity are described in the offers a refinement that improves animal welfare (OECD 2002a).
OECD guidelines (OECD 2008) and are essentially identical to those At the time of preparing this paper, there are no validated (e.g. OECD
called for under the Toxic Substances Control Act (TSCA) (TSCA 2016), test guidelines), internationally accepted, animal-free alternatives to the
Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) (FIFRA acute oral toxicity animal study that regulatory bodies accept. Based on
1996), and REACH regulations. The AOT tests, including the limit test, their common use in cytotoxicity assessments, the 3T3 (mouse fibro-
fixed-dose procedure, toxic class method, and up-and-down methods blasts) neutral red uptake (NRU) and the NHK (human keratinocytes)
(OECD 2002a; OECD 2002b; OECD 2008, respectively), each represent a NRU in vitro methods have been evaluated as potential alternatives to
more simplified study design compared to the original animal test AOT testing (Creton et al., 2010; Schrage et al., 2011; OECD 2010).
method (OECD 401, which was deleted in 2002) as a means of mini- However, these methods were found to not be sufficiently accurate as
mizing animal use. stand-alone test methods but recommended to be incorporated as part of
 GHS provides an internationally compatible system to classify and a weight of evidence approach for the selection of starting doses for
communicate physical, health, and environmental hazards of a sub- rodent AOT tests (Creton et al., 2010). (Quantitative) structure activity
stance for the protection of humans and the environment. Several relationship – or (Q)SAR – models have also not been sufficiently
toxicological endpoints are presented in the GHS regulation to enable developed or validated to enable them to be used as stand-alone alter-
proper hazard classification, including acute toxicity by the oral, natives to animal testing or to classify and waive/not test in the case of
dermal, and/or inhalation (gases, vapors, dusts & mists) route. There are REACH. However, (Q)SAR information can be used to supplement
five GHS categories for acute toxicity (Category 1– 5), which are banded experimental test data as part of a weight of evidence or an Intelligent
based on the dose or concentration required to produce a severe toxic Testing Strategy (ITS) approach (ECHA, 2008; Creton et al., 2010).
effect or death in 50% of the exposed population (i.e., LD50), with AOT in silico model development is aligned with the 3Rs mission to
Category 1 chemicals being the most toxic (see Table 1). These five acute replace existing methods that require laboratory animals. An AOT in
toxicity classification categories have corresponding pictograms, signal silico model offers an animal-free way to elucidate a compound’ s acute
words, and hazard statements, which are used for hazard communica- hazards to fulfill testing requirements, classification/labelling, or
tion on safety data sheets and chemical labels (UN GHS 2005). It should transportation purposes. Fundamental to the success of a global AOT in
be noted that not all classification categories are adopted in all regions in silico model is a sufficiently representative, large and high-quality
the world. Regulation (EC) 1272/2008 on classification, labelling and database and algorithms which have the capability to make reliable
packaging of substances and mixtures (CLP Regulation, EU 2008) has predictions for a broad range of chemical structures. (In the case of a
adopted Categories 1– 4, whereas category 5 substances, with a low statistical QSAR, the model itself would be derived from the database
Table 1
GHS classification criteria for AOT.
 Acute Toxicity Category 1 Category 2 Category 3 Category 4 Category 5 Not classified (NC)
 Oral (mg/kg) LD50 ≤ 5 5 < LD50 ≤ 50 50 < LD50 ≤ 300 300 < LD50 ≤ 2000 2000 < LD50 ≤ 5000 5000 < LD50

J. Bercu et al.
using an algorithm, but the manner in which any (Q)SAR makes pre- 2. Methodology
dictions of chemical hazard may be considered an algorithm, with data
not seen during the model development procedure required for external 2.1. (Q)SAR models
validation of the final model.) A reliable AOT in silico model could
complement an existing laboratory study to further reduce animals or There are two commonly used (Q)SAR methodologies referred to as
refine existing procedures. For example, an in silico AOT model can assist expert rule-based and statistical-based (Myatt et al., 2017). Leadscope
in predicting the starting dose for the OECD 420 AOT test (the only AOT (an Instem company) has recently developed and made available a first
test with a non-lethal endpoint), enabling the minimum number of an- generation of (Q)SAR models covering both methodologies to predict
imals to be used and avoid lethality. Another example is if the LD50 is GHS categories for rat acute oral toxicity (Leadscope 2020). Both
predicted to be > 2000 mg/kg, the limit dose can be utilized as the methodologies use a database of over 15,000 chemicals with rat AOT
starting dose with greater confidence, eliminating the need for lower results from a number of sources including the Registry of Toxic Effects
doses to be tested and reducing the number of animals used. In addition in Chemical Substances, ECHA, EU’ s Joint Research Council’ s Acu-
to use in regulatory requirements, classification and labelling, and toxBase, National Library Medicines (NLM) Hazardous Substances Data
transportation needs, a reliable AOT in silico tool has potential utility in Bank, OECD (eChemPortal), PAI (NICEATM) and TEST (NLM Chem-
early stages of research and development as an alternative to in vivo IDplus) (RTECS 2011; Kleinstreuer et al., 2018).
testing for assessing the likelihood of acute oral toxicity for a given A series of individual models have been developed from this com-
chemical series to guide subsequent testing strategies and compound bined dataset and used to predict GHS categories (1– 5 and NC). These
design. individual statistical models or sets of expert alerts predict whether a
 If an alternative model predicts AOT as reliably as an in vivo study, chemical is below a specified LD50 threshold corresponding to the GHS
the alternative method should be preferred and supported. When eval- cut-off values. The statistical-based models use a Partial Logistic
uating an alternative method, it should also be understood that the in Regression algorithm that incorporates structural features and calcu-
vivo AOT test itself has a variable response (Pham et al., 2020). Vari- lated physico-chemical properties. Whilst the models have undergone
ability, i.e. differences in the GHS class observed for the same chemical, subsequent development, the models build upon the approach previ-
has been observed in animal studies with 18%– 25% of studies ously reported in the literature (Yang 2005). For the expert rule-based
(depending on the route of exposure) on the same compound resulting in models, a set of 2867 structural alerts were encoded that will predict
a different GHS category (Allen et al., 2019) and even more-so (25– 27% whether a chemical is below a specified GHS threshold. These models
variability; Karmaus 2018) in test sets currently under investigation as are then used within a decision tree to compute a GHS category (Myatt
alternatives to the AOT test. In silico models should not show variability et al., 2019).
for the same compound, but their accuracy or apparent accuracy will This decision tree approach is outlined in Fig. 1 where for each in-
necessarily be limited by the variability in the experimental data used dividual methodology a GHS category is predicted, as well as an overall
for training and/or testing. Still, if experimental endpoint values used GHS category prediction derived from the individual methodologies. In
for training or testing were derived from multiple test results per Fig. 4, a chemical is predicted to be GHS category 3 using the expert rule-
chemical, the variability in the endpoint data could be reduced from the based approach and GHS category 4 using the statistical-based meth-
variability in single test results, potentially allowing in silico predictions odology. For the expert rule-based method, a set of alerts predicts
to be more reliable than individual test results, but not more reliable than whether the chemical’ s LD50 is below the 5 mg/kg threshold. Since it
the endpoint values seen during training. Therefore, it is expected that was not predicted to be below this threshold, a second alert set is used to
there will be an acceptable limit on the accuracy of in silico predictions as determine whether the chemical is below the 50 mg/kg threshold.
has been observed with AOT responses in animal studies. Again, the prediction was negative; however, a third set of alerts pre-
 (Q)SAR2 in silico models are increasingly being considered to predict dicted the chemical was below the 300 mg/kg threshold. Therefore, it
specific toxicological endpoints, such as LD50, based on the chemical was predicted to be between 50 and 300 mg/kg and hence assigned to
structure alone (Lapenna et al., 2010; Drwal et al., 2014; NASEM 2015; GHS category 3. A similar process was performed using a series of
Kleinstreuer et al., 2018). The purpose of this paper is to explore the use statistical-based models as shown in Fig. 1. In this case, the overall
of in silico models to advance the 3Rs for AOT. This paper will assess in prediction was category 4 (LD50 in the range of 300– 2000 mg/kg). The
silico models against chemicals such as pharmaceuticals, pharmaceutical most conservative value (GHS category 3) was used as the final
intermediates, plant protection products, plant protection product in- consensus model from the two methodologies.
termediates, metabolites, and starting materials, along with specialty The models allow for inspection of the underlying model informa-
chemicals submitted by manufacturers to determine their performance tion, such as feature weightings, to support an expert review. In addi-
compared to animal models. The results will guide the use and appli- tion, it is possible to review analogs in the database to provide additional
cation of in silico models within the framework of existing regulations supportive evidence, as shown in Fig. 2.
such as REACH, GHS, and transportation. Specifically, the following Collaborators were given access to the acute toxicity (Q)SAR models
paper outlines a cross-industry collaboration where each organization from Leadscope (Leadscope acute rat oral QSAR (v1) and alerts (v1)
collected historical AOT experimental data and ran AOT models over [System: Leadscope Model Applier v2.4]) to use in this exercise. Each
these chemicals. Each collaborator shared the experimental and pre- collaborator collected historical information on chemicals where a rat
dicted results and an analysis of all results was performed to understand AOT had been performed, with a Klimisch score of 1 or 2 (Klimisch et al.,
the AOT model’ s performance across different methodologies, across 1997) where possible, along with information on the study protocol,
different chemical sectors and of the consensus results. In addition, an study parameters and results (for the chemicals from the plant protec-
expert review of experimentally classified category 1 and category 2 tion product sector, 24% of compounds were retrieved from the Pesti-
results was performed to understand how such a review would support cide Properties Database (Lewis et al., 2016)). In some cases, a GHS
the overall workflow. category was derived and in other cases an LD50 value or range was
 identified. The chemicals were then loaded into the (Q)SAR software
 and prediction results were generated. The software calculated one of
 the following 8 values for each test chemical: Category 1, Category 2,
 2 The term “ (Q)SAR” is as an acronym for computational models that predict Category 3, Category 4, Category 5, Not Classified (NC), Out-of-Domain,
a biological response (such as acute toxicity) based on the chemical structure of or Indeterminate. The software may generate an out-of-domain result
the test molecule. It refers to both quantitative and non-quantitative structure- where a chemical is sufficiently different from the training set examples
activity relationships by placing the “ Q” in brackets. to make a reliable prediction or where the model’ s features do not

J. Bercu et al.
 Fig. 1. Illustration of how a prediction, based on two methodologies, are computed.
 Fig. 2. Analogs of the test chemical with known GHS categories derived from in vivo data.
overlap with features in the test chemical. The software may also rules:
generate an indeterminate prediction where there is conflicting infor-
mation, such as where the influence of substituents around a chemical •When an in vivo LD50 range was provided that spans multiple GHS
class is not fully understood. Any chemical where it was determined to categories (except for > 2000 mg/kg since the 5000 mg/kg dose is
be part of the training set was removed. This information was then often only used when it can be justified)
transferred to Excel spreadsheets along with relevant supporting infor- •In cases where it was possible to identify whether a chemical was
mation on the studies. To avoid sharing any potentially confidential present in the underlying model’ s database from the software output
information on the individual chemicals, all information that could
provide any chemical identification was removed. However, a reference In some cases, the individual collaborators provided both LD50 and
identifier was requested for each chemical in case questions needed to be GHS category results, in others only LD50 values or ranges were pro-
resolved later. vided. The following rules were adopted to consistently process the data:
2.2. Curating and combining the results •When only LD50 values were provided, a GHS category correspond-
 ing to the LD50 value or range was computed
 Each collaborator shared their in vivo results and predictions, as •When both an LD50 and GHS category were provided then the GHS
shown in Fig. 3. Initially, the individual results were analyzed to remove category was used when justified by the collaborator
entries that could not be used in this exercise, based on the following

J. Bercu et al.
 Fig. 3. Combining the results from multiple companies.
 Fig. 4. Number of chemicals for each experimental in vivo GHS category.
 •When an experimental value of > 2000 mg/kg was used, a “ Category (2) The proportion of compounds correctly predicted or one category
 5 or Not Classified” entry was used more conservative (for example, if the in vivo GHS category was 3,
 then a prediction of GHS 2 or 3 would be a match)
2.3. Generating summary statistics
 Two additional summary statistics were computed to assess the ac-
 The results were consolidated (as shown in Fig. 3), and a series of curacy of the models.
summary statistics were generated for the entire dataset as well as
subsets including collections from the pharmaceutical industry, plant (3) The proportion of compounds correctly predicted (for example, if
protection product industry and other chemical industries. These sum- the in vivo GHS category was 3, then only a prediction of GHS 3
mary statistics use an assessment of whether the experimental in vivo would be a match)
GHS category exactly matched the predicted GHS category. In cases (4) The proportion of compounds correctly predicted or one category
where the experimental category was assigned to the category “ Category higher/lower (for example, if the in vivo GHS category was 3, then
5 or Not Classified” , a correct match was recorded if the prediction was a prediction of GHS 2, 3 or 4 would be a match)
Category 5 or Not Classified.
 A series of summary statistics were calculated to support an assess- For each of these statistics, an overall assessment (i.e., the proportion
ment of whether the (Q)SAR test is fit-for-purpose for classification and across all test compounds) as well as a balanced assessment (based on
labeling, that is it predicts either the correct or a more potent category. the average proportion for each experimental in vivo GHS category) was
This analysis was performed on both the entire data set as well as subsets calculated. Whilst the values derived from the overall assessment are
of the data as explained below. more intuitive, the fact that the dataset was skewed towards a higher
 proportion of low toxicity chemicals (see below) makes the latter values
 (1) The proportion of compounds correctly or more conservatively more appropriate to consider.
 classified (for example, if the in vivo GHS category was 3, then a In addition, a baseline was computed using a random model (i.e., a
 prediction of GHS 1, 2 or 3 would be a match) random uniformly distributed assignment to category 1 through 5 and

J. Bercu et al.
not classified) and the same balanced summary statistics generated. This experimental in vivo GHS category is shown in Table 3. Two summary
was used for comparison purposes. statistics that help to understand whether the model is fit-for-purpose for
 classification and labelling are presented: (1) the percentage of correctly
2.4. Expert review predicted chemicals or chemicals predicted to be in a more conservative
 GHS category and (2) the percentage of correctly predicted chemicals or
 An additional manual assessment of experimentally determined chemicals predicted in an adjacent more conservative category. Two
category 1 or 2 chemicals that were predicted by the (Q)SAR models to additional summary statistics were calculated to help understand the
be in a less potent category was performed. This assessment used both accuracy of the model: (1) the percentage of correctly predicted chem-
information generated by the software (e.g., analogs, feature weight- icals and (2) the percentage of correctly predicted chemicals or chem-
ings) and any other information that would have been generated, icals predicted in an adjacent category. The inconclusive results were
including any in vitro assay results indicating a chemical’ s mechanism/ not used in calculating the summary statistics.
mode of action (MoA). The analysis was then revised based on any The data collected reflects the typical distribution of GHS categories
modified results from this expert review. within corporate collections and as such it is highly imbalanced and
 weighted towards the less toxic compounds. Therefore, an overall
 balanced assessment of the 4 summary statistics was calculated along-
3. Results side a baseline (represented by a random model). The balanced sum-
 mary statistics were computed by averaging the values for each
 Results were provided from 3M, Abbvie, Bristol Myers Squibb (BMS), category, shown in Table 3, apart from the "Cat 5. or NC values", with the
DSM, Genentech, Gilead Sciences, GlaxoSmithKline (GSK), Johnson and averages reported in Table 4. This information was not used in this
Johnson (J& J), Syngenta and Vertex. Information on 2568 chemicals assessment since this category spans two experimental categories.
was provided and, after processing the results, 2290 chemicals were The supplemental material contains analogous information to
used in the analysis. Given that the identities of the chemicals were not Tables 2– 4 for the assessment of statistical-based and the expert rule-
shared, it is not possible to determine whether any of the chemicals based methodologies (supplemental tables S1-S6) as well as the three
provided were duplicates; however, since these chemicals represent industrial sectors analyzed: pharmaceutical, plant protection products
proprietary lead compounds, candidate active ingredients, in- and other chemicals (Supplemental tables S8-S18). As previously dis-
termediates, etc. from different companies, as well as additional mar- cussed for analysis of the consensus model on the combined dataset, due
keted plant protection products and metabolites from a single database to the skewed nature of the datasets towards low toxicity chemicals, the
(Lewis et al., 2016), we can reasonably assume there is limited overlap balanced statistics presented therein provide valuable insight into the
because of the diverse proprietary chemical space being assessed. Any predictive performance of the different types of models on different
chemical where it was determined to be part of the training set was kinds of chemicals. Table S7 summarizes the results for different (Q)SAR
removed. Fig. 4 visually shows the number of chemicals in each of the methodologies, statistical-based and expert rule-based, along with the
experimental in vivo GHS categories. As previously noted, a category consensus from the two methodologies. The same summary statistics
“ Cat. 5 or NC” was created for chemicals where the experimental LD50 were calculated over all the data (i.e., these values are not balanced).
result was specified as > 2000 mg/kg. Table S19 summarizes the performance of the consensus model across
 A summary of how the Leadscope consensus model predicted the the different sectors: pharmaceutical sector, plant protection products
experimental in vivo GHS categories is shown in Table 2. The seven sector and other chemical sectors.
experimental categories used in this analysis are listed vertically along Supplemental tables S20, S21, and S22 show a series of experimental
with the six predicted categories (cat. 1– 5 and NC), shown horizontally. in vivo category 1 or 2 chemicals from the pharmaceutical industry, plant
Counts of the number of chemicals are shown in the table. To illustrate, protection product industry and broader chemical industry that are
there were 8 chemicals that had experimental in vivo values placing predicted as a less conservative category. For example, a chemical
them in category 1. Five of these 8 were predicted by the consensus whose experimental in vivo result is GHS category 1 yet the prediction is
model as category 1, 2 were predicted as category 2 and the remaining 1 either category 2, 3, 4, 5 or NC. An assessment of other information that
was predicted as category 5. The total value of 2181 results is less than would be available for these chemicals is also provided, including other
the 2290 chemicals analyzed since 109 predictions were inconclusive test results, information on chemical analogs as well as other informa-
(approximately 5% were either out-of-domain or indeterminate pre- tion from within the deployed models. Based on this information a
dictions). From this table, it can be seen that 95% of chemicals were determination was made as to whether the chemical would have been
either correctly predicted or were assigned to a more conservative correctly categorized based on an expert review of the totality of the
category. However, the skewed nature of this dataset, i.e. the higher information available. Using this information, Tables 5 and 6 illustrate
percentage of low toxic compounds, means that a balanced assessment how a combination of using the (Q)SAR models in addition to an expert
was also required (see below). review would modify the prediction results for experimental in vivo GHS
 An assessment of the performance of the consensus model for each
Table 2
Table showing counts of how the consensus model predicts for the different GHS categories.
 Predictedc
 Experimental Cat. 1 Cat. 2 Cat. 3 Cat. 4 Cat. 5 NC Total
 Cat. 1 5a 2 0 0 1 0 8
 Cat. 2 5 18a 5 2 2 1 33
 Cat. 3 1 29 52a 40 2 2 126
 Cat. 4 4 43 115 260a 38 8 468
 Cat. 5 1 15 54 106 59a 12 247
 Cat. 5 or NCb 3 48 164 343 128a 23a 709
 NC 9 32 119 227 116 87a 590
 Total 28 187 509 978 346 133 2181
 a Indicates where a correct prediction is made.
 b Where chemicals were identified as > 2000 mg/kg they were place in category “ Cat. 5 or NC” and not in Cat.5 or NC.
 c Not including inconclusive predictions.

J. Bercu et al.
Table 3
Breakdown of the results across different categories.
 Fit-for-purposeb Accuracyc

Source PDF: /instem.com/www.instem.com/docs/sot-2021/New-ICH-M7-BM-Model.pdf

 Regulatory Toxicology and Pharmacology 109 (2019) 104488
 Contents lists available atScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepage: www.elsevier.com/locate/yrtph
Transitioning to composite bacterial mutagenicity models in ICH M7 (Q)SART
analyses
Curran Landrya, Marlene T. Kima, Naomi L. Kruhlaka, Kevin P. Crossb, Roustem Saiakhovc,
Suman Chakravartic, Lidiya Stavitskayaa,∗
aUS Food and Drug Administration, Center for Drug Evaluation and Research, 10903 New Hampshire Ave, Silver Spring, MD, 20993, USA
bLeadscope Inc., 1393 Dublin Road, Columbus, OH, 43215, USA
cMulticase Inc., 23811 Chagrin Boulevard, Suite 305, Beachwood, OH, 44122, USA
ARTICLE INFO ABSTRACT
Keywords: The International Council on Harmonisation (ICH) M7(R1) guideline describes the use of complementary
Bacterial mutagenicity (quantitative) structure-activity relationship ((Q)SAR) models to assess the mutagenic potential of drug im-
Computational toxicology purities in new and generic drugs. Historically, the CASE Ultra and Leadscope software platforms used two
Genotoxicity different statistical-based models to predict mutations at G-C (guanine-cytosine) and A-T (adenine-thymine)
In vitro sites, to comprehensively assess bacterial mutagenesis. In the present study, composite bacterial mutagenicity
Regulatory review models covering multiple mutation types were developed. These new models contain more than double the
QSAR number of chemicals (n = 9,254 and n = 13,514) than the corresponding non-composite models and show
Structure-activity relationship
ICH M7 better toxicophore coverage. Additionally, the use of a single composite bacterial mutagenicity model simplifies
Ames impurity analysis in an ICH M7 (Q)SAR workflow by reducing the number of model outputs requiring review. An
Drug external validation set of 388 drug impurities representing proprietary pharmaceutical chemical space showed
 performance statistics ranging from of 66–82% in sensitivity, 91–95% in negative predictivity and 96% in
 coverage. This effort represents a major enhancement to these (Q)SAR models and their use under ICH M7(R1),
 leading to improved patient safety through greater predictive accuracy, applicability, and efficiency when as-
 sessing the bacterial mutagenic potential of drug impurities.
1. Introduction where the most commonly used combination of tests comprises the
 bacterial reverse mutation assay, the mouse lymphoma assay, thein
 The bacterial reverse mutation assay is designed to detect andvitrochromosomal aberration assay, and thein vivomicronucleus assay
classify mutagens. Specifically, the test uses several auxotrophic strains(Gatehouse, 2012; Stavitskaya et al., 2015). The test battery is intended
ofSalmonella entericaserovar Typhimurium andEscherichia colito de- to identify genotoxic substances that exhibit a greater likelihood of
tect point and frame-shift mutations, which include substitution, ad-subsequently causing carcinogenicity in humans.
dition, or deletion of one or more DNA base pairs (Ames et al., 1973A pivotal study conducted by;Ashby and Tennant (1988)showed
Green et al., 1976;Maron and Ames, 1983). The principle of the bac-that although not all carcinogens are genotoxic, many genotoxic che-
terial reverse mutation assay is to detect mutagens through the rever-micals are carcinogenic in rodents. This was later confirmed by
sion of auxotrophic bacteria to wild type in the presence of the testKirkland et al. (2005), who examined the correlation between carci-
substance. This assay can be conducted inSalmonella entericaTyphi-nogenicity and genotoxicity in at least one of the three assays
murium strains TA98, TA100, TA1535, TA1537 (or TA97, or TA97a),(Ames + mouse lymphoma assay,in vitromicronucleus assay, andin
and TA102 (orE. coliWP2uvrAwith or without pKM101) (ICH, 2011). vitrochromosomal aberration assay). The authors found that 93% of the
The bacterial reverse mutation assay is one of the most widely usedexamined carcinogens had positive results in one or more genotoxicity
components of the International Council on Harmonisation (ICH) S2assays. Furthermore, the results showed that the Ames test had the best
genotoxicity test battery to assess the safety of pharmaceuticals prior tospecificity, at 74%, for predicting the outcome of the rodent carcino-
clinical exposure (ICH, 2011). The battery includes multiple assays togenicity 2-year bioassay when compared to the other genotoxicity as-
detect mutagenic, clastogenic and aneugenic effectsin vitroandin vivo, says, making it the most promising early-screening assay. Early

Source PDF: /instem.com/www.instem.com/docs/sot-2021/Implementation-of-in-silico-toxicology-protocols.pdf

 Contents lists available at ScienceDirect
 Computational Toxicology
 journal homepag e: www.sci encedirect.co m/journal/ computatio nal-toxicol ogy
Implementation of in silico toxicology protocols within a visual and
interactive hazard assessment platform
Glenn J. Myatta,\*, Arianna Bassan b, Dave Bower a, Candice Johnson a, Scott Miller a,
Manuela Pavan b, Kevin P. Cross a
aInstem, 1393 Dublin Road, Columbus, OH 43215, USA
bInnovatune, Via Giulio Zanon 130/D, 35129 Padova, Italy
ARTICLE INFO ABSTRACT
Edited by Dr. Mark Cronin. Mechanistically-driven alternative approaches to hazard assessment invariably require a battery of tests,
 including both in silico models and experimental data. The decision-making process, from selection of the
Keywords: methods to combining the information based on the weight-of-evidence, is ideally described in published
In silico toxicology guidelines or protocols. This ensures that the application of such approaches is defendable to reviewers within
Visual framework regulatory agencies and across the industry. Examples include the ICH M7 pharmaceutical impurities guideline
ICH M7 and the published in silico toxicology protocols. To support an efficient, transparent, consistent and fully docu-
Pharmaceutical impurities mented implementation of these protocols, a new and novel interactive software solution is described to perform
Genetic toxicology such an integrated hazard assessment based on public and proprietary information.
Skin sensitization
Introduction The ICH M7 guideline describes how both experimental data
 alongside computational toxicology results are used to assess the po-
 In silico toxicology (or computational toxicology) is being used tential for DNA-reactive mutagenicity, as shown in Fig. 1. The guideline
directly or as part of the weight-of-evidence (WoE) for an increasing uses this information to assign an impurity to one of five classes (shown
number of regulatory and industrial applications. This is driven by the in Table 1), which in turn supports whether an impurity needs to be
need to (1) fill data gaps for chemicals in commerce with limited in- controlled further or if additional testing is required. To support the
formation, (2) improve the efficiency of the discovery process for assessment of classes 1, 2 and 5, it is important to identify any bacterial
chemical products, (3) support the replacement, reduction, and refine- mutagenicity and carcinogenicity data available for any of the impu-
ment of animal use (3Rs), and (4) support regulatory guidelines where in rities. The guideline also identifies chemical classes representing high
silico approaches are defined as acceptable approaches [1]. One such potency mutagenic carcinogens (termed “cohorts of concern”) which
regulatory guideline is the International Committee for Harmonization need to be handled separately as part of any risk assessment. These
(ICH) M7 guideline “Assessment and Control of DNA Reactive (Muta- cohorts of concern include aflatoxin-like-, N-nitroso-, and alkyl-azoxy
genic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic compounds. In the absence of any adequate experimental data, a
Risk” [2]. This guideline includes a computational toxicology option as a computational assessment based on two complementary methodologies
regulatory accepted test to predict the bacterial reverse mutation assay is recommended. One methodology should be an expert rule-based
(often referred to as the Ames test) [3]. This fast computational test is technology and the second should be a statistical-based technology.
included for several reasons. Firstly, for many of these impurities there An expert review of all the information is prudent to assess the relevance
may be insufficient amounts of the test material available for performing and reliability of the both the experimental data as well as the compu-
an actual Ames test. This may require synthesizing the chemical tational results [4,5,7,8]. In addition, an expert review can support the
(including actual or potentially present impurities) which would sub- class assignment for inconclusive computational results and even refute
stantially add to the time and cost of performing such an assessment. In the results given sufficient evidence, such as proprietary results for
addition, such models have been shown to be sufficiently accurate, chemicals analogs. The principles and procedures for performing and
especially when coupled with an expert review, and they support the documenting this process have been published by a working group
desired high-throughput assessment of the impurities [4–6]. including both regulators and industry [5].
 \*Corresponding author.
https://doi.org/10.1016/j.comtox.2021.100201
Received 16 September 2021; Received in revised form 23 October 2021; Accepted 26 October 2021

G.J. Myatt et al.
Fig. 1. Combining information on experimental data and computational toxi-
cology results to support the ICH M7 class assignment. Bacterial mutagenicity
and carcinogenicity data available for the target impurity are identified and
combined with predictions. Statistical- and expert rule-based methods are
applied for a computational toxicology assessment of mutagenicity. Predictions
can identify high potency mutagenic carcinogens (cohorts of concern).
 Fig. 2. A hazard assessment framework for in silico toxicology protocols.
Table 1 effects/mechanisms are used to support the assessment of one or more
ICH M7 Hazard Classification. toxicological endpoints; for example, this construct can be applied to
 Class Definition assess the activation of the Nrf2-ARE pathway (the mechanism) within
 1 Known mutagenic carcinogen the prediction of skin sensitization in human (the endpoint). Guidelines
 2 Known mutagen with unknown carcinogenic potential for an expert review of the experimental and in silico results along with
 3 Alerting structure, unrelated to the structure of the drug substance; no how the information may be combined are described within the pro-
 mutagenicity data
 4 Alerting structure, same alert in related compounds which have been tested tocols. The procedure for documenting the entire decision-making pro-
 and are non-mutagenic cess, along with any expert review, is also described in the protocols.
 5 No structural alerts, or alerting structure with sufficient data to demonstrate Hence, these protocols support the adoption of in silico approaches
 lack of mutagenicity or carcinogenicity within a well-defined hazard assessment framework by ensuring such
 methods are performed in a consistent, transparent, and reproducible
 The ICH M7 guideline is a widely adopted example of an approach to quality-driven manner.
hazard assessment based on the integration of a battery of both exper- Due to the complexity of these novel assessments described in such
imental in vitro and in vivo data alongside in silico results coupled with an protocols, an interactive and visual software application for performing
expert review. This type of integrated assessment is becoming increas- a hazard assessment is essential. This type of solution should support
ingly common in approaches that support a more mechanistically-driven both the integration of the relevant experimental data and in silico
and animal-free assessment. Initiatives such as the Adverse Outcome predictions as well as the assessment of the reliability of the combined
Pathways (AOPs), Integrated Approaches to Testing and Assessment information. It should also steer the integration of all the available in-
(IATA), New Approach Methodologies (NAMs), and Defined Approaches formation based on the rules and principles described in the protocols.
(DAs) are advancing and documenting the state of the science to enable The tool should also provide the ability to perform an expert review of
these future alternative and integrated approaches [9– 13]. the experimental data and/or in silico results at the same time as
 Experimental data generated using accepted protocols, such as the allowing any reviewer to assess the overall process of combining the
OECD test guidelines [14], supports the use of this data across different information. All expert review and any resulting changes should be
regulatory authorities and industry. The development of equivalent documented along with the entire decision-making process.
protocols for the use of in silico methods would similarly support The following paper outlines a proposal for an interactive and visual
adoption of these methods, whether as a standalone alternative method solution to this problem and discusses its implementation within the
or in combination with experimental results. The in silico protocols Leadscope computational toxicology solution. This includes the devel-
would build on work documenting best practices in computational opment of a visual and interactive hazard assessment platform in rela-
toxicology, such as the OECD validation principles [15], and the tion to the ICH M7 framework [4,5], the genetic toxicology in silico
described approaches to defining the battery of mechanisms and asso- protocol [16], and the skin sensitization in silico protocol [17]. The
ciated tests to support an integrated assessment. paper covers how the content, including databases containing historical
 A working group of over 70 organizations is currently generating toxicity information and computation models, are developed. It explains
such in silico toxicology protocols. This includes a framework outlining how the results from such database searches and in silico model appli-
the components for any protocol [1] along with protocols for specific cations are integrated within a visual platform and how such a platform
toxicology endpoints. To date, protocols for genetic toxicology [16] and may be interrogated, and expert review performed and documented.
skin sensitization [17] have been published with many protocols and The paper also presents information on the validation of the models and
position papers currently progressing. These protocols outline a series of includes four case studies illustrating applications of such a platform.
defined toxicological effects or mechanisms that ideally should be
assessed based upon available experimental data and/or in silico results. Methods
The protocols discuss the selection of such approaches, how to assess the
reliability of the information provided, and how to combine the avail- Overview
able information to establish an overall hazard assessment and associ-
ated level of confidence based on the WoE. The rules and principles The implementation of an integrated hazard assessment platform
underpinning this WoE process are provided within the protocol. This is supporting the application of in silico toxicology protocols [1] is sum-
illustrated conceptually in Fig. 2, showing how a series of toxicological marized in Fig. 3. The visual hazard assessment platform ideally queries

G.J. Myatt et al.
Fig. 3. Overview of the implementation of the visual hazard assessment platform. NTP refers to the National Toxicology Program online databases and CPDB refers
to the Carcinogenicity Potency Database, QSAR refers to Quantitative Structure-Activity Relationships.
both the toxicity databases as well as applies in silico models to support Studies can vary significantly in the level of detail provided in
the assessment of individual effects or mechanisms. Indeed, the platform describing the methodology used in identifying, verifying, and repre-
uses both experimental and in silico results for each effect/mechanism senting the chemical substances of primary interest being reported on. In
defined in the protocol or guideline. To support access to the experi- the best-case scenarios, an author will report three types of identifica-
mental results, the platform searches a database of historical toxico- tion for substances: typed identification numbers, tradenames or sys-
logical studies linked to chemical structures. Public sources of tematic names, and a structural representation. In the worst-case
toxicology data are used to populate this database. The database also scenarios, an author may only provide a synonym or codename for a
supports the generation of in silico models based on different method- substance, which, in some cases makes it impossible to determine any
ologies. In addition, these models are refined and annotated through chemical structure representation. In each case encountered the infor-
access to the literature and other online databases which enrich the mation regarding the substance identification is vetted and cross-
models with mechanistic interpretations. Once experimental and/or in compared to ensure agreement. If a conflict arises in the cross-
silico results for individual effects/mechanisms have been identified and comparison efforts, the context of the study is taken into consideration
reviewed, endpoints are then calculated based on the input along with to provide guidance in correctly identifying substances. For example, an
the rules and principles documented in the protocol. The visual platform examination of the totality of the information supports any resolution
interactively supports interrogation of the results and performing an where different or incomplete stereochemistry is provided.
expert review. The following sections outline how such toxicity data- As part of the content building, information on both the study design
bases and in silico models are developed within Leadscope computa- and results needs to be included in the database to support transparency
tional toxicology solutions, how these resources are integrated within and expert review. The underlying information is not always in an
the platform, and how this platform can be interrogated. How such a electronic form that is suitable for processing automatically. In certain
platform has been developed to support the ICH M7 guideline as well as cases, it is necessary to enter the information by hand into an electronic
the two published protocols [16,17] is specifically discussed. representation. Where it is in an electronic form, it is possible to develop
 customized applications to read the content directly into the database.
Toxicity database An essential process, irrespective of whether the step is performed
 manually or automatically, is to map the data elements described in the
 As illustrated in Fig. 3, there are many public sources of toxicity source material onto standardized terms. The Toxicity Markup Language
study information. These include online databases (such as the National or ToxML is a standardized organization of toxicity study design and
Toxicology Program [NTP] [18]), secondary sources of compiled in- results supported by controlled vocabularies that ensures the creation of
formation (such as the Carcinogenicity Potency Database [CPDB] [19]) a harmonized database [20].
as well as individual study records contained in publications or regula- A process of grading (i.e., creating an overall call for a specific
tory submissions. Information on both the tested chemicals and the toxicological effect or mechanism) is possible once the chemical struc-
toxicity study design and results are converted into an electronic data- ture registration process and the content processing is completed and the
base with information integrated for each compound. The following harmonized study records are linked to these chemicals. As an example,
reviews the process of creating this content. an overall assessment for bacterial mutagenicity would include an ex-
 It is important that all studies for the same chemical are linked to the amination of the test and study calls for all entries matching each
same electronic depiction of the chemical structure. This is achieved by registered chemical. This process uses a series of rules to assess the
comparing each chemical (test article) to the existing database. Based on different study results, such as whether an individual study source is
this comparison, the test article is either registered as a new chemical trusted or authoritative and if the study is compliant with accepted test
and given a new Leadscope ID or it is linked to a previously registered guidelines. In cases where the results are conflicting across the different
chemical. It can be challenging when only a chemical name has been studies, the WoE needs to be considered to derive an overall assessment.
reported, especially when the chemical has been referred to by different Table 2 summarizes the Leadscope database content used in the
names. When a chemical structure is displayed within the source ma- current platform and how such content maps onto the effects/mecha-
terial, the depiction of its stereochemistry as well as aromaticity and nisms within the three implemented frameworks (ICH M7, the genetic
tautomerism are considered as part of this matching. To support the toxicology in silico protocol, and the skin sensitization in silico protocol).
computational modelling, mixtures and salt forms are often linked to the
modellable forms of the chemical, referred to as the SAR form.

G.J. Myatt et al.
Table 2 database previously described. The models are based on a number of
Databases used to support the hazard assessment platform calculated descriptors: (1) pre-defined structural features [22], (2)
 Database Sources Mapped to effects/ Hazard calculated physico-chemical properties, (3) chemical scaffolds auto-
 mechanisms assessment matically identified to map onto a disproportionate numbers positive or
 framework negative examples [23], and (4) significant active structural features
 Carcinogenicity CCRIS, CDER, CFSAN- Carcinogenicity ICH M7 extracted from the literature. Having selected an appropriate subset of
 PAFA, CPDB, DSSTox these descriptors, a computational model based on the partial logistic
 DBCAN, IARC, regression algorithm [24] is applied to encode the relationship between
 ISSCAN, NTP, RTECS the descriptors and the toxicological response. The models are further
 Genetic CCRIS, CDER, CFSAN- Bacterial mutation ICH M7,
 toxicology OFAS, CFSAN- PAFA, genetic refined and then validated based on cross-validation and external vali-
 CPDB, EPA-Genetox, toxicology dation wherever possible. The models generate a probability of a posi-
 Submissions from Mouse Lymphoma Genetic tive outcome, and a final prediction is made using defined cut-off values.
 organizations, toxicology For example, when the bacterial mutation model calculates a probability
 Japanese NIHS, NTP, Chromosome Genetic greater than 0.6 a positive assignment is made, a probability less than
 Tokyo Eiken, aberration in vitro toxicology
 Publications, RTECS Micronucleus in Genetic 0.4 is assigned to be negative, and those predictions between 0.4 and 0.6
 vitro toxicology are assigned as indeterminates. The implementation of the models per-
 Chromosome Genetic forms an additional key step to assess whether the test chemical is within
 aberration in vivo toxicology the applicability domain of the model, i.e., whether there is an increased
 Micronucleus in Genetic
 vivo toxicology reliability because of the overlap with similar training set examples as
 Skin Publications, Protein reactivity Skin well as features used in the model.
 sensitization ICCVAM, NICEATM, sensitization There are three types of QSAR models used within this platform: (1)
 ECHA Activation of Nrf2- Skin “ single statistical” models (using the methodology discussed as in the
 ARE sensitization case of the bacterial mutation QSAR model), (2) “ balanced statistical”
 Expression of co- Skin
 stimulatory and sensitization models (used in cases where the toxicity response is skewed as in the
 adhesion molecules case of the in vivo micronucleus QSAR model), (3) “ categorical statisti-
 Reaction domain Skin cal” models (used when the response is ordinal as in the case of models
 sensitization related to skin sensitization).
 Rodent local lymph Skin The balanced statistical approach uses a series of models that are
 node proliferation sensitization
 Rodent Skin based on subsets from the training set, where each set is over or under-
 maximization sensitization sampled to create more even distribution of positive and negative ex-
 Human skin Skin amples. Training set examples from the underrepresented positive or
 sensitization sensitization negative class will be present in more than one subset. When making a
Legend: CCRIS - Chemical Carcinogenesis Research Information System; CDER - prediction, the test chemical will be run through all models and an
US FDA CDER (Center for Drug Evaluation and Research) product approval re- overall prediction calculated based on the combined results.
views; CFSAN- PAFA - US FDA CFSAN (Center for Food Safety and Applied When the toxicological response outcome is a categorical value
Nutrition) PAFA (Priority-based Assessment of Food Additives) database; based on either the severity of the outcome or the toxic dose, a series of
CFSAN-OFAS - Genetic toxicity database from the US FDA CFSAN (Center for models are built and incorporated within a decision tree. For example,
Food Safety and Applied Nutrition) reviews; CPDB - Carcinogenicity Potency the toxicological outcome for a Local Lymph Node Assay (LLNA) model
DataBase; DSSTox DBCAN - Distributed Structure-Searchable Toxicity (DSStox) is strong/extreme (where the effect concentration (EC31) is less than 1,
Public Database Network: DBCAN: EPA Water Disinfection By-Products with
Carcinogenicity Estimates; EPA-Genetox - Mutagenicity test data from the US moderate (1 ≤ EC3 < 10), weak (10 ≤ EC3 ≤ 100), or non-sensitizer
EPA; IARC - International Agency for Research on Cancer and Carcinogenicity (EC3 > 100). A series of individual models are built based on these
classification; ISSCAN - Chemical carcinogens: structures and experimental data cut-off values. In this case, three binary models are built to predict each
from Instituto Superiore di Sanita; Japanese NIHS - National Institute of Health of these categories; for example, an individual model predicts whether a
Sciences of Japan (Publicly release class A chemicals); NTP – National Toxi- chemical has an EC less than 1 (strong/extreme sensitizers), and two
cology Program; RTECS - Registry of Toxic Effects of Chemical Substances; other models predict the moderate and weak categories. The results
ICCVAM - Interagency Coordinating Committee on the Validation of Alternative from each of the models are then combined within a decision tree (as
Methods; NICEATM - The NTP Interagency Center for the Evaluation of Alter- illustrated in Fig. 4) to calculate the final category.
native Toxicological Methods; ECHA - European Chemicals Agency Besides QSAR, a second methodology referred to as expert rule-based
In silico models is developed in the Leadscope computational toxicology solution [21].
 This is based on a series of structural alerts that encode features that
 Both the ICH M7 guideline and the in silico toxicology protocols activate and deactivate the toxicity. Such alerts are derived from expert
recommend using multiple computational methodologies, including knowledge embedded in the literature and/or extracted from toxicity
statistical-based and expert rule-based, since multiple concurring com- databases. Structural alerts are ideally accompanied by a monograph
plementary methods increase the reliability of the prediction results [1]. describing the relevance of the moiety in the context of the endpoint of
Methodologies to profile chemicals into different toxicologically rele- interest, such as any mechanistic rationale, as well as all examples from
vant categories and read-across approaches also provide key informa- the database to support a contextual assessment. The identification of
tion in any hazard assessment framework. The following section outlines the series of expert alerts encoded within the Leadscope computational
the computational models used within the hazard assessment platform. toxicology solution [21] can be assisted by specific informatics capa-
 Statistical-based or Quantitative Structure-Activity Relationship bilities, such as clustering [25] and identification of significant chemical
(QSAR) models are developed within the Leadscope predictive data scaffolds [23].
miner [21]. A number of these models predict a binary outcome, for The application of expert alerts to any test chemical will result in a
example, the bacterial mutation statistical QSAR model predicts prediction (such as positive, negative, or indeterminate) alongside a
whether a chemical is mutagenic or non-mutagenic based on predictions
made using Ames test data. These models use a training set of chemicals 1 EC3 value: the amount of a chemical that is required to elicit a three-fold
and toxicological data (response) extracted from the toxicological increase in lymph node cell proliferative activity (SI ≥ 3)

G.J. Myatt et al.
Fig. 4. Decision tree for calculation of LLNA severity as implemented in the Leadscope computation toxicology solution Leadscope Model Applier. The platform
provides a means to explore such decision tree that combines the underlying “ categorical statistical” models.
confidence score based on the toxicological response value’ s precision Table 3 summarizes the different computational models that are
derived from the matching examples. Since a prediction is being made, it incorporated within the Leadscope hazard assessment platform,
is essential to understand the applicability domain through a compari- including which effects or mechanisms they map on to.
son of the test chemical to the underlying reference set of compounds
supporting the alerts. Visual hazard assessment platform
 In silico profilers also make use of mechanism-based structural alerts
[26]; however, they do not directly predict a toxicological outcome but An interactive graphical user interface of the hazard assessment
place a chemical into a category to support either an assessment or an framework has been developed by Leadscope to support the ICH M7
expert review. Several profilers have been incorporated within the guideline, the genetic toxicology in silico protocol and the skin sensiti-
Leadscope platform, including carcinogenicity cohorts of concern [2] zation in silico protocol. The platform includes nodes representing the
and reaction domains [27] to support the assessment of skin defined effects or mechanisms, displayed as gray nodes within Fig. 5.
sensitization. The relationship of these effects/mechanisms to one or more toxico-
 Finally, read-across is used to predict a toxicological outcome for a logical endpoints is also displayed. These endpoints are shown as nodes
given chemical (target) based on the toxicological response from one or and are colored blue, as illustrated in Fig. 5.
more sufficiently similar analogs (source). A read-across tool has been Wherever possible, each effect/mechanism is linked to results from
incorporated within the Leadscope computational toxicology solution to applying computational models and database searches and each node
provide the opportunity to include such a prediction for the different summarizes these results. This information is accompanied by a score
effects or mechanisms. The tool identifies similar chemicals based on a reflecting the reliability of the information, using the Reliability Score
series of different approaches, including structural similarity or a pre- (RS) value described in Myatt et al., 2018 (summarized in Table 4); a
defined chemical category. The tool supports the interactive explora- confidence score tied to the endpoint assessment is also included [1]. For
tion and refinement of the source chemicals, including the addition of example, in Fig. 5, the bacterial mutation experimental data is shown as
proprietary examples, which can be documented in the tool. It also helps positive with a reliability score of RS5 and there are two in silico pre-
formulate how the toxicity data on the source chemicals is read-across diction results, one positive statistical-based result and one equivocal
onto the target. Frameworks, such as the Read-Across Assessment expert rule-based result (both assigned a reliability score of RS5). This
Framework or RAAF [28], are incorporated within the platform to information is automatically combined into a positive assessment for the
support the complete expert review and documentation of the read- bacterial mutation effect/mechanism, with a reliability score of RS5.
across study. This assessment together with other information is then used to derive
Table 3
Summary of models incorporated into the current hazard assessment platform.
 Hazard assessment framework Effect/mechanism Computational models Type of model References
 ICH M7 Carcinogenicity Cohort of concern v1 Profiler
 ICH M7, genetic toxicology Bacterial mutation Bacterial mutation v2 Statistical-based [33]
 ICH M7, genetic toxicology Bacterial mutation Bacterial mutation v7 Expert rule-based
 Genetic toxicology Mouse Lymphoma MLA Activated v2; MLA unactivated v2 Statistical-based
 Genetic toxicology Chromosome aberration in vitro CA CHL v2 Statistical-based
 Genetic toxicology Chromosome aberration in vivo In vivo CA v2 Statistical-based
 Genetic toxicology Micronucleus in vivo In vivo micronucleus mouse v2 Statistical-based [34]
 Skin sensitization Protein reactivity DPRA v2 Statistical-based
 Skin sensitization Activation of Nrf2-ARE KeratinoSens v2 Statistical-based
 Skin sensitization Expression of co-stimulatory and adhesion molecules h-CLAT v2 Statistical-based
 Skin sensitization Reaction domain Reaction domain v2 Profiler
 Skin sensitization Rodent local lymph node proliferation LLNA Statistical-based
 Skin sensitization Rodent local lymph node proliferation LLNA Expert rule-based
Legend: ICH M7 - International Committee for Harmonization (ICH) M7 guideline; CA – Chromosome aberration; CHL - Chinese Hamster Lung cells; DPRA - Direct
Peptide Reactivity Assay; h-CLAT – Human Cell Line Activation Test; LLNA - Local Lymph Node Assay

G.J. Myatt et al.
 Fig. 5. On demand details available for nodes representing the effects/mechanisms and the derived endpoints.
Table 4 an assessment for the Gene Mutation endpoint (positive) along with a
Reliability Scores confidence score (low). The rules/principles for combining these types
 of information and generating the final assessment together with cor-
 Reliability Klimish Description Summary responding confidence are detailed in the protocols. It should be noted
 Score Score that the overall reliability score is associated with a single effect/
 RS1 1 Data reliable Well documented and accepted mechanism based on available experimental and/or in silico data for that
 without study or data from the effect/mechanism. The confidence score is associated with a single
 restriction literature endpoint; however, it is based on the propagation of information from
 Performed according to valid
 and/or accepted test guidelines all related effects/mechanisms as well as other endpoints.
 (e.g., OECD) Each node in the Leadscope platform is interactive: by clicking on
 Preferably performed any node further information is shown, as illustrated in Fig. 5. For
 according to good laboratory example, by clicking on the box annotated with a “ 1′′, information on the
 practices (GLP) individual studies from the toxicology database search is displayed. This
 RS2 2 Data with Well documented and
 restriction sufficient includes a summary of the results and a link to the full study report. It is
 Primarily not performed possible to select which of the studies, based on a review of the study
 according to GLP adequacy, to include in the current assessment. A default overall
 Partially complies with test assessment and reliability score based on the studies identified is shown;
 guideline
 RS3 – Expert review Read-across however, it is possible to update both values following an expert review
 Expert review of in silico result of the data. It is also possible that a proprietary study has been run on the
 (s) and/or Klimisch 3 or 4 data chemical, and this study may be included in the expert review; it can
 RS4 – Multiple thus be integrated and documented in the assessment by summarizing
 concurring the results and uploading the full study report into the platform.
 prediction results
 RS5 – Single acceptable Further details on the predictions are also available by clicking on
 in silico result the box annotated with a “ 2” (Fig. 5). This includes an explanation of the
 RS5 3 Data not reliable Inferences between the model results and access to structurally similar analogs. The protocols
 measuring system and test provide guidelines for elements to consider as part of an expert review of
 substance
 Test system not relevant to the in silico results. These guidelines are also incorporated into the
 exposureMethod not platform, as shown in Fig. 6. An inspection of any of these guideline
 acceptable for the endpoint elements may: (1) increase in the prediction’ s reliability, (2) result in no
 RS5 4 Data no Not sufficiently documented increase in the prediction’ s reliability, (3) refute the prediction, or (4)
 assignable for an expert review provide no additional supporting information. Fig. 6 shows how, for
 Lack of experimental details
 Referenced from short abstract each of the elements of an expert review, it is possible to view contextual
 or secondary literature information to support and document such an examination.
 The platform allows for the integration of a read-across study in an
 assessment for any of the effects/mechanism. A node is linked to the
 read-across tool that will both perform and document the read-across

G.J. Myatt et al.
 Fig. 6. In silico expert review checklist and accompanying contextual information.
study. It is also possible to add the results from models not directly corresponding platform is shown in Fig. 8 and includes experimental
incorporated within the platform or a read-across assessment performed data and in silico prediction models/profilers for bacterial mutation and
outside the platform. The details of these external results can be added, carcinogenicity. This information is combined, based on a series of rules
including meta information, modelling approach or any other parame- (shown in Fig. 9), to generate an overall ICH M7 class designation (as
ters. The full in silico report can also be uploaded into the platform to shown in Table 1) along with supporting information on the reliability
provide full transparency. and confidence of the information. It should be noted, there is no in-
 The platform collects all the information tied to both experimental termediate assessment of individual effects/mechanisms as the ICH M7
and in silico results, and any expert review that modified the individual class designation is based on the outcome of the available two nodes, i.
assessments or reliability scores; an overall assessment for the effect/ e., bacterial mutation and carcinogenicity. Fig. 10 shows the complete
mechanisms is then automatically derived as shown in Fig. 5 (annotated genetic toxicology hazard assessment framework, and Fig. 11 the com-
with “ 3” ). Additional details on the rules and principles that were used plete skin sensitization hazard assessment framework as they are
to derive these values can be inspected and potentially modified based implemented within the platform.
on a documented expert review. Fig. 7 shows how the reliability score
for the bacterial mutation experimental data was changed from RS5 to Results
RS1, after an expert review concluded the results warrant the highest
reliability score. Tables 5 and 6 summarize both the database content and the in silico
 Fig. 5 also shows how this information associated with effects/ models’ performance that are used within the hazard assessment plat-
mechanisms is, in turn, used to make an assessment for one or more form. The platform currently comprises three finalized frameworks: the
derived endpoints alongside a confidence level. The rule/principles for ICH M7 hazard assessment framework, complete genetic toxicology
deriving this call as documented in the protocol are available for in- hazard assessment framework [16], and the complete skin sensitization
spection, i.e., clicking on the node annotated with “ 4” in Fig. 5. In a hazard assessment framework [17].
similar manner, it is also possible to revise the assessment based on a
documented expert review.
 For the specific case of a complete ICH M7 hazard assessment, the
 Fig. 7. Interactively modify the results.

G.J. Myatt et al.
 Fig. 8. Implementation of the ICH M7 hazard assessment framework within the platform.
 Fig. 9. Rules for deriving the ICH M7 classification.
Case studies of the individual models (i.e., expert rule- and statistical-based systems)
 are summarized in a table alongside the ICH M7 class assignment, cor-
Overview responding confidence and additional supportive evidence and com-
 ments (Fig. 12). More specifically, no experimental bacterial
 The only required information to initiate a hazard assessment is the mutagenicity data nor carcinogenicity data are available from the
electronic record of the chemical structure. This can be either uploaded Leadscope databases, and the two complementary (Q)SAR methodolo-
from a file, such as a MOL file or SMILES string, identified through a gies provide negative predictions for bacterial mutagenicity. This in-
database search or drawn within Leadscope’ s structure drawing package formation is automatically combined into a Class 5 assignment with a
or elsewhere. The following case studies illustrate how the platform default medium level confidence. Such confidence level is justified by
(implemented in the Leadscope model applier v3.1) described in this the absence of any expert review at this initial stage of the protocol
paper can be used to assess four chemicals. workflow.
 To increase the confidence level, an expert review can be conducted,
Case study 1: ICH M7 assessment of 2-bromo-5-acetamidobenzoic acid and it is guided by the workflow encoded in the ICH M7 protocol in the
 Leadscope Model Applier. Fig. 13 illustrates the different steps of this
 Upon running the ICH M7 protocol for the target impurity (2-bromo- workflow. The expert review of the statistical-based model confirms the
5-acetamidobenzoic acid), available experimental data and the outcome negative prediction of the target impurity based on the following

G.J. Myatt et al.
 Fig. 10. Implementation of the genetic toxicology in silico protocol’ s hazard assessment framework within the platform.
elements: a) a low positive prediction probability (PPP = 0.173); b) a such mutagenic risk is equal to 8.1% (6% by Dobo et al. [6]). The expert
good coverage of the structure of the target impurity; c) negative fea- review can thus conclude that the target impurity is predicted as not
tures providing higher contributions to the prediction leading to a clear mutagenic, i.e., negative for bacterial in vitro mutagenicity (Ames test),
negative call; d) no concern from the features associated with positive and the confidence of the prediction is increased to a high level. As such,
contributions to the prediction since these features are represented in the target impurity 2-bromo-5-acetamidobenzoic acid is assigned to the
experimentally negative compounds (e.g., Acedoben and Acetanilide); ICH M7 Class 5, and a standard report is generated including all the
e) good accuracy of analogs’ predictions that supports the reliability of considerations of the expert review that were duly mapped throughout
the prediction. The review of the expert rule-based system also confirms the ICH M7 protocol workflow.
the negative outcome, given the absence of structural alerts of potential
concern for mutagenicity, and supporting evidence coming from the Case study 2: ICH M7 assessment of 1-chloro-2-nitrobenzene
experimental data for the closer analogs in the alert reference set, which
are negative and correctly predicted as negative by the model. The ICH M7 hazard assessment performed for 1-chloro-2-nitroben-
 For this molecule, the expert review notes that the potential meta- zene identifies available experimental data in the Leadscope toxicity
bolism of the target impurity should be also considered since the database: positive bacterial mutagenicity and carcinogenicity data.
chemical contains an aromatic amide functional group, that may be These data are organized by the tool in the summary table illustrated in
bioactivated to a primary aromatic amine [29]. Certain primary aro- Fig. 14. The target impurity is preliminarily assigned to the ICH M7 Class
matic amines are mutagenic, and the position of ring substituents in- 1 by the standardized workflow, which automatically sets a high con-
fluences the chemical’ s mutagenic potential [30]. Ahlberg and co- fidence for the assignment.
workers analyzed a series of functional groups in different positions An expert review of the positive data is conducted to confirm the
relative to the amino group to determine whether they are potentially overall positive outcome as illustrated in Fig. 15. By clicking on the
activating [30]. In the case of the target impurity, the carboxylic acid in experimental data bacterial mutagenicity node, the data used in the
the meta position and the bromine in the para position are not consid- assessment can be inspected alongside all the mutagenicity studies
ered activating. Therefore, even if the chemical undergoes metabolic identified for the target impurity. An expert review of the available data
activation resulting in a primary aromatic amine, the metabolite is un- determines that there is clear evidence of the mutagenic activity of the
likely to be mutagenic given the presence of these two ring substituents. target impurity and therefore the experimental data reliability is
 The expert review considerations confirm the negative outcome increased from a reliability score of RS5 to a reliability score of RS1. To
provided by the statistical- and expert rule-based models and the call’ s further confirm the positive outcome, the in silico predictions for 1-
reliability score is increased for the individual models to RS3 (prediction chloro-2-nitrobenzene are analyzed. The two complementary (Q)SAR
with expert review). The outcome of the two models can then be com- methodologies provide consistent positive outcomes. The positive pre-
bined to derive an overall assessment of the target impurity. For the diction given by the statistical model is driven by the correspondence of
combination of the two negative results, the current expert review also target impurity with an experimentally positive training compound; the
considers the low risk of missing a mutagenic impurity according to the model features identified by the model adequately cover the structure of
analysis published by Amberg et al. [4]. This analysis using a large the target impurity, with the aromatic nitro and cyclic nitro moieties
bacterial mutation data set shows that, when both statistical-based and providing higher contribution to the prediction. This results in a clear
expert rule-based methods generate a negative (in domain) assessment, positive call. The expert rule-based methodology further supports the

G.J. Myatt et al.
 Fig. 11. Implementation of the skin sensitization in silico protocol’ s hazard assessment framework within the platform.
Table 5 considerations, the current expert review concludes that the target im-
Summary of the results from the database purity 1-chloro-2-nitrobenzene is mutagenic, i.e., positive for bacterial
 mutation (Ames test) with a high confidence.
 Database Mapped to effects/ Number of Number Number The positive experimental carcinogenicity data and corresponding
 mechanisms chemicals of studies of tests studies identified by the tool for the target impurity are next inspected
 Carcinogenicity Rodent 5,700 18,084 27,099 together with all the carcinogenicity studies. Such studies provide clear
 carcinogenicity evidence of the carcinogenic activity of 1-chloro-2-nitrobenzene and
 Genetic Bacterial mutation 12,694 41,914 288,280 therefore the experimental data reliability is increased from a reliability
 toxicology Mouse Lymphoma 5,921 11,764 16,227
 Chromosome 5,660 6,575 11,794 score of RS5 to a reliability score of RS1.
 aberration in vitro The ICH M7 protocol workflow allows for the combination of the
 Micronucleus in vitro 1,298 794 1,065 positive bacterial mutation and carcinogenicity results that lead to an
 Chromosome 1,026 2,679 3,054 ICH M7 Class 1 assignment with high-confidence for the target impurity
 aberration in vivo
 Micronucleus in vivo 4,078 9,148 12,010 1-chloro-2-nitrobenzene. Experimental data, in silico predictions and
 Skin Protein reactivity 271 considerations of the expert review are all structured in a standardized
 sensitization Activation of Nrf2- 281 report to be shared with third parties. A reviewer can then use the
 ARE detailed information organized in such report to formulate an inde-
 Expression of co- 239 pendent assessment.
 stimulatory and
 adhesion molecules
 Reaction domain 458
 Rodent local lymph 2,176 3,266 1,893 Case study 3: Genetic toxicology assessment of m-xylylenediamine
 node proliferation
 Rodent 54 When the genetic toxicology protocol is applied on m-xylylenedi-
 maximization
 Human skin 151 amine, the tool performs database searches and run in silico models for
 sensitization each effect/mechanism defined by this hazard assessment framework
 [16], summarized in Fig. 15.
positive outcome because of the identification of the aromatic nitro The protocol window, as shown in Fig. 17, then guides the expert
structural alert of potential concern for mutagenicity. In addition, the review of the experimental data and in silico results as gathered by the
target impurity belongs to the alert reference set and this steers the tool for each effect/mechanism.
positive expert rule-based prediction. Based on the above First, the genetic mutation potential is assessed by considering
 available information for bacterial mutation and mammalian gene

G.J. Myatt et al.
Table 6
Summary of in silico performance results
 Effect/ Computational Type of Type of Count Concordance Sensitivity Specificity PPV NVP Comment
 mechanism models model validation (%) (%) (%) (%) (%)
 Bacterial Bacterial Statistical Cross 9,254 85 85 86 88 83
 mutation mutation v2 validation
 (5%)
 Bacterial Bacterial Statistical External 388 83 82 83 56 95 Reported in [33]
 mutation mutationv2 validation
 Bacterial Bacterial Expert rules Internal 11,528 87 87 88 89 86
 mutation mutation v7 validation
 Mouse MLA Activated Statistical Cross 675 76 75 76 72 80
 Lymphoma v2 validation
 (5%)
 Mouse MLA Statistical Cross 752 76 79 74 73 80
 Lymphoma unactivated v2 validation
 (5%)
 Chromosome In Vitro Chrom Statistical Cross 874 77 80 74 78 76
 aberration in Ab CHL v2 validation
 vitro (3%)
 Chromosome In Vivo Chrom Statistical Cross 285 77 80 74 78 76
 aberration in Ab Comp v2 validation
 vivo (2%)
 Micronucleus in In vivo Statistical Cross 1001 76 75 76 60 87 3 sub-models
 vivo micronucleus validation
 mouse v2 (5%)
 Micronucleus in In vivo Statistical External 42 80 67 84 57 89 91% coverage; Reported
 vivo micronucleus validation in [34]
 mouse
 Protein reactivity DPRA v2 Statistical Cross 176 87 93 71 90 79 Categorical model. The
 validation sensitivity and specificity
 of the DPRA categorical
 model was calculated
 based on the binary values
 of positive and negative,
 where positive reactivity
 values are defined as a
 mean % depletion >
 6.38% (low, moderate and
 high reactivity), and the
 no or minimal reactivity
 class (mean % depletion
 <= 6.37%) is negative.
 Activation of KeratinoSens v2 Statistical Cross 234 78 83 71 79 76
 Nrf2-ARE validation
 (10%)
 Expression of co- h-CLAT v2 Statistical Cross 179 75 76 72 91 46 4 sub-models
 stimulatory validation
 and adhesion
 molecules
 Rodent local LLNA v2 Statistical Cross 843 80 85 73 82 77 Categorical model. The
 lymph node (categorical) validation sensitivity and specificity
 proliferation of the LLNA categorical
 model was calculated
 based on the binary values
 of positive and negative,
 where positive values are
 defined as a EC3 % <=
 100% (weak, moderate
 and strong/extreme
 sensitizers), and the non-
 sensitizers (EC3 % >
 100%) are negative.
 Rodent local LLNA v2 Expert rules Internal 843 85 80 92 93 77 The sensitivity and
 lymph node validation specificity of the LLNA
 proliferation categorical model was
 calculated based on the
 binary values of positive
 and negative, where
 positive values are defined
 as a EC3 % <= 100%
 (weak, moderate and
 strong/extreme
 sensitizers), and the non-
 sensitizers (EC3 % >
 100%) are negative.

G.J. Myatt et al.
 Fig. 12. Summary of the preliminary results of the ICH M7 hazard assessment of 2-bromo-5-acetamidobenzoic acid.
 Fig. 13. ICH M7 hazard assessment workflow for 2-bromo-5-acetamidobenzoic acid.
mutation. The negative bacterial mutation experimental data is in the protocol workflow into an overall negative mouse lymphoma
inspected. It indicates clear evidence of non-mutagenic activity for the assessment with a reliability score of RS1.
target chemical according to studies compliant with the OECD 471 The bacterial mutation and the mouse lymphoma assessments are
guideline’ s requirements [3]. As such, the default reliability score of RS5 used to derive the overall negative gene mutation potential. Given the
is increased to RS1. To further confirm the negative outcome, the in silico reliability scores that have been set during the expert review, the con-
predictions for m-xylylenediamine are next analyzed. The two comple- fidence of this negative result is automatically assigned by the protocol
mentary (Q)SAR methodologies provide consistent negative outcomes. to “ High” [16].
Expert review sets the reliability score of these in silico results to RS3, Next the clastogenicity / aneugenicity in vitro endpoint (see Fig. 17)
whereas the overall negative bacterial mutation assessment that also is assessed by inspecting the available chromosome aberration in vitro
accounts for the available experimental data (RS1) can be associated experimental data and in silico results. An expert review confirms the
with an RS1 score. initial negative assessment that is associated with an RS1 score. These
 For the mouse lymphoma assessment, the protocol shows that results are used to derive the negative assessment for the clastogenicity /
negative experimental data are available and in silico predictions are aneugenicity in vitro endpoint with a medium confidence due to the lack
generated; an expert review concludes that there is sufficient evidence to of information on micronucleus in vitro.
increase the reliability of the experimental data to RS1 whereas the in The next step consists in reviewing the assessment of the clastoge-
silico predictions can be associated with an RS3 score. This is combined nicity/aneugenicity in vivo potential (see Fig. 16). No experimental

G.J. Myatt et al.
 Fig. 14. Summary of the preliminary results of the ICH M7 hazard assessment of 1-chloro-2-nitrobenzene.
 Fig. 15. ICH M7 hazard assessment workflow for 1-chloro-2-nitrobenzene.
evidence is available for m-xylylenediamine concerning chromosome results increases the experimental reliability score to RS1 and overturns
aberration in vivo, whereas the in silico model (i.e., a statistical-based the out-of-domain prediction into a negative prediction with RS5 score
system) provides an out-of-domain result triggered by the absence of (based on the good coverage of the structure of m-xylylenediamine in
similar structures in the training set. Expert review of this prediction addition to the lack of any reactive potential). A negative assessment
suggests that it is feasible to overturn the out-of-domain outcome into a with RS1 score is then set for the micronucleus in vivo assessment. The
negative call based on the following elements: a) a low positive pre- results for chromosome aberration in vivo (negative RS5) and micronu-
diction probability (PPP = 0.124); b) a good coverage of the m-xylyle- cleus in vivo (negative RS1) are combined for the negative assessment of
nediamine structure; c) negative features contributing to the prediction the overall clastogenicity / aneugenicity in vivo potential, with corre-
leading to a clear negative call; d) no features associated with a positive sponding “ Medium” confidence set by the tool according to the protocol
contribution. Since only one model is used to predict chromosome ab- rules [16].
erration in vivo without any sufficiently similar analogs, the reliability The clastogenicity / aneugenicity in vitro (negative, medium confi-
score of the chromosome aberration in vivo prediction is set to RS5. An dence) and in vivo (negative, medium confidence) sub-endpoints are
expert review of the micronucleus in vivo experimental data and in silico then combined into a single clastogenicity / aneugenicity endpoint. This

G.J. Myatt et al.
 Fig. 16. Summary of the preliminary results of the genetic toxicology protocol of m-xylylenediamine.
is assessed as negative with a “ Medium” confidence as proposed by the an increase in the reliability of the experimental h-CLAT data. An
tool. explanation of the positive statistical prediction shows that the result is
 Finally, all the results of the sub-endpoints (see Fig. 16) are auto- within the applicability domain of the model with 10 structural features
matically combined into a single overall negative assessment of “ Genetic mapping to the Bis-GMA structure and a predicted probability of 0.79.
Toxicity” with a medium confidence score, confirmed by the expert The feature weighting, relevancy of the model descriptors and suffi-
review. ciency of training set data are evaluated as part of the expert review
 The genetic toxicology assessment is then saved in a report sum- process as prompted by the checklist of items to consider for an expert
marizing the results alongside the elements considered in the expert review, Fig. 20.
reviews. The report consists of a single editable word document The feature which contributes most to the positive weight is the
including an executive summary (covering materials and methods used acrylate group. The activity of the training set examples cannot be
for the prediction of each effect/mechanism; any rules and principles explained through any moiety other than the (meth)acrylate feature,
used to combine the information; results for the individual effects/ and a potential reaction mechanism could be postulated based on this
mechanisms and associated reliability scores; results for the endpoints feature. This supports the relevancy of the structural moiety and an
and associated confidence scores) and the hazard assessment framework increased prediction reliability. Of note however, the training set data
view (broken down into a series of graphs) and any comments included are non-aromatic structures, which lack diversity and the influence of
in the assessment. In addition, a zip file containing an appendix of in- the aromatic ring on the activity cannot be adequately assessed. Overall,
formation is created including full study and in silico reports for each the expert review of the statistical model’ s positive prediction confirms
individual prediction. such result given the weighting of features and the mechanistic basis
 which could be attributed to the acrylate feature. The level of evidence
 supports an increase in reliability to an RS3 level for this positive result.
Case study 4: Skin sensitization assessment of bis-GMA Since the RS3 reliability is higher than the RS5 reliability of the exper-
 imental data, the positive prediction is used in the assessment. The
 The following case study describes the assessment of bis-GMA (CAS “ Events in Dendritic cells” endpoint automatically changes to a positive
1565– 94-2) using the implemented version of the skin sensitization outcome, with a medium confidence level. Fig. 21 shows a manual
protocol. The software returns an assessment of positive with high override of the result, the associated documentation and the updated
confidence for skin sensitization in humans. The main endpoints, “ skin “ Events in Dendritic cells” node.
sensitization in rodents” and “ skin sensitization in vitro” are assessed as After working through the assessment of the “ Events in Dendritic
positive with a high and low confidence levels, respectively. Fig. 18 cells” endpoint and understanding the negative experimental results and
provides an overview of the workflow used to derive the overall the evidence presented by the in silico methods, one of the two following
assessment. approaches could be taken. Either, an evaluation of in vitro endpoints
 The “ Explain” function is used to understand the basis for the posi- which were predicted by in silico models (“ Covalent interaction with
tive prediction and the confidence level. It is important to explore any skin proteins” and “ Events in Keratinocytes” ) could be continued or any
experimental data or in silico predictions that are used in the assessment additional experimental data could be assessed. The latter approach is
and how reliable the data are. In silico predictions are used to assess all adopted here since a high-quality experimental result would be suffi-
the relevant mechanisms/effects while experimental data are available cient for a regulatory assessment, particularly where any conflicting
for the LLNA and h-CLAT assessments. A high-level overview shows that information can be explained. An assessment of the rodent LLNA result is
the h-CLAT experimental data disagrees with the positive LLNA exper- made based on an experimental study linked to the test structure con-
imental assessment. Further, the positive h-CLAT statistical model tained in the database. This study indicates weak sensitization potential
outcome does not support the negative experimental h-CLAT assess- and is assigned a reliability score of 1. Clicking on the experimental
ment. Both these results are initially assigned a reliability score of RS5 results for the LLNA returns the studies that are available for expert
and the ‘Expression of co-stimulatory and adhesion molecules’ endpoint review, including their references and a comment field to document any
is left unassigned given conflicting assessments of the same reliabilities, findings, Fig. 22.
Fig. 19. This prompts an expert review. The LLNA statistical model (predicting weak potency class) and
 The skin sensitization protocol outlines factors that could lead to expert alert results (acrylates and methacrylate alert matched) further
false negative results in the h-CLAT experimental system and discusses support the positive assessment and potency classification. The assess-
the exclusion of chemicals with a Log P value > 3.5 from the applicability ment of the “ skin sensitization in rodents” endpoint is therefore assessed
domain of the h-CLAT test [17,31]. The calculated ALogP value of bis- as positive (weak potency) with high confidence. At this point, there
GMA (3.73) marginally falls within this range and a false negative does not appear to be conflicting evidence across the framework and the
experimental result is suspected. This non-applicability of the experi- positive “ sensitization in vitro” assessment supports the overall positive
mental system is reflected in the test guideline [32] and does not support

G.J. Myatt et al.
Fig. 17. Genetic toxicology assessment workflow for m-xylylenediamine. The original genetic toxicology assessment is reviewed by inspecting and analyzing: i)
genetic mutation potential; ii) clastogenicity in vitro; iii) clastogenicity in vivo.
outcome with high confidence for skin sensitization hazard of Bis-GMA. reactivity with a probability of 0.921 which is driven primarily by the
However, it is prudent to review the statistical models predicting protein acrylate feature (see Fig. 23).
reactivity and activation of the Nrf2-ARE pathway (keratinocyte acti- The result is similar to the Keratinosens™ statistical model’ s pre-
vation) to confirm that there is no conflicting information. Similar to the diction (see Table 6 for additional information on the models). In both
h-CLAT assessment, clicking on the nodes which contains the model cases, the expert review supported an increase in reliability to an RS3
predictions returns the explain model, find analogs and expert review level and the “ covalent interaction with skin proteins” and the “ Events in
fields (see Fig. 17). The DPRA model predicts moderate protein Keratinocytes” endpoints are both assessed as positive with medium

G.J. Myatt et al.
 Fig. 18. Workflow to derive an assessment of skin sensitization using the implemented skin sensitization protocol.
confidence. Together with the h-CLAT assessment, the “ skin sensitiza- silico toxicology protocols, the platform provides an approach that is
tion in vitro” endpoint is assessed as positive, with medium confidence defendable to colleagues and peers. By incorporating transparent met-
and the system explains that this assessment is based on an encoded rule rics of reliability, relevance and confidence, the approach supports many
of “ at least two positive assessments aligned” for the “ sensitization in different applications, from regulatory submissions to screening chem-
vitro” endpoint. The “ sensitization in vitro” assessment supports the final icals, that tolerate differing levels of uncertainty. The visual platform is
assessment of positive, with high confidence which is based primarily on transparent, clearly showing the steps in the assessment process, with
LLNA results as the key study. details available at any stage on demand. The ability to interact with the
 platform supports a thorough expert review. Such a review may modify
Discussion any conclusions on any of the effects, mechanisms, or derived toxico-
 logical endpoints. The deviations from the default assessment (described
 The tool described in this paper addresses many critical issues that in the protocols) are recorded along with their justifications. Automat-
enable the use of integrated approaches for toxicological hazard ically documenting this entire process, including the source materials
assessment to be used across industrial and regulatory applications. By (experimental and in silico results) and the entire decision-making pro-
being based on commonly agreed principles and procedures, such as in cess, and tracking the expert review rationale ensure the results are

G.J. Myatt et al.
Fig. 19. Assessment node for the “ Expression of co-stimulatory adhesion molecules” based on the h-CLAT method. Clicking on the result allows access to underlying
information.
 Fig. 20. Access to underlying information allows an expert review to be performed in a consistent manner.
effectively, completely, and consistently reported. At the same time, this available databases and in silico models. The framework described also
automatic approach avoids any transcription errors. The documentation supports the integration of proprietary experimental data to use in the
supports an outside review and would also enable a third party to repeat assessment of individual effects/mechanisms. Proprietary experimental
the process. The visual platform is based on a consistent organization data on chemicals analogs can also be utilized in this framework, by
and will support many different regulatory guidelines and protocols. introducing a read-across prediction and as part of an expert review of
Hence, this approach may be easily adopted when new regulations and the in silico results. In addition, proprietary models can also be used to
in silico toxicology protocols are developed. assess individual effects/mechanisms. This may be helpful when the test
 As mentioned earlier, an expert review plays an important role chemical falls outside the chemical space from which the public models
throughout the assessment process which is invariably subjective. To were built. When such assessments are performed for external groups,
mitigate this concern, a series of guidelines and case studies were such as regulatory agencies, it may be necessary to disclose the model’ s
introduced and detailed in different in silico protocol-related papers training set to be transparent. Such disclosure often makes the use of
[1,4,5,16,17]. Using these commonly agreed principles for performing these proprietary models impractical. There are also approaches that
such a review, a more consistent approach will support the application avoid the use of proprietary data directly yet incorporate knowledge
of in silico assessments across different regulatory frameworks and derived from proprietary database, such as the SAR fingerprinting
jurisdictions. approach. [30]
 This paper has outlined a series of case studies based on publicly

G.J. Myatt et al.
 Fig. 21. Influence of expert review findings on the assessment of “ Events in Dendritic cells” endpoint.
 Fig. 22. Review of experimental study and assessment of reliability.
Conclusion successful application of such approaches to hazard assessment will
 require the adoption of quality-driven standards and processes. Tools
 The integrated assessment of toxicological endpoints, where a bat- that support such assessments in an efficient, transparent, defendable,
tery of experimental and in silico methods are combined, is important to and repeatable manner, such as the visual and interactive platform
current and future toxicological hazard assessments. It provides a more described in this paper, will be essential to support hazard assessment
mechanistically interpretable approach that also supports the 3Rs. The based on these new methods.

G.J. Myatt et al.
 Fig. 23. A categorical model is used to predict the protein reactivity of Bis-GMA.
CRediT authorship contribution statement [4] A. Amberg, R.V. Andaya, L.T. Anger, C. Barber, L. Beilke, J. Bercu, D. Bower,
 A. Brigo, Z. Cammerer, K.P. Cross, L. Custer, K. Dobo, H. Gerets, V. Gervais,
 Glenn J. Myatt: Conceptualization, Writing – original draft, Writing S. Glowienke, S. Gomez, J. Van Gompel, J. Harvey, C. Hasselgren, M. Honma,
 C. Johnson, R. Jolly, R. Kemper, M. Kenyon, N. Kruhlak, P. Leavitt, S. Miller,
– review & editing, Funding acquisition. Arianna Bassan: Writing – W. Muster, R. Naven, J. Nicolette, A. Parenty, M. Powley, D.P. Quigley, M.
original draft, Writing – review & editing. Dave Bower: Investigation, V. Reddy, J.C. Sasaki, L. Stavitskaya, A. Teasdale, A. Trejo-Martin, S. Weiner, D.
Writing – review & editing. Candice Johnson: Writing – original draft, S. Welch, A. White, J. Wichard, D. Woolley, G.J. Myatt, Principles and procedures
 for handling out-of-domain and indeterminate results as part of ICH M7
Writing – review & editing. Scott Miller: Software, Writing – review & recommended (Q)SAR analyses, Regul. Toxicol. Pharmacol. 102 (2019) 53– 64,
editing. Manuela Pavan: Writing – original draft, Writing – review & https://doi.org/10.1016/j.yrtph.2018.12.007.
editing. Kevin P. Cross: Conceptualization, Software, Writing – review [5] A. Amberg, L. Beilke, J. Bercu, D. Bower, A. Brigo, K.P. Cross, L. Custer, K. Dobo,
 E. Dowdy, K.A. Ford, S. Glowienke, J. Van Gompel, J. Harvey, C. Hasselgren,
& editing. M. Honma, R. Jolly, R. Kemper, M. Kenyon, N. Kruhlak, P. Leavitt, S. Miller,
 W. Muster, J. Nicolette, A. Plaper, M. Powley, D.P. Quigley, M.V. Reddy, H.-
 P. Spirkl, L. Stavitskaya, A. Teasdale, S. Weiner, D.S. Welch, A. White, J. Wichard,
 G.J. Myatt, Principles and procedures for implementation of ICH M7 recommended
Declaration of Competing Interest (Q)SAR analyses, Regul. Toxicol. Pharmacol. 77 (2016) 13– 24, https://doi.org/
 10.1016/j.yrtph.2016.02.004.
 The authors declare that they have no known competing financial [6] K.L. Dobo, N. Greene, C. Fred, S. Glowienke, J.S. Harvey, C. Hasselgren, R. Jolly,
interests or personal relationships that could have appeared to influence M.O. Kenyon, J.B. Munzner, W. Muster, R. Neft, M. Vijayaraj Reddy, A.T. White,
 S. Weiner, In silico methods combined with expert knowledge rule out mutagenic
the work reported in this paper. potential of pharmaceutical impurities: an industry survey, Regul. Toxicol.
 Pharmacol. 62 (3) (2012) 449– 455, https://doi.org/10.1016/j.yrtph.2012.01.007.
 [7] C. Barber, A. Amberg, L. Custer, K.L. Dobo, S. Glowienke, J. Van Gompel,
Acknowledgement S. Gutsell, J. Harvey, M. Honma, M.O. Kenyon, N. Kruhlak, W. Muster,
 L. Stavitskaya, A. Teasdale, J. Vessey, J. Wichard, Establishing best practise in the
 Research reported in this publication was supported by the National application of expert review of mutagenicity under ICH M7, Regul. Toxicol.
 Pharmacol. 73 (1) (2015) 367– 377, https://doi.org/10.1016/j.yrtph.2015.07.018.
Institute of Environmental Health Sciences of the National Institutes of [8] M.W. Powley, (Q)SAR assessments of potentially mutagenic impurities: a
Health under Award Number R44ES026909. The content is solely the regulatory perspective on the utility of expert knowledge and data submission,
responsibility of the authors and does not necessarily represent the Regul. Toxicol. Pharmacol. 71 (2) (2015) 295– 300, https://doi.org/10.1016/j.
 yrtph.2014.12.012.
official views of the National Institutes of Health. [9] C.M. Ellison, P. Piechota, J.C. Madden, S.J. Enoch, M.T.D. Cronin, Adverse
 Outcome Pathway (AOP) informed modeling of aquatic toxicology: QSARs, read-
 across, and interspecies verification of modes of action, Environ. Sci. Technol. 50
References (7) (2016) 3995– 4007, https://doi.org/10.1021/acs.est.5b05918.
 [10] T.M. Martin, D.M. Young, C.R. Lilavois, M.G. Barron, Comparison of global and
 [1] G.J. Myatt, E. Ahlberg, Y. Akahori, D. Allen, A. Amberg, L.T. Anger, A. Aptula, mode of action-based models for aquatic toxicity, SAR QSAR Environ. Res. 26 (3)
 S. Auerbach, L. Beilke, P. Bellion, R. Benigni, J. Bercu, E.D. Booth, D. Bower, (2015) 245– 262, https://doi.org/10.1080/1062936X.2015.1018939.
 A. Brigo, N. Burden, Z. Cammerer, M.T.D. Cronin, K.P. Cross, L. Custer, [11] OECD, Guidance Document on the Reporting of Defined Approaches to Be Used
 M. Dettwiler, K. Dobo, K.A. Ford, M.C. Fortin, S.E. Gad-McDonald, N. Gellatly, within Integrated Approaches to Testing and Assessment, OECD Environment,
 V. Gervais, K.P. Glover, S. Glowienke, J. Van Gompel, S. Gutsell, B. Hardy, J. Health and Safety Publications, Paris, 2016. doi:10.1787/9789264274822-en.
 S. Harvey, J. Hillegass, M. Honma, J.-H. Hsieh, C.-W. Hsu, K. Hughes, C. Johnson, [12] OECD, Guidance Document on the Reporting of Defined Approaches and Individual
 R. Jolly, D. Jones, R. Kemper, M.O. Kenyon, M.T. Kim, N.L. Kruhlak, S.A. Kulkarni, Information Sources to be Used within Integrated Approaches to Testing and
 K. Kümmerer, P. Leavitt, B. Majer, S. Masten, S. Miller, J. Moser, M. Mumtaz, Assessment (IATA) for Skin Sensitisation, OECD Environment, Health and Safety
 W. Muster, L. Neilson, T.I. Oprea, G. Patlewicz, A. Paulino, E. Lo Piparo, Publications, Paris, 2016. doi:10.1787/9789264279285-en.
 M. Powley, D.P. Quigley, M.V. Reddy, A.-N. Richarz, P. Ruiz, B. Schilter, [13] US EPA, Integrated Risk Information System, US EPA. (2021). https://www.epa.go
 R. Serafimova, W. Simpson, L. Stavitskaya, R. Stidl, D. Suarez-Rodriguez, D. v/assessing-and-managing-chemicals-under-tsca/strategic-plan-reduce-use-ver
 T. Szabo, A. Teasdale, A. Trejo-Martin, J.-P. Valentin, A. Vuorinen, B.A. Wall, tebrate-animals-chemical (accessed April 9, 2021).
 P. Watts, A.T. White, J. Wichard, K.L. Witt, A. Woolley, D. Woolley, C. Zwickl, [14] OECD, OECD Test Guidelines for Chemicals, (2021). https://www.oecd.org/chem
 C. Hasselgren, In silico toxicology protocols, Regul. Toxicol. Pharmacol. 96 (2018) icalsafety/testing/oecdguidelinesforthetestingofchemicals.htm (accessed April 22,
 1– 17, https://doi.org/10.1016/j.yrtph.2018.04.014. 2021).
 [2] ICH, M7 (R1) Assessment and control of DNA reactive (mutagenic) impurities in [15] OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity
 pharmaceuticals to limit potential carcinogenic risk, European Medicines Agency, Relationship [(Q)SAR] Models, OECD Environment, Health and Safety
 2017. https://database.ich.org/sites/default/files/M7\_R1\_Guideline.pdf. Publications, Paris, 2007. doi:10.1787/9789264085442-en.
 [3] OECD, Test No. 471: Bacterial Reverse Mutation Test, OECD Publishing, Paris, [16] C. Hasselgren, E. Ahlberg, Y. Akahori, A. Amberg, L.T. Anger, F. Atienzar,
 1997. Doi:10.1787/9789264071247-en. S. Auerbach, L. Beilke, P. Bellion, R. Benigni, J. Bercu, E.D. Booth, D. Bower,

G.J. Myatt et al.
 A. Brigo, Z. Cammerer, M.T.D. Cronin, I. Crooks, K.P. Cross, L. Custer, K. Dobo, structural features by reassembling medicinal chemistry building blocks, J. Med.
 T. Doktorova, D. Faulkner, K.A. Ford, M.C. Fortin, M. Frericks, S.E. Gad-McDonald, Chem. 47 (24) (2004) 5984–5994, https://doi.org/10.1021/jm0497242.
 N. Gellatly, H. Gerets, V. Gervais, S. Glowienke, J. Van Gompel, J.S. Harvey, [25] P. Blower, K. Cross, M. Fligner, G. Myatt, J. Verducci, C. Yang, Systematic analysis
 J. Hillegass, M. Honma, J.H. Hsieh, C.W. Hsu, T.S. Barton-Maclaren, C. Johnson, of large screening sets in drug discovery, Curr. Drug Discovery Technol. 1 (2004)
 R. Jolly, D. Jones, R. Kemper, M.O. Kenyon, N.L. Kruhlak, S.A. Kulkarni, 37–47, https://doi.org/10.2174/1570163043484879.
 K. Kümmerer, P. Leavitt, S. Masten, S. Miller, C. Moudgal, W. Muster, A. Paulino, [26] S.J. Enoch, M.T.D. Cronin, C.M. Ellison, The use of a chemistry-based profiler for
 E. Lo Piparo, M. Powley, D.P. Quigley, M.V. Reddy, A.N. Richarz, B. Schilter, R. covalent DNA binding in the development of chemical categories for read-across
 D. Snyder, L. Stavitskaya, R. Stidl, D.T. Szabo, A. Teasdale, R.R. Tice, A. Trejo- for genotoxicity, Altern. Lab. Anim. 39 (2) (2011) 131–145, https://doi.org/
 Martin, A. Vuorinen, B.A. Wall, P. Watts, A.T. White, J. Wichard, K.L. Witt, 10.1177/026119291103900206.
 A. Woolley, D. Woolley, C. Zwickl, G.J. Myatt, Genetic toxicology in silico protocol, [27] A.O. Aptula, D.W. Roberts, Mechanistic applicability domains for nonanimal-based
 Regul. Toxicol. Pharmacol. 107 (2019) 104403, https://doi.org/10.1016/j. prediction of toxicological end points: general principles and application to
 yrtph.2019.104403. reactive toxicity, Chem. Res. Toxicol. 19 (2006) 1097–1105, https://doi.org/
[17] C. Johnson, E. Ahlberg, L.T. Anger, L. Beilke, R. Benigni, J. Bercu, S. Bobst, 10.1021/tx0601004.
 D. Bower, A. Brigo, S. Campbell, M.T.D. Cronin, I. Crooks, K.P. Cross, T. Doktorova, [28] ECHA, Read-Across Assessment Framework (RAAF), 2017. doi:10.2823/619212.
 T. Exner, D. Faulkner, I.M. Fearon, M. Fehr, S.C. Gad, V. Gervais, A. Giddings, [29] B. Testa, J.M. Mayer, Hydrolysis in Drug and Prodrug Metabolism: Chemistry,
 S. Glowienke, B. Hardy, C. Hasselgren, J. Hillegass, R. Jolly, E. Krupp, L. Lomnitski, Biochemistry, and Enzymology, 1st ed., John Wiley & Sons, Ltd, 2003. doi:
 J. Magby, J. Mestres, L. Milchak, S. Miller, W. Muster, L. Neilson, R. Parakhia, 10.1002/9783906390444.
 A. Parenty, P. Parris, A. Paulino, A.T. Paulino, D.W. Roberts, H. Schlecker, R. Stidl, [30] E. Ahlberg, A. Amberg, L.D. Beilke, D. Bower, K.P. Cross, L. Custer, K.A. Ford,
 D. Suarez-Rodrigez, D.T. Szabo, R.R. Tice, D. Urbisch, A. Vuorinen, B. Wall, J. Van Gompel, J. Harvey, M. Honma, R. Jolly, E. Joossens, R.A. Kemper,
 T. Weiler, A.T. White, J. Whritenour, J. Wichard, D. Woolley, C. Zwickl, G.J. Myatt, M. Kenyon, N. Kruhlak, L. Kuhnke, P. Leavitt, R. Naven, C. Neilan, D.P. Quigley,
 Skin sensitization in silico protocol, Regul. Toxicol. Pharmacol. 116 (2020) D. Shuey, H.-P. Spirkl, L. Stavitskaya, A. Teasdale, A. White, J. Wichard, C. Zwickl,
 104688, https://doi.org/10.1016/j.yrtph.2020.104688. G.J. Myatt, Extending (Q)SARs to incorporate proprietary knowledge for
[18] NTP, National Toxicology Program U.S. Departament of Health and Human regulatory purposes: a case study using aromatic amine mutagenicity, Regul.
 Services - Data & Resources, (2021). https://ntp.niehs.nih.gov/data/index.html. Toxicol. Pharmacol. 77 (2016) 1–12, https://doi.org/10.1016/j.
[19] NIH, Download Carcinogenic Potency Database (CPDB) Data, (2021). https yrtph.2016.02.003.
 ://www.nlm.nih.gov/databases/download/cpdb.html (accessed June 11, 2021). [31] O. Takenouchi, M. Miyazawa, K. Saito, T. Ashikaga, H. Sakaguchi, Predictive
[20] D. Bower, K. Cross, G. Myatt, Organisation of Toxicological Data in Databases, in: performance of the human Cell Line Activation Test (h-CLAT) for lipophilic
 D. Neagu, A.-.N. Richarz (Eds.), Big Data in Predictive Toxicology, Royal Society of chemicals with high octanol-water partition coefficients, J. Toxicol. Sci. 38 (4)
 Chemistry, Cambridge, 2020, pp. 108–165, https://doi.org/10.1039/ (2013) 599–609, https://doi.org/10.2131/jts.38.599.
 9781782623656-00108. [32] OECD, Test No. 442E: In Vitro Skin Sensitisation: In Vitro Skin Sensitisation assays
[21] Leadscope, Instem - Computational Toxicology, 2021. https://www.instem.com/so addressing the Key Event on activation of dendritic cells on the Adverse Outcome
 lutions/insilico/computational-toxicology.php. Pathway for Skin Sensitisation, OECD Publishing, Paris, 2018. doi:10.1787/
[22] G. Roberts, G.J. Myatt, W.P. Johnson, K.P. Cross, P.E. Blower, LeadScope: software 9789264264359-en.
 for exploring large sets of screening data, J. Chem. Inf. Comput. Sci. 40 (2000) [33] C. Landry, M.T. Kim, N.L. Kruhlak, K.P. Cross, R. Saiakhov, S. Chakravarti,
 1302–1314, https://doi.org/10.1021/ci0000631. L. Stavitskaya, Transitioning to composite bacterial mutagenicity models in ICH
[23] K.P. Cross, G. Myatt, C. Yang, M.A. Fligner, J.S. Verducci, P.E. Blower, Finding M7 (Q)SAR analyses, Regul. Toxicol. Pharmacol. 109 (2019) 104488, https://doi.
 discriminating structural features by reassembling common building blocks, org/10.1016/j.yrtph.2019.104488. In press.
 J. Med. Chem. 46 (22) (2003) 4770–4775, https://doi.org/10.1021/jm0302703. [34] J.W. Yoo, N.L. Kruhlak, C. Landry, K.P. Cross, A. Sedykh, L. Stavitskaya,
[24] C. Yang, K. Cross, G.J. Myatt, P.E. Blower, J.F. Rathman, Building predictive Development of improved QSAR models for predicting the outcome of the in vivo
 models for protein tyrosine phosphatase 1B inhibitors based on discriminating micronucleus genetic toxicity assay, Regul. Toxicol. Pharmacol. 113 (2020)
 104620, https://doi.org/10.1016/j.yrtph.2020.104620.

Source PDF: /instem.com/www.instem.com/docs/sot-2021/in-silico-toxicology-protocol.pdf

 Regulatory Toxicology and Pharmacology 96 (2018) 1–17
 Contents lists available at ScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepage: www.elsevier.com/locate/yrtph
In silico toxicology protocols T
Glenn J. Myatta,∗ , Ernst Ahlbergb, Yumi Akahoric, David Allend, Alexander Amberge,
Lennart T. Angere, Aynur Aptulaf, Scott Auerbachg, Lisa Beilkeh, Phillip Bellioni,
Romualdo Benignij, Joel Bercuk, Ewan D. Boothl, Dave Bowera, Alessandro Brigom,
Natalie Burdenn, Zoryana Cammerero, Mark T.D. Croninp, Kevin P. Crossa, Laura Custerq,
Magdalena Dettwilerr, Krista Dobos, Kevin A. Fordt, Marie C. Fortinu,
Samantha E. Gad-McDonaldv, Nichola Gellatlyn, Véronique Gervaisw, Kyle P. Gloverx,
Susanne Glowienkey, Jacky Van Gompelz, Steve Gutsellf, Barry Hardyaa, James S. Harveybb,
Jedd Hillegassq, Masamitsu Honmacc, Jui-Hua Hsiehdd, Chia-Wen Hsuee, Kathy Hughesﬀ ,
Candice Johnsona, Robert Jollygg, David Joneshh, Ray Kemperii, Michelle O. Kenyons,
Marlene T. Kimee, Naomi L. Kruhlakee, Sunil A. Kulkarniﬀ , Klaus Kümmererjj, Penny Leavittq,
Bernhard Majerkk, Scott Masteng, Scott Millera, Janet Moserll,mm , Moiz Mumtaznn,
Wolfgang Musterm, Louise Neilsonoo, Tudor I. Opreapp, Grace Patlewiczqq, Alexandre Paulinorr,
Elena Lo Pipaross, Mark Powleyee, Donald P. Quigleya, M. Vijayaraj Reddytt,
Andrea-Nicole Richarzuu, Patricia Ruiznn, Benoit Schilterss, Rositsa Seraﬁ movavv,
Wendy Simpsonf, Lidiya Stavitskayaee, Reinhard Stidlkk, Diana Suarez-Rodriguezf,
David T. Szaboww, Andrew Teasdalexx, Alejandra Trejo-Martink, Jean-Pierre Valentinyy,
Anna Vuorineni, Brian A. Wallzz, Pete Wattsaaa, Angela T. Whitebb, Joerg Wichardbbb,
Kristine L. Wittg, Adam Woolleyccc, David Woolleyccc, Craig Zwicklddd, Catrin Hasselgrena
a Leadscope, Inc., 1393 Dublin Rd, Columbus, OH 43215, USA
b Predictive Compound ADME & Safety, Drug Safety & Metabolism, AstraZeneca IMED Biotech Unit, Mölndal, Sweden
c Chemicals Evaluation and Research Institute, 1-4-25 Kouraku, Bunkyo-ku, Tokyo 112-0004 Japan
d Integrated Laboratory Systems, Inc., Research Triangle Park, NC, USA
e Sanoﬁ , R&D Preclinical Safety Frankfurt, Industriepark Hoechst, D-65926 Frankfurt am Main, Germany
f Unilever, Safety and Environmental Assurance Centre, Colworth, Beds, UK
g The National Institute of Environmental Health Sciences, Division of the National Toxicology Program, Research Triangle Park, NC 27709, USA
h Toxicology Solutions Inc., San Diego, CA, USA
i DSM Nutritional Products, Kaiseraugst, Switzerland
j Alpha-PreTox, via G.Pascoli 1, 00184 Roma, Italy
k Gilead Sciences, 333 Lakeside Drive, Foster City, CA, USA
l Syngenta, Product Safety Department, Jealott's Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, UK
m Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Switzerland
n National Centre for the Replacement, Reﬁ nement and Reduction of Animals in Research (NC3Rs), Gibbs Building, 215 Euston Road, London NW1 2BE, UK
o Janssen Research & Development, 1400 McKean Road, Spring House, PA 19477, USA
p School of Pharmacy and Chemistry, Liverpool John Moores University, Liverpool, L3 3AF, UK
q Bristol-Myers Squibb, Drug Safety Evaluation, 1 Squibb Dr, New Brunswick, NJ 08903, USA
r Elanco Animal Health, Schwarzwaldallee 215, 4058 Basel, Switzerland
s Pﬁ zer Global Research & Development, 558 Eastern Point Road, Groton, CT 06340, USA
t Global Blood Therapeutics, South San Francisco, CA 94080, USA
u Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, 170 Frelinghuysen Rd, Piscataway, NJ
08855, USA
v Gad Consulting Services, 4008 Barrett Drive, Suite 201, Raleigh, NC 27609, USA
w Biologie Servier, 905 Route de Saran, 45520 Gidy, France
x Defense Threat Reduction Agency, Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD 21010, USA
y Novartis Pharma AG, Pre-Clinical Safety, Werk Klybeck, CH-4057, Basel, Switzerland
z Janssen Pharmaceutical Companies of Johnson & Johnson, 2340 Beerse, Belgium
aa Douglas Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, CH-4057 Basel / Basel-Stadt, Switzerland
 ∗ Corresponding author.
 E-mail address: gmyatt@leadscope.com (G.J. Myatt).
https://doi.org/10.1016/j.yrtph.2018.04.014
Available online 17 April 2018
0273-2300/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
bb GlaxoSmithKline Pre-Clinical Development, Park Road, Ware, Hertfordshire, SG12 0DP, UK
cc National Institute of Health Sciences, Tokyo, Japan
dd Kelly Government Solutions, Research Triangle Park, NC 27709, USA
ee FDA Center for Drug Evaluation and Research, Silver Spring, MD 20993, USA
ﬀ Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, ON, K1A 0K9, Canada
gg Toxicology Division, Eli Lilly and Company, Indianapolis, IN, USA
hh Medicines and Healthcare Products Regulatory Agency, 151 Buckingham Palace Road, London, SW1W 9SZ, UK
ii Vertex Pharmaceuticals Inc., Discovery and Investigative Toxicology, 50 Northern Ave, Boston, MA, USA
jj Institute for Sustainable and Environmental Chemistry, Leuphana University Lüneburg, Scharnhorststraße 1/C13.311b, 21335 Lüneburg, Germany
kk Shire, Industriestrasse 20, 1221, Wien, Austria
ll Chemical Security Analysis Center, Department of Homeland Security, 3401 Ricketts Point Road, Aberdeen Proving Ground, MD 21010-5405, USA
mm Battelle Memorial Institute, 505 King Avenue, Columbus, OH 43210, USA
nn Agency for Toxic Substances and Disease Registry, US Department of Health and Human Services, Atlanta, GA, USA
oo British American Tobacco, Research and Development, Regents Park Road, Southampton, Hampshire, SO15 8TL, UK
pp Translational Informatics Division, Department of Internal Medicine, Health Sciences Center, The University of New Mexico, NM, USA
qq U.S. Environmental Protection Agency, National Center for Computational Toxicology, Research Triangle Park, NC 27711, USA
rr SAPEC Agro, S.A., Avenida do Rio Tejo, Herdade das Praias, 2910-440 Setúbal, Portugal
ss Chemical Food Safety Group, Nestlé Research Center, Lausanne, Switzerland
tt Merck Research Laboratories, West Point, PA 19486, USA
uu European Commission, Joint Research Centre, Directorate for Health, Consumers and Reference Materials, Chemical Safety and Alternative Methods Unit, Via Enrico
Fermi 2749, 21027 Ispra, VA, Italy
vv European Food Safety Authority, Via Carlo Magno 1A, 43126 Parma, Italy
ww RAI Services Company, 950 Reynolds Blvd., Winston-Salem, NC 27105, USA
xx AstraZeneca, Macclesﬁ eld, Cheshire, UK
yy UCB Biopharma SPRL, Chemin du Foriest, B-1420 Braine-l’Alleud, Belgium
zz Colgate-Palmolive Company, Piscataway, NJ 08854, USA
aaa Bibra, Cantium House, Railway Approach, Wallington, Surrey, SM6 0DZ, UK
bbb Bayer Pharma AG, Investigational Toxicology, Muellerstr. 178, D-13353 Berlin, Germany
ccc ForthTox Limited, PO Box 13550, Linlithgow, EH49 7YU, UK
ddd Transendix LLC, 1407 Moores Manor, Indianapolis, IN 46229, USA
ARTICLE INFO ABSTRACT
Keywords: The present publication surveys several applications of in silico (i.e., computational) toxicology approaches
In silico across diﬀ erent industries and institutions. It highlights the need to develop standardized protocols when con-
In silico toxicology ducting toxicity-related predictions. This contribution articulates the information needed for protocols to sup-
Computational toxicology port in silico predictions for major toxicological endpoints of concern (e.g., genetic toxicity, carcinogenicity,
Predictive toxicology acute toxicity, reproductive toxicity, developmental toxicity) across several industries and regulatory bodies.
QSAR Such novel in silico toxicology (IST) protocols, when fully developed and implemented, will ensure in silico
Expert alert toxicological assessments are performed and evaluated in a consistent, reproducible, and well-documented
Expert review
 manner across industries and regulatory bodies to support wider uptake and acceptance of the approaches. The
 development of IST protocols is an initiative developed through a collaboration among an international con-
 sortium to reﬂ ect the state-of-the-art in in silico toxicology for hazard identiﬁ cation and characterization. A
 general outline for describing the development of such protocols is included and it is based on in silico predic-
 tions and/or available experimental data for a deﬁ ned series of relevant toxicological eﬀ ects or mechanisms. The
 publication presents a novel approach for determining the reliability of in silico predictions alongside experi-
 mental data. In addition, we discuss how to determine the level of conﬁ dence in the assessment based on the
 relevance and reliability of the information.
1. Introduction  emergency situations where rapid understanding of potential tox-
 icological consequences from exposure is needed in the absence of
 In silico toxicology (IST) methods are computational approaches that existing toxicological testing data;
analyze, simulate, visualize, or predict the toxicity of chemicals. IST  cases where there is only a limited supply of a test material avail-
encompasses all methodologies for analyzing chemical and biological able;
properties generally based upon a chemical structure that represents  scenarios where there are challenges to conduct laboratory studies;
either an actual or a proposed (i.e., virtual) chemical. Today, in silico  instances where synthesis of a complex test material is not feasible;
approaches are often used in combination with other toxicity tests; and
however, the approaches are starting to be used to generate toxicity  situations where a less time-consuming and less expensive high-
assessments information with less need to perform any in vitro or in vivo throughput approach than an experimental test is needed.
studies depending on the decision context. IST uses models which can
be encoded within software tools to predict the potential toxicity of a IST methods are one approach to generating additional information
chemical and in some situations to quantitatively predict the toxic dose for complementing and ultimately enhancing the reliability or sup-
or potency. These models are based on experimental data, structure- porting a risk assessment, including an understanding of the structural
activity relationships, and scientiﬁ c knowledge (such as structural and/or mechanistic basis that may contribute ideas for the rational
alerts reported in the literature). design of new chemicals, development of a testing strategy or an overall
 There are a number of diﬀ erent situations where in silico methods weight-of-evidence evaluation. IST inherently supports the principle of
serve an important role in the hazard assessment of existing chemicals the 3Rs (replacement, reﬁ nement and reduction) relating to the use of
or new substances under development that would beneﬁ t from the animals in research (Russell and Burch, 1959; Ford, 2016). Table 1
development of in silico toxicology protocols. These include: outlines ﬁ fteen speciﬁ c uses of IST to illustrate the diversity of
 2

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Table 1
Applications of in silico toxicology.
 In silico toxicology application Discussion
 1. Alternative to test data. The use of non-animal alternative methods including in silico approaches, may substitute for other types of
 tests in regulatory submissions in certain cases. Acceptable alternative methods for ﬁ lling data gaps are
 outlined in Annex XI of the European Union's REACH regulation (EU, 2006). In the United States, Frank R.
 Lautenberg Chemical Safety for the 21st Century Act revised the Toxic Substances Control Act (TSCA) to
 include predictive models and expert review as part of an overall assessment (TSCA, 2016). The United States
 Food and Drug Administration (US FDA) Center for Devices and Radiological Health (CDRH) issued a guidance
 for industry and FDA staﬀ . This guidance is on the use of International Standard ISO 10993–1 for biological
 evaluation of medical devices and indicates in the absence of experimentally derived carcinogenicity
 information, structure activity relationship modeling for these materials may be needed (CDRH, 2016). The
 FDA draft guidance on Electronic Nicotine Delivery Devices (ENDS) also discusses the use of computational
 toxicology models in the absence of toxicological data for potential toxicants created by the aerosolization
 process (PMTA/FDA, 2016). When chemicals with limited toxicity data are required to be classiﬁ ed and
 labeled for shipping or other purposes, in silico toxicology provides an alternative method for quickly ﬁ lling
 the data gaps in the toxicity/safety information, such as predictions of acute toxicity to support assignment to
 the Globally Harmonized System of Classiﬁ cation and Labelling category (Freidig et al., 2007; ECHA, 2015).
 2. As part of the weight-of-evidence in regulatory submissions. There are currently several regulatory frameworks where only speciﬁ c laboratory tests for an endpoint of concern
 may be submitted (such as for drugs or food additives). However, in such cases,in silico predictions can be submitted
 alongside standard toxicological data to complement the assessment. This may includein silico assessments provided
 as supporting data or adjuncts to the primaryin vivo or in vitro studies to give a mechanistic understanding of the
 observed results and/or allow a better deﬁ nition of experimental needs. Additionally,in silico methods may be used
 to guide or prioritize in vitro testing (EU, 2012). The European Union's Cosmetics Regulation ( EU, 2009a)prohibits
 the use of animal testing for products or ingredients and a complete marketing ban of such products tested as a whole
 or containing tested ingredients. This requires the use ofalternative methods, such as IST, in the assessment of new
 cosmetics ingredients. In a recent memorandum, the European Commission's Scientiﬁ c Committee for Consumer
 Safety (SCCS), which is responsible for the risk assessmentof cosmetic ingredients, acknowledged the importance and
 limitations of in silico methods; the SCCS recommended thatin silico methods be used either for internal decision
 making or as part of a weight-of-evidence (WOE) approach to estimate toxicity risks before embarking on any
 experimental testing (SCCS, 2016).
 3. Mixtures assessment. Most exposures are not to a single chemical but rather to complex mixtures of chemicals that may be found in food,
 beverages, the environment, cigarette smoke, electronic nicotine delivery systems (ENDS)aerosols,botanicaldrugs
 or natural products. In certain situations, it may be possible to usein silico methods to assess individual components
 since today's in silico analysis can only be performed on discrete identiﬁ able chemicals. While preliminary analytical
 work is required to identify all chemicals in the mixture above appropriate Analytical Evaluation Thresholds (AET)
 (Ball and Norwood, 2012), leveraging in silico approaches may avoid having to synthesize or purify each of the
 potentially large number of mixture components to perform standard toxicological tests (Mumtaz et al., 2010).
 Careful consideration is required for mixtures when there aremultiplechemicalsforinter actions, such as synergistic
 or additive eﬀ ects that may have the same, similar or diﬀ erent mechanisms of action (MOA).
 4. Assessment of impurities and degradation products. Chemicals, such as pharmaceuticals or plant protection products, may contain low levels of impurities produced
 during manufacturing and degradation. Many such substances, when presentat levels above accepted thresholds,
 need to be assessed. In most cases, mutagenicity evaluation of the impurity under question is required as a ﬁ rst step
 of the risk assessment. (Harvey et al., 2017) The ICH M7 guideline provides speciﬁ c recommendations for assessing
 drug impurities (ICH M7, 2017(R1)), including the use of two complementary computational toxicology
 methodologies (i.e., statistical-based and expert rule-based models) to predict bacterial mutagenicity.
 5. Residues of plant protection products. Residues of plant protection products may be evaluated as a part of residue deﬁ nition for dietary risk
 assessment of plant protection products (EU, 2009b). In this context, in silico methods provide a useful
 alternative approach. (EFSA, 2016)
 6. Assessment of extractables and leachables. Medical devices, such as inhaled aerosols, food-contact substances, and consumer product packaging materials
 may pose a risk for human health due to release of potentially harmful chemicals that are used in the
 production of the components (Bossuyt et al., 2017). These include plasticizers, copolymers, vulcanization
 additives, etc. for which toxicological data is often lacking but where a risk assessment must be performed. A
 migration or leachables study supports the discovery, identiﬁ cation, and quantiﬁ cation of any leachables. An
 in silico toxicological assessment, in certain situations, can provide suﬃ cient data for the risk assessment.
 7. Workers' safety and occupational health. Chemicals used in the manufacture of a product are assessed for mutagenicity, carcinogenicity, skin and respiratory
 sensitization, irritation (skin, eye and respiratory), and reproductive and developmental toxicity and possibly acute
 toxicity. In silico assessments make it possible to estimate the potential toxicity of chemicals and adopt proper
 engineering controls and personal protective equipment usage to protect workers who could be exposed to these
 substances during production, transfer, storage, and delivery processes (EU, 2006). In silico approaches have been
 utilized to assess these major toxicological endpoints in the occupational safety setting.In silico methods to predict
 respiratory sensitization potential of industrial chemicals have recently been reviewed bySeed and Agius (2017).
 8. Metabolite analysis. Metabolites can present an increased or decreased risk of local or systemic toxicity compared with the parent
 chemical (Mumtaz and Durkin, 1992). While reactive or toxic metabolites may be formed by an organism,
 their identiﬁ cation, separation as well as possible synthesis for testing purposes may be challenging. In silico
 methods provide a practical alternative approach to understanding the safety proﬁ les of this potentially large
 number of chemicals as well as to support the prediction of metabolites.
 (continued on next page)
 3

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Table 1 (continued)
 In silico toxicology application Discussion
 9. Ecotoxicology. Various chemicals are discharged into the environment that may cause harm. Furthermore, the parent compounds
 can be transformed by hydrolysis, redox-reactions, or photolysis into numerous additional chemicals. IST methods
 often provide the most practical approach to assess the potential eﬀ ects on the environment and wildlife species of
 the many chemicals that are discharged. Prediction of physicochemical parameters supports assessment of potential
 environment exposure to the chemical (e.g., persistence and distribution). As an example,Chen et al., 2015
 describes the use of in silico assessment of potentially hazardous contaminants present in water.
 10. Green chemistry and safer alternatives. In silico methods can play an important role when identifying alternative chemicals that may have a safer
 proﬁ le than existing chemicals (Rastogi et al., 2014). This includes, for example, alternatives for use in
 manufacturing processes, alternative packaging/delivery materials and the use of speciﬁ c additives. In silico
 methods can provide insights about structural features responsible for the toxicity of diﬀ erent groups of
 chemicals and thereby allow for the rational design of intrinsically safer chemicals.
 11. Selection of product development candidates. In early product discovery or development, many thousands of compounds may be evaluated. In silico methods may
 provide a helpful approach to selecting candidates, since in silico methods are inexpensive, rapid to perform, and
 high throughput. In addition,in silico methods can suggest which molecular substructures (toxicophores) are
 responsible for the predicted toxic activity, thereby supporting the optimization of future compounds (Hillisch et al.,
 2015; Myattetal.,2016 ). Later in the product development process, a smaller number of chemicals may be selected
 as candidates to take forward for further development; in normal situations, preference would be given to the
 candidate(s) with the most advantageous safety proﬁ le(s) (Myattetal.,2016 ).
 12. Emergency response situations. When one or more chemicals are unexpectedly released into the environment (e.g., the West Virginia chemical spill
 (NTP, 2016)) or into a production process, it is important to quickly evaluate the potential eﬀ ects on humans,
 wildlife, and the environment. In such emergency situations the toxicological proﬁ le of the released chemicals needs
 to be established as quickly as possible to support the proper emergency response and to protect emergency services
 staﬀ and bystanders (Hochstein et al., 2008; Schilter et al., 2014). In such a limited timeframe and in the absence of
 previously generated data, in silico approaches may be a practical option for rapid hazard identiﬁ cation.
 13. Prioritizing testing of chemicals. In silico approaches can help prioritize in vitro and in vivo toxicology testing, based upon the chemical's
 exposure and prediction of toxicity; they are an important aspect of the work at several organizations such as
 the US EPA, National Toxicology Program, Environment and Climate Change Canada and ECHA (Schwetz,
 1995). In silico methods may be used to prioritize (based on potential toxicological liabilities) the order in
 which a series of toxicological studies will be performed (Myatt et al., 2016).
 14. Rationalization of in vivo or in vitro study results. As mentioned previously in the description of the in silico application titled “As part of the weight-of-evidence in
 regulatory studies”, results from quantitative structure-activity relationship (QSAR) models (toxicophore
 information, chemical fragments or physicochemical properties) may be used in conjunction with biological
 data to infer a mechanism of action (MOA), molecular initiating event (MIE), or mode of toxicity as part of an
 adverse outcome pathway (AOP) (Martin et al., 2015; Ellison et al., 2016). Information from in silico methods
 can also be used to tailor an in vivo study, e.g., by inclusion of additional endpoints. When existing
 experimental data on a compound are equivocal or when not all relevant safety information are available or
 accessible, in silico data may be used as additional information as part of the weight-of-evidence approach in
 reaching a more informed decision (Kruhlak et al., 2012).
applications that currently can beneﬁ t from in silico methods. Stanton 2007). Other initiatives include the North American Free Trade
and Kruszewski (2016) quantiﬁ ed the beneﬁ ts of using in silico and Agreement pesticides Quantitative Structure-Activity Relationship
read-across methods where they determined that the approach used (QSAR) guidance (NAFTA, 2012), considerations on the use of in silico
across two voluntary high-production-volume (HPV) chemical pro- approaches for assessing cosmetics ingredients (Amaral et al., 2014),
grams for 261 chemicals obviated the use of 100,000–150,000 test European Food Safety Agency report (EFSA, 2014), European Chemi-
animals and saved 50,000,000 US$ to 70,000,000 US$. cals Agency REACH supporting documentation (ECHA, 2008; ECHA,
 The increased interest and acceptance of in silico methods for reg- 2016; 2017b), Organization for Economic Co-operation and Develop-
ulatory data submission and chemicals evaluation is driving the adop- ment (OECD) documentation (OECD, 2007; OECD, 2014; OECD, 2015),
tion of its use for regulatory purposes. Several guidance documents and the ICH M7 guideline (previously mentioned) along with com-
have been drafted to improve standardization, harmonization, and plementary peer reviewed publications outlining the process for im-
uptake of in silico methods by regulatory authorities including the plementation of such computational assessments (e.g., Amberg et al.,
International Council for Harmonization (ICH) M7 guideline (assess- 2016; Barber et al., 2015; Powley, 2015; Schilter et al., 2014). Certain
ment and control of DNA reactive (mutagenic) impurities in pharma- projects have provided substantial guidance on the documentation of
ceuticals to limit potential carcinogenic risk) (ICH M7, 2017(R1)), the the models and prediction results (JRC, 2014; Patlewicz et al., 2016)as
European Union's Registration, Evaluation, Authorization, and restric- well as principles and workﬂ ows to support safety assessments (Bassan
tion of Chemicals (REACH) regulation (EU, 2006; ECHA, 2008; ECHA, and Worth, 2008; ECHA, 2015; Worth et al., 2014; Berggren et al.,
2015), European Food Safety Authority (EFSA) residue guidance (EFSA, 2017; Amaral et al., 2017).
2016), Canada's Chemicals Management Plan (CMP) assessments for These prior initiatives provide a robust foundation for the current
new and existing substances under the Canadian Environmental Pro- project to establish the IST protocols described here; however, several
tection Act, 1999 (CEPA 1999) (Canada, 2016), and the Toxic Sub- issues have hindered the general acceptance and use of in silico methods
stances Control Act (TSCA) (TSCA, 2016). A number of national and on a larger scale. In particular, there remains a lack of generally ac-
international initiatives have focused on developing speciﬁ c documents cepted procedures for performing in silico assessments for the tox-
supporting the use of in silico tools. The OECD has published a series of icological endpoints. The lack of such procedures or protocols has led to
(Quantitative) Structure-Activity Relationship (Q)SAR validation prin- inconsistency in the application and use of in silico tools across diﬀ erent
ciples that are discussed in detail in Section 2.3.2 (OECD, 2004; OECD, organizations, industries, and regulatory agencies (e.g., searching
 4

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
databases, applying predictive models and alerts, performing an expert 2. In silico toxicology protocols
review/assessment, documenting and communicating the results and
associated uncertainties). The use of traditional experimental evidence 2.1. Overview
coupled with in silico information to support hazard identiﬁ cation and
risk assessment also varies both across, and often within, organizations. Each IST protocol describes the prediction process in a consistent,
Although not always, such ad hoc approaches may be time-consuming transparent, and well-documented manner. This includes re-
and the results poorly accepted. Standardization of protocols will en- commendations on how to:
hance the acceptability of the methods and their results by end users.
Additionally, there are misconceptions about when in silico predictions 1) plan the in silico analyses including identifying what toxicological
are appropriate to use as well as a lack of deﬁ ned consensus processes eﬀ ects or mechanisms to predict (Section 2.2), what in silico meth-
for interpreting the result(s) of such predictions (Bower et al., 2017; odologies to use (Section 2.3.1), and other selection criteria for the
SCCS, 2016). Some scientists view in silico methods as a “black box”in silico methods (Section 2.3.2),
that inhibits their ability to critically assess the predictions and their 2) conduct the appropriate individual software predictions (Section
reliability (Alves et al., 2016). Others lack expertise to interpret the 2.3.3) and further database searches (Section 2.5),
results of in silico predictions, and some have an unrealistic expectation 3) perform and document the in silico analysis (Sections 2.6 and 2.7)
that an in silico prediction can always provide an unerring deﬁ nitive including expert review (Section 2.4), and
assessment. 4) report and share the information and assessment results, including
 Standardization of in silico tool use and interpretation of results information about uncertainties (Section 2.9).
would greatly reduce the burden on both industry and regulators to
provide conﬁ dence in or justiﬁ cation for the use of these approaches. Section 2.8 provides a template for the individual IST protocols for
The objective of developing IST protocols is to deﬁ ne in silico assess-major toxicological endpoints. IST protocols could be applicable for use
ment principles so the results can be generated, recorded, commu- with several in silico programs, including diﬀ erent in silico models and
nicated, archived and then evaluated in a uniform, consistent and re- databases.
producible manner. Incorporating these principles routinely into the
use of in silico methods will support a more transparent analysis of the 2.2. Toxicological eﬀ ects and mechanisms
results and serves to mitigate “black box” concerns.1 This approach is
similar to guideline studies that provide a framework for the proper In an experimental approach, hazard is evaluated based on speciﬁ c
conduct of toxicological studies and assurance in the validity of the observations (toxicological eﬀ ects) during toxicity studies. Often, toxi-
results (such as OECD Guidelines for the Testing of Chemicals) (OECD, city of a chemical involves a biological event: a non-speciﬁ c or speciﬁ c
2017). The development of these protocols is driven by consensus interaction with a vital biological structure, which causes sequential
amongst leading scientists representing industry, private sector and perturbation of a physiological pathway at a cellular, tissue, organ and/
governmental agencies. Consequently, this project provides an im- or system level, leading to a toxicological eﬀ ect observed at the or-
portant step towards a quality-driven science for IST or good in silico ganism level. Experiments evaluating the potential of a chemical to
practice. cause such a biological event (e.g., in vitro analysis of speciﬁ c interac-
 Herein, we provide a framework to develop a series of procedures tion with a cellular receptor or inhibition of an enzyme or non-speciﬁ c
for performing an in silico assessment to foster greater acceptance. cytotoxicity), may support hazard assessment and provide information
These IST protocols are being created for a number of toxicological about the mechanism of toxicity. Such an approach is utilized in the
endpoints (e.g., genetic toxicity, carcinogenicity, acute toxicity, re- Adverse Outcome Pathway (AOP), where identiﬁ cation of a molecular
productive toxicity, developmental toxicity) as well as other related initiating event supports assessment of the related adverse outcome at
properties (e.g., biodegradation and bioaccumulation) that could im- the organism level (Bell et al., 2016; OECD, 2016a; OECD, 2016b). A
pact the chemical hazard classiﬁ cation. Throughout this publication, computational approach to hazard assessment may address the two
these toxicological and related endpoints are referred to as “major complementary levels of hazard identiﬁ cation in a similar way (i.e.,
endpoints” and the protocols are referred to as IST protocols. These predicting the resulting manifestation (eﬀ ect) or the molecular per-
protocols will support the assessment of hazards and in some cases the turbation (mechanism) that led to the toxicological eﬀ ect).
prediction of quantitative values, such as a No Observed Adverse Eﬀ ect Each IST protocol deﬁ nes a series of known toxicological eﬀ ects and
Levels (NOAELs); however, these protocols do not deﬁ ne how a risk mechanisms relevant to the assessment of the major toxicological
assessment will be performed. This publication outlines the components endpoint. For example, in the reproductive toxicity IST protocol, the list
of an IST protocol, including schematics to describe how a prediction of toxicological eﬀ ects/mechanisms may include reduced sperm count,
could be performed, approaches to assess the reliability and conﬁ dence androgen signaling disruption in vitro, and so on. Within each IST
of the results, and items that may be considered as part of an expert protocol, these eﬀ ects/mechanisms may be species and/or route of
review. This publication also outlines the process for creating the IST administration speciﬁ c.
protocols through an international consortium comprising re- Fig. 1 outlines a general approach to performing an in silico as-
presentatives across regulatory agencies, government research agen- sessment. For each toxicological eﬀ ect/mechanism, relevant informa-
cies, diﬀ erent industrial sectors, academia and other stakeholders. tion (as deﬁ ned in the IST protocol) is collected, including any available
Speciﬁ c endpoint-dependent considerations will be described in future experimental data as well as in silico predictions. The experimental data
separate publications and IST protocols (developed as a result of this and/or in silico results are then analyzed and an overall assessment of
process) will also be published for widespread use and for incorporation the toxicological eﬀ ect or mechanism is generated alongside a relia-
into diﬀ erent technology platforms. bility score (deﬁ ned in Section 2.6.2) that reﬂ ects the quality of the
 results. The assessment results and reliability scores for a range of re-
 levant toxicological eﬀ ects/mechanisms are then used to support a
 hazard assessment within the hazard assessment framework.
 1 It should be noted that black box models may be acceptable in certain situations, such 2.3. In silico predictions
as compound ﬁ ltering and virtual screening, as long as they show acceptable performance 2.3.1. In silico methodologies
in validation studies; however, for most applications the acceptance of this class of models
is low. Several organizations develop and make available computer
 5

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Fig. 1. Overview of the IST protocol framework, showing how experimental data or in silico model(s) for each deﬁ ned toxicological eﬀ ect/mechanism are assessed
and used to support a hazard assessment. (Note Eﬀ ect/Mechanism N is used to illustrate that there can be any number of eﬀ ects/mechanisms in each protocol).
software packages for predicting toxicity or physicochemical properties coeﬃ cient [log P]), electronic and topological descriptors (e.g.,
of query chemical(s). These systems generally contain one or more quantum mechanics calculations), or chemical structure-based de-
models, where each model predicts the compound's putative tox- scriptors (e.g., the presence or absence of diﬀ erent functional
icological eﬀ ect or mechanism of action. For example, a model may groups) are generated and used to describe the training set com-
predict the results for bacterial gene mutation using data generated pounds. The model encodes the relationship between these de-
from the bacterial reverse mutation test or Ames test. These models may scriptors and the (toxicological) response. After the model is built
be revised over time as more data become available, structure-activity and validated (OECD, 2007; Myatt et al., 2016), it can be used to
relationships are better characterized, and any data set used is updated. make a prediction. The (physico)chemical descriptors incorporated
Each new or updated model is given a diﬀ erent version number because into the model are then generated for the test compound and are
the results from diﬀ erent model versions may vary and it is important to used by the model to generate a prediction. This prediction is only
track the source of the results (Amberg et al., 2016). accepted when the test compound is suﬃ ciently similar to the
 All IST protocols will identify the toxicological eﬀ ects or mechan- training set compounds (i.e., it is considered within the applicability
isms to be predicted as discussed in Section 2.2. These predictions may domain of the QSAR model, often considering the signiﬁ cance of
be dichotomous (e.g., predict mutagenic or non-mutagenic com- descriptors) (Netzeva et al., 2005; Carrió et al., 2014; Patlewicz
pounds), quantal (e.g., Globally Harmonized System [GHS] Classiﬁ ca- et al., 2016). This applicability domain analysis may be performed
tion and Labeling2 scheme) or quantitative/continuous (e.g., prediction automatically by some software to determine whether the training
of median toxic dose [TD50] values). The speciﬁ c IST protocols will set compounds share similar chemical and/or biological properties
detail the type of prediction(s) ideally generated. with the test chemical.
 The major in silico prediction methodologies include the following:  Expert rule-based (or expert/structural alerts). This metho-
 dology uses structural rules or alerts to make predictions for speciﬁ c
  Statistical-based (or QSAR). This methodology uses a mathema- toxicological eﬀ ects or mechanisms of toxicity. These rules are de-
 tical model that was derived from a training set of example che- rived from the literature or from an analysis of data sets generated
 micals. The training set includes the chemicals that were found to be by scientists. Structural alerts are deﬁ ned as molecular substructures
 positive and negative in a given toxicological study (e.g., the bac- that can activate the toxicological eﬀ ect or mechanism. The rules
 terial reverse mutation assay) or to induce a continuous response may also encode situations where the alert is deactivated. Expert
 (e.g., NOAEL in teratogenicity) that the model will predict. As part rule-based models often include a description of the toxic me-
 of the process to generate the model, physicochemical property- chanism and examples from the literature or other reference sources
 based descriptors (e.g., molecular weight, octanol water partition to justify the structural alert. A positive prediction is generally made
 when a structural alert is present (without deactivating structural
 features or properties) in the test compound. When no alerts are
 2 A chemical is assigned to a category (e.g., 1, 2, 3, 4, or 5) based on distinct ranges of triggered for a test chemical, a negative prediction may be generated
quantitative values (e.g., LD50). Examples of such ranges include LD50 < 5 mg/kg (i.e., for well investigated endpoints; however, additional analysis is
category 1) or 50–300 mg/kg (i.e., category 3). generally required to make this assessment as discussed further in
 6

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
 Section 2.4.3. 3. Chemical space. Often, in silico models will only make predictions
  Read-across: Read-across uses data on one or more analogs (the for speciﬁ c classes of chemicals, the so called “applicability do-
 “source”) to make a prediction about a query compound or com- main”. The chosen in silico model(s) may report the applicability
 pounds (the “target”). Source compounds are identiﬁ ed that have a domain assessment to demonstrate its proﬁ ciency for this class of
 structurally or toxicologically meaningful relationship to the target compounds. Vice versa, only models are ideally chosen where the
 compound, often underpinned by an understanding of a plausible query compound is in the applicability domain (Netzeva et al., 2005;
 biological mechanism shared between the source and target com- Carrió et al., 2014; Patlewicz et al., 2016).
 pounds. The toxicological experimental data from these source 4. Model combinations. Complementary or independent in silico
 compounds can then be used to “read-across” to the speciﬁ c target models may be selected, as concurring results increase the reliability
 compound(s). Read-across is an intellectually-derived endpoint- of the prediction (as discussed in Section 2.6.2).
 speciﬁ c method that provides justiﬁ cation for why a chemical is 5. Supporting an expert review. For QSAR models, tools to help the
 similar to another chemical (with respect to chemical reactivity, expert review (see Section 2.4) include the ability to allow ex-
 toxicokinetics, mechanism/mode of action, structure, physico- amination of the descriptors and weightings used in the model,
 chemical properties, and metabolic proﬁ le) (Wu et al., 2010; underlying training set data, and how the applicability domain as-
 ECETOC 2012; Patlewicz et al., 2013a; b; OECD 2014; Blackburn sessment was deﬁ ned. For expert rule-based systems, this could in-
 and Stuard, 2014; Patlewicz (2014); Patlewicz et al., 2015; Schultz clude how the alert was deﬁ ned (including any factors that activate
 et al., 2015; Ball et al., 2016; ECHA 2017b). or deactivate the alert), any mechanistic understanding associated
  Other approaches: In certain cases, other in silico methodologies with the alert, citations, and any relevant known examples of
 may be appropriate. Examples include the use of molecular dy- alerting chemicals.
 namics (e.g., simulating interactions of a query chemical with a
 metabolic enzyme) and receptor binding as an indication of a pos- Read across may be used when there are experimental data from
 sible Molecular Initiating Event (e.g., estrogen receptor-ligand high quality databases for one or more substances which are similar
 docking). enough to the target chemical of interest. The Read-Across Assessment
 Framework (RAAF), or similar published and established frameworks,
 Each IST protocol will include an assessment of key computational may be used to document the read-across assessment and to support its
aspects and speciﬁ c issues to consider. For example, when performing scientiﬁ c plausibility (ECHA, 2017b; Patlewicz et al., 2013b; Blackburn
read-across, issues such as the data quality of the source compound(s), and Stuard, 2014; Schultz et al., 2015; Patlewicz et al., 2015). The
how to perform an assessment of non-reactive chemical features and OECD has also produced guidance on the process of grouping chemicals
selection of grouping approaches used to form categories will be dis- and other considerations as part of a read-across assessment (OECD,
cussed to ensure source compound(s) are suﬃ ciently similar, both 2014), and ECHA has generated guidelines on the process of performing
chemically and biologically, for the endpoint being considered. a valid read-across assessment (ECHA, 2008).
 Each methodology has its strengths and weaknesses, which often
depend on the type of toxicological eﬀ ect or mechanism being pre-
dicted. This will be discussed in the individual IST protocols. In addi- 2.3.3. Running the in silico models
tion, there may be cases of unique or novel compounds for which it is All in silico systems require an electronic representation of the
not possible to make a prediction or for which conﬁ dence in the pre- chemical structure and any errors in this representation will result in
dictions is so low as to render it meaningless or unhelpful. invalid predictions. Therefore, it is important to ensure that the che-
 mical structure is properly curated and entered following conventions
2.3.2. In silico methods selection criteria set out by the model's developer, including appropriate representations
 In silico methods selection may include the following ﬁ ve con- for tautomers, aromaticity, salt forms, stereochemistry, charges, and
siderations: speciﬁ c functional groups (e.g., nitro or carboxylic acid groups). It is
 possible that diﬀ erent formats (i.e., SMILES vs. MOL ﬁ les) may be
 1. Relevant toxicological eﬀ ects or mechanisms. As discussed in processed diﬀ erently. It is also important to verify that the software
 Section 2.2, each IST protocol will deﬁ ne a series of toxicological correctly interprets the structural representation during processing,
 eﬀ ects or mechanisms relevant to a speciﬁ c endpoint and appro- particularly for complex molecules. For some types of chemicals, in si-
 priate in silico models need to be selected that predict these speciﬁ c lico models may not be applicable due to the structural representation
 eﬀ ects or mechanisms. or the unsuitability of the experiment assay for the speciﬁ c chemical
 2. Model validity. Best practices for validation of (Q)SAR in silico class. Some in silico models cannot distinguish cis- and trans-isomers.
 models have been documented in a number of publicationsExamples include non-discrete chemical substances, UVCBs (unknown/
 (Cherkasov et al., 2014; Raies and Bajic, 2016; Myatt et al., 2016),variable composition, complex reaction products and biologicals), me-
 and models built using these best practices may be preferred. The tals, inorganics, polymers, mixtures, organometallics and nano-mate-
 OECD has published a series of validation principles for in silico rials (Mansouri et al., 2016).
 models (OECD, 2004; OECD, 2007) and valid statistical-based or Some models, such as statistical-based models, allow for prediction
 expert rule-based in silico methods. Such (Q)SAR methods have: 1) a settings to be adjusted or turned oﬀ (e.g., they report “positive” when a
 deﬁ ned endpoint; 2) an unambiguous algorithm; 3) a deﬁ ned do- value is greater than a predetermined threshold). The settings are ide-
 main of applicability; 4) appropriate measures of goodness-of-ﬁ t, ally selected in a way that does not compromise the model's validity
 robustness and predictivity; and 5) a mechanistic interpretation, if (such as changing the validation statistics of the model) and appro-
 possible. Any in silico model must include documentation that sup- priately reported.
 ports an assessment of the model's scientiﬁ c validity, including the A thorough documentation of all selected models and computer
 toxicological eﬀ ect or mechanism being predicted, version number, software packages including, version numbers, and any parameters set,
 type of methodology, training set size and content, as well as any is needed as part of the materials and methods in suﬃ cient detail to
 predictive performance information. Validation performance is assess and potentially repeat the analysis (discussed in Section 2.9). In
 documented in report formats such as the QSAR Model Reporting addition, the results need to be presented in enough detail to fully
 Format (QMRF) (JRC, 2014). The level of adherence to the OECD understand how they were generated and to critically assess the ﬁ nd-
 principles and the performance statistics need to be appropriate for ings.
 the purpose of the assessment.
 7

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
2.4. In silico expert review Tables 2 and 3). This review may include knowledge from proprietary
 information available within an organization from the testing of related
2.4.1. Overview chemicals.
 As with in vitro or in vivo study data, in silico predictions may be When an expert review assesses multiple predictions from diﬀ erent
critically assessed and an expert review of the output is often prudent in silico systems, it is important to justify how they complement each
(Dobo et al., 2012; Sutter et al., 2013). Frameworks for conducting an other with regard to the training set (i.e., the use of relevant guideline
expert review ensure that it is performed in a consistent and transparent studies plus relevant chemical classes), methodology (e.g., expert rule-
manner. Examples of such a review framework include the Oﬃ ce of based vs. statistical-based vs. read-across), or QSAR descriptor sets.
Health Assessment and Translation (OHAT) systematic review and It is essential to document the reasoning and decisions of the expert
evidence integration (Rooney et al., 2014), weight-of-evidence assess- review steps so they can be retraced at any time, including the in-
ments (ECHA, 2017a), and Integrated Approaches to Testing and As- formation used as the basis for the review.
sessment (IATA) (OECD, 2016a; OECD, 2016b).
 The purpose of an in silico expert review is to evaluate the reliability 2.4.2. Expert review of statistical models
of the prediction. The outcome of the review provides information to An expert review of a statistical-based model involves a critical
include in the assessment of the toxicological eﬀ ect or mechanism. As assessment of how the model generated the prediction. This includes
part of this review, the expert might agree with, or refute, individual in examining the weightings of the model descriptors (e.g., structural
silico predictions. In addition, these reviews might support cases when a features or physicochemical properties related to toxicity), underlying
chemical is out of the applicability domain of the model, support the data, chemical space of the training set of the model, and the experi-
use of an equivocal prediction (i.e., there is evidence both for and mental results for analog compounds and model performance for these
against the supposition), or support cases where multiple predictions do analogs (e.g., nearest-neighbor list of compounds) (Amberg et al.,
not agree. A checklist of items to consider and report will help to ensure 2016). This may also incorporate an understanding of the mechanism of
such reviews are performed in a consistent manner (as illustrated in toxicity or knowledge of factors that activate or deactivate the toxicity.
Table 2
Checklist of elements to consider as part of an expert review of a QSAR model result.
 Expert review elements Considerations
 A. Inspection of model output  A review of the applicability domain information provided by the model's software
 might increase or decrease reliability in the prediction.
  The results of the QSAR model might include a score (e.g., a probability of a positiveoutcome). The prediction reliability may be increased where a score indicating a high
 likelihood can be justiﬁ ed through an expert review of the available information.
 B. Analysis of structural descriptors and corresponding training set data (see Note A)  As part of the process of building a QSAR model, structural descriptors are selected (often
 automatically) when there is a statistical association to the (toxicological) data to be
 predicted; however, the selected descriptorsmight not be biologically meaningful for the
 predicted toxicological eﬀ ect/mechanism, as discussed inPowley (2015). This assessment
 may be supported by inspecting the training set examples that match the descriptors
 wherever possible. An expert review may determine the result is incorrect if other structural
 moieties in the training set examples are more likely responsible for the biological activity,
 (i.e., the descriptors identiﬁ ed were coincidental and in fact irrelevant) (Amberg et al., 2016).
  Another scenario is when the structural descriptors map to experimental data that isincorrect and attributable to known problems with an assay. Again, these features may be
 discounted if they are not relevant to the toxicological eﬀ ect or mechanism and this may
 lead to a reversal of the overall assessment. For example, chemicals containing acid
 halides may give false positive results due to possible interaction with the solvent DMSO
 in the Ames assay (Amberg et al., 2015).
  Descriptors identiﬁ ed as signiﬁ cant by the model that are also present in the query compoundmay be associated with a biological mechanism. An expert review may evaluate whether the
 mechanism is plausible for the query compound, including potential metabolism
 consideration. For example, does the highlighted feature represent a known reactive group or
 a known toxicophore? This analysis may lead to an increase in prediction reliability.
  In some systems, it is possible to inspect the training set's experimental data andreferences for those examples that are primarily used in the prediction. An assessment of
 these full studies for these examples (as discussed in Section 2.5) could be used to justify
 an increase in the reliability of the prediction result.
  The structural diversity of the underlying chemicals for each signiﬁ cant descriptor may bereviewed as part of an expert review. Structural features that map to a large number of
 structurally diverse compounds would provide additional evidence that the toxicological
 eﬀ ects or mechanisms associated with the descriptor could be extrapolated across
 diﬀ erent chemical classes (increasing reliability in the prediction), whereas a structural
 feature whose underlying data constitutes a congeneric series might not, especially if the
 query compound is structurally distant (decreasing reliability in the prediction).
 C. Analysis of physicochemical descriptors used by model (see Note B)  Is there any supporting information from the literature or elsewhere to support any
 correlation between the physicochemical properties identiﬁ ed as signiﬁ cant by the
 model and the toxicological eﬀ ect/mechanism?
  An evaluation of the quality of the experimental data of the training set chemicals used forbuilding of the model (e.g., if a guideline study was used to generate these data) may
 increase the reliability of the prediction result.
 D. Assessment of other information  An evaluation of the performance of the model for structurally similar substances with
 known activity (selected by the user or provided by the system) might aﬀ ect the
 evaluation of the reliability of the prediction.
(Note A: items to consider when the QSAR model includes structure-based descriptions; Note B: items to consider when the QSAR model includes physicochemical
descriptors).
 8

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Table 3
Checklist of elements to consider as part of an expert review of results from expert rule-based.
 Expert 4review elements Considerations
 A. Alert score or qualitative output  The results from the alert system might include information related to the likelihood of a positive outcome (e.g.,
 precision of the alert). The reliability of the prediction may be increased when such a score can be justiﬁ ed through
 an expert review of the information provided.
 B. Justiﬁ cation of negative prediction  Additional considerations may be important where no alerts are identiﬁ ed in the test chemical. Such analysis may
 focus on similar analogs as well as other chemicals containing the diﬀ erent structural elements of the test chemical
 to verify there is no potential toxicity attributable to these fragments, such as additional reactive features. Such
 analysis may be used to evaluate the reliability of the negative prediction.
  If a negative prediction has a structure of concern, a further inspection of the rules may determine why the compoundwas not included to elucidate the underlying cause for ﬁ ring no alert. Is the prediction really negative, equivocal, or
 not in of the applicability domain of the model?
 C. Reliability of the mechanism of toxicity  Although the presence of a structural alert increases the potential of the chemical to exert a toxicological eﬀ ect or
 mechanism, this eﬀ ect may depend on other features of the molecule. If a mechanism of toxicity is proposed for the
 structural alert, then an expert may assess the plausibility of the mechanism for the query compound. For example,
 the presence of other substituents in the molecule may impact the activity, potentially deactivating the alerting
 structure. This may include metabolism considerations.
 D. Inspection of chemicals and experimental data
 matching the alert  The reliability of the prediction can be assessed by the quality of the experimental data of the reference setsubstances used to make the prediction (e.g., if a guideline study to generate these data).
  The structural diversity of the matching chemical may also be considered. For example, alerts that match diversestructures may increase the reliability over alerts where the matching chemicals are from a tight congeneric series.
 This is especially true when the reference set examples are structurally dissimilar from the query chemical.
  Review of the scientiﬁ c literature to support the alert to understand the strengths and limitations of the experimental
 data supporting it.
The items described in Table 2 provide a checklist of elements to con- study types and speciﬁ c result(s) corresponding to the identiﬁ ed tox-
sider as part of any QSAR expert review to ensure such a review is as icological eﬀ ects or mechanisms, as discussed in Section 2.2. To illus-
objective as possible, transparent and based on a consistent set of trate, in the assessment of the toxicological eﬀ ect/mechanism bacterial
considerations. An expert review may increase the reliability of statis- gene mutation (part of the genetic toxicity IST protocol), the overall
tical model results based on one or more elements deﬁ ned in Table 2. mutagenic or non-mutagenic results from a bacterial reverse mutation
 Individual IST protocols will outline speciﬁ c points to consider assay may be used. A more complex example is in the assessment of the
when performing an expert review, such as how the similarity of ana- toxicological eﬀ ect/mechanism of sperm morphology (part of the re-
logs could be assessed. productive IST protocol). Here, speciﬁ c results from potentially dif-
 ferent study types, such as one- or two-generation reproductive studies,
2.4.3. Expert review of expert rule-based (structural) alert systems repeated dose toxicity studies or segment I (fertility) studies, and pos-
 An expert review of the results from an expert rule-based alert sibly also from diﬀ erent species (rat, mouse, rabbit) will be applicable.
system may involve inspection of the underlying information as well as The selection of experimental study types need focus on those that
external knowledge. Special emphasis needs to be placed on the as- have general value based on scientiﬁ c justiﬁ cation. This includes study
sessment of chemicals where no alerts are identiﬁ ed in the expert alert types that have widespread use in risk assessments, regulatory accep-
system. When no alert is ﬁ red (i.e., it is not predicted active), it is often tance and that follow internationally recognized test guidelines. In
not reported if the prediction is negative, equivocal, or out of the ap- addition, other types of data may be considered relevant on a case-by-
plicability domain of the model and often no prediction is generated. case basis. Numerous guidance documents discuss acceptable studies,
An expert review may increase the reliability of the results based on one their relevancy, and their use in hazard identiﬁ cation, hazard char-
or more elements deﬁ ned in Table 3. acterization and risk assessment. These include guidance documents
 from the ICH (ICH, 2017), OECD ( OECD, 2017), European Food Safety
2.4.4. Read-across expert review Authority (EFSA) (EFSA, 2017), Scientiﬁ c Committee on Consumer
 Read-across contains an expert assessment by its nature: it requires Safety (SCCS) (SCCS, 2017), REACH/ECHA (ECHA, 2008; ECHA,
expert judgment of the analogs, their data and extrapolation to the 2015), United States Environmental Protection Agency (EPA) Oﬃ ce of
query chemical. For example, read-across assessments performed and Chemical Safety and Pollution Prevention (OCSPP, 2017), and National
documented according to the RAAF (i.e., following the detailed RAAF Institute of Environmental Health Sciences (NIEHS) (NIEHS, 2017)
Assessment Elements), or similar frameworks, as discussed earlier, in- guidance documents. Such guidance documents provide a useful basis
corporate an expert review as part of the assessment. This type of as- for test considerations but may not always be harmonized across leg-
sessment includes a strong justiﬁ cation for biological plausibility of any islation, industrial sector or geographical regions, as requirements may
analogs selected (including an assessment of the structural diﬀ erences diﬀ er across guidance documents.
and similarities to the target structure, and an analysis of potential The IST protocols will discuss how to assess and document the ex-
metabolism). It also includes an expert assessment when a read-across perimental data and uncertainties to ensure the proper justiﬁ cation of
prediction concludes there is an absence of eﬀ ects. In addition, an as- the experimental results’ reliability, including deﬁ ning what speciﬁ c
sessment of supporting evidence (including the reliability of the source elements or ﬁ elds are important to document. With older studies pre-
data), any weight-of-evidence considerations, and an assessment of any dating existing guidelines, it will often still be possible to perform an
possible bias in the selection of source chemicals is required. expert review to determine the adequacy of the data, but it will be
 important to document speciﬁ cally why the study results were con-
2.5. Assessment of available experimental data sidered acceptable or dismissed as unacceptable. The IST protocols will
 also provide recommendations on how to select a result when multiple
 Experimental data may have been previously generated and re- studies (with potentially conﬂ icting results) for the same eﬀ ect or me-
ported for a chemical being assessed, for example, in the literature or chanism are reported.
through a public or proprietary database. To support the identiﬁ cation Klimisch scores are a widely used approach adopted to support an
of experimental data, each IST protocol will identify a series of relevant assessment of experimental data reliability (Table 4; Klimisch et al.,
 9

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Table 4 IST protocols will document such procedures.
Summary of Klimisch scores for data reliability (adapted from Klimisch et al., There are multiple approaches to compile results. A cautious ap-
1997) (Note “restriction”, as part of scores 1 and 2, implies restricted quality). proach is to use the most conservative data or prediction for this as-
 Score Description Summary sessment. For example, when predicting the results of the bacterial
 reverse mutation test using two models, if either model's prediction
 1 Reliable without result is mutagenic then the overall assessment is mutagenic. Other
 restriction  Well documented and accepted study or datafrom the literatureoptions include a weight-of-evidence or consensus approach or selec-
  Performed according to valid and/or accepted tion of the prediction with the highest conﬁ dence (e.g., predictive
 test guidelines (e.g., OECD)
  Preferably performed according to good probability score and relevance of analogous structures). Speciﬁ c con-
 laboratory practices (GLP) siderations per endpoint may be addressed in the individual IST pro-
 2 Reliable with tocols and may be dependent on the problem formulation.
 restriction  Well documented and suﬃ cient Primarily not performed according to GLP
 3 Not reliable  Inferences between the measuring system Partially complies with test guideline2.6.2. Reliability scores
 and test substance Reliability, in this context, is deﬁ ned as the inherent quality of the
  Test system not relevant to exposure experimental study (Klimisch et al., 1997) and/or in silico analysis. It is
  Method not acceptable for the endpoint used to support any hazard assessment, in combination with other in-
  Not suﬃ ciently documented for an expertreview formation. A reliability score (RS) is associated with the toxicological
 4 Not assignable  Lack of experimental detailseﬀ ect or mechanism assessment (as shown in Fig. 1). As noted earlier,
  Referenced from short abstract or secondary when data from the literature or other sources are considered, Klimisch
 literature scores can be used to assess the reliability of the results. However, the
 Klimisch framework was never intended to assess the reliability of in
1997). The Klimisch score (1–4) is based on factors including whether silico predictions. It is also important to note that regardless of the
the test was compliant with the OECD principles of Good Laboratory approach taken, reliability assessments will contain subjective deci-
Practices (GLP) or Good In Vitro Methods Practices (GIVIMP) standards sions.
(OECD, 2016c), whether the data were generated using accepted test A number of general factors can aﬀ ect the reliability of in silico re-
guidelines, whether the data are available for independent inspection, sults:
and the quality of the report. ECHA uses this score, for example, as part
of its data submission process (ECHA, 2011), and there are tools to  Multiple in silico results: Combining results from multiple com-
support the assignment of Klimisch scores (ECVAM, 2017; Schneider plementary or independent in silico tools which use diﬀ erent meth-
et al., 2009). Another approach to the assessment of the reliability of odologies or QSAR descriptors and/or training sets, has been shown
the experimental data is the Science in Risk Assessment and Policy to improve overall sensitivity, but it can lower speciﬁ city by in-
(SciRAP) application, a web-based reporting and evaluation resource creasing false positive rates (Myatt et al., 2016). In the case of
created to help understand how academic toxicity-related studies can quantitative predictions, such process are overly conservative esti-
be used as part of any regulatory assessment (Molander et al., 2014). An mates. Hence, consistency across several diﬀ erent models can in-
approach proposed by EFSA is a detailed analysis of diﬀ erent para- crease the reliability of the results.
meters of the study (e.g. statistical power; veriﬁ cation of measurement  Expert review: A plausible and well-documented read-across(consistent with the RAAF or similar frameworks) may be accep-
methods and data; control of experimental variables that could aﬀ ect table as part of a REACH regulatory submission as an alternative to
measurements; universality of the eﬀ ects in validated test systems using experimental data. A structured expert review is implicit in any
relevant animal strains and appropriate routes of exposure, etc.) with read-across assessment (as discussed in Section 2.4.4). Similarly, an
detailed documentation of the process (EFSA, 2011). explicit expert review (following the elements described in Sections
 2.4.2 and 2.4.3) of the in silico predictions can improve the relia-
2.6. Combined assessment of experimental data and in silico predictions bility of the ﬁ nal results, especially for negative predictions (Dobo
 et al., 2012).
2.6.1. Toxicological eﬀ ect or mechanism assessment To generate an overall reliability score for assessments based on
 Reliable data, generally deﬁ ned by Klimisch scores 1 or 2 reviewed experimental data and/or in silico predictions, the Klimisch score has
by an expert (see Table 4), is ideally used for the toxicological e ﬀ ect or been adapted (as shown in Fig. 2) to include an assessment of in silico
mechanism (shown in Fig. 1) whenever available 3. In the absence of prediction results.
adequate experimental data, results from one or more in silico models Experimental data assigned a Klimisch score of 1 or 2 is assigned a
can be used to support assessment of the toxicological eﬀ ect or me- score of RS1 and RS2, respectively, in this revised scheme. In silico re-
chanism. When multiple in silico model results, from potentially dif- sults are not assigned a score of RS1 or RS2 since adequate experi-
ferent methodologies, or QSAR models using diﬀ erent descriptors and/ mental data is preferred over in silico predictions. Since in silico results
or training sets, are generated per toxicological eﬀ ect or mechanism, may be used directly as part of certain regulatory submissions, whereas
the individual results need to be compiled to provide one overall as- experimental data with a Klimisch score of 3 or 4 would not (or only as
sessment, as shown in Fig. 1. This assessment may take into con- supporting data under REACH, for example), the next two categories
sideration information from any expert review of the in silico results, as (RS3 and RS4) represent, in part, in silico predictions. The following
certain results may need to be refuted. Similarly, when there are data may be acceptable as part of a regulatory submission: (1) an adequately
assigned Klimisch 3 or 4 and/or there are in silico results, this in- performed read-across prediction (EU, 2006), or (2) an expert review of
formation needs to be compiled into an overall assessment. Individual in silico and/or other experimental data (ICH M7, 2017(R1); EU 2006);
 they are assigned a reliability score of RS3. A score of RS4 would be
 3 As mentioned in Section 2.5, where high quality experimental data are available (as assigned when two or more predictive models are available that are
shown in Fig. 1), it may not be necessary to run in silico models. However, generating in complementary, with concurring results (with no expert review), and
silico predictions for chemicals with known values is sometimes performed to verify ex- no supporting literature data are available. Examples include those
perimental results because an unexpected positive or negative experimental result in a
physical assay may be explained by the presence of an active impurity or to provide predictive models that use either substantially diﬀ erent QSAR de-
additional weight-of-evidence or for other reasons. scriptors and/or QSAR training sets or diﬀ erent in silico methodologies.
 10

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
 Fig. 2. Reliability of toxicity assessments based on computational models and experimental data.
If two or more in silico model results do not agree, then an expert review 2.6.3. Worked examples
would be required to assess the results. This review might increase the Three examples from Amberg et al. (2016) illustrate how the fra-
conﬁ dence in the assessment, resulting in an increased reliability score mework described in this publication can be used for determining a
of RS3. A single acceptable (as discussed in Section 2.3.2) in silico model toxicological eﬀ ect or mechanism assessment and reliability score,
result, without further expert review, is aﬀ orded the same reliability based on experimental data and/or in silico predictions. Assessing re-
score of RS5 as an actual test result of lowest reliability (Klimisch 3 or liability is an initial step in the overall assessment of hazard, where it
4). The in silico result is placed in the same category as low reliability will be combined with other information, including an evaluation of the
data because such models inform decisions based on a series of com- relevance of the information, to support decision making.
pounds or trends. However, this reliability score may be increased In the example in Fig. 3, no experimental data were identi ﬁ ed. Two
following expert review. This reliability score closely follows the ICH in silico models were run; the statistical-based model prediction was
M7 guideline, where submissions corresponding to reliability scores negative and the expert rule-based alert prediction was negative. The
RS1-RS4 would be accepted according to the guideline. In addition to initial score would be RS4 based on multiple concurring prediction
this score, it may be helpful to document any additional considerations results; however, an expert review was performed on the results from
that may be important to the overall assessment. Individual IST pro- both methodologies and the negative result was conﬁ rmed with in-
tocols may deviate from this scheme with appropriate justiﬁ cation. creased reliability. The review concluded there were no potentially
 reactive features in the chemical. This resulted in a negative overall
 Fig. 3. Determining the bacterial gene mutation assessment and reliability score for two concurring in silico results with expert review.
 11

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
 Fig. 4. Determining the bacterial gene mutation assessment and reliability score for two concurring in silico results with no expert review.
assessment and a reliability score of RS3 (as a result of the expert re- (e.g., genetic toxicity). Other identiﬁ ed toxicological eﬀ ects or me-
view increasing the reliability). chanisms are associated with toxicological endpoints as shown in Fig. 6.
 In the example in Fig. 4, no experimental data were identi ﬁ ed. Two For example, the mammalian gene mutation (eﬀ ect/mechanism 2) is
in silico models were run; the statistical model prediction was positive also relevant to the assessment of gene mutations (endpoint 1) and
and the expert alert prediction was positive. No expert review of the clastogenicity (endpoint 2) is another endpoint to be used in the as-
results was performed. The overall assessment was therefore positive sessment of genetic toxicity (a major toxicological endpoint). Fig. 6 also
and a reliability score of RS4 was assigned as a result of two concurring includes another example to illustrate how this scheme might be used to
positive predictions using complementary in silico methodologies but assess male reproductive toxicity.
without expert review. The hazard assessment framework scheme for each IST protocol will
 In the example in Fig. 5, no experimental data were identi ﬁ ed. Two contain diﬀ erent numbers of toxicological endpoints as needed to
in silico models were run; the statistical model prediction was positive support the assessment of each major toxicological endpoint in a
and the expert alert prediction was negative. An expert review was complete and transparent manner.
performed on the results from both methodologies, refuting the statis- It is noteworthy that only the toxicological endpoints required to
tical model's positive prediction. This review was based on an analysis support a particular problem formulation need to be assessed. For ex-
of the test chemical's potential to react with DNA and the highlighted ample, in certain applications only an assessment of gene mutation may
structural feature was determined to be irrelevant for the mechanism of be needed (i.e., it may not be necessary to compute clastogenicity or the
interaction with DNA. This resulted in a negative overall assessment genetic toxicity major toxicological endpoint).
and a reliability score of RS3 (as a result of the expert review increasing
the reliability). 2.7.2. Relevance
 Relevance, in this context, is deﬁ ned as the scientiﬁ c predictivity of
2.7. Hazard assessment framework the each toxicological eﬀ ect or mechanism for the purpose of assessing
 a speciﬁ c toxicological endpoint. As shown in Fig. 6, the assessment of
2.7.1. Toxicological endpoints toxicological endpoints may be based on the associated toxicological
 Fig. 6 illustrates a general scheme for the prediction of a major eﬀ ects or mechanisms. To support a transparent overall analysis, the
toxicological endpoint. In this scheme, the speciﬁ c toxicological eﬀ ects relevance of the toxicological eﬀ ect/mechanism information in support
or mechanisms are used to support the assessment of a series of tox- of the assessment of the associated toxicological endpoint will be de-
icological endpoints. These toxicological endpoint assessments are, in ﬁ ned in the IST protocols. This relevance will be based on the collective
turn, used in the overall assessment of the major toxicological endpoint. experience of the consortium and available validation information.
In Fig. 6,e ﬀ ect/mechanism 1 is identiﬁ ed as being relevant to an as-
sessment of a speciﬁ c toxicological endpoint (Endpoint 1). For example, 2.7.3. Toxicological endpoint assessment
bacterial gene mutation (eﬀ ect/mechanism 1) is relevant to the as- The assessment of each toxicological endpoint (as shown in Fig. 6)is
sessment of gene mutation (endpoint 1). Endpoint 1 is, in turn, one of a function of all associated toxicological eﬀ ects or mechanisms and, in
the endpoints that are relevant to the major toxicological endpoint some cases, other toxicological endpoints. For example, in Fig. 6,
 Fig. 5. Determining the bacterial gene mutation assessment and reliability score where there is no experimental data available and conﬂ icting in silico results.
 12

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
 Fig. 6. Hazard assessment framework.
bacterial gene mutation and mammalian gene mutation (toxicological endpoint has to support the decision context(s), regulatory framework
eﬀ ects or mechanisms) are associated with gene mutation, whereas and the type of product being assessed. Minimum conﬁ dence scores for
gene mutation and clastogenicity (both toxicological endpoints) are regulatory purposes may need to be set; however for other applications,
associated with genetic toxicity. Rules or general principles for com- the use of these scores may be based on the individual organization's
bining all associated results for each endpoint will be deﬁ ned in therisk tolerance or based on the context, a decision on the maximum
upcoming IST protocols. For example, a rule may state that if one of the permitted eﬀ ort to be expended (since higher conﬁ dence score may be
associated eﬀ ects/mechanisms is positive then the endpoint assessment generated with additional resources), or an organization's internal
is positive. These rules or principles will take into consideration how policy for using the conﬁ dence scores for speciﬁ c tasks.
combinations of diﬀ erent toxicological eﬀ ects/mechanisms are eval-
uated to generate an assessment for any toxicological endpoint which 2.7.5. Expert review of toxicological endpoints
may include a sequence of steps and incorporate Boolean logic. In certain situations, an expert review of the toxicological endpoint
 assessment and/or conﬁ dence may be warranted, and speciﬁ c points to
2.7.4. Toxicological endpoint conﬁ dence consider as part of such an expert review will be detailed in the in-
 Conﬁ dence, in this context, is deﬁ ned as a score that combines the dividual IST protocols. This review may take into consideration the
reliability and relevance of the associated toxicological eﬀ ects or me- context of the assessment, that is, the type of product being assessed
chanisms. This is an additional score associated with toxicological and any potential regulatory framework. It may be helpful to document
endpoints. The score may, in some cases, use other toxicological end- any additional considerations concerning the assessment and con-
point conﬁ dence scores (as shown in Fig. 6). This score will also take ﬁ dence to support an overall assessment.
into consideration the completeness of the information available; for
example, the conﬁ dence score may be lowered when information on an 2.8. In silico toxicology protocol components
eﬀ ect or mechanism is missing. It will also include complementary ef-
fects or mechanisms that need to be considered. This score will be Ongoing eﬀ orts are concentrated on the development of individual
generated based on a series of general principles and/or rules deﬁ ned in IST protocols for major endpoints including genetic toxicity, carcino-
each IST protocol. Each protocol will outline the diﬀ erent conﬁ dence genicity, acute toxicity, repeated dose toxicity, reproductive toxicity,
values to generate, such as high, medium or low. and developmental toxicity. Table 5 outlines proposed common com-
 A conﬁ dence score is one of the most important items to generate. ponents for these IST protocols.
Diﬀ erent decision contexts tolerate a diﬀ erent level of conﬁ dence in the
assessment result as exempliﬁ ed in the following two scenarios. 2.9. Reporting formats
1) Scenario 1. The decision is to prioritize a large number of chemicals Standardized reporting of the results and expert review is good
 to screen as part of product development. In this scenario, selecting scientiﬁ c practice and assures that when such information is commu-
 a small subset of compounds using in silico methods supports stra- nicated to regulatory authorities, it is complete, consistent and trans-
 tegic resource utilization with the eventual goal of reducing overall parent; this may avoid requests for additional information and maintain
 costs. a consistent, expedient, and streamline regulatory review process.
2) Scenario 2. A regulatory submission for a new cosmetic ingredient is Table 6 outlines a proposed structure for the report format.
 being prepared based on results from in silico methods. The proposed report format is more comprehensive than existing
 data formats by including information on overall assessment and expert
 Although in both scenarios, toxicological endpoint assessments reviews. For example, the “QSAR prediction reporting format” (QPRF;
generated at the highest level of conﬁ dence would be preferable, JRC, 2014) could be used to report the individual model results (as
Scenario 1 could still make beneﬁ cial use of lower conﬁ dence predic- shown in Section D of Table 6), or “QSAR model reporting format”
tions because the safety consequences of a false negative is lower than (QMRF) can be used to report the QSAR model's details (as shown in
in Scenario 2. Therefore, a risk assessment which takes into account the Section H of Table 6).
acceptable tolerance for a wrong prediction can be used to evaluate the The new proposed report format collects enough details on how the
necessity for high conﬁ dence. predictions were generated to enable another expert to repeat the
 The assignment of the conﬁ dence score for each toxicological process. It is also important that the reasoning and decisions of the
 13

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Table 5
Common components of an IST protocol (IATA = Integrated Approaches to Testing and Assessment; AOP = Adverse Outcome Pathways).
 Introduction  Describe the major toxicological endpoint being assessed
  Outline the general hazard assessment framework, including how a series of toxicological eﬀ ects or mechanisms arerelated to one or more endpoints
 In silico methodologies and models  Identify toxicological eﬀ ects or mechanisms that might realistically be predicted Provide citations to any applicable AOPs or IATAs used
  Deﬁ ne what in silico methodologies are appropriate to use
  Specify additional considerations as to what constitutes an acceptable model
 Experimental data  Deﬁ ne speci Discuss issues to be considered as part of any read-across analysisﬁ c study types and result(s) relevant to each toxicological eﬀ ect or mechanism
  Deﬁ ne and justify the relevance of the information to the assessment of the toxicological endpoint (deﬁ ned in thehazard assessment framework)
  Deﬁ ne speciﬁ c factors to consider when assessing the results and documenting the reliability of any available data or
 reference speciﬁ c test guideline(s)
 Toxicological eﬀ ects or mechanisms assessment and  Identify sources of data that may be considered
 reliability scores  Describe how each toxicological eﬀ ect or mechanism assessment may be generated from available experimentaldata and/or in silico prediction(s)
  Deﬁ ne additional items to consider as part of an expert reviewﬁ c issues to consider as part of the reliability score
 Toxicological endpoint assessment and conﬁ dence  Describe the toxicological endpoints that will be used as part of the hazard assessment framework Discuss any endpoint speci
  Describe the rules or principles for determining each endpoint assessment, based on the associated eﬀ ect/mechanisms or other endpoints
  Deﬁ ne the rules or principles for determining each toxicological endpoint conﬁ dence, based on the relevance andreliability (from associated eﬀ ects/mechanisms) or conﬁ dence (from associated endpoints)
 Reporting  Deﬁ ne a format for a report of the results, expert review and conclusions Identify points to consider as part of any expert review
 Other considerations  Case studies
expert review steps are transparently documented and can be retraced The development of IST protocols will support the use and adoption of
at any time, including the information used as their basis for conclu- in silico methods in the same manner in which in vitro and in vivo test
sions. guidelines support the use and adoption of those assays.
 Fig. 7 summarizes the steps to perform an in silico assessment con-
3. Summary and outlook sistent with the framework deﬁ ned in this publication. The key ele-
 ments needed for the development of IST protocols are outlined in this
 IST is poised to play an increasingly signiﬁ cant role in the assess- publication, including: 1) how to select, assess and integrate in silico
ment of chemicals in a range of chemical exposure scenarios that have predictions alongside experimental data for deﬁ ned toxicological ef-
the potential to impact public health. Thus, this is an opportune time for fects or mechanisms, including a new methodology for establishing the
the development of IST protocols. As expected, the quality and quantity reliability of this assessment, 2) a hazard assessment framework for
of experimental data will vary as will the available in silico methods. For systematic assessment of these toxicological eﬀ ects or mechanisms to
example, experimental data could be from a variety of sources, studies, predict speciﬁ c endpoints and assess the conﬁ dence in the results.
protocols and laboratories using or not using GLP standards. Similarly, Wherever possible, this is based on mechanistic knowledge on diﬀ erent
several in silico methods and approaches are available for assessment of biological levels of organization (Bell et al., 2016; OECD, 2016a; OECD,
toxicity. Thus, accepted selection criteria have to be deﬁ ned for ex- 2016b). Overall, the IST protocols will contain information to ensure
perimental data and in silico methods, for consistent and uniform use. predictions are performed in a consistent, repeatable, transparent and
Table 6
Elements of an in silico toxicology report (QMRF = QSAR Model Reporting Format).
 Section Content
 Title page - Title (including information on the decision context)
 - Who generated the report and from which organization
 - Who performed the in silico analysis and/or expert review, including their organization
 - Date when this analysis was performed
 - Who the analysis was conducted for
 Executive summary - Provide a summary of the study
 - Describe the toxicity or properties being predicted
 - Include a table or summary showing the following:
  The chemical(s) analyzed
  Summary of in silico results, reviewed experimental data and overall assessment for each toxicological eﬀ ect or mechanism
  Summary of toxicological endpoint assessment and conﬁ dence
 Purpose - Speciﬁ cation of the problem formulation Summary of supporting information
 Materials and methods - QSAR model(s), expert alerts, and other models used with version number(s) and any parameters set as part of the prediction (e.g., QMRF format)
 - Databases searched with version number(s)
 - Tools used as part of any read-across with version number(s)
 Results of Analysis - Details of the results and expert review of the in silico models and any experimental data, including results of the applicability domain analysis
 - Report of any read-across analysis, including source analogs and read-across justiﬁ cations
 Conclusion - Summarize the overall analysis including experimental data, in silico methods and expert review
 - Final prediction that is based on expert judgment
 References - Complete bibliographic information or links to this information, including test guidelines referred to in the experimental data, etc.
 Appendices (optional) - Full (or summary) study reports used or links to the report, detailed (or summary) in silico reports, reports on the models used (e.g., QMRF reports)
 14

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
 Fig. 7. Summary of the IST protocol process.
ultimately accepted manner and will include a checklist (as deﬁ ned in MHRA Disclaimer: Any opinions expressed in this document are the
Section 2.4) to guide an expert review of the information. Each in- author's and are not necessarily shared by other assessors at the
dividual IST protocol will address how predictions will be performed in Medicines and Healthcare products Regulatory Agency (MHRA). As
alignment with the framework discussed in this publication. These new such, they cannot be considered to be UK policy. The mention of
protocols will provide speciﬁ c guidance for each toxicological endpoint, commercial products, their sources, or their use in connection with
including situations where no AOP or IATA is currently available. These material reported herein is not to be construed as either an actual or
protocols build on and fully incorporate wherever possible the con- implied endorsement of such products by the UK's MHRA.
siderable work previously reported, such as the OECD validation prin- CDC/ATSDR Disclaimer: The ﬁ ndings and conclusions in this report
ciples (see Sections 2.3.2), IATAs (see Sections 2.2), AOPs (see Sections are those of the author(s) and do not necessarily represent the oﬃ cial
2.2), read-across frameworks (see Sections 2.3.2, 2.6.2), the Klimisch position of the Centers for Disease Control and Prevention or the
score (see Sections 2.5, 2.6.1, 2.6.2) and the QMRF/QPRF (see Sections Agency for Toxic Substances and Disease Registry. Mention of trade
2.3.2, 2.9). names is not an endorsement of any commercial product.
 The IST protocols do not deﬁ ne how a risk assessment will be per- EPA Disclaimer: The views expressed in this article are those of the
formed; they solely deﬁ ne the process which will lead to the prediction author(s) and do not necessarily reﬂ ect the views or policies of the U.S.
of the potential toxicity (hazard) of a chemical. Risk analysis depends Environmental Protection Agency. Mention of trade names or com-
on the exposure scenario, industry, regulatory framework and decision mercial products does not constitute endorsement or recommendation
context based on the level of tolerated uncertainty and is performed in for use.
the hands of an expert. EFSA Disclaimer: This paper reﬂ ects Dr. Seraﬁ mova's personal view
 The process of developing IST protocols requires an understanding and is not endorsed by the European Food Safety Authority.
of the best practices and science across various organizations, diﬀ erent Health Canada Disclaimer: The ﬁ ndings and conclusions in this re-
industries and regulatory authorities. To develop such protocols, an port are those of the author(s) and do not necessarily represent the
international consortium was established comprising regulators, gov- oﬃ cial position of Health Canada. The mention of commercial pro-
ernment agencies, industry, academics, model developers, and con- ducts, their sources, or their use should not be construed as either an
sultants across many diﬀ erent sectors. This consortium initially devel- actual or implied endorsement of such products by Health Canada.
oped the overall strategy outlined in this publication. WorkingDHS Disclaimer: The ﬁ ndings and conclusions in this report are
subgroups will develop individual IST protocols for major endpoints those of the author(s) and do not necessarily represent the oﬃ cial po-
including genetic toxicity, carcinogenicity, acute toxicity, reproductive sition of the U.S. Government, Department of Homeland Security
toxicity, and developmental toxicity. As each IST protocol is estab- (DHS), DHS Science and Technology (S&T) Homeland Security
lished, it will be reviewed internally within each organization and Advanced Research Projects Agency (HSARPA), or the Chemical
published. This process will evolve over time, as computational tech- Security Analysis Center (CSAC). In no event shall either the U.S.
nology progresses, as will the assays and other information relevant to Government, DHS, HSARPA, CSAC, or the author(s) have any respon-
assessing these major endpoints emerges. Hence, similar to other test sibility or liability for any consequences of any use, misuse, inability to
guidelines, the IST protocols will need to be periodically reviewed and use, or reliance upon the information contained herein, nor do any of
updated. The implementation of IST protocols will also require user- the parties set forth in this disclaimer warrant or otherwise represent in
friendly tools for performing such analyses and reporting the results, any way the accuracy, adequacy, eﬃ cacy, or applicability of the con-
education, as well as further collaboration with organizations to sup- tents hereof. The use of trade, ﬁ rm, company, or corporation names
port global adoption. including product descriptions does not constitute an oﬃ cial DHS en-
 dorsement of any service or product.
Disclaimer European Commission Disclaimer: The views expressed are solely
 those of the authors and the content does not necessarily represent the
 FDA CDER Disclaimer: The ﬁ ndings and conclusions in this manu- views or position of the European Commission.
script have not been formally disseminated by the FDA and should not
be construed to represent any agency determination or policy. The
mention of commercial products, their sources, or their use in con- Acknowledgements
nection with material reported herein is not to be construed as either an
actual or implied endorsement of such products by the Department of Research reported in this publication was supported by the National
Health and Human Services. Institute of Environmental Health Sciences of the National Institutes of
 NIEHS Disclaimer: The ﬁ ndings and conclusions in this report are Health under Award Number R43ES026909. The content is solely the
those of the author(s) and do not necessarily represent the oﬃ cial po- responsibility of the authors and does not necessarily represent the
sition of the National Institutes of Health (NIH), NIEHS. The mention of oﬃ cial views of the National Institutes of Health. The authors would
commercial products, their sources, or their use should not be con- like to thank Sachin Bhusari and George Pugh from the Coca-Cola
strued as either an actual or implied endorsement of such products by Company for their input into the manuscript.
the NIH/NIEHS.
 15

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
Transparency document Chen, B., Zhang, T., Bond, T., Gan, Y., 2015. Development of quantitative structure ac-
 tivity relationship (QSAR) model for disinfection byproduct research: a review of
 Transparency document related to this article can be found online at methods and resources. J. Hazard Mater. 299, 260–279. http://dx.doi.org/10.1016/j.
 jhazmat.2015.06.054.
http://dx.doi.org/10.1016/j.yrtph.2018.04.014 . Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden,
 J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer,
References R., Benigni, R., Yang, C., Rathman, J., Terﬂ oth, L., Gasteiger, J., Richard, A., Tropsha,
 A., 2014. QSAR modeling: where have you been? Where are you going to? J. Med.
 Chem. 57, 4977–5010. http://dx.doi.org/10.1021/jm4004285.
Alves, V., Muratov, E., Capuzzi, S., Politi, R., Low, Y., Braga, R., Zakharov, A.V., Sedykh, Dobo, K.L., Greene, N., Fred, C., Glowienke, S., Harvey, J.S., Hasselgren, C., Jolly, R.,
 A., Mokshyna, E., Farag, S., Andrade, C., Kuz’min, D., Fourches, D., Tropsha, A., Kenyon, M.O., Munzner, J.B., Muster, W., Neft, R., Reddy, M.V., White, A.T., Weiner,
 2016. Alarms about structural alerts. Green Chem. 18 (16), 4348–4360. http://dx. S., 2012. In silico methods combined with expert knowledge rule out mutagenic po-
 doi.org/10.1039/C6GC01492E. tential of pharmaceutical impurities: an industry survey. Regul. Toxicol. Pharmacol.
Amaral, R.T.D., Ansell, J., Aptula, N., Ashikaga, T., Chaudhry, Q., Hirose, A., Jaworska, 62, 449–455. http://dx.doi.org/10.1016/j.yrtph.2012.01.007.
 J., Kojima, H., Lafranconi, M., Matthews, E., Milstein, S., Roesler, C., Vaillancourt, E., ECETOC, 2012. ECETOC Technical Report No. 116: Category Approaches, Read-across.
 Verma, R., Worth, A., Yourick, J., 2014. Report for the International Cooperation on (Q)SAR. http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-116-
 Cosmetics Regulation. In Silico Approaches for Safety Assessment of Cosmetic Category-approaches-Read-across-QSAR.pdf.
 Ingredients. https://www.pharmamedtechbi.com/∼ /media/Supporting ECHA, 2008. Guidance on Information Requirements and Chemical Safety Assessment
 %20Documents/The%20Rose%20Sheet/36/ICCR%20In%20Silico%20Report.pdf . Chapter R.6: QSARs and Grouping of Chemicals. https://echa.europa.eu/
Amaral, R., Amores Da Silva, P., Ansell, J., Boisleve, F., Dent, M., Hatao, M., Hirose, A., documents/10162/13632/information\_requirements\_r6\_en.pdf/77f49f81-b76d-
 Kasai, Y., Kojima, H., Kern, P., Kreiling, R., Milstein, S., Oliveira, J., Richarz, A., 40ab-8513-4f3a533b6ac9.
 Taalman, R., Vaillancourt, E., Verma, R., Vieira, N.C., Weiss, C., 2017. Report for the ECHA, 2011. Guidance on Information Requirements and Chemical Safety Assessment
 International Cooperation on Cosmetics Regulation, Joint Regulators-Industry Chapter R.4: Evaluation of Available Information. https://echa.europa.eu/
 Working Group: Integrated Strategies for Safety Assessments of Cosmetic Ingredients documents/10162/13643/information\_requirements\_r4\_en.pdf/d6395ad2-1596-
 - Part I. http://www.iccr-cosmetics.org/ﬁ les/4715/0824/0761/ICCR-11\_JWG\_ 4708-ba86-0136686d205e.
 Integrated\_Strategies\_Integrated\_Strategies\_for\_Safety\_Assessments\_of\_Cosmetic\_ ECHA, 2015. Guidance on the Application of the CLP Criteria Guidance to Regulation
 Ingredients\_-\_Part\_I.pdf. (EC) No 1272/2008 on Classiﬁ cation, Labelling and Packaging (CLP) of Substances
Amberg, A., Harvey, J.S., Czich, A., Spirkl, H.-P., Robinson, S., White, A., Elder, D.P., and Mixtures. https://echa.europa.eu/documents/10162/23036412/clp\_en.pdf/
 2015. Do Carboxylic/Sulfonic acid halides really present a mutagenic and carcino- 58b5dc6d-ac2a-4910-9702-e9e1f5051cc5.
 genic risk as impurities in ﬁ nal drug products? Org. Process Res. Dev. 19, 1495–1506. ECHA, 2016. Practical Guide: How to Use and Report (Q)SARs. https://echa.europa.eu/
 http://dx.doi.org/10.1021/acs.oprd.5b00106. documents/10162/13655/pg\_report\_qsars\_en.pdf.
Amberg, A., Beilke, L., Bercu, J., Bower, D., Brigo, A., Cross, K.P., Custer, L., Dobo, K., ECHA, 2017a. Weight of Evidence. https://echa.europa.eu/support/registration/how-to-
 Dowdy, E., Ford, K.A., Glowienke, S., Gompel, J.V., Harvey, J., Hasselgren, C., avoid-unnecessary-testing-on-animals/weight-of-evidence.
 Honma, M., Jolly, R., Kemper, R., Kenyon, M., Kruhlak, N., Leavitt, P., Miller, S., ECHA, 2017b. Read-across Assessment Framework (RAAF). https://echa.europa.eu/
 Muster, W., Nicolette, J., Plaper, A., Powley, M., Quigley, D.P., Reddy, M.V., Spirkl, documents/10162/13628/raaf\_en.pdf.
 H.-P., Stavitskaya, L., Teasdale, A., Weiner, S., Welch, D.S., White, A., Wichard, J., ECVAM, 2017. ToxRTool - Toxicological Data Reliability Assessment Tool. https://eurl-
 Myatt, G.J., 2016. Principles and procedures for implementation of ICH M7 re- ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool/toxrtool-
 commended (Q)SAR analyses. Regul. Toxicol. Pharmacol. 77, 13–24. http://dx.doi. toxicological-data-reliability-assessment-tool.
 org/10.1016//j.yrtph.2016.02.004. EFSA, 2011. European Food Safety Authority; Submission of scientiﬁ c peer-reviewed
Ball, N., Cronin, M.T., Shen, J., Blackburn, K., Booth, E.D., Bouhifd, M., Donley, E., open literature for the approval of pesticide active substances under Regulation (EC)
 Egnash, L., Hastings, C., Juberg, D.R., Kleensang, A., Kleinstreuer, N., Kroese, E.D., No 1107/2009. EFSA J. 9, 1–49. http://dx.doi.org/10.2903/j.efsa.2011.2092.
 Lee, A.C., Luechtefeld, T., Maertens, A., Marty, S., Naciﬀ , J.M., Palmer, J., Pamies, D., https://www.efsa.europa.eu/en/efsajournal/pub/2092.
 Penman, M., Richarz, A.N., Russo, D.P., Stuard, S.B., Patlewicz, G., van Ravenzwaay, EFSA, 2014. Modern methodologies and tools for human hazard assessment of chemicals.
 B., Wu, S., Zhu, H., Hartung, T., 2016. Toward good read-across practice (GRAP) EFSA J. 12, 1–87. http://dx.doi.org/10.2903/j.efsa.2014.3638. 2014. https://www.
 guidance. ALTEX 33, 149–166. https://doi.org/10.14573/altex.1601251. efsa.europa.eu/de/efsajournal/pub/3638.
Ball, D., Norwood, D., 2012. Leachables and Extractables Handbook. Wiley, Pages, pp. EFSA, 2016. Guidance on the establishment of the residue deﬁ nition for dietary risk as-
 58–79. sessment: EFSA Panel on Plant Protection Products and their Residues (PPR). EFSA J.
Barber, C., Amberg, A., Custer, L., Dobo, K.L., Glowienke, S., Gompel, J.V., Gutsell, S., 14, 1–12. http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2016.4549/epdf.
 Harvey, J., Honma, M., Kenyon, M.O., Kruhlak, N., Muster, W., Stavitskaya, L., EFSA, 2017. European Food Safety Authority. http://www.efsa.europa.eu/.
 Teasdale, A., Vessey, J., Wichard, J., 2015. Establishing best practise in the appli- Ellison, C.M., Piechota, P., Madden, J.C., Enoch, S.J., Cronin, M.T.D., 2016. Adverse
 cation of expert review of mutagenicity under ICH M7. Regul. Toxicol. Pharmacol. outcome pathway (AOP) informed modeling of aquatic toxicology: QSARs, read-
 73, 367–377. http://dx.doi.org/10.1016/j.yrtph.2015.07.018. across, and interspecies veriﬁ cation of modes of action. Environ. Sci. Technol. 50,
Bassan, A., Worth, A.P., 2008. The integrated use of models for the properties and eﬀ ects 3995–4007. http://dx.doi.org/10.1021/acs.est.5b05918.
 of chemicals by means of a structured workﬂ ow. QSAR Comb. Sci. 27, 6–20. http:// EU, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council
 dx.doi.org/10.1002/qsar.200710119. of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and
Bell, S.M., Angrish, M.M., Wood, C.E., Edwards, S.W., 2016. Integrating publicly available Restriction of Chemicals (REACH). http://eur-lex.europa.eu/legal-content/EN/TXT/
 data to generate computationally predicted adverse outcome pathways for fatty liver. PDF/?uri=CELEX:02006R1907-20161011&from=EN.
 Toxicol. Sci. 150, 510–520. http://dx.doi.org/10.1093/toxsci/kfw017. EU, 2009a. Regulation (EC) No 1223/2009 of the European Parliament and of the Council
Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A.-N., Bois, F.Y., Exner, T., of 30 November 2009 on Cosmetic Products (Recast) (Text with EEA Relevance).
 Leite, S., Grunsven, L.A.V., Worth, A., Mahony, C., 2017. Ab initio chemical safety http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:
 assessment: a workﬂ ow based on exposure considerations and non-animal methods. L:2009:342:0059:0209:en:PDF.
 Computational Toxicology 4, 31–44. http://dx.doi.org/10.1016/j.comtox.2017.10. EU, 2009b. Regulation (EC) No 1107/2009 of the European Parliament and of the Council
 001. of 21 October 2009 Concerning the Placing of Plant Protection Products on the
Blackburn, K., Stuard, S.B., 2014. A framework to facilitate consistent characterization of Market and Repealing Council Directives 79/117/EEC and 91/414/EEC. http://eur-
 read across uncertainty. Regul. Toxicol. Pharmacol. 68, 353–362. http://dx.doi.org/ lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32009R1107.
 10.1016/j.yrtph.2014.01.004. EU, 2012. Guidance Document on the Assessment of the Equivalence of Technical
Bossuyt, M.V., Hoeck, E.V., Raitano, G., Manganelli, S., Braeken, E., Ates, G., Vanhaecke, Materials of Substances Regulated under Regulation (EC) No 1107/2009 SANCO/
 T., Miert, S.V., Benfenati, E., Mertens, B., Rogiers, V., 2017. (Q)SAR tools for priority 10597/2003 –rev. 10.1 13https://ec.europa.eu/food/sites/food/ﬁ les/plant/docs/
 setting: a case study with printed paper and board food contact material substances. pesticides\_guidance\_equivalence-chem-substances\_en.pdf.
 Food Chem. Toxicol. 102, 109–119. http://dx.doi.org/10.1016/j.fct.2017.02.002. Ford, K.A., 2016. Reﬁ nement, reduction, and replacement of animal toxicity tests by
Bower, D., Cross, K.P., Esche, S., Myatt, G.J., Quigley, D.P., 2017. In silico toxicology: an computational methods. ILAR J. 57 (2), 226–233. http://dx.doi.org/10.1093/ilar/
 Overview of toxicity databases, prediction methodologies, and expert review. In: ilw031.
 Richardson, R.J., Johnson, D.E. (Eds.), Computational Systems Pharmacology and Freidig, A., Dekkers, S., Verwei, M., Zvinavashe, E., Bessems, J., Sandt, J.V.D., 2007.
 Toxicology. Royal Society of Chemistry. http://dx.doi.org/10.1039/ Development of a QSAR for worst case estimates of acute toxicity of chemically re-
 9781782623731-00209. active compounds. Toxicol. Lett. 170, 214–222. http://dx.doi.org/10.1016/j.toxlet.
Canada, 2016. Chemicals Management Plan. Government of Canada. https://www. 2007.03.008.
 canada.ca/en/health-canada/services/chemical-substances/chemicals-management- Harvey, J., Fleetwood, A., Ogilvie, R., Teasdale, A., Wilcox, P., Spanhaak, S., 2017.
 plan.html. Management of organic impurities in small molecule medicinal products: deriving
Carrió, P., Pinto, M., Ecker, G., Sanz, F., Pastor, M., 2014. Applicability domain analysis safe limits for use in early development. Regul. Toxicol. Pharmacol. 84, 116–123.
 (ADAN): a robust method for assessing the reliability of drug property predictions. J. http://dx.doi.org/10.1016/j.yrtph.2016.12.011.
 Chem. Inf. Model. 54, 1500–1511. http://dx.doi.org/10.1021/ci500172z. Hillisch, A., Heinrich, N., Wild, H., 2015. Computational chemistry in the pharmaceutical
CDRH, 2016. Use of International Standard ISO 10993-1. Biological Evaluation of Medical industry: from childhood to adolescence. Chem. Med. Chem. 10, 1958–1962. http://
 Devices - Part 1: Evaluation and Testing within a Risk Management Process. Guidance dx.doi.org/10.1002/cmdc.201500346.
 for Industry and Food and Drug Administration Staﬀ June 16, 2016. https://www. Hochstein, C., Arnesen, S., Goshorn, J., Szczur, M., 2008. Selected resources for emer-
 fda.gov/downloads/medicaldevices/deviceregulationandguidance/ gency and disaster preparedness and response from the United States national library
 guidancedocuments/ucm348890.pdf. of medicine. Med. Ref. Serv. Q. 27, 1–20. http://dx.doi.org/10.1300/j115v27n01\_01.
 16

G.J. Myatt et al. Regulatory Toxicology and Pharmacology 96 (2018) 1–17
ICH M7, 2017. (R1). Assessment and Control of DNA Reactive (Mutagenic) Impurities in OECD, 2017. OECD Guidelines for the Testing of Chemicals. http://www.oecd.org/
 Pharmaceuticals to Limit Potential Carcinogenic Risk. http://www.ich.org/ chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm.
 ﬁ leadmin/Public\_Web\_Site/ICH\_Products/Guidelines/Multidisciplinary/M7/M7\_R1\_ Patlewicz, G., Ball, N., Booth, E.D., Hulzebos, E., Zvinavashe, E., Hennes, C., 2013a. Use
 Addendum\_Step\_4\_31Mar2017.pdf. of category approaches, read-across and (Q)SAR: general considerations. Regul.
ICH, 2017. International Council for Harmonisation of Technical Requirements for Toxicol. Pharmacol. 67, 1–12. http://dx.doi.org/10.1016/j.yrtph.2013.06.002.
 Pharmaceuticals for Human Use. http://www.ich.org/products/guidelines.html. Patlewicz, G., Roberts, D.W., Aptula, A., Blackburn, K., Hubesch, B., 2013b. Workshop:
JRC, 2014. JRC QSAR Model Database and QSAR Model Reporting Formats. https://eurl- use of “read-across ” for chemical safety assessment under REACH. Regul. Toxicol.
 ecvam.jrc.ec.europa.eu/laboratories-research/predictive\_toxicology/qsar\_tools/qrf Pharmacol. 65, 226–228. http://dx.doi.org/10.1016/j.yrtph.2012.12.004.
 and http://qsardb.jrc.it/qmrf/. Patlewicz, G., 2014. Read-across approaches - misconceptions, promises and challenges
Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating ahead. ALTEX 31, 387–396. http://dx.doi.org/10.14573/altex.1410071.
 the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Patlewicz, G., Ball, N., Boogaard, P., Becker, R., Hubesch, B., 2015. Building scientiﬁ c
 Pharmacol. 25, 1–5. http://dx.doi.org/10.1006/rtph.1996.1076. conﬁ dence in the development and evaluation of read-across. Regul. Toxicol.
Kruhlak, N.L., Benz, R.D., Zhou, H., Colatsky, T.J., 2012. (Q)SAR modeling and safety Pharmacol. 72, 117–133. http://dx.doi.org/10.1016/j.yrtph.2015.03.015.
 assessment in regulatory review. Clin. Pharmacol. Ther. 91, 529–534. http://dx.doi. Patlewicz, G., Worth, A.P., Ball, N., 2016. Validation of Computational Methods.
 org/10.1038/clpt.2011.300. Advances in Experimental Medicine and Biology Validation of Alternative Methods
Mansouri, K., Grulke, C.M., Richard, A.M., Judson, R.S., Williams, A.J., 2016. An auto- for Toxicity Testing. pp. 165–187. http://dx.doi.org/10.1007/978-3-319-33826-2\_6.
 mated curation procedure for addressing chemical errors and inconsistencies in PMTA/FDA, 2016. Premarket Tobacco Product Applications for Electronic Nicotine
 public datasets used in QSAR modelling. SAR QSAR Environ. Res. 27, 911–937. Delivery Systems. Guidance for Industry. U.S. Department of Health and Human
 http://dx.doi.org/10.1080/1062936x.2016.1253611. Services, Food and Drug Administration, Center for Tobacco Products May 2016.
Martin, T., Young, D., Lilavois, C., Barron, M., 2015. Comparison of global and mode of https://www.fda.gov/downloads/tobaccoproducts/labeling/
 action-based models for aquatic toxicity. SAR QSAR Environ. Res. 26, 245–262. rulesregulationsguidance/ucm499352.pdf.
 http://dx.doi.org/10.1080/1062936x.2015.1018939. Powley, M.W., 2015. (Q)SAR assessments of potentially mutagenic impurities: a reg-
Molander, L., Ågerstrand, M., Beronius, A., Hanberg, A., Rudén, C., 2014. Science in risk ulatory perspective on the utility of expert knowledge and data submission. Regul.
 assessment and policy (SciRAP): an online resource for evaluating and reporting in Toxicol. Pharmacol. 71, 295–300. http://dx.doi.org/10.1016/j.yrtph.2014.12.012.
 vivo(Eco)Toxicity studies. Hum. Ecol. Risk Assess. 21, 753–762. http://dx.doi.org/ Raies, A.B., Bajic, V.B., 2016. In silico toxicology: computational methods for the pre-
 10.1080/10807039.2014.928104. diction of chemical toxicity. Wiley Interdiscip. Rev. Comput. Mol. Sci. 6, 147–172.
Mumtaz, M.M., Suk, W.A., Yang, R.S.H., 2010. Introduction to mixtures toxicology and http://dx.doi.org/10.1002/wcms.1240.
 risk assessment. In: Mumtaz, M. (Ed.), Principles and Practice of Mixtures Toxicology. Rastogi, T., Leder, C., Kümmerer, K., 2014. Designing green derivatives of β -blocker
 Mumtaz. M (Ed.) Wiley-VCH, Weinheim, Germany, pp. 1–25. http://dx.doi.org/10. Metoprolol: a tiered approach for green and sustainable pharmacy and chemistry.
 1002/9783527630196.ch1. Chemosphere 111, 493–499. http://dx.doi.org/10.1016/j.chemosphere.2014.03.
Mumtaz, M.M., Durkin, P.R., 1992. A weight of evidence scheme for assessing interac- 119.
 tions in chemical mixtures. Toxicol. Indus. Health 8, 377–406. Rooney, A.A., Boyles, A.L., Wolfe, M.S., Bucher, J.R., Thayer, K.A., 2014. Systematic
Myatt, G.J., Beilke, L.D., Cross, K.P., 2016. In silico tools and their application. In: review and evidence integration for literature-based environmental health science
 Reedijk, J. (Ed.), Reference Module in Chemistry, Molecular Sciences and Chemical assessments. Environ. Health Perspect. 122, 711–718. http://dx.doi.org/10.1289/
 Engineering. Elsevier. https://doi.org/10.1016/B978-0-12-409547-2.12379-0. ehp.1307972.
NAFTA, 2012. TWG Quantitative Structure Activity Relationships [(Q)SAR] Guidance Russell, W.M.S., Burch, R.L., 1959. The Principles of Humane Experimental Technique.
 Document. https://archive.epa.gov/pesticides/news/web/pdf/qsar-guidance.pdf. Methuen, London.
Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., SCCS, 2016. Memorandum on the Use of in Silico Methods for Assessment of Chemical
 Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova- Hazard. SCCS/1578/16. https://ec.europa.eu/health/scientiﬁ c\_committees/
 Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, consumer\_safety/docs/sccs\_o\_200.pdf.
 D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C., 2005. Current status of SCCS, 2017. Scientiﬁ c Committee on Consumer Safety. https://ec.europa.eu/health/
 methods for deﬁ ning the applicability domain of (quantitative) structure-activity scientiﬁ c\_committees/consumer\_safety\_en.
 relationships. The report and recommendations of ECVAM workshop 52. ATLA 33, Schilter, B., Benigni, R., Boobis, A., Chiodini, A., Cockburn, A., Cronin, M.T., Piparo, E.L.,
 155–173. Modi, S., Thiel, A., Worth, A., 2014. Establishing the level of safety concern for
NIEHS, 2017. National Institute of Environmental Health Sciences. https://www.niehs. chemicals in food without the need for toxicity testing. Regul. Toxicol. Pharmacol.
 nih.gov/index.cfm. 68, 275–296. http://dx.doi.org/10.1016/j.yrtph.2013.08.018.
NTP, 2016. West Virginia Chemical Spill: NTP Studies. https://ntp.niehs.nih.gov/results/ Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-
 areas/wvspill/studies/. Ovaskainen, A., Hartung, T., Hoﬀ mann, S., 2009. “ToxRTool ”, a new tool to assess
OCSPP, 2017. Oﬃ ce of Chemical Safety and Pollution Prevention (OCSPP). https:// the reliability of toxicological data. Toxicol. Lett. 189, 138–144. http://dx.doi.org/
 www.epa.gov/aboutepa/about-oﬃ ce-chemical-safety-and-pollution-prevention- 10.1016/j.toxlet.2009.05.013.
 ocspp. Schultz, T., Amcoﬀ , P., Berggren, E., Gautier, F., Klaric, M., Knight, D., Mahony, C.,
OECD, 2004. The Report from the Expert Group on (Quantitative) Structure-activity Schwarz, M., White, A., Cronin, M., 2015. A strategy for structuring and reporting a
 Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs, No. 49 read-across prediction of toxicity. Regul. Toxicol. Pharmacol. 72, 586–601. http://dx.
 (ENV/JM/MONO(2004)24). http://www.oecd.org/oﬃ cialdocuments/ doi.org/10.1016/j.yrtph.2015.05.016.
 publicdisplaydocumentpdf/?cote=env/jm/mono(2004)24&doclanguage=en . Schwetz, B., 1995. Use of mechanistic and pharmacokinetic data for risk assessment at the
OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure-activity national Institute of environmental health sciences (NIEHS). Toxicol. Lett. 79, 29–32.
 Relationships [(Q)SAR] Models, OECD Environment Health and Safety Publications http://dx.doi.org/10.1016/0378-4274(95)03354-n.
 Series on Testing and Assessment No. 69 (ENV/JM/MONO(2007)2. http://www. Seed, M.J., Agius, R.M., 2017. Progress with Structure-Activity Relationship modelling of
 oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure- occupational chemical respiratory sensitizers. Curr. Opin. Allergy Clin. Immunol. 17,
 activity-relationship-q-sar-models-9789264085442-en.htm . 64–71. http://dx.doi.org/10.1097/ACI.0000000000000355.
OECD, 2014. Guidance on Grouping of Chemicals, Second Edition. OECD Environment Stanton, K., Kruszewski, F.H., 2016. Quantifying the beneﬁ ts of using read-across and in
 Health and Safety Publications Series on Testing & Assessment. No. 194 ENV/JM/ silico techniques to fulﬁ ll hazard data requirements for chemical categories. Regul.
 MONO(2014)4. http://www.oecd.org/oﬃ cialdocuments/ Toxicol. Pharmacol. 81, 250–259. http://dx.doi.org/10.1016/j.yrtph.2016.09.004.
 publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en. Sutter, A., Amberg, A., Boyer, S., Brigo, A., Contrera, J.F., Custer, L.L., Dobo, K.L.,
OECD, 2015. Fundamental and Guiding Principles for (Q)SAR Analysis of Chemical Gervais, V., Glowienke, S., Gompel, J.V., Greene, N., Muster, W., Nicolette, J., Reddy,
 Carcinogens with Mechanistic Considerations, Monograph 229 (ENV/JM/MONO M.V., Thybaud, V., Vock, E., White, A.T., Müller, L., 2013. Use of in silico systems and
 (2015)46), Series on Testing and Assessment No. 229. http://www.oecd.org/ expert knowledge for structure-based assessment of potentially mutagenic impurities.
 oﬃ cialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)46& Regul. Toxicol. Pharmacol. 67, 39–52. http://dx.doi.org/10.1016/j.yrtph.2013.05.
 doclanguage=en. 001.
OECD, 2016a. Guidance Document on the Reporting of Deﬁ ned Approaches to Be Used TSCA, 2016. Toxic Substances Control Act (TSCA). https://www.congress.gov/bill/
 within Integrated Approaches to Testing and Assessment, Series on Testing & 114th-congress/senate-bill/697/all-info.
 Assessment No. 255, ENV/JM/MONO(2016)28. http://www.oecd.org/ Worth, A., Barroso, J., Bremer, S., Burton, J., Casati, S., Coecke, S., Corvi, R., Desprez, B.,
 oﬃ cialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)28& Dumont, C., Gouliarmou, V., Goumenou, M., Gräpel, R., Griesinger, C., Halder, M.,
 doclanguage=en. Janusch Roi, A., Kienzler, A., Madia, F., Munn, S., Nepelska, M., Paini, A., Price, A.,
OECD, 2016b. 6. Guidance Document on the Reporting of Deﬁ ned Approaches and Prieto, P., Rolaki, A., Schäﬀ er, M., Triebe, J., Whelan, M., Wittwehr, C., Zuang, V.,
 Individual Information Sources to Be Used within Integrated Approaches to Testing 2014. Alternative Methods for Regulatory Toxicology – a State-of-the-art Review.
 and Assessment (IATA) for Skin Sensitation. ENV/JM/MONO(2016)29. Series on JRC Report EUR 26797 EN. Publications Oﬃ ce of the European Union. http://
 Testing & Assessment No. 25. http://www.oecd.org/oﬃ cialdocuments/ publications.jrc.ec.europa.eu/repository/bitstream/JRC91361/echa\_jrc\_sla\_report\_
 publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29&doclanguage=en . public\_05-09-14\_withcover%20ipo.pdf.
OECD, 2016c. Draft Guidance Document on Good in Vitro Methods Practices (GIVIMP) Wu, S., Blackburn, K., Amburgey, J., Jaworska, J., Federle, T., 2010. A framework for
 for the Development and Implementation of in Vitro Methods for Regulatory Use in using structural, reactivity, metabolic and physicochemical similarity to evaluate the
 Human Safety Assessment. http://www.oecd.org/env/ehs/testing/OECD\_Draft\_ suitability of analogs for SAR-based toxicological assessments. Regul. Toxicol.
 GIVIMP\_in\_Human\_Safety\_Assessment.pdf. Pharmacol. 56, 67–81. http://dx.doi.org/10.1016/j.yrtph.2009.09.006.
 17

Source PDF: /instem.com/www.instem.com/docs/sot-2021/skin-sensitization-protocol.pdf

 Regulatory Toxicology and Pharmacology 116 (2020) 104688
 Contents lists available at ScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepage: www.elsevier.com/locate/yrtph
Skin sensitization in silico protocol T
Candice Johnsona,∗ , Ernst Ahlbergb, Lennart T. Angerc, Lisa Beilked, Romualdo Benignie,
Joel Bercuf, Sol Bobstg, David Bowera, Alessandro Brigoh, Sarah Campbelli, Mark T.D. Croninj,
Ian Crooksk, Kevin P. Crossa, Tatyana Doktoroval, Thomas Exnerl, David Faulknerm,
Ian M. Fearonn, Markus Fehro, Shayne C. Gadp, Véronique Gervaisq, Amanda Giddingsr,
Susanne Glowienkes, Barry Hardyl, Catrin Hasselgrenc, Jedd Hillegasst, Robert Jollyu,
Eckart Kruppv, Liat Lomnitskiw, Jason Magbyx, Jordi Mestresy, Lawrence Milchakz, Scott Millera,
Wolfgang Musterh, Louise Neilsonaa, Rahul Parakhiabb, Alexis Parentys, Patricia Parriscc,
Alexandre Paulinodd, Ana Theresa Paulinodd, David W. Robertsi, Harald Schleckeree,
Reinhard Stidlﬀ , Diana Suarez-Rodrigezgg, David T. Szabohh, Raymond R. Ticeii, Daniel Urbischjj,
Anna Vuorineno, Brian Wallx, Thibaud Weilerq, Angela T. Whiter, Jessica Whritenourkk,
Joerg Wichardee, David Woolleyll, Craig Zwicklmm, Glenn J. Myatta
a Leadscope, Inc, 1393 Dublin Rd, Columbus, OH, 43215, USA
b Bioinformatics Department, The University of Uppsala, 752 36, Uppsala, Sweden
c Genentech, Inc., 1 DNA Way, South San Francisco, CA, 94080, USA
d Toxicology Solutions Inc., San Diego, CA, USA
e Alpha-PreTox, Via G.Pascoli 1, 00184, Roma, Italy
f Gilead Sciences, 333 Lakeside Drive, Foster City, CA, USA
g Toxsci Advisors LLC, 2016 Main Suite 1901, Houston, TX, USA
h Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoﬀ mann-La Roche Ltd, Grenzacherstrasse 124,
4070, Basel, Switzerland
i Nelson Laboratories, LLC, 6280 South Redwood Road, Salt Lake City, UT, 84123, USA
j School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, L3 3AF, UK
k British American Tobacco, Research and Development, Regents Park Road, Southampton, Hampshire, SO15 8TL, UK
l Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, CH-4057, Basel / Basel-Stadt, Switzerland
m Chemical Sciences Division, Lawrence Berkeley National Lab, USA
n whatIF? Consulting Ltd, The Crispin Burr Street, Harwell, OX11 0DT, UK
o DSM Nutritional Products, Kaiseraugst, Switzerland
p Gad Consulting Services, 4008 Barrett Drive, Suite 201, Raleigh, NC, 27609, USA
q Servier Group, 905 Route de Saran, 45520, Gidy, France
r GlaxoSmithKline, Park Road, Ware, Hertfordshire, SG12 0DP, United Kingdom
s Novartis Pharma AG, Pre-Clinical Safety, Werk Klybeck, CH-4057, Basel, Switzerland
t Bristol-Myers Squibb, Drug Safety Evaluation, 1 Squibb Dr, New Brunswick, NJ, 08903, USA
u Toxicology Division, Eli Lilly and Company, Indianapolis, IN, USA
v Sanoﬁ , Corporate HSE, Global Product Steweardship, Industriepark Hoechst, D-65926, Frankfurt Am Main, Germany
w Perrigo Israel Pharmaceuticals Ltd, Shoham, Israel
x Colgate-Palmolive Technology Center, 909 River Road, Piscataway, NJ, 08855, USA
y IMIM Hospital Del Mar Institute of Medical Research and University Pompeu Fabra, Doctor Aiguader 88, Parc de Recerca Biomèdica, 08003 Barcelona, Spain; and
Chemotargets SL, Baldiri Reixac 4, Parc Cientíﬁ c de Barcelona, 08028, Barcelona, Spain
z 3M Company, St. Paul, MN, USA
aa Broughton Nicotine Services, Oak Tree House, West Craven Drive, Earby, Lancashire, BB18 6JZ, UK
bb Church & Dwight Co., Inc, 469 North Harrison Street, Princeton, NJ, 08543, USA
cc Pﬁ zer Worldwide Research and Development, Sandwich, UK
dd ORO AGRI Europe, S.A, Palmela, Portugal
ee Bayer AG, Research & Development, Pharmaceuticals, Industrial Chemicals Toxicology & Genetic Toxicology, 42096, Wuppertal, Germany
ﬀ Safetree Consulting e.U, Vienna, Austria
gg FStox Consulting LTD, 2 Brooks Road, Raunds, Wellingborough, NN9 6NS, UK
hh PPG Industries, Pittsburgh, PA, 15146, USA
ii RTice Consulting, Hillsborough, NC, 27278, USA
jj BASF SE, Product Safety, Carl-Bosch-Strasse 38, 67056, Ludwigshafen Am Rhein, Germany
 ∗ Corresponding author.
 E-mail address: cjohnson@leadscope.com (C. Johnson).
https://doi.org/10.1016/j.yrtph.2020.104688
Available online 01 July 2020
0273-2300/ © 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/BY-NC-ND/4.0/).

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
kk Pﬁ zer Inc., Drug Safety Research and Development, Eastern Point Road, Groton, CT, 06340, USA
ll ForthTox Limited, PO Box 13550, Linlithgow, EH49 7YU, UK
mm Transendix LLC, 1407 Moores Manor, Indianapolis, IN, 46229, USA
ARTICLE INFO ABSTRACT
Keywords: The assessment of skin sensitization has evolved over the past few years to include in vitro assessments of key
In silico events along the adverse outcome pathway and opportunistically capitalize on the strengths of in silico methods
In silico toxicology to support a weight of evidence assessment without conducting a test in animals. While in silico methods vary
Computational toxicology greatly in their purpose and format; there is a need to standardize the underlying principles on which such
Computational toxicology protocols models are developed and to make transparent the implications for the uncertainty in the overall assessment. In
(Q)SAR this contribution, the relationship between skin sensitization relevant eﬀ ects, mechanisms, and endpoints are
Expert alerts built into a hazard assessment framework. Based on the relevance of the mechanisms and eﬀ ects as well as the
Expert review
Skin sensitization strengths and limitations of the experimental systems used to identify them, rules and principles are deﬁ ned for
Deﬁ ned approach deriving skin sensitization in silico assessments. Further, the assignments of reliability and conﬁ dence scores that
Integrated approaches to testing and reﬂ ect the overall strength of the assessment are discussed. This skin sensitization protocol supports the im-
assessment (IATA) plementation and acceptance of in silico approaches for the prediction of skin sensitization.
Extractables and leachables
1. Introduction the European Union; and Section 4(h) (Reduction of Testing in Verte-
 brates) of the Toxic Substances Control Act (TSCA) in the United States.
 Allergic contact dermatitis (ACD) is a common skin condition that These regulations either prohibit the use of animal testing or only allow
results from the induction of a dermal immunological response after animal testing if results obtained by alternative methods are not suﬃ -
repeated exposure to a skin-sensitizing substance. ACD poses a sig- cient to assess the sensitizing potential of a chemical. The “ 3Rs” to-
niﬁ cant public and occupational health concern, and much eﬀ ort has gether with the need for higher throughput and more mechanistically
been dedicated to the identiﬁ cation and classiﬁ cation of skin sensiti- informative methods, continue to drive the development of non-animal
zers. Historically, assessors have relied on human (Human repeat insult methods. In this regard, in silico, in chemico, and in vitro methods in
patch tests (HRIPT) and Human maximization tests (HMT)) or animal concert play an integral role in the hazard assessment of skin sensiti-
testing, the latter commonly using guinea pig (Guinea pig maximization zation.
(GPMT) and Buehler tests(BT))(Organisation for Economic Co-opera- In silico models, along with in vitro tests, have been and continue to
tion and Development (OECD), 1992) and mouse models (Local lymph be developed for predicting the outcome of the four key events (KEs)
node assay (LLNA))(OECD, 2010a) to identify potential skin sensitizers. described in the OECD adverse outcome pathway (AOP) for skin sen-
The guiding principles of the “ 3Rs” (replacement, reduction, and re- sitization (OECD, 2014). It is generally accepted that the skin sensi-
ﬁ nement) as applied to animal research(RUSSELL and BURCH, 1959) tizing hazard of a chemical can be eﬀ ectively assessed through the in-
have inﬂ uenced the implementation of regulations, such as the 7th tegration of non-animal approaches (Kleinstreuer et al., 2018; OECD,
amendment of the Cosmetic Directive (Council Directive 76/768/EEC 2017). However, there may be data gaps that are generated through the
of 1976-07-27; Cosmetics Regulation: REGULATION (EC) No. 1223/ exclusion of chemicals that do not meet the physicochemical property
2009), European substances legislation No. 1907/2006 (Registration, requirements for the in vitro tests, and in silico methods that could be
Evaluation, Authorization and Restriction of Chemicals (REACH)) in used to ﬁ ll such gaps may lack transparency as they are sometimes
 Fig. 1. A generic hazard assessment framework that shows the relationship between the key components of the protocol.
 2

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
viewed as “ black box” tools. There is also no consensus on how to in- outcome and are reﬂ ected in the AOP for skin sensitization (Myatt et al.,
tegrate in vitro data and/or in silico predictions for these events with 2018). The mechanisms and eﬀ ects are assessed based on in silico or
existing in vivo data. existing experimental data. Each mechanism/eﬀ ect assessment is as-
 The protocol detailed in this publication outlines a framework in signed a reliability score (RS) which reﬂ ects the inherent quality of the
which in silico methods could be applied and integrated with existing in assessment Section 2 of the Supplementary Material (SM2). The re-
vivo and in vitro experimental data to identify potential skin sensitizers, levance (scientiﬁ c predictivity) of the eﬀ ect/mechanism is also as-
and to provide consensus on the development of models and the in- sessed. Rules and principles are used to combine the mechanisms/ef-
terpretation of model results. In silico methods are likely to play an fects to derive an assessment of non-apical endpoints (i.e., endpoint 1
important role in understanding the hazard and risk associated with and 2 in Fig. 1) that are relevant for sensitization. The non-apical
chemicals (Myatt et al., 2018). Assessing sensitization is a necessary endpoint assessment is assigned a conﬁ dence score, which is a reﬂ ec-
component of classiﬁ cation and labelling, workers’ safety and occupa- tion of the reliability, relevance, and completeness of the assessment.
tional health (where ~20– 30% of compounds may be sensitizers), Non-apical endpoints are combined via rules and principles to derive an
regulation of cosmetics and other industrial chemicals as well as pro- overall assessment for skin sensitization (the apical endpoint) with an
duct discovery. Previous studies have evaluated the potential use of in associated conﬁ dence score. The framework is designed to derive an
silico tools to predict sensitization hazard or potential (Roberts and assessment for hazard, with risk being outside the scope of the protocol.
Aptula, 2014; Roberts et al., 2006). However, there remains a need for Fig. 2 shows the hazard assessment framework for sensitization and the
in silico guidelines and the deﬁ nition of principles and procedures that relationships between the following endpoints:
are speciﬁ c to the prediction of skin sensitization relevant mechanisms.
To this end, this skin sensitization protocol has been developed based • Covalent interaction with skin proteins
on the experience of a cross-industry consortium comprising 39 dif- • Events in keratinocytes
ferent organizations and represents a consensus of how to use in silico • Events in dendritic cells
methods to predict skin sensitization. • Skin sensitization in vitro (deﬁ ned approach)
 • Skin sensitization in rodent lymphocytes
1.1. Hazard assessment framework (HAF) • Skin sensitization in rodents
 • Skin sensitization in humans (weight of evidence)
 Fig. 1 provides a representation of a generic hazard assessment A comprehensive and mechanistic assessment for skin sensitization
framework. The hazard assessment framework deﬁ nes the relationship includes the four KEs described in the AOP as well as available in vivo
between mechanisms and eﬀ ects that are relevant for the prediction of data and other supporting elements (OECD, 2014). A mechanistic un-
skin sensitization. The mechanisms and eﬀ ects are molecular pertur- derstanding of the sensitizing process is detailed within the AOP for
bations and manifestations, respectively, that lead to the adverse
Fig. 2. The hazard assessment framework describing the in silico components relevant for skin sensitization. In silico models could be developed for any eﬀ ect or
mechanism within grey boxes.
 3

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
skin sensitization and becomes necessary in the development of this considerations (targets on the surface of a protein are more easily ac-
framework. In order for a chemical to exert a sensitizing eﬀ ect, a series cessible than those in folds), and the microenvironment (hydrophilic or
of well-deﬁ ned stages/events occur that lead to the development of hydrophobic) (OECD, 2014). The formation of this complex is critical
eﬀ ector T cells (as opposed to regulatory T cells, which lead to tolerance for the activation of the immunological cells that are responsible for
(OECD, 2014). A chemical's ability to induce each KE is critical in- sensitization.
formation that is used in the development of the HAF. Sensitization is
acquired through two distinct phases. During the initial induction 1.1.2. Key event (KE) 2: Events in keratinocytes
phase, the immune system is primed through dendritic cell presentation It is accepted that interactions with the hapten lead to the mod-
of the sensitizing chemical to naïve T-cells. The induction phase occurs ulation of inﬂ ammation-related pathways and oxidative stress response
upon ﬁ rst contact with the sensitizer and a physiological response is pathways in keratinocytes (OECD, 2014)(Fig. 3).
typically mild or absent. Upon re-exposure to the same sensitizer, the Nuclear factor (erythroid-derived 2)-like 2 (Nrf2) is a transcription
primed immune system is activated and an inﬂ ammatory response oc- factor that trans-locates into the nucleus of keratinocytes and binds to
curs. This phase is called the elicitation or challenge phase and results antioxidant/electrophile response elements (ARE). This in turn, in-
in the manifestation of the symptoms associated with ACD: the ap- itiates the transcription of genes related to oxidative stress responses,
pearance of rashes, blisters, and welts. A comprehensive assessment of such as NADPH-quinone oxidoreductase 1 (NQ01) and glutathione S-
the skin sensitization potential of a chemical includes the four KEs that transferase (GST). Nrf2 is repressed and controlled by the Kelch-like
are described in the induction phase (OECD, 2014). ECH-associated protein 1 (Keap1), which facilitates the ubiquitination
 and degradation of Nrf2. Keap1 is a cysteine (thiol) rich protein which
1.1.1. Key event (KE) 1: Molecular initiating event (MIE) – covalentcan be modiﬁ ed by electrophiles (haptens) and oxidants. This mod-
interaction with skin proteins iﬁ cation to Keap1 induces conformational changes in the protein that
 The MIE for acquiring skin sensitization is the covalent binding of releases bound Nrf2, allowing it to bind AREs and promote the ex-
an electrophilic chemical to a nucleophilic protein, typically the thiol pression of cyto-protective mechanisms (OECD, 2014). In addition,
group of cysteine or the primary amine group of lysine (Fig. 3). The interaction of the hapten with keratinocytes stimulates the production
interaction of the sensitizer (hapten) with the protein leads to the for- of pro-inﬂ ammatory cytokines such as IL-18 (Natsch, 2010). The release
mation of a stable hapten-protein conjugate. While a hapten-bound of cytokines by keratinocytes (among other factors) plays a role in
protein may result from direct interaction of the protein with an elec- stimulating the maturation of dendritic cells (Sumpter et al., 2019).
trophile, some chemicals require either metabolic (pro-haptens), or
abiotic transformation through oxidation (pre-haptens) prior to com- 1.1.3. Key event 3: Events in dendritic cells
plexing with dermal proteins. The hapten-protein interaction depends Langerhans cells and dermal DCs are responsible for the presenta-
on the number of available nucleophilic target residues, steriction of the protein-hapten complex to naïve T-cells in the lymph node
 Fig. 3. Adverse Outcome Pathway (AOP) for skin sensitization. MIE-molecular initiating event, KE (1– 4) - Key Events 1– 4.
 4

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
during the induction phase (Fig. 3). Following the uptake of the protein- prediction of the adverse outcome in humans. The incorporation of
hapten conjugate, DCs process and present these peptide fragments in lines of evidence that may not directly relate to sensitization; such as
the context of major histocompatibity complex (MHC) molecules to skin irritation, means that the protocol takes the form of an integrated
naïve T cells. Matured DCs migrate to the dermis and to the lymph node approach to testing and assessment (IATA).
under the inﬂ uence of cytokines and chemokines that are secreted by
keratinocytes and ﬁ broblast in the dermis (OECD, 2014; Sumpter et al., 1.2. Integrated approach to testing and assessment (IATA)
2019). During maturation, cell surface markers, adhesion molecules,
cytokines, and chemokines are upregulated. The upregulation of co- Given the deﬁ nition of an AOP for skin sensitization and the
stimulatory adhesion molecules (e.g., CD54, CD86) ensures that pro- availability of historical data, the endpoint is eﬀ ectively predicted using
fessional antigen presenting cells develop and initiate an immune re- an IATA. Limited data for the KEs along the AOP have restricted the
sponse. When there is a lack of co-stimulation, T-cell anergy (a state in development and applicability of in silico models to predict these end-
which the lymphocytes remain hypo-responsive after encounter with points while in vitro testing is mainly used to derive an assessment of the
antigen) and a lack of sensitization may result (OECD, 2014; Vocanson activation of KEs along the AOP pathways. This may change in the
et al., 2009). future, as more data become available and more robust in silico models
 can be developed. Nonetheless, through an integrated scheme, the
1.1.4. Key event 4: Events in lymphocytes overall endpoint of ‘ skin sensitization in humans’ is assessed as a
 Presentation of the fragmented peptide complex within the MHC to function of the activity at each KE, with additional evidence from either
naïve T-cells results in their activation. This leads to the diﬀ erentiation existing data or in silico predictions of in vivo responses and metabolic
and proliferation of memory T-cells. Memory T-cells migrate to the biotransformation. Previous research has focused on developing such
dermis and also circulate throughout the body. Upon re-exposure to the schemes and these non-animal integrated strategies are receiving in-
same hapten, the memory T-cells are activated (elicitation phase) and terest from regulatory authorities. The publication of the ‘ Interim
the immune response is triggered; the result is the manifestation of Science Policy: Use of Alternative Approaches for Skin Sensitization as a
ACD, an irreversible immunologic response (OECD, 2014). Replacement for Laboratory Animal Testing’ is an example of regulators
 KE 1– 4 can be used to assess the ‘ skin sensitization in vitro endpoint’ , adopting this more integrated approach (EPA, 2018). Additional non-
which in turn can be extrapolated to the ‘ skin sensitization in humans’ animal assessment strategies are currently being developed and vali-
endpoint as shown in Fig. 2. These in vitro endpoints can also be pre- dated, and more approaches may be adopted for regulatory purposes in
dicted by in silico models as outlined in the HAF (Fig. 2) and described the future (Kleinstreuer et al., 2018). While several integrated ap-
in Section 2. proaches invoke the AOP and integrate the KEs to derive an overall
 The availability of in vivo (usually rodent) data is relevant to the assessment of skin sensitization, it has been argued that failure or
overall assessment of ‘ skin sensitization in humans’ and facilitates the ability to sensitize could be explained by (in)suﬃ cient activity in the
development of in silico methods to predict the results. KE 4 (lympho- ‘ covalent interaction with skin proteins’ endpoint, and the evaluation of
cyte activation and proliferation) can be measured with an in vivo subsequent KEs is less important (Roberts and Aptula, 2008). To this
mouse model and the adverse outcome (e.g., erythema) can be assessed end, the authors believe that a HAF that can facilitate multiple ap-
in guinea pigs. The events in lymphocytes (when assessed in mice) and proaches is necessary. The ideal framework should be generic enough
the guinea pig assessments can be combined to provide an overall as- to facilitate possible variations in analysis while maintaining a high
sessment of ‘ skin sensitization in rodents’ . Skin irritation may be a level of reproducibility and transparency. Rules and principles for
confounding factor and so is also considered at this point. An overall combining results for each endpoint are deﬁ ned in this protocol. These
assessment of ‘ skin sensitization in humans’ can be determined through rules will set the foundation for the reproducibility and ﬂ exibility of the
the integration of the ‘ skin sensitization in vitro’ and ‘ skin sensitization framework presented here.
in rodents’ endpoints. Historical human test data may also be available
and in silico models can be developed to facilitate its prediction. This 1.3. Deﬁ ned approaches
information also propagates into the ‘ skin sensitization in humans’
endpoint. Previous approaches have incorporated rules that connect various
 The HAF consists of evaluation of KE1-4 via in vitro or in vivo testing, aspects of the toxicological pathway to skin sensitization. The “ 2 out of
physio-chemical properties, and human data (Fig. 2). The assumption is 3” integrated testing strategy approach to skin sensitization hazard
made that all chemicals are capable of dermal penetration as a con- identiﬁ cation proposed by BASF uses a data interpretation procedure
servative measure (Fitzpatrick et al., 2017). The endpoints in the fra- (DIP) that labels a chemical as a sensitizer or non-sensitizer based on
mework may be informed through available data, in silico predictions, the concordant reactivity of the chemical in two in vitro tests for KE1 -
or data acquired through conducting a test. The protocol deﬁ nes gen- KE3 (Urbisch et al., 2015). Several other integrated strategies have been
eral rules and principles for integrating data towards an overalldeveloped to assess either hazard or potency (Section 1 of the
Table 1
Sources of data for the development of in silico methods.
 Database Description
 NTP-ICE Integrated Chemical Environment (ICE), an open access database with results from NTP Interagency Center for the Evaluation of Alternative Toxicological
 Methods (NICEATM)
 SkinSensDB SkinSensDB is a collection of data from published literature to facilitate the development of AOP-based computational prediction methods(Wang et al., 2017)
 ECHA-CHEM European Chemicals Agency (ECHA) database is an open access database containing data for chemicals manufactured and imported in Europe. Although the
 summaries are publicly available, extracting data in large amounts requires special consideration as the studies are proprietary
 TOXNET-HSDB Hazardous Substances Data Bank (HSDB) is an open source database that provides information on human exposure to potentially hazardous chemicals
 EURL-ECVAM- The European Union Reference Laboratory for alternatives to animal testing database service on alternative methods to animal experimentation is an open
 DB-ALM access database, containing information on percutaneous absorption
 CosIng European Commission database of current and historical data for cosmetic substances and ingredients
 RIFM The Research Institute For Fragrance Materials (RIFM) monographs contain human health and toxicological data for fragrance and ﬂ avor raw materials.
 Proprietary Databases generated within a speciﬁ c institution. Structure activity relationship (SAR) ﬁ ngerprints
 Literature Manual curation of peer-reviewed articles and published training sets such as Cronin and Basketter (1994)
 5

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
supplementary materials and described in detail elsewhere (OECD, 2.1. Covalent interaction with skin proteins, KE1
2017)). Each approach addresses particular elements of the AOP. At the
time of this manuscript, no single approach is viewed as being superior In silico and or experimental assessments for whether a given com-
to the others and selected approaches vary based on the availability of pound will participate in covalent interactions with skin proteins are
computational tools and data. primarily generated based on understanding of metabolism, reaction
 domain assignment and protein reactivity.
2. In silico methodologies and models 2.1.1. Dermal metabolism
 The allergenic potential of a chemical may be increased or de-
 Historically, in silico models have focused on the prediction of an- creased through metabolic pathways or abiotic oxidation; these factors
imal data (particularly the LLNA), and few have considered the rest of are important for predicting a chemical's potential to induce dermal
the mechanisms established in the AOP. Therefore, it is necessary to sensitization. Metabolic detoxiﬁ cation takes place in two phases, which
examine how in silico tools could be developed to model mechanisms may or may not occur simultaneously. Phase II metabolism appears to
related to the KEs described earlier. be more abundant and active in the skin than in the liver, although
 Depending on the availability of high-quantity data, diﬀ erent types Phase I enzymes – though not dominant – are inducible in the skin
of in silico models can be developed. Table 1 provides a list of data (Dumont et al., 2015). Given di ﬀ erences in expression proﬁ les between
sources. Larger amounts of data, preferably with a strong mechanistic the liver and skin, the potential use of liver metabolic data to predict
understanding of a speciﬁ c toxicological process, can support many metabolites in the skin will necessitate strategies for accounting for the
diﬀ erent types of models. Datasets that cover a broad chemical space diﬀ erences in the expression of isoenzymes between liver and skin
can support the development of global Quantitative Structure-Activity (Madden et al., 2017). 2 One strategy for predicting metabolic activation
Relationship ((Q)SAR) models, provided that the descriptors are re- towards sensitization in dermal tissues is to derive alerts to indicate if a
levant and mechanistically-related to the endpoint that is being pre- chemical may be a pro-hapten. This approach is currently limited by the
dicted (Roberts et al., 2007). Where data are sparse, generated with size of the databases of pro-haptens and a general lack of skin speciﬁ c
diﬀ erent protocols, or generated through multiple mechanistic path- data (although knowledge has been gained through experience over the
ways (as may be the case in human studies), methods such as expert- years). Currently, it appears that the range of structural features that
alerts or read-across may be more appropriate.1 Statistical models may are activated towards sensitization via metabolic pathways is small.
also be developed; however, these models are potentially limited by a Given the absence of skin-speciﬁ c metabolic data, it is challenging to
smaller applicability domain. On the other hand, the mechanistic un- deﬁ nitively conclude on the topic. Natsch and Haupt (2013) in-
derstanding and classiﬁ cation of chemicals into a mechanistic domain vestigated the activation of pro-haptens by rat liver S9 fractions in the
means that local QSAR modeling may be a feasible approach for as- KeratinoSens™ assay, and identiﬁ ed phenolic and alkoxy groups at-
sessing events related to the sensitizing endpoint. One of the earliest tached to a benzene ring, some aromatic amines, and conjugated dienes
attempts to develop a local mechanism-based QSAR model to predict in or in conjunction with six-membered ring as structural features that
EC3 concentrations in the LLNA, used the Relative Alkylation Index may require pro-activation to behave as haptens in the assay (Natsch
(RAI, a function of electrophilic reactivity, lipophilicity, and dose) and Haupt, 2013; Basketter et al., 2005; Basketter et al., 2005). The
(Roberts et al., 1991; Roberts and Williams, 1982). Subsequently, sev- features identiﬁ ed do not represent a comprehensive and thoroughly
eral Quantitative Mechanistic Models (QMM) have been developed deﬁ ned list of features that undergo metabolic transformation leading
with the goal of identifying physicochemical and other descriptors that to sensitization.
contribute to a mechanistic understanding of an endpoint of interest
(Aptula and Roberts, 2006; Roberts and Aptula, 2014; Roberts et al.,2.1.2. Reaction domain
2011; Roberts and Andreas, 2009). The rest of Section 2 discusses the Existing mechanistic information on hapten-protein interactions has
mechanisms or eﬀ ects that could be predicted and which types of in been used to construct in silico models for predicting sensitization po-
silico methodologies could facilitate the predictions. On a general note, tential based on a compound's structure and known – or predicted –
in silico methods typically derive structure activity relationships (SAR) reaction chemistry. The mechanisms for forming protein-hapten com-
for organic salts by using the structure of the freebase. In cases where a plexes involve the interaction between an electrophilic chemical
metallic fragment will be removed in the generation of the freebase to (hapten) and the nucleophilic moiety on a skin protein (generally thiol
derive the SAR form of the structure, the potential hazard posed by the or primary amine groups). Common mechanisms by which the sensi-
metal should be considered. In the area of skin sensitization, removing tizer (hapten) may bind to the protein are: Michael addition, acylation,
nickel fragments may lead to an underestimation of hazard for struc- Schiﬀ base formation, unimolecular nucleophilic substitution (SN1),
tures that contain them. To more accurately facilitate predictions in bimolecular nucleophilic substitution (SN2), or nucleophilic aromatic
these cases, the metal may be attached to the ligand, or the metal may substitution (SNAr). Within each of these mechanistic domains, there
be kept unattached in the training set. The model builder may also are mechanistic alerts and structural alerts. Structural alerts are deﬁ ned
decide to remove the salt structure entirely from the training set; as molecular substructures that can activate the toxicological eﬀ ect or
thereby, excluding the metal from the applicability domain of the mechanism (Myatt et al., 2018). Structural alerts that are characterized
model. by a common reaction site are deﬁ ned as mechanistic alerts (Aptula and
 The following sections describe general considerations for building Roberts, 2006; Enoch et al., 2008; Roberts et al., 2015). Structural and
in silico models based on the available chemistry, biology, and testing mechanistic data do not always suggest a toxic eﬀ ect, however – some
data. Section 1 of the supplementary material provides a detailed de- structural features, such as steric hinderance, have been found to mi-
scription of the experimental data that are relevant for assessing skin tigate toxicity by decreasing the ability of the hapten to covalently bind
sensitization. Methods to assess the reliability of the data as well as in to proteins – and these features may improve an in silico model by
silico predictions have been previously described by (Myatt et al., 2018) providing this additional information.
and are summarized in Section 2 of the supplemental material. Classiﬁ cation of mechanistic and structural alerts within mechan-
 istic domains allows for local QSAR modelling within each domain
 1 The reader is referred to (Myatt et al., 2018) for a more general discussion 2 The supplemental material provides a brief summary of the diﬀ erences
on these methods. between skin and liver metabolic enzymes with relevance to humans.
 6

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
(OECD, 2011), provided that one has the relevant quantitative in- modiﬁ cation of the cysteine-rich Keap1 protein could be used to de-
formation describing the protein-hapten bond. To this end, the fol- velop mechanistically-relevant QSAR models. There may be limitations
lowing physical-chemical property descriptors are commonly used to in predicting compounds which preferentially bind hard nucleophiles
predict interactions between haptens and proteins: Molecular weight such as lysine since the in vitro tests predicting KE2 rely on the cysteine-
(MW), Log P, solubility, rotational bonds, electronic and topological dependent modiﬁ cation of Keap1. Therefore, false negative predictions
descriptors (e.g., quantum mechanics calculations), or chemical struc- may be more common for compounds that react via acyl transfer,
ture-based descriptors (e.g., the presence or absence of diﬀ erent func- within the domain of Schiﬀ base formers, including short chain alde-
tional groups) (OECD, 2011). The factors constituting an acceptable hydes, and longer chain saturated alkanals. Other electrophiles that
and validated model have been described in previous work (Myatt prefer hard nucleophiles may also produce false negative predictions
et al., 2018). However, it must be noted that due to the expert nature of (Urbisch et al., 2015). This could be a potential issue in read-across
deriving structural alerts based on reaction chemistry, existing in silico analysis and should be addressed during an expert review.
tools can only incorporate our current knowledge of protein-hapten In silico prediction of KeratinoSens™ and LuSens (in vitro test
reaction chemistry (rather than the quantiﬁ cation of a physical or methods for assessing ARE activation in keratinocytes) data yields di-
biological process), and that future models could be improved as we chotomous (either positive or negative) test results (OECD, 2018b).
increase our mechanistic understanding of these processes. QSARs on However, integrated assessments of potency may require continuous
the other hand, are not limited by current knowledge of mechanistic data input such as EC1.5 (the lowest concentration inducing a 1.5-fold
processes and the combined use of structural alerts and QSARs may add change in luciferase activity), IC50 (concentration for 50% reduction of
value to the analysis. viability) and EC3 values (concentration with 3 fold luciferase induc-
 tion) (Natsch et al., 2015).
2.1.3. Protein reactivity
 Protein reactivity has been studied using model nucleophiles to 2.3. Events in dendritic cells, KE3
assess protein-chemical interaction in in chemico assays. While the
binding mechanism between the protein and the chemical could be Dendritic cell activation is similar to keratinocyte activation in that
described based on reaction chemistry as discussed in the previous predictions can be made on the levels of protein and gene expression.
section, any in silico tools (either statistical or expert rule-based) de- Methods have been validated for measuring the expression of speciﬁ c
veloped based on in chemico assay data will be limited in their ability to cell surface markers which contribute to T cell activation and pro-
predict sensitization due to pro-activation. To overcome this limitation, liferation. Published databases may contain data for dendritic cell gene
predictions based on reaction chemistry, protein reactivity, and dermal expression of co-stimulatory and adhesion molecules (cell surface
metabolism should be considered in concert to generate an overall as- markers: CD54 and CD86) and Interleukin-8 (IL-8) (Nukada et al., 2011;
sessment (described in Section 3.1). Urbisch et al., 2015).
 While protein reactivity measurement is feasible across all reaction As noted for the KE2 endpoint, care must be taken when integrating
domains described in section 2.1.2, experimental results show that testing data from the various in vitro assays into KE3 in silico models due
within the domain of Schiﬀ base formers there is a lower correlation to diﬀ erences in the types of data that may be produced by diﬀ erent
between the in chemico-based DPRA model and in vivo and human data assays. The continuous data outcomes predicted for these assays, such
(Urbisch et al., 2015). While Schi ﬀ base formation may be theoretically as the EC150 and EC200 values from the h-CLAT assay; the CV70 and the
feasible, the abundance of water within the peptide reactivity testing EC150 in the U-SENS™ assay could be used in integrated strategies to
environment may limit some reactions. As such, peptide reactivity was predict potency. These and other in silico predictions of the Ind-IL8LA
found to correlate poorly with the potency of aldehydes, as Schiﬀ base (induced interleukin-8 luciferase activity) could be used to support the
formation may be limited under testing conditions in the DPRA (Natsch hazard assessment; however, since a statistically-derived experimental
et al., 2015). Further analysis revealed that more potent Schiﬀ base variable (conﬁ dence interval) is needed to determine a positive call, a
formers (atranol, chloratranol, and salicylaldehyde) are reactive under more practical approach may be to dichotomize the assay results and
physiological conditions (Natsch et al., 2012). However, the LLNA EC3 make binary predictions.
values of Schiﬀ base formers are well correlated (R2 = 0.95) with a Often, it is helpful to build models that use threshold values to
combination of logP and a reactivity parameter based on substituent convert continuous data into dichotomous (yes or no) values. For any of
constants (Roberts et al., 2006). Diﬀ erential reactivity within a me- the in vitro or in chemico test methods that are used to assess a KE along
chanistic domain is an issue that could become relevant in the devel- the AOP, using threshold values, in silico predictions could generate
opment of in silico models, and particularly in those that use read-dichotomous predictions of KE activity using these in vitro or in chemico
across. Such instances may not be unique to the protein reactivity test endpoints.
mechanism but may require examination across all toxicological end-
points. 2.4. Events in human lymphocytes, KE4
2.2. Events in keratinocytes, KE2 The lack of standardized data makes in silico predictions of in vitro T
 cell activation and proliferation challenging. A paucity of data for this
 A comprehensive prediction of keratinocyte activation covers events endpoint is not surprising, however, as the value of predicting this key
on several levels of biological organization and includes the expression event remains in question, and the signiﬁ cance of an in vitro estimate of
of biochemical, genomic, and proteomic pathways, and quantiﬁ es the KE4 can only be speculated at this time. It is possible that the magni-
release of pro-inﬂ ammatory mediators that stimulate dendritic cells in tude of the T cell responses at KE4 may be the key event that allows us
KE3 (OECD, 2014). Validated protocols are established for assessing the to make distinctions between diﬀ erent potency classes in vitro (OECD,
induction of ARE dependent pathways, and, as such, the development 2014), but the issue has not been settled. Consequently, only the in vivo
of in silico models can be considered for this assessment. However, the Local Lymph Node Assay has been accepted as a standardized method
breadth of information and data describing other pathways could be for assessing this endpoint.
informative and may drive the development of in silico models to pre-
dict additional pathways in the future. 2.5. Events in rodent lymphocytes, KE4
 Statistical modelling is feasible; however, the availability of data is a
critical factor inﬂ uencing the success of measures to implement models The LLNA is the only standardized in vivo method used to measure
based on AOP in vitro tests. Descriptors relating to the covalentthe proliferation of lymphocytes in response to immune system priming
 7

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
by a test chemical as well as the potency of the chemical as a skinpositive prediction of the “ skin sensitization in humans” endpoint is
sensitizer. The results of the assay are reported as the concentration of low. The reverse may also be considered: If SItest >SISLS and an
the chemical needed to induce T-cell proliferation by a pre-chosen alerting structure exists for sensitization; then the chemical may be
factor (usually 3, 1.6, or 1.8 times the baseline amount as assessed by suspected to be a true positive (Basketter et al., 2009). The con ﬁ dence
the stimulation index (SI))(OECD, 2010b; 2010a, 2018a). The LLNA has could be adjusted accordingly based on the weight of evidence pre-
been used extensively, and it is quite feasible to build in silico models sented. This sort of analysis would be considered with a low reliability
using statistical and rule-based methods due to the ready availability of LLNA study which may have been conducted at irritant concentrations.
data, although, the majority of such data is proprietary. While the Generally, the LLNA test is preceded by dose ﬁ nding range studies and
publicly-available LLNA data could facilitate statistical modeling, the minimally irritating to not irritating concentrations are tested.
model coverage may be reduced for industrial applications. However, Some LLNA protocols (LLNA-DA, and LLNA-BrdU-ELISA) use non-
the combined use of statistical modeling and structural alert deﬁ nitions radioactive methods to quantify lymphocyte proliferation. Results from
could be a strategy to overcome this limitation. these protocols could be combined in training sets that would facilitate
 The irritation potential of a chemical could be a confounding factor binary level predictions; however, varying criteria for predicting a po-
in the experimental LLNA, and the issue of irritation translates into in sitive call may complicate the prediction of a meaningful continuous SI
silico assessments. Training set examples and analogs under con- or ECX value (where x is 3, 1.6, or 1.8 depending on the LLNA protocol
sideration for read-across should be examined for their irritation po- used) from such a dataset and would require a valid strategy for in-
tential. Studies indicate that non-sensitizing irritants (such as surfac- tegrating the data. Another relevant issue with LLNA datasets that
tants) could be overestimated by the LLNA, leading to false positive arises in the curation process is the comparison and combination of SI
results (Ball et al., 2011; OECD, 2010a). While this is certainly the caseand EC3 values for tests conducted in diﬀ erent vehicles. While it seems
for sodium lauryl sulfate (SLS), chloroform/methanol, Triton X-100, logical that vehicle eﬀ ects are normalized in the derivation of the SI and
oxalic acid, methyl salicylate, and nonanoic acid, analysis of chemicals EC3 values, there are mechanisms that could lead to enhanced bioa-
known to be skin irritants has not validated this generalization across vailability depending on the choice of vehicle. The rapid evaporation of
the entire class of non-sensitizing irritants(Ball et al., 2011). Most non- acetone, for example, may result in volatilization of the test chemical
sensitizing irritants are negative in the LLNA and those that are positive and decreased bioavailability; whereas dimethyl sulfoxide (DMSO)
may produce borderline results (with few exceptions). For example, the could potentially enhance penetration. Diﬀ ering results may be ob-
sensitization hazard of SLS is derived from a clear dose-response curve tained between two LLNA tests using diﬀ erent vehicles and this could
that is indicative of a positive LLNA result; however, when a weight-of- inﬂ uence hazard assessment (Hoﬀ mann, 2015). In some cases, vehicle
evidence (WoE) approach is used, the interpretation of the LLNA results eﬀ ects may lead to the assignment of a chemical to two neighboring
may be reversed. There is no evidence that SLS is a skin sensitizer in potency classes (Anderson et al., 2011; Basketter et al., 2001; Dumont
humans despite exposure; albeit limited, it lacks a structural alert for et al., 2016; Hoﬀ mann, 2015). This inherent variability in the LLNA
sensitization and is a strong irritant (Basketter et al., 2009). Hence, data (not exclusively caused by diﬀ erent vehicles) is translated to in
Basketter et al., 2009 have suggested that for the SI results obtained for silico predictions. When combining multiple data sources, the most
SLS in the LLNA (SISLS), a WoE approach could be developed around the conservative SI and ECx values could be adopted, unless there is com-
false positive result to implement this approach in a general sense. pelling evidence that the vehicle is potentiating or attenuating the ef-
Using SLS as reference for a test chemical with unknown skin sensiti- fect of the test chemical. A less conservative, but valid, approach is to
zation hazard, irritant potential and SI predictions (SItest); if theuse the mean, or median values, among other valid approaches
SItest <SISLS and no structural alert exists of sensitization, then the (Hoﬀ mann et al., 2018).
LLNA prediction could be a suspected false positive and conﬁ dence in a
 Fig. 4. The hazard assessment framework annotated with sections that discuss the assessment and conﬁ dence score of each endpoint.
 8

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
2.6. Skin sensitization in rodents Fig. 4 shows the hazard assessment framework annotated with refer-
 ences to where each of the following sections applies.
 The skin sensitization in rodent endpoint is evaluated through the
use of the GPMT and the BT method. Guinea pigs were historically used 3.1. Covalent interaction with skin proteins assessment
to assess skin sensitization. Similar to the LLNA, while public data are
available, much of the GPMT and BT data are proprietary. The data that Assessment of the ‘ covalent interaction with skin proteins’ endpoint
exist could facilitate statistical modeling, the derivation of expert alerts, includes consideration of metabolic transformation, reaction chemistry,
and read-across. and DPRA/ADRA predictions. Fig. 5 shows how rules could be made
 around the available information to derive an overall prediction of
2.7. Skin sensitization in humans hazard. If an experimental result is positive for the methods assessing
 KE1 (DPRA/ADRA), then a positive assessment of the ‘ covalent inter-
 Historical data exist for this endpoint and, based on data quantity, action with skin proteins’ is warranted. However, the reliability of the
expert-alert derivation and read-across may be preferable to statistical prediction, as assessed by the scheme presented in Table 6 of the sup-
methods. In silico predictions could be useful for the prediction of di- plementary material and described in (Myatt et al., 2018), varies de-
chotomized results of positive/negative. Potency predictions could be pending on the quality of the information presented and this has an
challenging based on data availability. Evidence to support human inﬂ uence on the conﬁ dence score. The quality and reliability of an in
predictions includes clinical data (DPT) and usage/occupational ex- silico DPRA/ADRA prediction could be assessed according to the expert
posure data (Api et al., 2017). Further, the integration of the ‘ skinreview criteria described in (Myatt et al., 2018). Additional con-
sensitization in vitro’ and the ‘ skin sensitization in rodents’ endpoints, siderations for both experimental (test article) and in silico (training set
along with any direct human evidence, are considered together as examples and analogs) results include situations in which DPRA/ADRA
weight of evidence for the prediction of the ‘ skin sensitization in hu- could lead to a false positive result due to oxidizing properties of the
mans’ endpoint. test chemical, which can lead to peptide dimerization. An expert review
 could inform on whether or not this is likely and if the assessment and
3. Endpoint assessment and conﬁ dence conﬁ dence score need adjustment. Assessments of negative DPRA/
 ADRA results vary based on consideration of the metabolic potential of
 The protocol details the integration of data with diﬀ erent reli- the chemical together with knowledge of reaction chemistry. In general,
abilities and relevance. Further, there may be cases in which informa- when the chemical is expected to be out of the metabolic domain of the
tion that is critical to an assessment is missing. This section outlines the DPRA/ADRA then precedence is given to clearly-deﬁ ned knowledge of
rules/principles that could be applied when deriving an assessment and reaction chemistry (including mitigating factors, such as sterics) in the
its associated conﬁ dence based on the totality of evidence presented. overall assessment of the ‘ covalent interaction with skin proteins’
 Fig. 5. Decision tree showing how an
 overall assessment and conﬁ dence score
 could be derived for the covalent interac-
 tion of skin proteins. The conﬁ dence scores
 are based on RS1 experimental data: as-
 suming relevant data and high reliability,
 and, in practice, conﬁ dence scores may
 need to be adjusted based on reliability
 scores, SM Table 8. \*If a pro-reactivity do-
 main is assigned and the metabolic site
 (determined using structural alerts for skin
 metabolism) coincides with the pro-re-
 activity domain center then the reversal in
 assessment occurs. If the metabolic site and
 the reactivity domain center do not align
 then the assessment is inconclusive. §§The
 inconclusive result is applicable in situa-
 tions where structural alerts could be used
 to determine if a structure is expected to
 undergo metabolism but not identify the
 metabolites. In this case, since the reactivity
 of the metabolite cannot be conﬁ rmed, a
 conclusion cannot be made on the assess-
 ment. If the reactivity of the metabolites
 could be predicted then the ﬁ nal assessment
 depends on the metabolite reactivity.
 9

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
endpoint. If the reaction chemistry indicates a mechanism leading to model nucleophile may lead to false negative predictions, although this
sensitization; particularly if the mechanism requires pro-activation then occurs to a lesser extent in the ADRA than in the DPRA (Fujita et al.,
the overall assessment of the ‘ covalent interaction with skin proteins’ is 2019).
positive based on reaction chemistry knowledge, but the conﬁ dence is In cases where the DPRA/ADRA result is positive, but no mechan-
medium. If the test article is out of the metabolic domain, negative in istic alert can be assigned, it is worth considering whether mechanistic
DPRA/ADRA and no mechanistic alert could be identiﬁ ed in theknowledge could be provided by the protein reactivity results particu-
structure of the test chemical based on reaction chemistry, then the larly when close analogs point to the same structure-activity relation-
DPRA/ADRA result is inconclusive as it cannot be said that the overall ship. Fig. 5 shows the ‘ covalent interaction with skin proteins’ endpoint
assessment is either negative or positive. However, if metabolism is not and the conﬁ dence score decision tree based on RS1 data. The con-
predicted to occur and the chemical is considered within the metabolic ﬁ dence scores are expected to vary based on reliability and relevance;
domain of the DPRA/ADRA, then the negative result should be given as such, there are several possible permutations of the decision tree.
consideration in the overall assessment. A negative DPRA/ADRA pre- These general “ rules” are expanded to provide a sense of the conﬁ dence
diction (within the DPRA/ADRA metabolic domain) and a positive assigned to assessments with varying reliabilities and relevance, Sup-
mechanistic alert lead to a negative overall assessment, with a medium plementary Material, section 4 (SM 4).
conﬁ dence level, given that the DPRA/ADRA result is experimental and
the positive mechanistic alert introduces some uncertainty. An expert 3.2. Events in keratinocytes
review would consider whether or not the test chemical is within the
Schiﬀ base reaction domain. In these cases a negative DPRA/ADRA The conﬁ dence score obtained for the activation of the events in
result may be mechanistically justiﬁ able due to the protein-hapten in- keratinocytes towards skin sensitization varies based on the Log Kow of
teraction being unfavorable under the test conditions as a result of the the chemical. If there is a positive prediction (RS1, experimental) and
abundance of water; particularly for chemicals that are indicated as less the Log Kow is < 5, then the result is assigned a high conﬁ dence. If the
potent sensitizers by other methods. In this case, the overall assessment Log Kow is greater than 5, then the conﬁ dence is medium for a positive
could be considered positive (after expert review) with a low con- result and low for a negative prediction, since limited information is
ﬁ dence. This positive result is based on giving greater precedence to the available for such chemicals (OECD, 2018b). Regardless of Log Kow
mechanistic alert within this domain, and the decreased relevance of values, negative results could be further assessed based on the occur-
the DPRA/ADRA due to the diﬀ erential reactivity of chemicals within rence of metabolism and the chemical mechanism of action.
the Schiﬀ base domain. Further, co-elution of the test article with the A metabolic alert (indicative of an expected metabolic
Fig. 6A. Decision trees showing how an overall assessment and conﬁ dence score could be derived for the ‘ events in keratinocytes’ . The conﬁ dence scores here are
based on RS1 experimental data: assuming relevant data and high reliability, and, in practice, conﬁ dence scores may need to be adjusted based on reliability scores.
 10

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
transformation) along with a negative RS1/2 experimental or RS3 in assays. Where a false negative seems likely, the low conﬁ dence is ap-
silico result, could indicate reduced relevance of the in vitro assays propriate. In cases where the analogs are true negatives, the conﬁ dence
predicting KE2 in this case – possibly because limited metabolic com- score could be increased to a medium level and this reﬂ ects that while
petency of the cells used in the assay are responsible for a false nega- uncertainty is somewhat reduced, there is not absolute certainty in the
tive. Therefore, the overall assessment would be negative but with a assessment. Within any other domain, a negative KE2 prediction is
low conﬁ dence score. If there is no biochemical transformation pre- considered with high conﬁ dence, given RS1/2 data. Varying reli-
dicted, then the chemical mechanism of action could be considered. A abilities of the data could change the conﬁ dence scores in Fig. 6A and B
negative assessment for a chemical within the acyl transfer domain and (see SM Table 9).
Schiﬀ Base domain is conservatively assigned a low conﬁ dence score
based on the preference of chemicals within these domains for the ly-
sine instead of the cysteine moiety (representing decreased relevance). 3.3. Events in dendritic cells
It is worth mentioning that some chemicals within these domains are
accurately predicted as true negatives and a review of the relevance is An overall assessment of the events in dendritic cells could be made
necessary to assign a higher conﬁ dence. Such a review might include an based on the h-CLAT (Fig. 7), U-SENS ™ or IL-8 Luc assays (Fig. 8). A
examination of close analogs (or the test structure if data is available) positive response from these assays typically translates to a positive
for their assessment in the DPRA/ADRA and or an animal model. If overall call for the events in dendritic cells with high conﬁ dence in the
close analogs are positive in the DPRA/ADRA and the lysine moiety; but activation of the dendritic cells towards sensitization, but an expert
not cysteine, is implicated for covalent modiﬁ cation then the relevance reviewer would be needed to adjust overall calls and conﬁ dence scores
of the KE2 assays for predicting the test structure may be challenged. for certain chemical classes, structural features, and physical-chemical
However, if cysteine modiﬁ cation is apparent in the DPRA/ADRA properties. For example: some chemical classes, such as surfactants,
(positive for covalent interaction with skin proteins), it is more diﬃ cult may lead to false positive results in the U-SENS™, and a negative result
to challenge the relevance of the KE2 assays on that basis and con- for a chemical that has a Log Kow greater than 3.5 is considered in-
ﬂ icting information is presented by the two KEs. The analogs may be conclusive for the h-CLAT. The pro/pre-hapten status of the test che-
further assessed and screened for existing animal data and/or in silico mical is also relevant in each of the three assays. Negative results for
predictions of the LLNA or GPMT. This serves the purpose to assess the structures in which a site of metabolism leading to sensitization has
likelihood of a false negative prediction of the test structure by the KE2 been identiﬁ ed are accepted with a medium level conﬁ dence from the
 h-CLAT, U-SENS™ and IL-8 assays. In cases where there are no
Fig. 6B. Decision trees showing how an overall assessment and conﬁ dence score could be derived for the ‘ events in keratinocytes’ . The conﬁ dence scores here are
based on RS1 experimental data: assuming relevant data and high reliability, and, in practice, conﬁ dence scores may need to be adjusted based on reliability scores.
 11

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
Fig. 7. Decision tree showing how an overall assessment and conﬁ dence score could be derived for the ‘ events in dendritic cells’ based on the h-CLAT assay. The
conﬁ dence scores here are based on RS1 experimental data: assuming relevant data and high reliability, and, in practice, conﬁ dence scores may need to be adjusted
based on reliability scores.
additional parameters confounding the prediction, then the conﬁ dence with skin proteins’ , ‘ Events in keratinocytes’ , and ‘ Events in dendritic
level is high for the negative predictions from the h-CLAT, U-SENS™, cells’ (KEs in the AOP) are considered. These KEs are assessed based on
and IL-8 Luc assays. knowledge of reaction chemistry and mechanistic understanding that is
 not explicitly considered within the “ AOP 2 out of 3” approach. Similar
3.4. Skin sensitization in vitro to the “ AOP 2 out of 3,” an overall assessment of hazard for the ‘ skin
 sensitization in vitro’ endpoint is determined based on a 2 out of 3
 Integrating data to derive an overall assessment for the ‘ skin sen- consensus among the endpoints. If outcomes (in silico/experimental) are
sitization in vitro’ endpoint that correlates with the in vivo endpoint is an available for only two endpoints, and they have aligned outcomes, the
active area of research. A number of deﬁ ned approaches (DA) which overall assessment of the endpoint is based on the concordant assess-
use varying DIPs have been developed to determine an overall assess- ments and the lower conﬁ dence score propagates. The adoption of the
ment of skin sensitization using non-animal/in-vitro/in silico models. lower conﬁ dence score reﬂ ects a conservative view of the assessment at
Any of the DAs described in Section 1 may be adopted here. There has this stage of the analysis. However, if the conﬁ dence scores have the
been regulatory acceptance of the “ AOP 2 out of 3” approach and the same value for non-concordant assessments, then the overall prediction
KE3/1 sequential testing strategy (STS) as alternatives to the LLNA for for the ‘ skin sensitization in vitro’ endpoint is inconclusive. Where there
regulatory submission to the United States Environmental Protection are two concordant assessments, and the non-concordant assessment
Agency (US EPA) (EPA, 2018). Here, we discuss how to derive an occurs with high conﬁ dence, then the overall conﬁ dence could be
overall assessment and conﬁ dence when the “ AOP 2 out of 3” approach lowered by one level. Table 2 provides examples showing the derivation
is used within the framework presented in this protocol. of the overall assessment and the rationale for the ﬁ nal conﬁ dence
 The “ AOP 2 out of 3” uses the outcome of three individual assays score. An alternative point of view suggests that the assays that predict
that map to three KEs to derive a ﬁ nal assessment; however, within the the ‘ events in keratinocytes’ , and ‘ events in dendritic cells’ , are de-
framework presented the assay results are integrated and propagated to pendent on the ability of the test chemical to bind protein and therefore
the three endpoints related to each key event. The diﬀ erence between point to the activation of the molecular initiating event, ‘ covalent in-
the “ AOP 2 out of 3” and the approach used in the framework is subtle, teraction with skin proteins’ . In this point of view, any improvement in
but is worth mention. The “ AOP 2 out of 3” approach considers thepredictive performance that results from integrating the KEs across the
outcome of the experimental systems – DPRA, KeratinoSens™, and h- AOP is a result of reducing the inﬂ uence of technical limitations of each
CLAT – but within the framework presented, the: ‘ Covalent interaction of the assays (Roberts, 2018; Roberts and Grace, 2018).
 12

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
 Fig. 8. Decision tree showing how an
 overall assessment and conﬁ dence score
 could be derived for the ‘ events in dendritic
 cells’ based on the U-SENS™ and IL-8 Luc
 assay data. The conﬁ dence scores here are
 based on RS1 experimental data: assuming
 relevant data and high reliability, and, in
 practice, conﬁ dence scores may need to be
 adjusted based on reliability scores, SM
 Table 10.
 The discussion thus far has focused on assessing hazard from in vitro with the in vivo studies except in unique cases; for example, when
data, but there are also existing strategies for predicting potency in metabolism is thought to inﬂ uence the outcome. As such, no change in
humans from in vitro data based on the DAs described in Section 1 and conﬁ dence (reliability and relevance of the prediction) is expected due
reviewed in (Kleinstreuer et al., 2018). The Artiﬁ cial Neural Network to the extrapolation of in vitro hazard.
Model for Predicting LLNA EC3 (Shiseido); Bayesian Network DIP (BN-
ITS-3) for Hazard and Potency Identiﬁ cation of Skin Sensitizers (P&G); 3.6. Skin sensitization in rodent lymphocytes
Sequential Testing Strategy (STS) for Sensitizing Potency Classiﬁ cation
Based on in Chemico and In Vitro Data (Kao); and ITS for Sensitizing A negative result in the LLNA is propagated to the skin sensitization
Potency Classiﬁ cation Based on In Silico; In Chemico, and In Vitro Data in rodent lymphocytes endpoint with high conﬁ dence. A weak sensi-
(Kao) were found to predict potency class equally well, or better than tizer may require investigation of the skin irritation potential of the
the LLNA. Similar to the earlier discussion on hazard, the DAs for as- chemical, particularly if the result is derived from a lower-reliability
sessing potency use biological assay outcomes (mechanisms/eﬀ ects study that may not have considered irritation prior to designing the
assessment within the HAF e.g. DPRA, KeratinoSens™, h-CLAT) as test. The skin irritation potential will be determined through a HAF that
endpoints and may integrate the information with in silico methods to will be published in a separate protocol. Positive results due to con-
determine a potency class. Within the HAF presented, the assay out- founding factors from irritants usually result in a low-level increase in
comes (in vitro/in silico eﬀ ects/mechanisms assessment) are interpreted lymphocytes which could be misinterpreted as a weak sensitizing re-
in the context of their toxicological signiﬁ cance and integrated to de- sponse. In cases where a chemical is found to have a strong skin irri-
termine a toxicological endpoint according to the rules and principles tation potential and is a weak sensitizer and the inﬂ uence of irritation
outlined in previous sections. The overall assessments of the KE end- cannot be ruled out, a positive assessment with low conﬁ dence could be
points may substitute for the outcome of the individual test methods in assigned to the ‘ Events in rodent lymphocytes’ endpoint (Fig. 9).
data interpretation procedures.
 3.7. Skin sensitization in rodents
3.5. Skin sensitization in vitro to skin sensitization in human extrapolation
 This endpoint integrates guinea pig (GPMT and BT) and mouse
 Extrapolation of in vitro skin sensitization results to human skin (LLNA) data. In the absence of LLNA data, the endpoint could be de-
sensitization predictions is necessary to satisfy the European Union's termined through the scheme shown in Fig. 10. If guinea pig tests are
7th Amendment of the Cosmetic Directive and REACH regulations not conducted according to standard protocols, irritation could become
which require and prefer the use of non-animal test methods for as- a confounding factor in the interpretation of the guinea pig test results
sessing the human skin sensitization endpoint. The deﬁ nition of the and inﬂ uence the relevance of the study (OECD, 1992). Freund's com-
AOP and the mechanistic information provided by the assays that map plete adjuvant (FCA) is used to maximize the guinea pig response;
to the AOP allow the human hazard identiﬁ ed for the ‘ skin sensitization however, FCA may also lower the irritation threshold. The implication
in vitro’ outcome to be propagated to the human endpoint. The re- is that concentrations that were identiﬁ ed as non-irritating and suitable
levance of the integrated in vitro battery of tests is equally weighted for the challenge reaction might in fact produce an irritant response.
 13

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
 Further, a hyperirritable state may be induced by the test article during
 the induction phase that is not represented in the control, unless a
 suitably irritating surrogate is used to induce the hyperirritable state in
 the controls (Kligman and Basketter, 1995; OECD, 1992). An irritant
 eﬀ ect cannot be distinguished from an allergic response by visual ex-
 amination. As such, post challenge examination is helpful in distin-
 guishing a sensitization response from an irritant eﬀ ect. Chemicals that
 are identiﬁ ed as irritants could be conﬁ dently predicted as non-sensi-
 tizers if observations of erythema dissipate within one day of challenge
 and/or there is a negative re-challenge test one week after the initial
 challenge (Kligman and Basketter, 1995). A positive result for a che-
 mical that is irritating but predicted to be a weak sensitizer is aﬀ orded a
 low conﬁ dence score if deviations from OECD, 1992 result in decreased
 reliability and relevance of the study as discussed above.
 When both guinea pig and mouse data are available and are con-
 cordant, then the result is translated to the ‘ skin sensitization in rodent’
 endpoint with exact or higher conﬁ dence scores being adopted. For
 example, if the LLNA is positive with medium conﬁ dence and the
 GPMT/BT is positive with low conﬁ dence, then the skin sensitization in
 rodent endpoint is assessed as positive with medium conﬁ dence. In
 cases where the data are discordant, the strategy for deriving an overall
 assessment may vary case-to-case. A high reliability guinea pig test has
 an advantage over the LLNA because it includes both induction and
 challenge phases, and is as such, more representative of the entire
 sensitization process. However, in contrast to the LLNA, the guinea pig
 test results are based on a qualitative measure and a subjective end-
 point. Potency is better assessed through the LLNA since it is derived
 from dose-response relationships and the read-out is quantitative;
 nonetheless, some chemical classes are over-classiﬁ ed in the LLNA. It is
 valuable to consider how the challenge reaction aﬀ ects interpretation of
 an assessment. It could be argued that the LLNA is an assay and non-
 speciﬁ c reactions can occur that may or may not relate to allergenic
 potential (respiratory sensitizers test positive in the LLNA, for example)
 while the dermal challenge in the guinea pig tests lends more con-
 ﬁ dence that any observations of sensitization are speciﬁ c to the skin. A
 default principle that could be adopted is to evaluate the ‘ skin sensiti-
 zation in rodent’ endpoint based on either the LLNA or GPMT/BT as-
 sessment with the higher conﬁ dence score and conservatively decrease
 the score by one level to reﬂ ect any uncertainty. For example, an LLNA
 that is assessed as positive with medium conﬁ dence, and a GP test that
 is negative with low conﬁ dence, would lead to a ‘ skin sensitization in
 rodent’ assessment as positive with low conﬁ dence. In these circum-
 stances, a review of the predictions is prudent and the assessment and
 conﬁ dence scores may be adjusted based on the review.
 3.8. Skin sensitization in rodents to skin sensitization in human
 extrapolation
 There are two schools of thought on rodent-to-human extrapolation
 that draw from a two diﬀ erent perspectives on risk assessment: one is
 that LLNA potency categories and EC3 values correlate well with
 human potency categories and NOEL values, and could therefore be
 used as a surrogate for the NOEL and for direct prediction of human
 potency class (Basketter et al., 2005). Alternatively, a safety factor may
 be incorporated based on the interspecies variation that may occur
 between the mouse and humans; although, this factor could be lowered
 in cases where a better correlation may be expected (e.g., based on
 existing human data for a close analogue) (Roberts and Api, 2018).
 Roberts and Api (2018), have deﬁ ned alerts for cases where the LLNA is
 not a good predictor of human potency. Guinea pig tests also provide
 relevant information on hazard and potency. However, tests that use
 adjuvant and intradermal routes of exposure (GPMT) present a chal-
 lenge for interpreting human potency, and in those situations potency
 estimation via the BT may be more relevant. The data however, could
 serve in a weight-of-evidence case for potency determination through
 interpretation and comparison of diﬀ erent test results and also with
 14

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
 Fig. 9. Decision tree showing how an
 overall assessment and conﬁ dence score
 could be derived for the “ Events in rodent
 lymphocytes” based on the LLNA. The con-
 ﬁ dence scores here are based on RS1/2 ex-
 perimental data (except in the case of \*):
 assuming relevant data and high reliability,
 and, in practice, conﬁ dence scores may
 need to be adjusted based on reliability
 scores. \*Concentrations tested in the LLNA
 are either non-irritating or mildly irritating.
 The low conﬁ dence score reﬂ ects the non-
 speciﬁ c increase in lymphocyte prolifera-
 tion that could occur with irritants.
Fig. 10. Decision tree showing how an overall assessment and conﬁ dence score could be derived for the ‘ skin sensitization in rodents’ endpoint based on guinea pig
tests. The conﬁ dence scores here are based on RS1 experimental data (except in the case of \*): assuming relevant data and high reliability, and, in practice, conﬁ dence
scores may need to be adjusted based on reliability scores. \*GPMT/BT challenge concentrations are non-irritating; however, deviations from OECD 406 may reduce
the relevance of the study and decrease the conﬁ dence in the endpoint.
 15

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
known benchmark chemicals (Kimber et al., 2001). these criteria when assessing the HMT and HRIPTs. The exposure sce-
 narios in the HMT and HRIPT may not represent real-world exposure
 because the test chemical is applied under occlusive conditions and the
3.9. Skin sensitization in humans outcomes can be viewed as subjective because an observer grades the
 skin reaction.
 The ‘ skin sensitization in humans’ endpoint could be evaluated
through several other endpoints such as the ‘ skin sensitization in vitro’
endpoint (section 3.5), the ‘ skin sensitization in rodent endpoints’ 4. Case studies
(section 3.8), or through the integration of the ‘ skin sensitization in
vitro’ , ‘ skin sensitization in rodents’ , and human assessments, combined The case studies demonstrate the interpretation of results when a
with supporting data from non-standard endpoints such as photo- series of statistical models ((Q)SARs), structural alerts, or read-across
allergy. A positive HMT/HRIPT is indicative of adverse outcome in are used to ﬁ ll data gaps for eﬀ ects and mechanisms that are included in
humans and can potentially be used to assign a potency class. In the the hazard assessment framework. The studies demonstrate how aspects
absence of reliable studies, other sources of evidence may be sought. of the rules and principles are implemented to derive an assessment,
The ﬁ rst line of evidence arises from the toxicological relationships that reliability score and conﬁ dence score, when the assessment is made
could be drawn from the chemical's structure. The presence of a using either or both existing experimental data or in silico methods.
structural alert for sensitization in humans provides evidence for the
elicitation of the adverse outcome. Structural alerts and diagnostic
patch testing with positive incidences in greater than 1% of the popu- 4.1. Case 1a: Compound with conﬂ icting data (“skin sensitization in vitro”
lation (considered to be high incidence) in relation to low usage volume endpoint determination)
(a measure of exposure) provides evidence for the skin sensitization
potential of a chemical, although it does not provide a deﬁ nite assess- An assessor needed to determine the hazard associated with a
ment (Api et al., 2017). If a compound has no structural alerts and compound. The compound was predicted to be reactive towards pro-
diagnostic patch testing data indicate < 1% frequency, the overall teins via an Acyl or SN2 reaction, and could be assigned to a reaction
evidence may together indicate a negative assessment, especially if the domain based on reaction chemistry alerts. Data that was generated
use volume is high. It is important to note that the indication of a 1% based on OECD TG 442C (DPRA) was available for the compound. The
incidence rate is based on expert opinion and as such is not meant to data indicated that the compound was negative for protein reactivity.
represent a rule that requires strict compliance. Many combinations of Based on adherence to the test guideline, a reliability score of RS1 was
scenarios are possible. assigned to the study. In silico tools (statistical results (QSAR) and
 In cases where human and in vitro/in vivo sensitization assessments alerts) were available for the DPRA prediction, and these predictions
do not align, additional information could be gathered from the ‘ skin were also negative. The statistical model and the alerts both had a re-
sensitization in vitro’ and/or ‘ skin sensitization in rodents’ endpoints to liability score of RS5. In silico assessments of dermal metabolism were
build a weight of evidence case. There are many permutations of assay negative after an expert review. The review increased the reliability of
results at this level but some general guidance can be provided to the the dermal metabolism alert from RS5 to RS3. The overall assessment
evaluator towards an overall assessment. It is generally recommended for ‘ covalent interaction with skin proteins’ was negative; however, the
that the assessments that are assigned more frequently should be pro- conﬁ dence was assigned as medium, based on the conﬂ icting me-
pagated to the overall human endpoint. However, if reliable human chanistic/reaction chemistry alert for protein reactivity, Fig. 11a.
data (RS1/2) is available, then the assessment of this data is given There is experimental data for the KeratinoSens™ assay which is
priority in the decision-making process. Table 13 of the supplementary aﬀ orded a positive assessment with a reliability score of RS1 (the study
material expands on the principles to derive an overall assessment adhered to OECD TG 442D), so the overall assessment for the ‘ Events in
given in vitro and rodent evidence. Due to ethical concerns, human Keratinocytes’ KE is positive with high conﬁ dence. Experimental data is
testing is no longer considered appropriate for most compounds, so not available for the ‘ Events in Dendritic Cells’ KE. The assessor would
much of the human data is older, or based on clinical reports, and may like to use the “ 2 out of 3” approach and is faced with two conﬂ icting
therefore lack information to assess its quality, necessitating the ﬁ lter of assessments based on in vitro data. A statistical model (QSAR) was used
expert opinion. Careful consideration is required in assessing con- to predict the results of the h-CLAT assay and the assessment is negative
ﬁ dence of the HMT and HRIPTs. For the HMT and, especially for the with a reliability score of RS3, after an expert review. The overall as-
HRIPT as used by the fragrance industry, low doses are often tested as sessment of the ‘ Events in Dendritic Cells’ KE is negative with a medium
the goal is to corroborate an animal study while trying to avoid sensi- conﬁ dence. Based on the two concordant assessments with aligned
tizing the subjects. Therefore, there can be quite a bit of uncertainty in a conﬁ dence scores (Negative, Medium conﬁ dence), and a third assess-
negative result because a higher test concentration could potentially ment that is conﬂ icting with high conﬁ dence (Positive, High con-
produce a positive result in humans. ﬁ dence), the overall assessment of in vitro skin sensitization endpoint is
 Table 3 shows factors to consider in assigning conﬁ dence to a negative with low conﬁ dence.
human study in general. There are however some speciﬁ c exceptions to
Table 3
Factors increasing and decreasing conﬁ dence in a human study (Schulz et al., 2010; Sibbald and Roland, 1998).
 Factors increasing conﬁ dence Factors decreasing conﬁ dence
 Objective clearly stated and linked to measured outcome Ambiguous objective, poorly linked to measured outcome
 Randomized controlled study Uncontrolled and not randomized (or case report)
 Randomized double-blind study No blinded control in study
 Study conducted long enough to observe the eﬀ ect Study duration too short to observe the eﬀ ect
 Control substance application matches test substance application and represents Control substance application does not match test substance application or does not
 the real-world exposure represents the real-world exposure scenario
 Outcome clearly deﬁ ned and measured through a quantitative endpoint Subjective outcome based on perception
 Statistical rationale behind determination of sample size No rationale behind sample size selection
 Description of study population available for review No description of study population available
 16

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
 Fig. 11a. Derivation of the ‘ skin sensitization in vitro’ endpoint using the “ AOP 2 out of 3” approach (Case 1a).
4.2. Case 1b: Compound with conﬂ icting data (‘skin sensitization in the Reconstructed Human Epidermis (RHE) test method. The assess-
humans’ endpoint determination) ment of skin irritation is positive with a score of RS1. The assessor
 conducts an expert review of the LLNA and suspects a false positive
 A further assessment was completed for the same compound as in LLNA result. The ‘ Events in Rodent Lymphocytes’ endpoint could be
Case 1a. This assessor has LLNA and GPMT data with conﬂ icting as- assigned as positive with low conﬁ dence; however, the negative in silico
sessments. The LLNA data is positive with an EC3 (%) value that in- results are more reliable and relevant in this situation and the negative
dicates weak sensitization. The study is assigned the lowest reliability assessment carries over to the ‘ Events in Rodent Lymphocytes’ endpoint
score of 5 based on signiﬁ cant deviations from OECD Test No. 429 that with medium conﬁ dence. The GPMT data is negative with a reliability
could alter both the reliability and relevance of the study. In silico as- of RS1 since the study adhered to OECD 406 and the irritant eﬀ ect was
sessments using expert alerts and statistical models are both negative. considered in the study design and interpretation of results. In silico
The weak sensitizing eﬀ ect and the mis-aligned in silico results prompt models agree with the experimental GPMT result. The overall assess-
the assessor to consider the irritation potential of the chemical.ment of ‘ Skin sensitization in rodents’ is negative with a high con-
Experimental data is available for the in vitro skin irritation test using ﬁ dence, Fig. 11b.
 Fig. 11b. Derivation of the ‘ Skin Sensitization in Rodents’ endpoint.
 17

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
 To further investigate the outcome in humans, the assessor con- considered in the case. A chemical is being screened for possible use in
ducted an in silico assessment using a set of alerts that were developed the cosmetics industry. It is expected to undergo metabolic transfor-
using HMT and HRIPT data as a reference database and no alerting mation leading to the formation of quinones, which have a high prob-
structure were found. No human study data were available; however, ability to react via Michael addition (MA). There are positive alerts for
DPT data were available and consecutive patients showed frequencies dermal metabolism and the site of metabolism coincides with a pro-MA
of 0% in a study. The absence of positive DPT results are indicative of reactivity alert. Negative DPRA data are available and the DPRA study
no sensitization in humans, although a conclusion cannot be made from is assigned a reliability score of 1 based on adherence to OECD TG
DPT data alone. 442C. However, based on knowledge that the compound contains a pro-
 Given the weight of evidence presented in Case 1a and 1b a ﬁ nal reactive feature that coincides with a site of metabolism, the relevance
determination of the ‘ skin sensitization in humans’ can be made. In this of the DPRA for testing the compound is challenged since any activity
case, a well conducted GPMT carried signiﬁ cant weight towards the that results from a metabolic transformation may be missed. The DPRA
negative sensitization assessment with high conﬁ dence; reﬂ ecting the test is considered not relevant for the compound tested, and the as-
high reliability and relevance of the information. Other evidence sup- sessment of the ‘ Covalent interaction with skin proteins’ is based on the
porting a negative assessment included a negative protein binding test assignment of a pro-reactive domain. Although there may be cases
which was reinforced by negative in silico models predictions of protein where a pro-reactive domain assignment does not lead to protein in-
binding; and negative (Q)SARs predicting the ‘ Events in dendritic cells’ , teraction due to deactivating features, a conservative approach to as-
and LLNA. The conﬂ icting piece of information presented by the LLNA sessing the endpoint given a pro-reactive feature is to assign a positive
study was viewed as less reliable and relevant information due pri- assessment with a lowered conﬁ dence. No other in vitro data are
marily to confounding irritant eﬀ ects in the study. A second piece of available for the compound. A (Q)SAR was developed based on pro-
conﬂ icting information was presented by the KeratinoSens™ experi- prietary data for the KeratinoSens™ assay. The test compound is as-
mental study. While no speciﬁ c explanation for this false positive was sessed as positive in the (Q)SAR, with the pro-MA feature identiﬁ ed as
determined, the body of negative evidence for the ‘ skin sensitization in signiﬁ cant by the model. After a review of the (Q)SAR prediction, the
vitro’ endpoint supports the negative assessment and the low conﬁ dence ‘ Events in Keratinocytes’ is assessed as positive with medium con-
reﬂ ects any uncertainty in the assessment of that endpoint. However, ﬁ dence. No data or models were available for the ‘ Events in dendritic
the ‘ sensitization in vitro’ assessment does not discredit the ‘ skin sensi- cells’ endpoint. Given the positive assessment for ‘ Covalent interaction
tization in rodents’ assessment. Since the in vitro and rodent endpoints with skin proteins’ and the ‘ Events in Keratinocytes’ , the overall as-
are both equally relevant when the in vitro endpoint is derived through sessment for the ‘ Skin Sensitization in vitro’ endpoint is made using the
adeﬁ ned approach, the endpoint that contains more reliable informa- “ 2 out of 3” approach. The overall assessment of the ‘ Skin Sensitization
tion contributes more to the overall conﬁ dence. The in vitro endpoint in vitro’ endpoint is positive with low conﬁ dence based on the two
does not introduce any uncertainty in the GPMT experimental ﬁ ndings, aligned positive assessments and the lower conﬁ dence score propa-
and taken together with the DPT data, the ﬁ nal conﬁ dence score is high gating to the endpoint, Fig. 12a. It is possible to extrapolate the existing
in this negative case, Fig. 11c. There may be instances where a higher hazard information to the ‘ Skin sensitization in humans’ endpoint and
level of conservatism is necessary than presented. In such instances, the assess it as positive with low conﬁ dence.
conﬁ dence score could be reduced to medium, although a change in the
assessment might be diﬃ cult to justify. 4.4. Case 2b: Pro/pre-hapten assessment example 2
4.3. Case 2a: Pro/pre-hapten assessment Consider an extension of the case presented in Section 4.3. LLNA
 data are not available for the test compound but are available for close
 Fig. 12a details the assessment for the mechanisms/e ﬀ ects that were analogs. In addition there is a low quality guinea pig test for the test
Fig. 11c. Derivation of the ‘ skin Sensitization in Humans’ endpoint from the weight of evidence presented from the ‘ Skin Sensitization skin in vitro’ and ‘ Skin
Sensitization in Rodents’ endpoints. DPT data is also used to support the overall assessment.
 18

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
 Fig. 12a. Derivation of the ‘ skin sensitization in vitro’ endpoint using the “ AOP 2 out of 3” approach (Case 2).
compound that indicates a positive sensitization response. Read-across endpoints both support that assignment of a positive hazard assessment
is performed using the LLNA data for the analogs. The analogs all for the ‘ skin sensitization in humans’ with medium conﬁ dence. Fig. 12b
contained the pro-reactive feature and formed a congeneric series that shows the ﬂ ow of information within the hazard assessment framework.
allowed interpolation of the LLNA EC3 value. The EC3 value was pre-
dicted to be 3.2%, indicative of a moderate sensitizer. The ‘ Events in 5. Reporting
rodent lymphocytes’ endpoint was assessed as positive with medium
conﬁ dence based on the read-across result. The guinea pig test is as- An important consideration towards in silico standardization, re-
signed a reliability score of RS5 based on deviations from OECD 406. A producibility and transparency is a consistent reporting format (Myatt
review of the study showed that for an induction concentration of 1%, et al., 2018). The general protocol (Myatt et al., 2018) describes a
the sensitization incidence is 100% suggesting that the compound could proposed reporting format that includes the elements that provide
be classiﬁ ed as a Category 1A sensitizer. After an expert review of the completeness of information. The report format is reproduced in
study, the reliability score is increased to RS3. The overall assessment of Table 4 with a minor modiﬁ cation for the skin sensitization endpoint. In
the ‘ Skin Sensitization in Rodents’ endpoint is assessed as positive, with addition to the description of models, databases, and tools that were
medium conﬁ dence based on the weight of evidence presented by the used, it is also recommended to describe any IATAs, DIPs or DAs that
LLNA read-across and guinea pig study. were used in deriving the overall assessment. The details that are sug-
 The ‘ Skin sensitization in rodents’ and the ‘ Skin sensitization in vitro’ gested should allow another expert to repeat the process and achieve
 Fig. 12b. Derivation of the ‘ Skin Sensitization in Humans’ using the “ AOP 2 out of 3” approach (Case 2).
 19

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
Table 4
Elements of an in silico toxicology report.
 Section Content
 Title page - Title (including information on the decision context)
 - Who generated the report and from which organization
 - Who performed the in silico analysis and/or expert review, including their organization
 - Date when this analysis was performed
 - Who the analysis was conducted for
 Executive summary - Provide a summary of the study
 - Describe the toxicity or properties being predicted
 - Include a table or summary showing the following:
 o The chemical(s) analyzed
 o Summary of in silico results, reviewed experimental data and overall assessment for each toxicological eﬀ ect or mechanism
 o Summary of toxicological endpoint assessment and conﬁ dence
 o Summary of supporting information
 Purpose - Speciﬁ cation of the problem formulation
 Materials and methods - QSAR model(s), expert alerts, and other models used with version number(s) and any parameters set as part of the prediction (e.g., QMRFa format)
 - Databases searched with version number(s)
 - Description of any IATAs, DIPs, DAs used
 - Tools used as part of any read-across with version number(s)
 Results of Analysis - Details of the results and expert review of the in silico models and any experimental data, including results of the applicability domain analysis
 - Report of any read-across analysis, including source analogs and read-across justiﬁ cations
 Conclusion - Summarize the overall analysis including experimental data, in silico methods and expert review
 - Final prediction that is based on expert judgment
 References - Complete bibliographic information or links to this information, including test guidelines referred to in the experimental data, etc.
 Appendices (optional) - Full (or summary) study reports used or links to the report, detailed (or summary) in silico reports, reports on the models used (e.g., QMRF reports)
 a QMRF – QSAR Model Reporting Format.
the same results. Further, the standardized report enables streamlined responsibility of the authors and does not necessarily represent the
and consistent review of regulatory submissions across industries and oﬃ cial views of the National Institutes of Health.
endpoints. Section 5 of the Supplementary Material (SM5) provides an
example of a report for sensitization hazard. Appendix A. Supplementary data
6. Conclusion Supplementary data to this article can be found online at https://
 doi.org/10.1016/j.yrtph.2020.104688.
 The skin sensitization in silico protocol presented here is the ﬁ rst
publication to outline a systematic assessment of skin sensitization Funding
based on both experimental data and in silico predictions. It includes a
HAF and provides general rules for the in silico toxicological assessment Research reported in this publication was supported by the National
of chemicals within the framework. The framework is transparent and Institute of Environmental Health Sciences of the National Institutes of
ﬂ exible as it does not require the generation of all endpoints to derive Health under Award Number R44ES026909. The content is solely the
an overall assessment of ‘ skin sensitization in humans’ and can ac- responsibility of the authors and does not necessarily represent the
commodate quantitative and qualitative predictions and/or experi- oﬃ cial views of the National Institutes of Health.
mental results. There are cases where extrapolation to the human
endpoint is possible and this has been described. The corresponding References
assessment of the conﬁ dence for all endpoints allows the protocol to be
used in a variety of use cases. For example, assessments with low Anderson, Stacey E., Siegel, Paul D., Meade, B.J., 2011. The LLNA: a brief review of recent
conﬁ dence scores may still have practical usage in screening or prior- advances and limitations. J. Allergy 2011, 424203.
itization use cases. In addition, the protocol highlights experimental Api, Anne Marie, Parakhia, Rahul, Oʼ Brien, Devin, Basketter, David A., 2017. Fragrances
 categorized according to relative human skin sensitization potency. Dermatitis:
approaches or in silico models that could be incorporated into the HAF Contact, Atopic, Occupational, Drug 28 (5), 299– 307.
in the near future. Expert review is a critical element in any such pro- Aptula, Aynur O., Roberts, David W., 2006. Mechanistic applicability domains for non-
cedure and items to consider as part of this review are listed to support animal-based prediction of toxicological end Points: general principles and applica-
 tion to reactive toxicity. Chem. Res. Toxicol. 19 (8), 1097– 1105.
a more consistent assessment. The standardization of the HAF for per- Ball, Nicholas, Stuart, Cagen, Carrillo, Juan-Carlos, Certa, Hans, Eigler, Dorothea, Emter,
forming in silico methods is designed to support increased use and ac- Roger, Frank, Faulhammer, Garcia, Christine, Graham, Cynthia, Haux, Carl, Kolle,
ceptance of in silico tools among regulatory agencies and industries Susanne N., Reinhard, Kreiling, Natsch, Andreas, Mehling, Annette, 2011. Evaluating
 the sensitization potential of surfactants: integrating data from the local lymph node
alike. assay, Guinea pig maximization test, and in vitro methods in a weight-of-evidence
 approach. Regul. Toxicol. Pharmacol. 60 (3), 389– 400.
Declaration of Competing Interest Basketter, D.A., Gerberick, G.F., Kimber, I., 2001. Skin sensitisation, vehicle eﬀ ects and
 the local lymph node assay. Food Chem. Toxicol. 39 (6), 621– 627.
 Basketter, David A., Clapp, Catherine, Jeﬀ eries, Donna, Saﬀ ord, Bob, Ryan, Cindy A.,
 The authors declare that they have no known competing ﬁ nancial Frank, Gerberick, Dearman, Rebecca J., Kimber, Ian, 2005. Predictive identiﬁ cation
 of human skin sensitization thresholds. Contact Dermatitis 53 (5), 260– 267.
interests or personal relationships that could have appeared to inﬂ u- Basketter, David A., McFadden, John F., Frank, Gerberick, Cockshott, Amanda, Kimber,
ence the work reported in this paper. Ian, 2009. Nothing is perfect, not even the local lymph node assay: a commentary and
 the implications for REACH. Contact Dermatitis 60 (2), 65– 69.
 Cronin, MTD, Basketter, DA, 1994. Multivariate QSAR analysis of a skin sensitization
Acknowledgements database. SAR QSAR Environ. Res. 2 (3), 159– 179. https://doi.org/10.1080/
 10629369408029901.
 Research reported in this publication was supported by the National Dumont, Coralie, Barroso, João, Matys, Izabela, Worth, Andrew, Casati, Silvia, 2016.
Institute of Environmental Health Sciences of the National Institutes of Analysis of the local lymph node assay (LLNA) variability for assessing the prediction
 of skin sensitisation potential and potency of chemicals with non-animal approaches.
Health under Award Number R43ES026909. The content is solely the Toxicol. Vitro 34, 220– 228.
 20

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
Dumont, Coralie, Prieto, Pilar, Asturiol, David, Worth, Andrew, 2015. Review of the – FCM, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing,
 availability of in vitro and in silico methods for assessing dermal bioavailability. Paris.
 Appl. in Vitro Toxicol. 1 (2), 147– 164. Oecd, 2018b. Test No. 442D. In: Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test
Enoch, S.J., Madden, J.C., Cronin, M.T.D., 2008. Identiﬁ cation of mechanisms of toxic Method, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing,
 action for skin sensitisation using a SMARTS pattern based approach. SAR QSAR Paris.
 Environ. Res. 19 (5– 6), 555– 578. Roberts, D.W., Aptula, A.O., 2014. Electrophilic reactivity and skin sensitization potency
Epa, 2018. Draft Interim Science Policy: Use of Alternative Approaches for Skin of S N Ar electrophiles. Chem. Res. Toxicol. 27 (2), 240– 246.
 Sensitization as a Replacement for Laboratory Animal Testing. Roberts, D.W., Aptula, A.O., Cronin, M.T.D., Hulzebos, E., Patlewicz, G., 2007. “Global
Fitzpatrick, Jeremy M., Roberts, David W., Patlewicz, Grace, 2017. What determines skin (Q)SARs for skin sensitisation– assessment against OECD principles‖ . SAR QSAR
 sensitization potency: myths, maybes and realities. The 500 molecular weight cut-oﬀ : Environ. Res. 18 (3– 4), 343– 365.
 an updated analysis. J. Appl. Toxicol. 37 (1), 105– 116. Roberts, D.W., Aptula, A.O., Patlewicz, G.Y., 2011. Chemistry-based risk assessment for
Fujita, Masaharu, Yamamoto, Yusuke, Wanibuchi, Sayaka, Katsuoka, Yasuhiro, Kasahara, skin sensitization: quantitative mechanistic modeling for the S N Ar domain. Chem.
 Toshihiko, 2019. The underlying factors that explain why nucleophilic reagents Res. Toxicol. 24 (7), 1003– 1011.
 rarely Co-elute with test chemicals in the ADRA. J. Pharmacol. Toxicol. Methods 96, Roberts, D.W., Fraginals, R., Lepoittevin, J.P., Benezra, C., 1991. Reﬁ nement of the re-
 95– 105. lative alkylation index (RAI) model for skin sensitization and application to mouse
Hoﬀ mann, Sebastian, 2015. LLNA variability: an essential ingredient for a comprehensive and Guinea-pig test data for alkyl alkanesulphonates. Arch. Dermatol. Res. 283 (6),
 assessment of non-animal skin sensitization test methods and strategies. ALTEX 387– 394.
 Alternat. Anim. Exp. 32 4 SE-Short communications. Roberts, D.W., Williams, D.L., 1982. The derivation of quantitative correlations between
Hoﬀ mann, Sebastian, Kleinstreuer, Nicole, Alépée, Nathalie, Allen, David, Api, Anne skin sensitisation and physio-chemical parameters for alkylating agents, and their
 Marie, Ashikaga, Takao, Clouet, Elodie, Cluzel, Magalie, Desprez, Bertrand, Gellatly, application to experimental data for sultones. J. Theor. Biol. 99 (4), 807– 825.
 Nichola, Goebel, Carsten, Kern, Petra S., Klaric, Martina, Kühnl, Jochen, Lalko, Jon Roberts, David W., 2018. Is a combination of assays really needed for non-animal pre-
 F., Martinozzi-Teissier, Silvia, Mewes, Karsten, Miyazawa, Masaaki, Parakhia, Rahul, diction of skin sensitization potential? Performance of the GARDTM (genomic allergen
 Vliet, Erwin van, Zang, Qingda, Petersohn, Dirk, 2018. Non-animal methods to pre- rapid detection) assay in comparison with OECD guideline assays alone and in
 dict skin sensitization (I): the cosmetics europe database. Crit. Rev. Toxicol. 48 (5), combination. Regul. Toxicol. Pharmacol. 98, 155– 160.
 344– 358. Roberts, David W., Api, Anne Marie, 2018. Chemical applicability domain of the local
Kimber, I., Basketter, D.A., Berthold, K., Butler, M., Garrigue, J.L., Lea, L., Newsome, C., lymph node assay (LLNA) for skin sensitisation potency. Part 4. Quantitative corre-
 Roggeband, R., Steiling, W., Stropp, G., Waterman, S., Wiemann, C., 2001. Skin lation of LLNA potency with human potency. Regul. Toxicol. Pharmacol. 96, 76– 84.
 sensitization testing in potency and risk assessment. Toxicol. Sci. 59 (2), 198– 208. Roberts, David W., Api, Anne Marie, Saﬀ ord, Robert J., Lalko, Jon F., 2015. Principles for
Kleinstreuer, Nicole C., Hoﬀ mann, Sebastian, Alépée, Nathalie, Allen, David, Ashikaga, identiﬁ cation of high potency category chemicals for which the dermal sensitisation
 Takao, Casey, Warren, Clouet, Elodie, Cluzel, Magalie, Desprez, Bertrand, Gellatly, threshold (DST) approach should not Be applied. Regul. Toxicol. Pharmacol. 72 (3),
 Nichola, Göbel, Carsten, Kern, Petra S., Klaric, Martina, Kühnl, Jochen, Martinozzi- 683– 693.
 Teissier, Silvia, Mewes, Karsten, Miyazawa, Masaaki, Strickland, Judy, Vliet, Erwin Roberts, David W., Aptula, Aynur O., 2008. Determinants of skin sensitisation potential.
 van, Zang, Qingda, Petersohn, Dirk, 2018. Non-Animal methods to predict skin J. Appl. Toxicol. 28 (3), 377– 387.
 sensitization (II): an assessment of deﬁ ned approaches. Crit. Rev. Toxicol. 48 (5), Roberts, David W., Aptula, Aynur O., Patlewicz, Grace, 2006. Mechanistic applicability
 359– 374. domains for non-animal based prediction of toxicological endpoints. QSAR analysis of
Kligman, A.M., Basketter, D.A., 1995. A critical commentary and updating of the Guinea the Schiﬀ base applicability domain for skin sensitization. Chem. Res. Toxicol. 19 (9),
 pig maximization test. Contact Dermatitis 32 (3), 129– 134. 1228– 1233.
Madden, J.C., Webb, S., Enoch, S.J., Colley, H.E., Murdoch, C., Shipley, R., Sharma, P., Roberts, David W., Andreas, Natsch, 2009. High throughput kinetic proﬁ ling approach for
 Yang, C., Cronin, M.T.D., 2017. In silico prediction of skin metabolism and its im- covalent binding to peptides: application to skin sensitization potency of Michael
 plication in toxicity assessment. Comput. Toxiol. 3, 44– 57. acceptor electrophiles. Chem. Res. Toxicol. 22 (3), 592– 603.
Myatt, G.J., Ahlberg, E., Akahori, Y., Allen, D., Amberg, A., Anger, L.T., Aptula, A., Roberts, David W., Grace, Patlewicz, 2018. Non-animal assessment of skin sensitization
 Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D., hazard: is an integrated testing strategy needed, and if so what should Be integrated?
 Brigo, A., Burden, N., Cammerer, Z., Cronin, M.T.D., Cross, K.P., Custer, L., Dettwiler, J. Appl. Toxicol. 38 (1), 41– 50.
 M., Dobo, K., Ford, K.A., Fortin, M.C., Gad-McDonald, S.E., Gellatly, N., Gervais, V., Russell, W.M.S., Burch, R.L., 1959. The Principles of Humane Experimental Technique.
 Glover, K.P., Glowienke, S., Van Gompel, J., Gutsell, S., Hardy, B., Harvey, J.S., Methuen & Co. Ltd, London.
 Hillegass, J., Honma, M., Hsieh, J.H., Hsu, C.W., Hughes, K., Johnson, C., Jolly, R., Schulz, K.F., Altman, D.G., Moher, D., 2010. CONSORT 2010 statement: updated guide-
 Jones, D., Kemper, R., Kenyon, M.O., Kim, M.T., Kruhlak, N.L., Kulkarni, S.A., lines for reporting parallel group randomised trials. BMJ 340 (mar23 1) c332– c332.
 Kümmerer, K., Leavitt, P., Majer, B., Masten, S., Miller, S., Moser, J., Mumtaz, M., Sibbald, B., Roland, M., 1998. Understanding controlled trials. Why are randomised
 Muster, W., Neilson, L., Oprea, T.I., Patlewicz, G., Paulino, A., Lo Piparo, E., Powley, controlled trials important? Br. Med. J. 316 (7126), 201.
 M., Quigley, D.P., Reddy, M.V., Richarz, A.N., Ruiz, P., Schilter, B., Seraﬁ mova, R., Sumpter, Tina L., Balmert, Stephen C., Kaplan, Daniel H., 2019. Cutaneous immune re-
 Simpson, W., Stavitskaya, L., Stidl, R., Suarez-Rodriguez, D., Szabo, D.T., Teasdale, sponses mediated by dendritic cells and mast cells. JCI Insight 4 (1).
 A., Trejo-Martin, A., Valentin, J.P., Vuorinen, A., Wall, B.A., Watts, P., White, A.T., Urbisch, Daniel, Mehling, Annette, Guth, Katharina, Ramirez, Tzutzuy, Honarvar,
 Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C., Hasselgren, C., 2018. In Naveed, Kolle, Susanne, Landsiedel, Robert, Jaworska, Joanna, Kern, Petra S., Frank,
 silico toxicology protocols. Regul. Toxicol. Pharmacol. 96. Gerberick, Natsch, Andreas, Emter, Roger, Ashikaga, Takao, Miyazawa, Masaaki,
Natsch, Andreas, 2010. The nrf2-keap1-ARE toxicity pathway as a cellular sensor for skin Sakaguchi, Hitoshi, 2015. Assessing skin sensitization hazard in mice and men using
 sensitizers— functional relevance and a hypothesis on innate reactions to skin sen- non-animal test methods. Regul. Toxicol. Pharmacol. 71 (2), 337– 351.
 sitizers. Toxicol. Sci. 113 (2), 284– 292. Vocanson, M., Hennino, A., Rozières, A., Poyet, G., Nicolas, J.F., 2009. Eﬀ ector and
Natsch, Andreas, Emter, Roger, Gfeller, Hans, Haupt, Tina, Ellis, Graham, 2015. regulatory mechanisms in allergic contact dermatitis. Allergy 64 (12), 1699– 1714.
 Predicting skin sensitizer potency based on in vitro data from KeratinoSens and ki- Wang, Chia Chi, Lin, Ying Chi, Wang, Shan Shan, Shih, Chieh, Lin, Yi Hui, Tung, Chun
 netic peptide binding: global versus domain-based assessment. Toxicol. Sci. 143 (2), Wei, 2017. SkinSensDB: a curated database for skin sensitization assays. J. Cheminf.
 319– 332.
Natsch, Andreas, Gfeller, Hans, Haupt, Tina, Brunner, Gerhard, 2012. Chemical reactivity Glossary
 and skin sensitization potential for benzaldehydes: can Schiﬀ base formation explain
 everything? Chem. Res. Toxicol. 25 (10), 2203– 2215.
Natsch, Andreas, Haupt, Tina, 2013. Utility of rat liver S9 fractions to study skin-sensi- Ac: Acylation
 tizing prohaptens in a modiﬁ ed KeratinoSens assay. Toxicol. Sci. 135 (2), 356– 368. ACD: Allergic Contact Dermatitis
Nukada, Yuko, Takao, Ashikaga, Sakaguchi, Hitoshi, Sono, Sakiko, Mugita, Nanae, ADRA: Amino Acid Derivative Reactivity Assay
 Hirota, Morihiko, Miyazawa, Masaaki, Ito, Yuichi, Sasa, Hitoshi, Nishiyama, Naohiro, AOP: Adverse Outcome Pathway
 2011. Predictive performance for human skin sensitizing potential of the human cell ARE: Antioxidant/electrophile response element
 line activation test (h-CLAT). Contact Dermatitis 65 (6), 343– 353. BT: Buehler test
Oecd, 1992. Test No. 406: Skin Sensitisation:OECD Guidelines for the Testing of CD54: Cluster of Diﬀ erentiation 54, a co-stimulatory adhesion molecule that is expressed
 Chemicals, Section 4. OECD Publishing, Paris. in dendritic cells
Oecd, 2010a. Test No. 429: Skin Sensitisation: Local Lymph Node Assay, OECD Guidelines CD86: Cluster of Diﬀ erentiation 86, a co-stimulatory adhesion molecule that is expressed
 for the Testing of Chemicals, Section 4. OECD Publishing, Paris. in dendritic cells
Oecd, 2010b. Test No. 442A: Skin Sensitization: Skin Sensitization: Local Lymph Node CV70: Concentration of test chemical yielding a cell viability of 70% in the U-SENS™
 Assay: DA, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, method
 Paris. DA: Deﬁ ned Approach
Oecd, 2011. Report of the expert consultation on scientiﬁ c and regulatory evaluation of DC: Dendritic cells
 organic chemistry mechanism-based structural alerts for the identiﬁ cation of protein- DIP: Data interpretation procedure
 binding chemicals. Series Test. Assess. 139. DPRA: Direct Peptide Reactivity Assay
Oecd, 2014. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent DSA05: Dose per skin area that produced a positive response in 5% of the tested popu-
 Binding to Proteins. OECD. lation
Oecd, 2017. Guidance Document on the Reporting of Deﬁ ned Approaches and Individual EC1.5: Lowest concentration inducing a 1.5-fold change in luciferase activity in the assays
 Information Sources to Be Used within Integrated Approaches to Testing and measuring KE2
 Assessment (IATA) for Skin Sensitisation. OECD. EC150: Eﬀ ective concentrations yielding a relative ﬂ uorescence intensity [RFI] of 150%
Oecd, 2018a. Test No. 442B: Skin Sensitization: Local Lymph Node Assay: BrdU-ELISA or for CD86 in the h-CLAT test
 21

C. Johnson, et al. Regulatory Toxicology and Pharmacology 116 (2020) 104688
EC200: Eﬀ ective concentrations yielding a relative ﬂ uorescence intensity [RFI] of 200% Keap1: Kelch-like ECH-associated protein 1
 for CD54 in the h-CLAT test LLNA: Local Lymph Node Assay
EC3: Eﬀ ective concentration of a test chemical that gives a stimulation index with a three- LOEL: Lowest observed eﬀ ect level
 fold increase over the vehicle control in the LLNA Log Kow: n-octanol/water partition coeﬃ cient
EC3: Concentration with 3 fold luciferase induction in the KeratinoSens™ test MA: Michael addition
GARD: Genomic allergen rapid detection MHC: Major histo-compatibility complex
GPMT: Guinea Pig Maximization test MIE: Molecular initiating event
GST: Glutathione S-transferase NOEL: No Observed Eﬀ ect Level
HAF: Hazard assessment framework NQ01: NADPH-quinone oxidoreductase 1
h-CLAT: Human Cell Line Activation test Nrf2: Nuclear factor (erythroid-derived 2)-like 2
HMT: Human Maximization Test OECD: Organization for Economic Co-operation and Development
HRIPT: Human Repeat Insult Patch Test QMM: Quantitative Mechanistic Models
hTCPA: human T cell priming assay (Q)SAR: (Quantitative) Structure-Activity Relationship
IATA: Integrated approach to testing and assessment RFI: Relative ﬂ uorescence intensity
IC: Induction concentration in GPMT SB: Schiﬀ base formation
IC50: Concentration for 50% reduction of viability in KeratinoSens™ test SI: Stimulation index
IL-18: Interleukin-18 SLS: Sodium lauryl sulfate
IL-8: Interleukin-8 SM: Supplementary material
IL-8 Luc: Interleukin-8 Reporter Gene Assay SN1: Unimolecular nucleophilic substitution
KE: Key Event SN2: Bimolecular nucleophilic substitution
KE1: Key event 1: Covalent interaction with skin proteins SNAr: Nucleophilic aromatic substitution
KE2: Key event 2: Events in keratinocytes STS: Sequential Testing Strategy
KE3: Key event 3: Events in dendritic cells U-SENS™: U937 cell line activation Test
KE4: Key event 4: Events in lymphocytes
 22

Source PDF: /instem.com/www.instem.com/docs/sot-2021/Genetic-toxicology-in-silico-protocol.pdf

 Regulatory Toxicology and Pharmacology 107 (2019) 104403
 Contents lists available at ScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepage: www.elsevier.com/locate/yrtph
Genetic toxicology in silico protocol☆ T
Catrin Hasselgrena,∗ , Ernst Ahlbergb, Yumi Akahoric, Alexander Ambergd, Lennart T. Angerd,
Franck Atienzare,1 , Scott Auerbachf, Lisa Beilkeg, Phillip Bellionh, Romualdo Benignii, Joel Bercuj,
Ewan D. Boothk, Dave Bowerl, Alessandro Brigom, Zoryana Cammerern, Mark T.D. Cronino,
Ian Crooksp, Kevin P. Crossl, Laura Custerq, Krista Dobor, Tatyana Doktorovas, David Faulknert,
Kevin A. Fordu, Marie C. Fortinv,w , Markus Frericksx, Samantha E. Gad-McDonaldy,
Nichola Gellatlyz, Helga Geretsaa, Véronique Gervaisab, Susanne Glowienkeac,
Jacky Van Gompelad, James S. Harveyae, Jedd Hillegassq, Masamitsu Honmaaf, Jui-Hua Hsiehag,
Chia-Wen Hsuah, Tara S. Barton-Maclarenai, Candice Johnsonl, Robert Jollyaj, David Jonesak,
Ray Kemperal, Michelle O. Kenyonr, Naomi L. Kruhlakah, Sunil A. Kulkarniai, Klaus Kümmereram,
Penny Leavittq, Scott Mastenf, Scott Millerl, Chandrika Moudgala, Wolfgang Musterm,
Alexandre Paulinoan, Elena Lo Piparoao, Mark Powleyap, Donald P. Quigleyl,
M. Vijayaray Reddyap, Andrea-Nicole Richarzaq, Benoit Schilterao, Ronald D. Snyderar,
Lidiya Stavitskayaah, Reinhard Stidlas, David T. Szaboat, Andrew Teasdaleau, Raymond R. Ticeav,
Alejandra Trejo-Martinj, Anna Vuorinenh, Brian A. Wallaw, Pete Wattsax, Angela T. Whiteae,
Joerg Wicharday, Kristine L. Wittf, Adam Woolleyaz, David Woolleyaz, Craig Zwicklba,
Glenn J. Myattl
a Genentech, Inc., 1 DNA Way, South San Francisco, CA, 94080, USA
b Predictive Compound ADME & Safety, Drug Safety & Metabolism, AstraZeneca IMED Biotech Unit, Mölndal, Sweden
c Chemicals Evaluation and Research Institute, 1-4-25 Kouraku, Bunkyo-ku, Tokyo, 112-0004, Japan
d Sanoﬁ , R&D Preclinical Safety Frankfurt, Industriepark Hoechst, D-65926, Frankfurt am Main, Germany
e UCB BioPharma SPRL, Chemin du Foriest, B-1420 Braine-l’Alleud, Belgium
f The National Institute of Environmental Health Sciences, Division of the National Toxicology Program, Research Triangle Park, NC, 27709, USA
g Toxicology Solutions Inc., San Diego, CA, USA
h DSM Nutritional Products, Kaiseraugst, Switzerland
i Alpha-PreTox, via G.Pascoli 1, 00184, Roma, Italy
j Gilead Sciences, 333 Lakeside Drive, Foster City, CA, USA
k Syngenta, Product Safety Department, Jealott's Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, UK
l Leadscope, Inc, 1393 Dublin Rd, Columbus, OH, 43215, USA
m Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoﬀ mann-La Roche Ltd, Grenzacherstrasse 124,
4070, Basel, Switzerland
n Janssen Research & Development, 1400 McKean Road, Spring House, PA, 19477, USA
 ☆ FDA CDER Disclaimer: This article reﬂ ects the views of the authors and should not be construed to represent FDA's views or policies. The mention of commercial
products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by
the Department of Health and Human Services.
 NIEHS Disclaimer: The ﬁ ndings and conclusions in this report are those of the author(s) and do not necessarily represent the oﬃcial position of the National
Institutes of Health (NIH), NIEHS. The mention of commercial products, their sources, or their use should not be construed as either an actual or implied endorsement
of such products by the NIH/NIEHS.
 MHRA Disclaimer: Any opinions expressed in this document are the author's and are not necessarily shared by other assessors at the Medicines and Healthcare
products Regulatory Agency (MHRA). As such, they cannot be considered to be UK policy. The mention of commercial products, their sources, or their use in
connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the UK's MHRA.
 Health Canada Disclaimer: The ﬁ ndings and conclusions in this report are those of the author(s) and do not necessarily represent the oﬃcial position of Health
Canada. The mention of commercial products, their sources, or their use should not be construed as either an actual or implied endorsement of such products by
Health Canada.
 European Commission Disclaimer: The views expressed are solely those of the authors and the content does not necessarily represent the views or position of the
European Commission.
 ∗ Corresponding author.
 E-mail address: hasselgren.catrin@gene.com (C. Hasselgren).
https://doi.org/10.1016/j.yrtph.2019.104403
Available online 11 June 2019
0273-2300/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license
(http://creativecommons.org/licenses/BY/4.0/).

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
o School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, L3 3AF, UK
p British American Tobacco, Research and Development, Regents Park Road, Southampton, Hampshire, SO15 8TL, UK
q Bristol-Myers Squibb, Drug Safety Evaluation, 1 Squibb Dr, New Brunswick, NJ, 08903, USA
r Pﬁ zer Global Research & Development, 558 Eastern Point Road, Groton, CT, 06340, USA
s Douglas Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, CH-4057, Basel / Basel-Stadt, Switzerland
t Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 70A-1161A, Berkeley, CA, 947020, USA
u Global Blood Therapeutics, 171 Oyster Point Boulevard, South San Francisco, CA, 94080, USA
v Jazz Pharmaceuticals, Inc., 200 Princeton South Corporate Center, Suite 180, Ewing, NJ, 08628, USA
w Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, 170 Frelinghuysen Rd, Piscataway, NJ, 08855, USA
x BASF SE, 67056, Ludwigshafen, Germany
y Gad Consulting Services, 4008 Barrett Drive, Suite 201, Raleigh, NC, 27609, USA
z National Centre for the Replacement, Reﬁ nement and Reduction of Animals in Research (NC3Rs), Gibbs Building, 215 Euston Road, London, NW1 2BE, UK
aa UCB BioPharma SPRL, Chemin du Foriest, B-1420, Braine-l’Alleud, Belgium
ab Servier Group, 905 route de Saran, 45520, Gidy, France
ac Novartis Pharma AG, Pre-Clinical Safety, Werk Klybeck, CH, 4057, Basel, Switzerland
ad Janssen Pharmaceutical Companies of Johnson & Johnson, 2340, Beerse, Belgium
ae GlaxoSmithKline Pre-Clinical Development, Park Road, Ware, Hertfordshire, SG12 0DP, UK
af Division of Genetics and Mutagenesis, National Institute of Health Sciences, Kanagawa, 210-9501, Japan
ag Kelly Government Solutions, Research Triangle Park, NC, 27709, USA
ah FDA Center for Drug Evaluation and Research, Silver Spring, MD, USA
ai Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, ON, K1A 0K9, Canada
aj Toxicology Division, Eli Lilly and Company, Indianapolis, IN, USA
ak Medicines and Healthcare Products Regulatory Agency, 10 South Colonnade, Canary Wharf, London, E14 4PU, UK
al Vertex Pharmaceuticals Inc., Predictive and Investigative Safety Assessment, 50 Northern Ave, Boston, MA, USA
am Institute for Sustainable and Environmental Chemistry, Leuphana University Lüneburg, Scharnhorststraße 1/C13.311b, 21335, Lüneburg, Germany
an Oro Agri Setúbal, Portugal
ao Nestlé Research Vers-chez-les-Blanc, Lausanne, Switzerland
ap Merck Research Laboratories, West Point, PA, 19486, USA
aq European Commission, Joint Research Centre (JRC), Ispra, Italy
ar RDS Consulting Services, 2936 Wooded Vista Ct, Mason, OH, 45040, USA
as Safetree Consulting e.U, Vienna, Austria
at PPG Industries, Pittsburgh, PA, 15146, USA
au AstraZeneca, Macclesﬁ eld, Cheshire, UK
av RTice Consulting, Hillsborough, NC, USA
aw Colgate-Palmolive Company, Piscataway, NJ, 08854, USA
ax Bibra, Cantium House, Railway Approach, Wallington, Surrey, SM6 0DZ, UK
ay Bayer AG, Pharmaceuticals Division, Investigational Toxicology, Muellerstr. 178, D-13353, Berlin, Germany
az ForthTox Limited, PO Box 13550, Linlithgow, EH49 7YU, UK
ba Transendix LLC, 1407 Moores Manor, Indianapolis, IN, 46229, USA
ARTICLE INFO ABSTRACT
Keywords: In silico toxicology (IST) approaches to rapidly assess chemical hazard, and usage of such methods is increasing
In silico in all applications but especially for regulatory submissions, such as for assessing chemicals under REACH as well
In silico toxicology as the ICH M7 guideline for drug impurities. There are a number of obstacles to performing an IST assessment,
Computational toxicology protocols including uncertainty in how such an assessment and associated expert review should be performed or what is ﬁ t
(Q)SAR for purpose, as well as a lack of conﬁ dence that the results will be accepted by colleagues, collaborators and
Expert alerts regulatory authorities. To address this, a project to develop a series of IST protocols for diﬀ erent hazard end-
Expert review points has been initiated and this paper describes the genetic toxicity in silico (GIST) protocol. The protocol
Genetic toxicology
 outlines a hazard assessment framework including key eﬀ ects/mechanisms and their relationships to endpoints
 such as gene mutation and clastogenicity. IST models and data are reviewed that support the assessment of these
 eﬀ ects/mechanisms along with deﬁ ned approaches for combining the information and evaluating the conﬁ dence
 in the assessment. This protocol has been developed through a consortium of toxicologists, computational sci-
 entists, and regulatory scientists across several industries to support the implementation and acceptance of in
 silico approaches.
1. Introduction health, as well as to support hazard and risk assessment activities or to
 prioritize chemicals for in vitro or in vivo testing. The ﬁ rst regulation to
 The use of computational methods to assess the biological properties formally include the use of in silico approaches to address information
of chemicals is well established in many diﬀ erent industry sectors in- requirements for the purposes of hazard identiﬁ cation and risk assess-
cluding the pharmaceutical, cosmetic, food, plant protection, biocides, ment was REACH (Registration, Evaluation, Authorisation and Re-
and general chemical industries (Marchant, 2012; Hasselgren et al., striction of Chemicals) (REACH, 2006). This regulation, which applies
2013). Computational methods are used during diﬀ erent stages of to chemicals manufactured or imported into the European Union where
product development for purposes such as optimizing potency towards their import or use is not covered by other speciﬁ ed legislation. In
a protein target, determining the reactivity of a chemical, predicting the addition, since 2014, with the implementation of the International
rate of transmembrane permeability, or predicting toxicological end- Council for Harmonisation of Technical Requirements for Pharmaceu-
points. In the ﬁ eld of toxicology, computational (in silico) methods are ticals for Human Use (ICH) M7 guideline (ICH, 2014; ICH, 2017),
widely used to predict toxicological eﬀ ects directly relevant to human regulatory authorities, such as the US Food and Drug Administration
 (US FDA), the Japanese Pharmaceutical and Medical Devices Agency
 (PMDA), and the European Medicines Agency (EMA) accept in silico
 1 Current address: L'Oréal S.A., 9 rue Pierre Dreyfus, 92110 Clichy, France.
 2

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
assessments of the mutagenic potential of drug impurities. The ICH M7
guideline represented a milestone for the regulatory acceptance of
computational methods for hazard assessment in pharmaceuticals and
the implementation of the guideline has inﬂ uenced the use of in silico
assessments for other applications, such as the risk assessment of ex-
tractables and leachables, both for pharmaceuticals and for other in-
dustries. Other examples include the revision of the Toxic Substances
Control Act (TSCA) to include predictive models and expert review as
part of an overall assessment as well as the US FDA Center for Devices
and Radiological Health (CDRH) issuing a guidance for industry on the
use of International Standard ISO 10993–1 for biological evaluation of
medical devices and indicating that in the absence of experimentally
derived carcinogenicity information, structure-activity relationship
modeling for these materials may be used (CDRH, 2016). These e ﬀ orts
and advancements in the adoption of in silico methods help to support Fig. 1. Basic decision scheme for genetic toxicity.
the replacement, reduction and reﬁ nement (3Rs) of animal testing and
are well aligned with the rapid screening approach that is common et al., 2012)) or panels of genes ( Li et al., 2015), or di ﬀ erential cyto-
practice in the early development of chemical products (Ford, 2016; toxicity using isogenic cell lines that have been knocked out for dif-
Stanton and Kruszewski, 2016). ferent DNA repair enzymes (Yamamoto et al., 2011) have also been
 In a previous publication (Myatt et al., 2018), the use of in silicoutilized. Later in the section, these types of methods are referred to as
toxicology was discussed in more detail and highlighted that the ﬁ eld is, “primary DNA damage”. Some of these test methods are no longer
to some extent, hampered by lack of clarity concerning appropriate commonly used due to limitations in sensitivity (UDS) or the lack of a
procedures related to the application, interpretation, and utilization of mechanistic underpinning (SCE), while others are generally used as
in silico approaches. To improve this situation, and to provide guidance, screens to generate complementary information to provide a weight-of-
an in silico toxicology (IST) protocol template has been designed. The evidence for mechanistic understanding.
general IST protocol, as well as the protocol for genetic toxicology (the Depending on the industry sector, slightly diﬀ erent combinations of
GIST protocol) described here, have been developed by an international tests may be required as outlined in published guidance documents to
consortium comprising over 50 organizations including industry, aca- support regulatory data requirements, such as the ICH S2 (R1) guidance
demia and government agencies, utilizing the extensive experience of for drugs (ICH, 2012), European Food Safety Authority (EFSA) gui-
its members and hence representing the state of the art of in silicodance (EFSA, 2011) for food and feed safety assessment, the REACH
toxicology application. guidance (ECHA, 2011) for registration of chemicals or ISO 10993 –1
 (CDRH, 2016) for evaluation of medical devices. In this publication, the
1.1. Genetic toxicology in silico protocol overview intention is not to adhere to any speciﬁ c guidance, but rather to base
 the assessments on a decision scheme (simple version shown in Fig. 1),
 Genetic toxicology (genotoxicity) concerns the eﬀ ects induced by outlining a strategy for assessing genotoxicity based on coverage of the
genetic alterations that may occur in somatic and/or germ cells fol- three major endpoints of genotoxicity as well as the generic term
lowing exposure to a chemical agent. Chemical agents can induce “primary DNA damage” , mentioned earlier in this section. Implicitly,
changes in DNA through direct or indirect interactions and the con- this leaves room for alternatives in terms of speciﬁ c study types. The
sequences of the genetic alterations may manifest as death and/or commonly used genetic toxicology studies and the respective mechan-
mutations in exposed cell populations. If somatic cells are aﬀ ected, this isms/eﬀ ects they identify are shown in Fig. 2.
might, for example, result in the development of cancer or neurode- The purpose of this GIST protocol is to outline the process for de-
generative diseases (OECD, 2015). Alternatively, if the damage occurs termining whether a chemical agent is genotoxic or not, as well as the
in germ cells, it might manifest as reproductive defects or heritable level of conﬁ dence related to the assessment. The process allows for the
changes that could eventually result in genetic diseases (OECD, 2015). potential inclusion of additional information based on the results of
 Genotoxicity testing for hazard identiﬁ cation and risk assessment is other test methods or other supporting information, such as a history of
designed to characterize the ability of a chemical agent to induce ge- safe use in food (Constable et al., 2007). The process of performing a
netic alterations (OECD, 2016a). A comprehensive assessment of gen- risk assessment of a chemical agent will depend on many factors, such
otoxicity incorporates a battery of tests that evaluate for: as the exposure conditions and in what context the agent is being in-
 vestigated. Deﬁ ned risk assessment is considered out of scope for the
 i. Gene mutation (mutagenicity): Permanent, transmissible changes in GIST protocol and should be performed in a situation dependent con-
 the DNA that result from the induction of DNA adducts, insertions, text although the GIST protocol can be used to support this activity.
 inversions, and small deletions.
 ii. Clastogenicity: Structural chromosomal damage leading to sections
 of a chromosome being duplicated, deleted, or rearranged. 2. In silico methodologies
iii. Aneugenicity: Numerical chromosomal abnormalities (aneuploidy)
 where an abnormal number of chromosomes is generated, often by 2.1. Data availability for in silico models
 disruption of the microtubule apparatus necessary for the orderly
 segregation of chromosomes during nuclear division. The general protocol paper outlined some of the in silico meth-
 odologies that can be used to generate predictions (Myatt et al., 2018).
 In addition to test methods that evaluate these endpoints, tests di- These include i) rule-based (or “expert”) systems that identify the
rectly detecting the presence of DNA damage (e.g., sister chromatid presence of a structural moiety, also referred to as a structural alert,
exchange (SCE) assay, alkaline comet assay), or the repair of certain that may indicate genotoxic potential, and ii) statistical (quantitative
types of DNA damage (e.g., unscheduled DNA synthesis (UDS) test) structure-activity relationship (QSAR) models that use a variety of
have been used. In addition, the upregulation of DNA repair enzymes molecular descriptors such as structural fragments or physicochemical
and related stress response pathways that focus on individual genes properties to predict activity. Here, these two types are collectively
(e.g., Gadd45a (Gentronix, 2018), p53 (Witt et al., 2017), ATAD5 (Fox referred to as “(Q)SAR” models. In addition, “read-across” (OECD,
 3

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 Fig. 2. Common genetic toxicology mechanisms/eﬀ ects and corresponding studies.
2014) is a methodology that utilizes experimental or computed prop- strains have sometimes been modeled separately (Stavitskaya, 2013)
erties, such as physicochemical properties, together with structural si- from data generated using the Salmonella strains TA98, TA100,
milarity and experimental data for structural analogs to extrapolate TA1535, and TA1537, because the mechanistic basis of mutation in-
from source chemical(s) to a target (query) chemical(s) (OECD, 2014). duction is diﬀ erent for these two groups. The second group have GC
 The types of in silico tools that can be developed for a speciﬁ c base pairs at the primary reversion site, which are not as sensitive to
endpoint are, to a great extent, driven by the availability (amount and detecting certain oxidizing mutagens, cross-linking agents and hy-
quality) of experimental data for model development, as well as the drazines which are instead better detected using the TA102 or E.coli
degree to which the chemicals of interest exert their toxicity via a strains which have an AT base pair at the primary reversion site (OECD,
common mechanism. In silico tools are most easily developed for end- 1997a). An additional consideration is the distribution of the biological
points with a well understood and similar mode of action for which a response. It is not unusual for the available data sets to be skewed so
large number of data points are available. Bacterial mutagenicity is a that one assay result classiﬁ cation (e.g., negative/positive) occurs much
relevant example of such an endpoint and consequently, in this area of more frequently. Usually, inactive compounds are more abundant but
genetic toxicology, the development of (Q)SAR models is the most es- regardless, the resulting imbalance will require speciﬁ c strategies to be
tablished, largely due to the realization of an electrophilic mode of applied during the modeling procedure to avoid unbalanced predictions
action for many genotoxic agents (Miller and Miller, 1981; Ashby and due to the prior probability resulting from the training set distribution.
Tennant, 1988) and the availability of a large data set (> 2000 com-
pounds). Conversely, endpoints with less data or studies where the 2.2. In silico tools
response can be due to several diﬀ erent mechanistic pathways (e.g.,
chromosome damage) are more challenging for (Q)SAR modeling. 2.2.1. Mutagenicity
 Table 1 lists an estimate of the number of compounds in the public 2.2.1.1. Bacterial mutagenicity. The majority of in silico tools developed
domain with associated genetic toxicology data, as published in the for this endpoint have been built using data generated in the bacterial
Leadscope toxicity database (Leadscope, 2018). In many cases, multiple reverse mutation (Ames) assay that relies primarily on Salmonella
results for the same assay and chemical will be available, sometimes typhimurium tester strains. Historically, this assay has been viewed as
with conﬂ icting results and/or conclusions. Private or commercial or- the “gold standard” of mutagenicity testing and the ﬁ rst SARs relating
ganizations may have access to additional compounds with experi-
mental data (e.g. from product development), which can be combined Table 1
with publicly available data. There are some factors to consider, with Estimated number of compounds with genetic toxicology data in the public
respect to using experimental data for modeling or read-across: (1) data domain.\*
conﬂ icts need to be resolved (this is not always possible) and experi- Assay/study type Number of compounds
mental protocols need to be examined to ensure that only data mea-
sured and interpreted under similar conditions are merged, and (2) Bacterial mutagenicity 10,440
chemical structures need to be accurate. The general protocol provides Chromosome aberration (in vitro) 1690
more speciﬁ c details regarding considerations when using data for Chromosome aberration (in vivo) 360
 Mammalian cell mutagenicity (in vitro) 2390
modeling or read-across (Myatt et al., 2018). In the case of genetic Micronucleus (in vitro) 290
toxicology, it may be relevant to look at modeling certain subsets of Micronucleus (in vivo) 1850
assay results. For example, data generated using the Escherichia coli (E. \* Number of compounds with at least one experimental result listed as part of
coli) WP2 uvrA pKM101 and Salmonella typhimurium TA102 (TA102) the Leadscope toxicity database (Leadscope, 2018).
 4

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
chemical structure and bacterial mutagenicity using data generated start of exposure e.g., (Sofuni et al., 1990; Galloway et al., 2011).
using the Ames assay were published in 1988 by Ashby and Tennant Currently, from a regulatory context data generated using these two cell
(1988). Several tools are available for modeling this endpoint, lines are interchangeable, as well as those using other mammalian cell
including expert rule-based systems and statistical models. The lines such as human peripheral blood lymphocytes, as long as the same
application of two complementary models, one rule-based and one protocol is followed (OECD, 2016b).
statistical-based model, is described and recommended in the ICH M7
guideline for the evaluation of potential mutagenic impurities in 2.2.2.2. In vitro micronucleus. Few data following a standardized
pharmaceuticals (ICH, 2017) and by EFSA for dietary risk assessment protocol are available in the public domain for this endpoint, due to
(EFSA, 2016). the relatively recent adoption of an OECD test guideline (number 487)
 for this assay (OECD, 2010; OECD, 2016i), and as a consequence,
2.2.1.2. Mammalian cell mutagenicity. Both statistical models and rule- statistical modeling would be limited with a narrow applicability
based systems utilizing mouse lymphoma assay (MLA) (L5178Y cells) domain. The derivation of expert/structural alerts is therefore the
data are available. Historically, application of these models often most promising in silico approach until more data are published;
resulted in many false positive predictions, which was in part due to however, the in vitro micronucleus (MN) assay is becoming more
some of the experimental data from which the models were derived widely used than the in vitro chromosomal aberration (CA) assay and
being liberally interpreted as evidence of mutagenicity. The criteria for it is assumed that the body of data will grow in the near future. It is also
interpretation of the experimental data were re-evaluated by Moore likely that some larger organizations have proprietary models for this
et al., 2003, 2006, resulting in more stringent criteria, which led toendpoint. For the currently available public data, the majority of
changes to some of the experimental conclusions. In this context, and to positive data have not been diﬀ erentiated between clastogenicity and
ensure the best possible predictive power, it is important for the aneugenicity with regard to mechanism of action. In individual cases,
compilation of training sets to take the most contemporary data read-across may be possible if suitable chemical analogs, e.g. as deﬁ ned
evaluation criteria into account. It should be noted that the currently in the Read-Across Assessment Framework (ECHA, 2008; ECHA, 2017),
available in silico MLA models do not provide informationare available.
diﬀ erentiating mutagenicity versus clastogenicity, and either or both
endpoints may be implicated in a positive response in this assay. 2.2.2.3. In vivo chromosomal aberration. The number of available data
 For assays using other mammalian mutagenicity cell lines, such as points is small (mainly bone marrow studies performed in rats), since
those detecting mutations at hypoxanthine-guanine phosphoribosyl this assay is often reserved for mechanistic investigations rather than as
transferase (HPRT), and at a transgene of xanthineguanine phosphor- a core genotoxicity assay, limiting the use of statistical models beyond
ibosyl transferase (XPRT) which are treated as equivalent by some in- speciﬁ c compound classes. The derivation of expert alerts and the
dustry sectors and regulatory agencies, there are currently not enough application of read-across, when experimental data for analogs are
data available to generate useful models, although these assays may be available, may be the most relevant methodologies for this endpoint.
referenced in expert systems. This also applies to in vivo mutagenicity
studies. In addition to referencing them in expert systems, such data can 2.2.2.4. In vivo micronucleus. Most publicly available data have been
be used for read-across to support a weight-of-evidence scenario, if they generated using bone marrow and/or peripheral blood studies
are available. It is expected that there will eventually be enough data performed in mice or bone marrow studies performed in rats. The
available in the public domain to support model development. diﬀ erence in experimental procedure is a result of the fact that
 micronucleated erythrocytes are removed from the blood by the
2.2.2. Clastogenicity spleen in rats but not mice (Dertinger et al., 2011b; Hayashi, 2016).
 Both statistical and rule-based tools for in vitro and in vivo clasto- The ability to use ﬂ ow cytometry to measure the frequency of
genicity are available as commercial and free tools. Clastogenicity can micronucleated erythrocytes has greatly increased test chemical
result from numerous and diverse mechanisms of action (Bender et al., throughput while making the data collected more robust (Hayashi,
1974; Snyder, 2000, 2010; Kaina, 2004). Furthermore, cytotoxicity can 2016). Furthermore, this approach has been used to evaluate
confound the results of in vitro clastogenicity assessments (Kirkland micronuclei in immature erythrocytes in the blood of rats (MacGregor
et al., 2007b; Parry et al., 2010b; Galloway et al., 2011; Honda et al.,et al., 2006). Data are available in su ﬃ cient amounts to build a
2018). Consequently, it is challenging to build highly predictive in silico statistical model although it will have a limited applicability domain.
models for this endpoint. In addition, the supporting datasets are mostly Rat and mouse MN data should be analyzed separately as the diﬀ erent
quite small and therefore the applicability domain of these models is species have diﬀ erences in responses (positive/negative) to some
often limited from a chemical space perspective. In general, the in silico chemical agents. Diﬀ erent strains, sexes, or administration routes are
models are better at identifying reactive compounds that damage DNA usually not separated as there is not enough data to support this and the
directly, thereby leading to clastogenicity, than they are at correctly individual datasets would be too small. Read-across or rule-based
predicting compounds involved in indirect, non-DNA-reactive eﬀ ects systems may better address diﬀ erences in response where such factors
leading to clastogenicity (e.g., oﬀ -target interactions disturbing cellular are thought to be important and if they can be related to certain
homeostasis or non-covalent intercalation between DNA base pairs). chemical classes in a systematic manner. Historically, the increases in
For better prediction of indirect, non-DNA-reactive eﬀ ects, supple- MN formation in vivo have not been evaluated to determine if the
mental structural similarity searching or the use of speciﬁ c models for response is due to clastogenicity or aneugenicity. Any in silico model
the prediction of oﬀ -targets known to be involved in clastogenic eﬀ ects built using these data will therefore not be speciﬁ c as to the nature of
can provide additional important information (Olaharski et al., 2009; the type of chromosomal changes, but rather the endpoint of the assay,
Hsu et al., 2018). MN formation per se. In some cases, the mechanism of MN formation
 can be inferred by combined interpretation with other assay results or
2.2.2.1. In vitro chromosomal aberration. The majority of available data speciﬁ c staining techniques (e.g., kinetochore staining) (Hennig et al.,
in the public domain has been generated using Chinese hamster ovary 1988) or more recently based on size distribution using ﬂ ow cytometric
(CHO) or Chinese hamster lung (CHL) cell lines. Initially, in the 1980's, methods (Torous et al., 1998b). A speci ﬁ c example of a combined
it appeared that the two cell lines diﬀ ered signiﬁ cantly in theirinterpretation would be a negative prediction for in vitro or in vivo CA
sensitivity with respect to identifying genotoxic compounds but but a positive in vitro or in vivo MN prediction, leading to an overall
subsequent in-depth comparisons demonstrated that the apparent prediction that clastogenic eﬀ ects are unlikely but that aneugenic
diﬀ erences were due simply to when cells were sampled after the eﬀ ects are possible. It is clear that interpretation of experimental data
 5

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
could lead to such a conclusion and although more uncertain, one could assessment of genotoxicity (Wilde et al., 2017; Bryce et al., 2018); and
in theory interpret the in silico results in a similar manner. It would also those that integrate DNA damage response into an overall assessment of
be possible to look at the most similar examples in the training set and toxicity using high throughput transcriptomic proﬁ ling to derive points
in a read-across approach determine what the mechanism might be. of departure for risk assessment (Farmahin et al., 2017; Mav et al.,
Since the availability of data is scarce and because the underlying 2018). Any potential tools built using any of these test methods will not
mechanism has rarely been determined in historical data, this is not be further discussed in the GIST protocol as they tend to be less ex-
usually feasible with the current body of data available for modeling. tensively validated, even though they may be useful in some cases.
 Once such test methods are accepted by the wider community and their
2.2.3. Aneugenicity use is justiﬁ ed through validation exercises, in silico methods built from
 Historically, data generated using the in vitro and in vivo MN assays their data should be formally incorporated within this framework.
were not routinely evaluated in such a way that the mechanism of MN
formation could be determined. Consequently, it is often not possible to 2.3. Applying in silico tools
diﬀ erentiate an aneugen from a clastogen when evaluating the majority
of data published. The number of data points available for modeling The practical aspect of applying in silico tools was discussed in detail
where the mechanism has been unambiguously determined is small and in the general protocol (Myatt et al., 2018) including how to select
would therefore not support statistical modeling. The limited data for models based on their performance, applicability domain, and model
this endpoint could be suitable for read-across if analogs with me- complementarity as well as the factors to consider when running che-
chanistic information could be found, or for deriving expert alerts. micals through the models such as ensuring chemical drawing con-
Although not regularly reported, an increase in the number of mono- ventions are adhered to that follow any requirements of the model
nucleated cells with micronuclei can indicate an aneugenic mode of developer. Hence, this will not be further discussed here. Criteria for
action (Rosefort et al., 2004) and could be used to diﬀ erentiate between the selection of suitable in silico methodologies, as well as reporting
the two modes of action. With new automated methods (Torous et al., strategies were also detailed in the general strategy paper (Myatt et al.,
1998a; Dertinger et al., 2011a) for identifying and scoring micronuclei 2018).
becoming more widely used and special methods for diﬀ erentiation
between chromosome fragments (clastogenicity) and whole chromo- 2.4. Expert review of in silico tools
somes (aneugenicity) like kinetochore staining or analysis of the mi-
cronuclei size, it is anticipated that this situation will improve in the The application of in silico tools for hazard identiﬁ cation may in-
next few years. volve an expert review of both the models and the predictions. It is
 important to determine that the models were built according to ac-
2.2.4. Other endpoints cepted criteria (Myatt, 2016) and using relevant training datasets. The
 Other methods relevant for experimental genotoxicity testing have endpoint training data used will dictate what can be predicted. For
not yet been generally accepted for making regulatory decisions and/or example, if only compounds tested in E. coli uvrA pKM101 and S. ty-
the data generated by these test methods are not available in suﬃ cient phimurium TA102 are used to build a bacterial mutagenicity model,
amounts to build reliable in silico models (see (Mahadevan et al., 2011; then the output is only relevant for these strains and may not be ex-
Zeiger et al., 2015; Dearﬁ eld et al., 2017)). These test methods include trapolated to predict the outcome of a full OECD guideline compliant
those that evaluate the upregulation of speciﬁ c DNA damage response bacterial reverse mutation assay which requires at a minimum the in-
elements such as GADD45A (Knight et al., 2009; Hughes et al., 2012), clusion of ﬁ ve bacterial strains. Ideally, to ensure that the data origi-
H2AX (Kim et al., 2011; Mishima, 2017), ATAD5 (Fox et al., 2012), and nates from comparable protocols, only experimental data generated
TP53 (Clewell et al., 2014; Witt et al., 2017) using reporter genes, or using guideline-compliant conditions should be used. However, in
multiple DNA damage response elements evaluated using targeted practice, pragmatic approaches may need to be considered to ensure
transcriptomic platforms (Aubrecht and Caba, 2005; Sakai et al., 2014; that the models cover a wide chemical space without any unnecessary
Li et al., 2015, 2017; Corvi et al., 2016); those that evaluate the dif- compromise to data quality. As an illustration, experimental data from
ferential responses in wild-type and isogenic DNA repair deﬁ cient DT- an assay involving a limited number of bacterial strains are often in-
40 cells (Yamamoto et al., 2011; Nishihara et al., 2016) or TK6 cells cluded in model building if the compound is shown to be mutagenic in
(Saha et al., 2018); those that integrate multiple endpoints into anat least one strain and as long as the other experimental conditions
Table 2
In vitro genetic toxicology assays.
 OECD Test Guideline Name Endpoint Comments
 471 (OECD, 1997a) Bacterial reverse mutation test (Ames) Gene mutation
 473 (OECD, 2016c) In vitro mammalian chromosomal aberration test Clastogenicity
 476 (OECD, 2016f) In vitro mammalian cell gene mutation test (HPRT/XPRT) Gene mutation
 487 (OECD, 2016i) In vitro mammalian cell micronucleus test Clastogenicity/Kinetochore staining or MN sizing required to diﬀ erentiate
 Aneugenicity between clastogenicity and aneugenicity
 490 (OECD, 2016k) In vitro mammalian cell gene mutation tests using Gene mutation/ Mutant colony sizing may diﬀ erentiate clastogenic and
 thymidine kinase gene (MLA/TK6) Clastogenicity mutagenic events
 472 (OECD, 2015) Genetic toxicology: Escherichia coli, reverse assay Gene mutation Deleted by OECD (integrated into OECD 471)
 479 (OECD, 1986a) Genetic toxicology: In vitro sister chromatid exchange assayChromosome aberrations Deleted by OECD
 in mammalian cells
 480 (OECD, 1986a) Genetic toxicology: Saccharomyces cerevisiae, gene mutation Gene mutation Deleted by OECD
 assay
 481 (OECD, 1986b) Genetic toxicology: Saccharomyces cerevisiae, mitotic Mitotic recombination Deleted by OECD
 recombination assay
 482 (OECD, 1986c) Genetic toxicology: DNA damage and repair, unscheduled Unscheduled DNA synthesis Deleted by OECD
 DNA synthesis in mammalian cells in vitro
 In vitro comet assay DNA damage Not listed by the OECD but commonly used in the
 pharmaceutical sector
 6

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
adhere to established guidelines. The justiﬁ cation for this is that the test identiﬁ ed data as well as its relevance to any of the mechanistic as-
guidelines only require one strain to be positive for the test article to be sessments related to the major genotoxicity endpoints. In the general
considered mutagenic. For compounds to be considered negative, it is protocol publication (Myatt et al., 2018), we proposed to assess data
preferable to have negative data from all recommended strains (OECD, quality using Klimisch scores (Klimisch et al., 1997) as this is a widely
1997a). This is often not available and a certain degree of compromise accepted methodology used by ECHA, for example, in the Read-Across
in both the number of strains and data quality is usually accepted. For Assessment Framework (ECHA, 2017) and can readily be generated
the purpose of this protocol, assessing the underlying data used in the using the ToxRTool (European Commission, 2018b). Klimisch scores
model building is an important component of assigning a reliability rank data from 1 to 4, depending on how the experiment was conducted
score to the prediction, which will be discussed further in section 3.1. (and reported), taking into consideration for example, whether the
 experiment was compliant with Good Laboratory Practice (GLP) and
3. Laboratory data whether details of the experiment are available for review. These scores
 provide a consistent and reproducible way to classify the reliability of
3.1. Experimental assays and studies the test results.
 An expert review of any identiﬁ ed experimental dataset may be
 Tables 2 and 3 list in vitro and in vivo assays, respectively, that are performed to assign the appropriate Klimisch score. A detailed de-
frequently used to assess genotoxicity, as well as annotation of the scription of how this can be performed has been published by ECHA
mechanism(s) each assay may identify. For detailed descriptions of the (ECHA, 2011). For assays relating to genotoxicity that are mentioned in
experimental protocols, the OECD test guidelines may be consulted. this GIST protocol, the experimental conditions can be examined in
Test methods that are no longer supported by the OECD are listed in the relation to the relevant OECD test guideline. For test methods that no
tables but are not discussed further. Data that were generated in the longer have a current OECD test guideline, such as the in vitro SCE assay
past using such assays may be considered appropriate for use only if no in mammalian cells, a historical version of the guideline can be used to
additional or higher relevance data are available. In general, it is also determine whether the experimental conditions were relevant at the
important when using historical data to evaluate if the relevant reg- time of data generation. Although these data are considered of lower
ulatory guidelines or data requirements have changed since the data relevance as use of the assay was discontinued for being scientiﬁ cally
were generated, so that they can be assigned a contemporary quality questioned, there are situations where no other experimental data are
score. For example, changes to some of the in vitro protocols used in the available and they may be used in a weight-of-evidence scenario.
pharmaceutical sector were made after a 2006 EURL ECVAM workshop In addition, historical data that were generated under conditions
(Kirkland et al., 2007a) on assessing false positive rates of mammalian described in a previous version of a test guideline can be used if the data
in vitro tests. The new protocols introduced requirements for p53 were generated and reported in such a way that they can be re-eval-
competent cell lines and lowering of maximum tested concentrations, uated in accordance with the current guideline version and best prac-
amongst other things, to reduce the number of unnecessary follow-up in tice (e.g., as described by the International Workshop on Genotoxicity
vivo studies (Kirkland and Fowler, 2010; Parry et al., 2010a; Fowler Testing (Kirkland, 1994; Kirkland, 2000; Kirkland, 2003; Kirkland
et al., 2012a, 2012b). A number of these recommendations were et al., 2007c; Kirkland et al., 2007d; Kirkland et al., 2011; Martus et al.,
adopted under the OECD test guideline revisions performed in 2015)). This may not always be possible, and an expert review will
2014–2016. determine what reliability can be assigned on a case-by-case basis
 considering the particular chemical class, as well as the experimental
3.2. Expert review of experimental data details.
 In situations where multiple experimental results are available for a
 An important step in any hazard identiﬁ cation process is a search test substance, diﬀ erent scenarios can be envisaged. If several experi-
for existing experimental data from endpoint-relevant in vitro and in ments are found for the same assay that were performed in diﬀ erent
vivo assays. In this context, it is pertinent to assess the quality of any laboratories or under slightly diﬀ erent (guideline compliant)
Table 3
In vivo genetic toxicology assays.
 OECD Test Name Endpoint Comments
 Guideline
 474 (OECD, 2016d) Mammalian erythrocyte micronucleus Clastogenicity/aneugenicity Kinetochore staining or MN sizing required to diﬀ erentiate clastogens
 test and aneugens
 475 (OECD, 2016e) Mammalian bone marrow chromosome Clastogenicity
 aberration test
 478 (OECD, 2016g) Genetic toxicology: Rodent dominant Chromosome aberration by Germ cell assay
 lethal test clastogenicity/aneugenicity (gene
 mutations)
 483 (OECD, 2016h) Mammalian spermatogonial Clastogenicity Germ cell assay
 chromosome aberration test
 485 (OECD, 1986e) Genetic toxicology: Mouse heritable Clastogenicity/aneugenicity Not updated in 2014–16 revisions
 translocation assay
 486 (OECD, 1997b) Unscheduled DNA synthesis (UDS) test Unscheduled DNA synthesis Not updated in 2014–16 revisions
 with mammalian liver cells in vivo
 488 (OECD, 2013) Transgenic rodent somatic and germ Gene mutation Somatic and male germ cell assays. Many data reported before adoption
 cell gene mutation assays of the OECD test guideline so critical data review essential
 489 (OECD, 2016j) In vivo mammalian alkaline comet DNA damage Many data reported before adoption of the OECD test guideline so critical
 assay data review essential
 484 (OECD, 1986d) Genetic toxicology: Mouse spot test Gene mutation Deleted by OECD
 Pig-a Gene mutation No guideline adopted by the OECD but is under review for inclusion
 (Gollapudi et al., 2015). This assay may be considered as having high
 relevance in the pharmaceutical sector due to the inclusion in the ICH M7
 Note 3 as a follow-up assay to a positive bacterial mutagenicity result
 7

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
experimental conditions, the data with the best Klimisch score may be example, be subject to further testing or used under strictly controlled
given stronger weight. In cases where there are multiple conﬂ icting conditions. In contrast, a positive result in an in vitro CA test can be de-
results with the same Klimisch score and it cannot be determined risked or conﬁ rmed by performing an in vivo CA study. From a 3Rs point
through the expert review if one result is more reliable than the other of view, it is desirable to perform in vivo testing as a last resort and to
(s), the results can either be considered unusable for hazard identiﬁ - incorporate genotoxicity testing into general toxicology testing that
cation if they are of low quality, or a conservative approach might be may be required for other purposes. Generally, all in vivo studies are
taken where the occurrence of a positive result takes precedence. It considered to have high relevance with respect to an overall assessment
would be critical in this situation to scrutinize the experimental con- of genotoxicity. Conversely, tests where the OECD test guideline has
ditions in detail, taking into account factors such as compound purity, been deleted (OECD, 2016a) have been assigned a low relevance. The
potential cytotoxicity, solvent eﬀ ects, etc. Alternatively, a weight-of- relevance score is to some extent more subjective than the reliability
evidence approach could be taken with the ﬁ nal call being dependent score as diﬀ erent organizations and industry sectors, as well as reg-
on the judgement of a subject matter expert. ulatory agencies working to the data requirements of diﬀ ering regula-
 There are particular elements, or ﬁ elds, relating to the experiment tions, may apply diﬀ erent criteria in this respect. Even within an or-
that are essential to review and document when assessing data. This ganization, diﬀ erent toxicologists may have individual preferences and
practice supports an eﬃ cient and thorough review of the ﬁ nal assess- experiences inﬂ uencing their choice of assays. Furthermore, the che-
ments and ensures that the review is conducted in a consistent way. mical agent (with its physicochemical properties and structural aspects)
Table 4 lists the relevant ﬁ elds for any in vitro test and shows an ex- may dictate which assays are relevant. This protocol reﬂ ects a general
ample of an Ames assay result for ﬂ uorobenzene. Table 5 lists the re- view of assay relevance, but it is recognized that there could be situa-
quired ﬁ elds for an in vivo MN assay with bosutinib as an example. tions where an expert review may justify a diﬀ erent interpretation. The
 In addition to understanding the quality of the data, which relates to suitability of diﬀ erent assays in terms of follow-up actions, mechanisms
the technical aspect of the information, the scientiﬁ c relevance to the identiﬁ ed by each assay and many other aspects have been reviewed
toxicological endpoint result needs to be determined. “Relevance” was and discussed in a publication by the Health and Environmental Sci-
deﬁ ned in the general protocol (Myatt et al., 2018) and relates to the ences Institute (HESI) In Vitro Genetic Toxicity Testing Review Sub-
predictivity of a speciﬁ c toxicological eﬀ ect or mechanism (gene mu- group (Dearﬁ eld et al., 2011).
tation, clastogenicity, aneugenicity) to the toxicological endpoint
(genotoxicity). As an example, the bacterial mutagenicity assay is
considered highly relevant with respect to genotoxicity, whereas an in 3.3. Sources of genetic toxicology data
vitro CA test may be considered to have lower relevance (Custer, 2015).
The rationale is related to how these tests are managed in a practical Table 6 provides a non-exhaustive list of available sources of genetic
setting, where a bacterial mutagenicity assay is often not followed up toxicology data. There are also databases that comprise several sources.
with in vivo testing and, in many industries, a positive result in this Individual databases will support diﬀ erent types of queries such as
assay is often considered suﬃ cient to stop the development of a can- various identiﬁ ers (Chemical Abstracts Service registration number
didate active substance. Other industry sectors may adopt a diﬀ erent (CASRN), synonym or chemical name) and/or chemical structure. If
level of concern and a manufacturing chemical intermediate might, for possible, it is desirable to know the batch of compound that was tested,
 as well as the associated characterization data, as it is relevant to know
Table 4
Relevant ﬁ elds to document for an in vitro assay. Example: bacterial mutation assay.
 Compound identiﬁ er: CASRN: 462-06-6; Fluorobenzene; Benzene, ﬂ uoro-
 Compound purity: Not reported
 Compound solubility: Not reported
 Solvent DMSO
 Study call: Positive
 Title: Genetic Toxicity Evaluation of Fluorobenzene in Salmonella/E.coli Mutagenicity Test or Ames Test. Study 639736
 Reference: https://tools.niehs.nih.gov/cebs3/ntpViews/?activeTab=summary&studyNumber=639736
 Study type: Bacterial mutagenicity (Ames)
 Source: National Toxicology Program
 Species: Salmonella typhimurium
 Strain/Cell type (Number TA98 (N = 6); TA100 (N = 7); TA1535 (N = 2)
 of tests):
 Metabolic activation Absent (N = 4); Present (N = 11)
 (Number of tests):
 Metabolic activation Aroclor 1254 treated rat liver S-9 fraction (30%), Aroclor 1254 induced hamster liver S-9 fraction (30%)
 system:
 Dose summary: 0–1666 μ g/plate; 0–750 μ g/plate
 Toxicity: No cytotoxicity reported
 Method: Pre-incubation; Plate test – vapor from liquid
 Controls used: Strain (wo Positive controls: TA98 (4-Nitro-o-phenylenediamine/2-aminoanthracene), TA100 (Sodium azide/2-aminoanthracene), TA1535 (Sodium azide/2-
 S9/w S9): aminoanthracene)
 Control values within Yes
 historical ranges:
 OECD test guideline: 471
 Current guideline No, insuﬃ cient bacterial tester strains used
 compliance
 Study Report: https://tools.niehs.nih.gov/cebs3/ntpViews/?activeTab=detail&studyNumber=639736&reportFormat=XLS
 GLP compliance: No
 Year conducted 1991
 Klimisch score: 3
 Rationale for reliability Not tested up to guideline recommended concentrations and insuﬃ cient number of strains and not tested according to GLP
 incl. deﬁ ciencies:
 8

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
Table 5
Relevant ﬁ elds to document for an in vivo assay: example in vivo micronucleus.
 Compound identiﬁ er: CASRN: 918639-08-4; Bosutinib
 Compound purity: 99.49%
 Compound solubility: Not reported
 Study call: Negative
 Title: SKI-606: Single dose oral (gavage) bone marrow micronucleus study in male mice
 Reference: http://www.accessdata.fda.gov/drugsatfda\_docs/nda/2012/203341Orig1s000PharmR.pdf#page=151
 Study type: In vivo clastogenicity assay in rodent (micronucleus assay)
 Source: FDA CDER
 Species (Number of CD mouse (N = 6)
 subjects):
 Target cell/organ: Bone marrow
 Sex: Male
 TK parameters Dose 2000 mg/kg, Cmax = 9811 ± 3998 ng/mL, tmax = 2.0 h, AUC0-24 = 172495 ± 26050 ng\*hr/mL
 No observed adverse event 2000 μ g/kg (from single dose toxicology study) (unit as stated in original report); it is assumed that this is an error and the correct level is
 level: 2000 mg/kg
 Dose summary: 0, 500, 1000, and 2000 mg/kg as 10 ml/kg, single oral dose
 Number of days of 1
 treatment:
 Timepoints for tissue 24 h; 48 h
 harvesting:
 Controls used: Positive control: Cyclophosphamide (50 mg/kg)
 Current guideline Yes (OECD 474)
 compliance
 Study Report: http://www.accessdata.fda.gov/drugsatfda\_docs/nda/2012/203341Orig1s000PharmR.pdf
 GLP compliance: Yes
 Year conducted 2003
 Klimisch score: 1
 Rationale for reliability incl.
 deﬁ ciencies:
the purity of the tested chemical. The presence of potential impurities is Structure searches should be performed with care, considering factors
important; even small quantities of a mutagenic impurity may result in such as stereochemistry, tautomerism, salt form, and counter ions, for
a false positive result. This type of information is not always available example. It might be necessary to search for both the parent compound
in public databases but can often be found in corporate databases. and alternative forms when searching for a particular chemical if it is
Table 6
Some sources of genetic toxicology data (non-exhaustive list).\*
 Database Description
 ATSDR Open access database from the Agency for Toxic Substances and Disease Registry (ATSDR) includes toxicological proﬁ les for the hazardous substances
 including genotoxicity (ATSDR, 2018)
 CCRIS Chemical Carcinogenesis Research Information System (CCRIS), open access database covering chemical carcinogens and genotoxicants, including
 structures and experimental data, covering the period 1985–2011 (CCRIS, 2011)
 Drugs@FDA Open access database from US FDA CDER product approval reviews (FDA, 2018)
 EPA Comptox Dashboard Open access Distributed Structure-Searchable Toxicity (DSSTox) Database Network from the United States Environmental Protection Agency (EPA)
 including content from other sources (e.g., CPDB, ISSCAN, Tox21 and ToxCast) (DSSTox 2018, EPA 2018, EPA 219)
 ECHA Open access European Chemicals Agency (ECHA) database containing experimental data and read across results for chemicals manufactured and
 imported in Europe as regulated by the REACH guidance (ECHA, 2018)
 ELSIE The Extractables and Leachables Safety Information Exchange (ELSIE) database is a collection of experimental data shared and accessed by
 consortium members (ELSIE, 2018)
 EURL ECVAM Open access Genotoxicity & Carcinogenicity consolidated database containing available genotoxicity and carcinogenicity data for Ames positive
 compounds (European Commission, 2018a)
 GENE-TOX GENE-TOX provides genetic toxicology (mutagenicity) test data from expert peer review of open scientiﬁ c literature for more than 3000 chemicals
 from the EPA (GENE-TOX, 1998). GENE-TOX covers the years 1991–1998.
 IPS INCHEM Open access International Program on Chemical Safety search for variety of summary documents (INCHEM, 2015)
 IRIS Open access data from the EPA in support of human health risk assessment, focusing on hazard identiﬁ cation and dose-response assessment (IRIS,
 2015)
 ISSCAN Open access database on chemical carcinogens, including structures and experimental data from Istituto Superiore di Sanità (Benigni et al., 2008)
 ISSMIC Open access database on in vivo micronucleus mutagenicity results from Istituto Superiore di Sanità (Benigni and Bossa, 2008; Benigni et al., 2012)
 JECDB Open access Japanese Existing Chemical Data Base (JECDB) containing high production volume chemicals (JECDB, 2018)
 Leadscope Commercial genetic toxicity databases from numerous sources (including US FDA CDER product approval reviews, FDA CFSAN, National Toxicology
 Program (NTP), CCRIS) as well as ongoing data harvesting from the literature (Leadscope, 2018)
 NTP – CEBS Chemical Eﬀ ects in Biological Systems (CEBS). Open access database of NTP results (NTP, 2018)
 PAN Open access Pesticide Action Network (PAN) Pesticide Database (PAN, 2018)
 PharmaPendium Commercial preclinical toxicity and clinical safety data from FDA and EMA approval documents (PharmaPendium, 2018)
 RTECS Registry of Toxic Eﬀ ects of Chemicals (RTECS). Commercial database available through third parties (e.g., Leadscope) (Sweet et al., 1999; RTECS,
 2018)
 TOXNET/ChemIDPlus Open access on-line toxicity search system from the US National Library of Medicine with access to archived versions of CCRIS and GENE-TOX
 (Wexler, 2001; TOXNET, 2018)
 OECD QSAR Toolbox Open access to database of genotoxicity as well as other toxicology data. (OECD, 2019)
 VITIC Commercial database from Lhasa Limited, including data from published and unpublished sources (VITIC, 2018)
\*Modiﬁ ed from Amberg et al. (2016)
 9

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
not known how the structures have been reported. It may be helpful to diﬀ erences in chemical space. In these cases it is relevant to search
perform a substructure search, which looks for compounds with open proprietary databases, as these often contain high quality sources of
substitution patterns, or a “family search” that will retrieve diﬀ erent information. From the documentation and reporting point of view, as
salt forms and also analogs with diﬀ erent chirality. Some databases well as for any regulatory submission, this may be an issue as a thor-
additionally provide regulatory authority classiﬁ cation with respect to ough expert review and ﬁ nal assessment needs to be documented and
mutagenicity and carcinogenicity; the International Agency for Re- disclosed to reviewers to enable their independent evaluation. Any
search on Cancer (IARC), will for example, provide carcinogenicity analogs or other relevant structures should preferably be included in
classiﬁ cation and ECHA provides carcinogenicity, mutagenicity or re- the ﬁ nal report for full transparency of the assessment.
productive toxicity (CMR) classiﬁ cations.
 4. Combined assessment of in silico predictions and experimental
3.4. Other data references data
 In addition to the above listed databases, other sources of data, such 4.1. Reliability score
as model training sets or other compilations of experimental data, can
be searched for supporting information. In some cases, substances may The general protocol (Myatt et al., 2018) provides detailed in-
have already undergone a risk assessment by a regulatory committee, formation on how to combine in silico predictions with experimental
and this output can be useful either directly or in a modiﬁ ed format in data, where these are available. The process will not be outlined in this
the hazard identiﬁ cation process. For instance, for the evaluation of publication, but involves expert review of the model(s), the prediction
bacterial mutagenicity the ICH M7 (R1) addendum (ICH, 2017) pro- (s) as well as a review on the quality of the experimental data. In
vides detailed information on risk assessment of a number of chemicals general, it is preferable that experimental data be of Klimisch score 1 or
and is applicable in the pharmaceutical sector. The addendum discusses 2, depending on the situation, to be considered of high enough quality
acceptable intakes of certain chemical residues or impurities that are to support decision making. It is recognized that this is not always
mutagens and/or carcinogens and that are common in pharmaceutical possible. However, depending on the use case, there could be situations
manufacturing. Another source is the “EURL ECVAM Genotoxicity & where expert review and data quality assessment is not feasible and a
Carcinogenicity Consolidated Database of Ames Positive Chemicals” lower level of conﬁ dence is acceptable, such as screening.
(European Commission, 2018a) (also listed in Table 6), which con- To enable a standardized method of performing an assessment of
tains > 700 unique chemical compounds that are bacterial mutagens experimental results and in silico results together, an extension to the
and have a variety of additionally reported in vitro and in vivo geno- Klimisch score (Table 7) has been introduced to allow scoring of in silico
toxicity and carcinogenicity data. This database contains an “overall components alongside experimental results using a Reliability Score
call” based on a set of deﬁ ned criteria for the reliability and quality of (RS) (Myatt et al., 2018). Experimental data of Klimisch score 1 and 2
the data when results from more than one source are available. are essentially unchanged in their original Klimisch description but are
 The Joint FAO/WHO Expert Committee on Food Additives (JECFA) referred to as RS1 and RS2. Furthermore, the lower quality Klimisch
maintains a database of ﬂ avors, food additives, contaminants, toxicants, categories 3 and 4 have been placed in the lowest RS category of 5. This
and veterinary drugs that have been reviewed with respect to human accommodates the use of in silico results of high quality in categories
safety (JECFA, 2018). Similarly, there are databases with food and RS3 and RS4, illustrating their higher acceptability in certain regulatory
ﬂ avor substances that are Generally Recognized as Safe (GRAS). The contexts, compared to low quality experimental data or single, lower
FDA maintains the Select Committee on GRAS Substances (SCOGS) quality in silico result (RS5). For genetic toxicology, this is of particular
(GRAS, 2018) database containing reports of the opinion and conclu- importance for both REACH and ICH M7 applications, for example.
sions on food substances while The Flavor and Extract Manufacturers
Association of the United States (FEMA) (FEMA, 2018b) maintains the 4.2. Toxicological eﬀ ect or mechanism assessment
FEMA GRAS lists (FEMA, 2018a) of GRAS ﬂ avors. In addition, organi-
zations like the National Institute for Occupational Safety and Health Toxicological eﬀ ects are deﬁ ned as observations derived from the
(NIOSH) (NIOSH, 2018), the Occupational and Health Administration experimental tests considered relevant for genetic toxicology (i.e., the
(OSHA) (OSHA, 2018), the OECD (OECD, 2018) and various regulatory in vitro and in vivo tests listed in Tables 2 and 3). An assessment will take
agencies have searchable data repositories that can be accessed. into account all of the experimental and in silico information available
 For many commercial organizations, it may be diﬃ cult to ﬁ nd for the query compound for each eﬀ ect separately, in a weight-of-evi-
structural analogs for proprietary compounds in public databases due to dence scenario. A simple hypothetical example is shown in Fig. 3.
Table 7
Reliability of toxicity assessments based on computational models and experimental data (Myatt et al., 2018).
 Reliability Score Klimisch Score Description Summary
 1 1 Data reliable without restriction  Well documented and accepted study or data from the literature
  Performed according to valid and/or accepted test guidelines (e.g., OECD)
 2 2 Data reliable with restriction  Well documented and suﬃ cient Preferably performed according to good laboratory practices (GLP)
  Primarily not performed according to GLP
 3 – Expert Review  Read-across Partially complies with test guideline
 4 – Multiple concurring prediction results Expert review of in silico result(s) and/or Klimisch 3 or 4 data
 5 – Single acceptable in silico result
 5 3 Data not reliable  Inferences between the measuring system and test substance
  Test system not relevant to exposure
  Method not acceptable for the endpoint
 5 4 Data not assignable  Lack of experimental details Not suﬃ ciently documented for an expert review
  Referenced from short abstract or secondary literature
 10

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 on the technical reliability of a result with the relevance of the assay
 from which it was derived, for predicting the toxicological endpoint
 being assessed. The determined conﬁ dence for each endpoint (gene
 mutation, clastogenicity, aneugenicity) will eventually propagate to the
 conﬁ dence for the overall call as to genotoxicity. As was discussed
 earlier with respect to relevance, the assigned conﬁ dence is somewhat
 subjective. To provide a starting point for how to combine terms, a set
 Fig. 3. Eﬀ ect assessment of the reverse bacterial mutation assay. of rules has been devised for combining results, based on a conservative
 approach for combining relevance and reliability for the most com-
 In this case, experimental data were found for the compound and it monly occurring components of genotoxicity hazard identiﬁ cation. This
was reported to be inactive in a limited (too few strains) bacterial re- rule set is available in the supplementary material of this publication
verse mutation test. After expert review, it was concluded that the assay and can be adapted to accommodate organizational preferences or
was run under appropriate conditions, but only in strains TA98 and other needs. There may be times when it is not desirable to perform a
TA100. The result is hence not suﬃ cient to support a full assessment of full evaluation of all the genotoxicity endpoints. For example, only
bacterial mutagenicity and a Klimisch score of 3 is assigned to the data, bacterial mutagenicity is required for an ICH M7 assessment. Subsets of
which results in a reliability score of RS5. Two complementary in silico the components can be used as appropriate in a situation dependent
models for bacterial mutagenicity (incorporating E. coli/S. typhimurium manner. A scheme including the genotoxic eﬀ ects and endpoints that
TA102 and additional Salmonella data into both models) were applied are amenable to the generation of in silico tools, using data currently
and the compound was predicted to be negative in both models. The available in the public domain, is shown in Fig. 5. It is possible that
individual models have initial reliability scores of RS5 but since they private organizations have additional types of data that could also be
concur, the combined score would be RS4. Further expert review used to generate in silico tools.
showed that the predictions were of good quality and there were, for An expert review of all the endpoint evaluations (described in sec-
example, no reactive features identiﬁ ed. The in silico predictions are tion 4.4) may be performed to balance the relevance of each assay call
therefore assigned a reliability score of RS3. The weight-of-evidence for to the overall genetic toxicology assessment. Depending on the use case,
this compound supports the assessment that the compound is not a the conﬁ dence required may vary. For situations where false negatives
bacterial mutagen and the overall reliability score for bacterial gene may be acceptable and not be associated with health consequences,
mutation is set to RS3. For comparison, if the experimental results had such as prioritization for more in-depth experimental testing, a lower
been reported as positive in one of the two strains, the Klimisch score level of conﬁ dence may be acceptable. However, in a human health
would still have been 3 and the initial reliability score would have been hazard identiﬁ cation and risk assessment situation, a more conservative
RS5. However, during expert review of the experimental data, it would view is taken and higher conﬁ dence is required. In the general protocol
have been appropriate to consider the result suﬃ cient for an assessment (Myatt et al., 2018), we outlined the general principles around the in-
of bacterial mutagenicity and to change the reliability score to RS3 as ﬂ uence that a particular level of conﬁ dence has.
one positive strain is considered enough to make a positive call for the
compound. 4.4. Expert review of combined endpoint assessments
4.3. Toxicological endpoint assessments The expert review of genotoxic eﬀ ects may include review of the in
 silico predictions and experimental data, as outlined earlier. The as-
 Combining the genotoxic eﬀ ect assessments that relate to a speciﬁ c sessments might involve an expert review to weigh the individual assay
genotoxic endpoint is required to generate an overall endpoint call. results and in silico predictions, as well as any other information, such as
Fig. 4 shows a continuation of the hypothetical example from Fig. 3 and experimental data for structural analogs or details that would inﬂ uence
illustrates the inclusion of a mammalian gene mutation result. the interpretation or translatability of a result. For example, a com-
 To perform this summary assessment, the concept of “Conﬁ dence” pound with antibacterial properties may be diﬃ cult to assay in a bac-
was introduced. Where “Reliability” relates to the quality of the ex- terial reverse mutation assay, due to the expected high cytotoxicity in a
perimental data or the in silico prediction and “Relevance” relates the bacterial reverse mutation assay, and therefore, mammalian cell sys-
assay to the mechanism or toxicological eﬀ ect, “Conﬁ dence” combines tems are usually recommended in these cases. Along similar lines, if
the two parameters in addition to assessing the completeness (or cov- such a compound is predicted with in silico tools to be negative in a
erage) of the information. It provides a method for merging information bacterial mutation test, even with high reliability, but predicted by an
 Fig. 4. Combining information to assess the “gene mutation” endpoint. \*The assignment of the “Conﬁ dence” is discussed in the following sections.
 11

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 Fig. 5. Current in silico components most relevant to genotoxicity.
in silico model to be a mammalian mutagen, the expert review may are discussed in the following section. For some of these, it is not
consider that the bacterial reverse mutation result may be misleading in possible to disclose the chemical structures as they are proprietary
the context of a combined “gene mutation” review. Even though the compounds. However, the included examples have been selected to
bacterial reverse mutation result would normally be considered to be of show various aspects of the GIST protocol; emphasizing how the var-
higher relevance due to the availability of more chemically diverse and ious model outputs and experimental data components can be ﬁ tted
abundant data for this endpoint, the mechanistic expert review could in into this framework without judging the validity of the generated
this case rank the mammalian in silico prediction higher. It may at this components.
point also be important to include information from primary DNA da-
mage experiments (or models) to determine the mechanism of action. 4.5.1. Toxicological eﬀ ect or mechanism examples
Table 8 includes some points to consider during an endpoint assess- 4.5.1.1. Acid chloride (bacterial gene mutation). Fig. 6 shows a case
ment. The expert review will also determine the level of conﬁ dence that study of an acid chloride impurity which is being assessed for bacterial
can be placed in the endpoint summary. gene mutation potential for ICH M7 risk assessment. No experimental
 data could be found for the compound and two in silico tools, one
4.5. Worked examples statistical- and one rule-based, were applied. The prediction from the
 statistical model indicates that the compound may be a bacterial
 A number of case studies that have been contributed by co-authors mutagen due to the presence of the acid halide functionality. The
Table 8
Some elements of a mechanism expert review.
 Expert review elements Considerations
 1. Chemical class assay response Information such as if the compound belongs to a chemical class that may not be suited to particular assays, for example
 antibiotics in the bacterial reverse mutation assay, or interaction of the test substance with selected vehicle. In such cases, the
 bacterial mutagenicity result might be considered inappropriate and results from mammalian cell assays should be used.
 2. Mode of action Combinations of assay results within a particular mechanistic class may provide information on the mode of action of a compound
 (e.g., diﬀ erent bacterial strains are speciﬁ c for diﬀ erent types of mutations).
 3. Alerts that predict a particular mechanism Some alerts may provide information on the mechanism through which a compound acts.
 12

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 consider if there are other reasons for the observation of mutagenic
 activity related to the experimental procedures and/or the test article.
 One of the more frequently occurring reasons for an unexpected
 positive response in bacterial mutagenicity assays is the presence of a
 potent mutagenic impurity in the test article. In this particular case, an
 aldehyde was identiﬁ ed as a degradation product in API X and shown to
 be mutagenic. Follow-up testing of puriﬁ ed API X found it to be non-
 mutagenic and the bacterial gene mutation assessment would at this
 point be updated from the “Indeterminate” to “Negative” with a
 reliability score of RS2. Since the formation of the degradant could be
 avoided by modiﬁ cation of the synthetic route, it had no direct bearing
 Fig. 6. Assessment of an acid chloride compound. on the classiﬁ cation of API X.
rule-based model gives an “Indeterminate” prediction and also 4.5.1.3. 3-Methyl-5-isothiazolamine (bacterial gene mutation). Fig. 8
highlights the acid halide functionality. Acid halides are a structural shows the assessment components for 3-methyl-5-isothiazolamine
alert class for bacterial mutagenicity that were discussed recently related to bacterial gene mutation. Experimental data were available
(Amberg et al., 2015) and it was shown that with the exception of in the public domain for this compound where it was reported to have
dimethylcarbamic chloride, the compounds tested and available for been tested in TA98, TA100, TA1535, TA1537 and TA1538 with and
model building were active in the bacterial reverse mutation assay due without metabolic activation using induced rat liver S9 and hamster
to a reaction between the DMSO solvent and the test agent. When liver S9 (Cameron et al., 1985). Further examination of the data
retested in other solvents, the majority of compounds show norevealed that the experiments were conducted under acceptable
mutagenic activity. Despite the positive and indeterminate in silico conditions and that the tested concentration range went to higher
predictions, each with a reliability of RS5, an expert review revealed levels than normally required by OECD TG 471, the test guideline for
that the underlying data for the statistical model supporting the the bacterial reverse mutation assay, but the compound was not tested
prediction are with high certainty false positives and the prediction in an E. coli or S. typhimurium TA102 strain, which is required to fulﬁ ll
was refuted. Expert review of the supporting text for the alert supports the current OECD test guideline. The standard maximum concentration
this outcome. The overall assessment of bacterial mutagenicity is usually set to 5 mg/plate and this study reported maximum
concludes that the compound is predicted to be inactive (negative) concentrations of 7.43 mg/plate. The data were initially assigned as
and the reliability score is set to RS3. The approach to this assessment is positive with a Klimisch score of 3, indicating that the experiment was
aligned with current ICH M7 guidance. partially compliant with guidelines. However, when assessing the
 individual bacterial strain concentration responses, the biological
4.5.1.2. Drug impurity - API X (bacterial mutation). There may berelevance of the data was further questioned as the compound was
situations when an expert review can give an indication that the only active in TA1538 (a strain not required by the OECD test
experimental results might not be correct. This is illustrated in the guideline) at concentrations higher than the guideline recommended
following example using Active Pharmaceutical Ingredient (API) X. API 5 mg/plate and only with hamster S9 metabolic activation. With rat S9
X was initially tested in the bacterial reverse mutation assay and found and at concentrations up to 5mg/plate, the compound was found to be
to have mutagenic activity. In contrast, as shown in Fig. 7, the in silico inactive. At this point, an expert review of the data indicated that as the
predictions from both the statistical and the expert alert models predict compound was negative at concentrations up to regulatory requirement
API X to be inactive in the bacterial reverse mutation assay. An expert of 5 mg/plate, the compound could potentially be viewed as negative
review of the information indicates that the models as well as the with a reliability score of RS5 as this cannot be increased, considering
predictions appear robust and the reliability score which initially is set that the compound was not tested in E. coli or S. typhimurium TA102
to RS4 due to two concurring models, is raised to RS3 after the expert strains. Additionally, there is discrepancy seen with the two metabolic
review. In cases where experimental data are positive and in silico activation systems. In silico methods were applied to further reﬁ ne the
predictions are negative, the conservative approach would be to accept hazard identiﬁ cation. When reviewing these results, the statistical
the positive experimental data, in which case the assessment would be model output from a Salmonella model classiﬁ ed the compound as out
positive with a reliability score of RS1, RS2, or RS5, depending on the of domain and the E. coli model predicted it to be negative. Review of
quality of the experimental data. However, if the scientiﬁ c review the E. coli model results indicated that the prediction was not supported
suggests that there is a valid reason to question the experimental result, by many analogs or structural descriptors and is mainly driven by
the initial assessment for the compound could be Indeterminate, given physicochemical properties. The expert alert model predicts 3-methyl-
the conﬂ icting results from the experimental and in silico outputs, 5-isothiazolamine to be positive for bacterial gene mutation. In this
although this outcome would not be acceptable as a ﬁ nal conclusion case, however, there are compounds in the reference set that contain
from a drug regulatory standpoint. A reliability score is not assigned if the thiazolamine functionality that the alert is based on, but they are
the assessment is considered indeterminate. Given that the structure of not necessarily isothiazolamines. Additionally, further review shows
API X is not predicted to be DNA reactive, it could be relevant tothat the majority of the reference structures also have other alerts such
 Fig. 7. The conﬂ icting in silico and experimental results of API X feeding into the overall bacterial gene mutation assessment.
 13

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 Fig. 8. Bacterial gene mutation assessment of 3-Methyl-5-isothiazolamine.
as aromatic nitro groups. At this point, there is contradictoryAn expert review was performed on the in silico model results and the
information to consider: the low reliability (RS5) experimental result review concluded that the predictions were well supported and there
indicating the compound is negative up to 5 mg/plate but active at was suﬃ cient evidence to increase the reliability to RS3 from the
higher concentrations, and the inconclusive in silico results. By formally individual models’ scores of RS5. A single statistical model predicted
following the proposed scheme, it would be acceptable to view the the compound as negative for mammalian gene mutations (built using
compound as negative, but with a reliability score of RS5, as the expert MLA training set data). An expert review was performed but the
review did not reveal evidence supporting a higher score. In aevidence concerning the prediction was not considered suﬃ cient to
conservative scenario, if this compound, for example, is an impurity raise the reliability score higher than RS5. The results from the bacterial
that has consequences for human safety, retesting the compound in a and mammalian gene mutation endpoints were used as part of the
guideline acceptable study would be preferred. Indeed, when 3-methyl- assessment of the overall gene mutation potential. The conﬁ dence was
5-isothiazolamine was retested according to the OECD guideline in a assigned as “Medium” as outlined in the suggested set of rules in the
full 5-strain bacterial reverse mutation assay, with and without induced supplementary information. It should be noted that this prediction itself
rat liver metabolic activation, the compound was found to be non- refers to the in vitro gene mutation response. In a scenario where this
mutagenic (Ahlberg et al., 2016). At this point, the assessment could be result would feed into a framework supporting overall genotoxic
updated with a “Negative” result with a reliability score of RS1potential, it would be pertinent to consider that certain aromatic
assigned. It should be noted that this assessment refers speciﬁ cally to amides and sulfonamides do not show activity in the bacterial assay
bacterial gene mutation and that any other available experimental data, due to the amide bond not being metabolized by S9, but may be active
such as MLA data, would be used to support the corresponding endpoint in an in vivo experiment.
they relate to, which may or may not diﬀ er from the bacterial
mutagenicity assessment. For a more comprehensive analysis of
potential genotoxicity, such data may need to be considered and 4.5.2.2. Plant protection product active ingredient metabolite assessment
follow-up testing may need to be performed. (genetic toxicology). A herbicide metabolite was assessed using in silico
 methods for genotoxicity. Experimental data generated on the active
 ingredient (AI) was available and the data conﬁ rmed that the AI has no
4.5.2. Toxicological endpoint examples genotoxic potential based on negative bacterial gene mutation, in vitro
4.5.2.1. Aromatic amide (gene mutation). Fig. 9 shows the assessment of mammalian gene mutation and in vitro CA assay results, as well as a
a compound containing an aromatic amide functionality. Bacterial gene negative in vivo CA study. The metabolite was noted to have high
mutation and mammalian gene mutation eﬀ ects/mechanisms were structural similarity to the AI. Fig. 10 shows the initial in silico
identiﬁ ed as relevant to the assessment of the gene mutationgenotoxicity assessment of the metabolite. The metabolite was
endpoint. Two independent and concurring in silico models were run predicted by two methodologies to be inactive in the bacterial reverse
to predict bacterial gene mutation, one expert rule-based and the mutation assay. It was out of domain for the mammalian gene mutation
second statistical-based, and both model predictions were negative. model as well as the in vivo CA model (however, the related endpoint
 Fig. 9. Gene mutation assessment of an aromatic amide compound.
 14

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
Fig. 10. The initial in silico genetic toxicology assessment for the plant protection product active ingredient metabolite. Note the change in assessment outcome for in
vitro CA before and after expert review. \*NA refers to “Not available” since these results were not possible to generate.
“in vivo MN prediction” was in domain). Two expert alert systems for in endpoint. Similarly, the in vitro and in vivo clastogenicity/aneugenicity
vitro CA induction were applied, one indicating that the compound has endpoints were considered to have low conﬁ dence related to the ne-
clastogenic potential due to the presence of a carboxylic acid related gative assessments as there was limited information available. The
alert, and the other that it does not. Expert review of the in silico results combination of these assessments resulted in the metabolite being
was performed by looking at speciﬁ c details of the alert and theconsidered to have low genotoxic potential but with a low conﬁ dence.
surrounding SAR. Suﬃ cient experimental data for analogs matching the It should be noted that this assessment did not take aneugenicity into
alert convinced the assessor that the alert could be dismissed, and the in account at all with the exception of a predicted negative in vivo mi-
vitro CA endpoint was set to negative with a reliability score of RS3, cronucleus result. This is an additional reason to consider this assess-
after the expert review. Expert review was also performed on the gene ment of being of low conﬁ dence.
mutation endpoints as well as the predicted in vivo MN results to Following the in silico assessment exercise, the metabolite was tested
conﬁ rm that these were of suﬃ cient quality. In the case of genein experimental assays for conﬁ rmation. The compound was tested in
mutation, the reliability score could be increased to RS3, but this was an OECD and GLP compliant bacterial reverse mutation assay and
not the case for the in vivo MN assessment and it remained at RS5. found to be inactive. It was also tested in an OECD and GLP compliant
 Following the suggested conservative scheme included in the sup- in vitro micronucleus assay and again, no activity was detected. Fig. 11
plementary material, for combining toxicological eﬀ ect outputs, the shows how these experimental data would inﬂ uence the assessment if
gene mutation endpoint was considered as negative with a low con- the protocol framework was applied. The increased reliability scores
ﬁ dence due to the lack of information on the mammalian gene mutation from the bacterial gene mutation and the in vitro micronucleus tests
 15

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
Fig. 11. Inﬂ uence of including the experimental results in genetic toxicology assessment for the plant protection AI metabolite. Diﬀ erences compared to Fig. 10 are
indicated in red text. \*NA refers to “Not available” since these results were not possible to generate. (For interpretation of the references to colour in this ﬁ gure
legend, the reader is referred to the Web version of this article.)
would result in high conﬁ dence in the individual endpoints as well as in 4.5.2.3. Plant protection product groundwater metabolite assessment
the overall genotoxicity assessment, which now would result in a ne- (genetic toxicology). Fig. 12 illustrates the in silico genotoxicity
gative outcome with medium conﬁ dence. It may appear surprising that assessment of a plant protection product AI metabolite with the
the conﬁ dence is only set at medium, despite highly reliable experi- potential to leach into the groundwater. The AI is categorized as an
mental results demonstrating no genotoxic activity. However, to dis- IARC Class 2 carcinogen and hence may bear risk to humans, and
tinguish from a situation where in vivo studies were also performed, the control strategies are required. It has, however, been shown
conﬁ dence cannot, in a general sense, be higher as there needs to be experimentally to be non-genotoxic and it is hypothesized that the
room to increase the weight-of-evidence by the inclusion of in vivo re- carcinogenicity is mediated through an endocrine disruption
sults or expert review. The addition of an in vivo negative outcome mechanism. The metabolite is a polar molecule containing functional
would have brought the conﬁ dence up to “high”. However, in this groups in a similar environment to the parent molecule. The bacterial
particular case, an expert opinion was included in the ﬁ nal outcome, gene mutation assessment was performed by read-across and the
which raised the conﬁ dence to high. Suﬃ cient experimental data were application of statistical models and expert alerts. The read-across
available for the parent AI in a full regulatory battery of in vitro and in exercise concluded that the metabolite is likely to be negative but the
vivo studies, showing that the AI had no genotoxic potential. The analysis was not considered robust due to the lipophilicity of the
structural similarity between the metabolite and the AI was high and metabolite being outside the range of the analogs. Therefore, the result
the available in vitro data for the metabolite showed similar responses, was set at RS5 even though read-across could technically be considered
therefore no further concern was raised about the in vivo activity of the an expert reviewed method and could therefore have been set to RS3
metabolite. Furthermore, it is recognized that there are diﬀ erent reg- directly with a more robust analysis. Two independent statistical
ulatory guidelines with respect to in vivo studies and that in some in- models for bacterial mutagenicity were applied, both indicating that
dustries, an in vivo test would not be required for a high conﬁ dence the metabolite was negative, and the reliability score was set to RS4 as
assessment. there were two concurring and independent models. The rule-based
 method highlighted an alert (positive, RS5), but after an expert review,
 16

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 Fig. 12. In silico assessment of a plant protection product metabolite.
the alert was dismissed, as the chemical environment of the alerting also generic in nature and hence not speciﬁ c to the structural en-
moiety was dissimilar between the training set examples and the vironment in the metabolite (or the AI, for that matter). Furthermore,
metabolite and the alert was therefore considered not relevant (the the parent AI triggered the same in silico response but had been con-
result from this model is considered negative with a reliability score of ﬁ rmed to show no clastogenic eﬀ ects in vivo. For the predicted positive
RS3). No in silico assessment was made for mammalian gene mutation outcome in the in vivo MN test, expert review suggested that the pre-
using computational models but comparison (read-across) with the dicted activity would be due to carbamate and simple substituted ac-
predicted genotoxicity proﬁ le of the parent molecule indicated that rylamide compounds, formed as downstream metabolites of the meta-
there should be no concern for mammalian mutagenicity. The call for bolite, rather than to the metabolite under review. For the analogs
the in vitro gene mutation endpoint was set to negative with medium investigated with experimental data, only the carbamates appeared to
conﬁ dence. truly ﬂ ag as being related to any activity. Due to the physicochemical
 In vitro CA was also investigated using read-across. The weight of properties of the metabolite, it was considered highly unlikely that
the evidence did not give a clear indication of potential for CA induc- these would form in vivo and hence, the in vivo alert was overruled. The
tion and was considered Indeterminate. Rule-based methods predicted summary assessment for the metabolite concluded that there was
the metabolite to be positive in the in vivo MN test and in the in vitro CA medium conﬁ dence that there was no gene mutation potential and low
assay. Both of these predictions were given reliability scores of RS5. conﬁ dence for the lack of clastogenic potential. Aneugenic eﬀ ects have
Expert review of the examples related to the in vitro CA prediction not been covered. The overall genetic toxicology assessment was
questioned the relevance as they did not bear strong structural simi- therefore set to negative with low conﬁ dence. After review of the
larity to the metabolite. The alert triggered in the in vitro CA model was submission, the regulatory authority also concluded that some
 17

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
experimental testing be conducted to speciﬁ cally ascertain the pre- oﬃ cial views of the National Institutes of Health.
dicted lack of genotoxic potential.
 Declaration of interests
5. Reporting
 Dr. Myatt reports grants from National Institutes of Health, during
 “Good in silico practice” requires a reproducible, transparent, and the conduct of the study; and it is disclosed that FDA's Center for Drug
standardized procedure and it is important to document the entire Evaluation and Research (CDER) and Leadscope Inc. are parties to a
process of performing the genetic toxicology assessment. This is com- formal Research Collaboration Agreement (RCA). Dr. Naomi Kruhlak is
parable to Good Laboratory Practice (GLP) documentation of in vitro or the FDA CDER Principal Investigator for this agreement, Dr. Stavitskaya
in vivo studies and will enable the results to be reviewed rapidly and contributes to this agreement and Dr. Kevin Cross is the Leadscope
thoroughly by, for example, regulatory agencies. The general protocol Principal Investigator for this agreement.
(Myatt et al., 2018) lists relevant types of information that should be
included in the report to ensure that the information is complete.Acknowledgements
Speciﬁ cally, chemical structures (including analogs in case of read-
across) and the models used need to be well documented. Research reported in this publication was supported by the National
 Institute of Environmental Health Sciences of the National Institutes of
6. Discussion Health under Award Number R43ES026909. The content is solely the
 responsibility of the authors and does not necessarily represent the
 The GIST protocol should be applied in a context-dependent manner oﬃ cial views of the National Institutes of Health.
and in accordance with relevant guidelines. For example, if the appli-
cation is for an ICH M7 assessment, then in addition to the re-Appendix A. Supplementary data
commendations provided in the guidance document, there are pub-
lications that provide more detailed procedures as well as caseSupplementary data to this article can be found online at https://
examples to illustrate best practices (Barber et al., 2015; Amberg et al.,doi.org/10.1016/j.yrtph.2019.104403.
2016). Similarly, there are, for example, guidelines for chemical re-
gistration through the REACH regulation (REACH, 2006; ECHA, 2008; References
ECHA, 2017) and Canada's Chemicals Management Program (Canada,
2016), the EFSA deﬁ nition of residue guidance (EFSA, 2016) and the Ahlberg, E., Amberg, A., Beilke, L.D., Bower, D., Cross, K.P., Custer, L., Ford, K.A., Van
Toxic Substances Control Act (TSCA) (TSCA, 2016). Gompel, J., Harvey, J., Honma, M., Jolly, R., Joossens, E., Kemper, R.A., Kenyon, M.,
 The protocol presented in this publication represents the current Kruhlak, N., Kuhnke, L., Leavitt, P., Naven, R., Neilan, C., Quigley, D.P., Shuey, D.,
 Spirkl, H.P., Stavitskaya, L., Teasdale, A., White, A., Wichard, J., Zwickl, C., Myatt,
state-of-the-art in in silico genetic toxicology. As new methods, both G.J., 2016. Extending (Q)SARs to incorporate proprietary knowledge for regulatory
experimental and computational, are developed and as new data be- purposes: a case study using aromatic amine mutagenicity. Regul. Toxicol.
come available, the recommendations presented herein will need to be Pharmacol. 77, 1–12. https://doi.org/10.1016/j.yrtph.2016.02.003.
 Amberg, A., Harvey, J.S., Czich, A., Spirkl, H.-P., Robinson, S., White, A., Elder, D.P.,
revised and updated. Additionally, it is important that the protocol 2015. Do carboxylic/sulfonic acid halides really present a mutagenic and carcino-
reﬂ ects current regulatory standards, data requirements, and changes as genic risk as impurities in ﬁ nal drug products? Org. Process Res. Dev. 19, 1495–1506.
these are revised. For example, as more in vitro micronucleus data are https://doi.org/10.1021/acs.oprd.5b00106.
generated with diﬀ erentiation of clastogenicity from aneugenicity me- Amberg, A., Beilke, L., Bercu, J., Bower, D., Brigo, A., Cross, K.P., Custer, L., Dobo, K.,
 Dowdy, E., Ford, K.A., Glowienke, S., Van Gompel, J., Harvey, J., Hasselgren, C.,
chanisms, statistical modeling may become an option for the separate Honma, M., Jolly, R., Kemper, R., Kenyon, M., Kruhlak, N., Leavitt, P., Miller, S.,
mechanistic endpoints. Since aneugenicity is generally considered to be Muster, W., Nicolette, J., Plaper, A., Powley, M., Quigley, D.P., Reddy, M.V., Spirkl,
a thresholded endpoint, this would involve an important change in the H.P., Stavitskaya, L., Teasdale, A., Weiner, S., Welch, D.S., White, A., Wichard, J.,
 Myatt, G.J., 2016. Principles and procedures for implementation of ICH M7 re-
current GIST protocol. Similarly, new assays will be accepted in a commended (Q)SAR analyses. Regul. Toxicol. Pharmacol. 77, 13–24. https://doi.org/
regulatory context. Also, there are considerable eﬀ orts underway to 10.1016/j.yrtph.2016.02.004.
develop and evaluate the impact that methods such as toxicogenomics, Ashby, J., Tennant, R.W., 1988. Chemical structure, Salmonella mutagenicity and extent
 of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals
ﬂ ow cytometric biomarker assays, and other mechanistic platforms (see tested in rodents by the U.S. NCI/NTP. Mutat. Res. 204, 17–115.
Section 2.2.4) can have on genotoxicity testing. As with any other ATSDR, 2018. ATSDR Toxic Substances Portal. https://www.atsdr.cdc.gov/substances/
protocol, it is therefore important to regularly revise the GIST protocol index.asp.
 Aubrecht, J., Caba, E., 2005. Gene expression proﬁ le analysis: an emerging approach to
to include new developments and remove outdated sections as appro- investigate mechanisms of genotoxicity. Pharmacogenomics 6, 419–428. https://doi.
priate. org/10.1517/14622416.6.4.419.
 Barber, C., Amberg, A., Custer, L., Dobo, K.L., Glowienke, S., Van Gompel, J., Gutsell, S.,
7. Conclusion Harvey, J., Honma, M., Kenyon, M.O., Kruhlak, N., Muster, W., Stavitskaya, L.,
 Teasdale, A., Vessey, J., Wichard, J., 2015. Establishing best practise in the appli-
 cation of expert review of mutagenicity under ICH M7. Regul. Toxicol. Pharmacol.
 Applying a standardized format for performing and reporting in si- 73, 367–377. https://doi.org/10.1016/j.yrtph.2015.07.018.
 Bender, M.A., Griggs, H.G., Bedford, J.S., 1974. Mechanisms of chromosomal aberration
lico assessments for hazard identiﬁ cation will enable a transparent and production III. Chemicals and ionizing radiation. Mutat. Res. Fund Mol. Mech.
consistent review of the results. This is beneﬁ cial both for organizations Mutagen 23, 197–212. https://doi.org/10.1016/0027-5107(74)90140-7.
and individuals performing such analyses as well as review boards and Benigni, R., Bossa, C., 2008. Structure alerts for carcinogenicity, and the Salmonella assay
regulatory agencies that consider such analyses. Along the same prin- system: a novel insight through the chemical relational databases technology. Mutat.
 Res. 659, 248–261. https://doi.org/10.1016/j.mrrev.2008.05.003.
ciples of standard practice for in vitro or in vivo data generation, the aim Benigni, R., Bossa, C., Richard, A.M., Yang, C., 2008. A novel approach: chemical rela-
is to foster the use of good in silico practices to promote the use of these tional databases, and the role of the ISSCAN database on assessing chemical carci-
methodologies to their full potential. nogenicity. Ann. Ist. Super Sanita 44, 48–56.
 Benigni, R., Bossa, C., Tcheremenskaia, O., Battistelli, C.L., Crettaz, P., 2012. The new
 ISSMIC database on in vivo micronucleus and its role in assessing genotoxicity testing
Funding strategies. Mutagenesis 27, 87–92. https://doi.org/10.1093/mutage/ger064.
 Bryce, S.M., Bernacki, D.T., Smith-Roe, S.L., Witt, K.L., Bemis, J.C., Dertinger, S.D., 2018.
 Investigating the generalizability of the MultiFlow (R) DNA damage assay and several
 Research reported in this publication was supported by the National companion machine learning models with a set of 103 diverse test chemicals. Toxicol.
Institute of Environmental Health Sciences of the National Institutes of Sci. 162, 146–166. https://doi.org/10.1093/toxsci/kfx235.
Health under AwardNumber R43ES026909. The content is solely the Cameron, T.P., Hughes, T.J., Kirby, P.E., Palmer, K.A., Fung, V.A., Dunkel, V.C., 1985.
 Mutagenic activity of 5 thiazole compounds in the Salmonella/microsome and mouse
responsibility of the authors and does not necessarily represent the
 18

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 lymphoma TK+/− assays. Mutat. Res. Genet. Toxicol. 155, 17–25. https://doi.org/ genotoxicity assays. I. Choice of cell type. Mutat. Res. Genet. Toxicol. Environ.
 10.1016/0165-1218(85)90020-5. Mutagen 742, 11–25. https://doi.org/10.1016/j.mrgentox.2011.10.014.
Canada, 2016. Canada's Chemicals Management Program. https://www.canada.ca/en/ Fowler, P., Smith, R., Smith, K., Young, J., Jeﬀ rey, L., Kirkland, D., Pfuhler, S.,
 health-canada/services/chemical-substances/chemicals-management-plan.html . Carmichael, P., 2012b. Reduction of misleading (“false”) positive results in mam-
CCRIS, 2011. Chemical Carcinogenesis Research Information System (CCRIS). https:// malian cell genotoxicity assays. II. Importance of accurate toxicity measurement.
 www.nlm.nih.gov/pubs/factsheets/ccrisfs.html. Mutat. Res. Genet. Toxicol. Environ. Mutagen 747, 104–117. https://doi.org/10.
CDRH, 2016. Use of International Standard ISO 10993-1. Biological Evaluation of Medical 1016/j.mrgentox.2012.04.013.
 Devices - Part 1: Evaluation and Testing within a Risk Management Process. Fox, J.T., Sakamuru, S., Huang, R., Teneva, N., Simmons, S.O., Xia, M., Tice, R.R., Austin,
Clewell, R.A., Sun, B., Adeleye, Y., Carmichael, P., Efremenko, A., McMullen, P.D., C.P., Myung, K., 2012. High-throughput genotoxicity assay identiﬁ es antioxidants as
 Pendse, S., Trask, O.J., White, A., Andersen, M.E., 2014. Proﬁ ling dose-dependent inducers of DNA damage response and cell death. Proc. Natl. Acad. Sci. U. S. A. 109,
 activation of p53-mediated signaling pathways by chemicals with distinct mechan- 5423–5428. https://doi.org/10.1073/pnas.1114278109.
 isms of DNA damage. Toxicol. Sci. 142, 56–73. https://doi.org/10.1093/toxsci/ Galloway, S., Lorge, E., Aardema, M.J., Eastmond, D., Fellows, M., Heﬂ ich, R., Kirkland,
 kfu153. D., Levy, D.D., Lynch, A.M., Marzin, D., Morita, T., Schuler, M., Speit, G., 2011.
Constable, A., Jonas, D., Cockburn, A., Davi, A., Edwards, G., Hepburn, P., Herouet- Workshop summary: top concentration for in vitro mammalian cell genotoxicity as-
 Guicheney, C., Knowles, M., Moseley, B., Oberdörfer, R., Samuels, F., 2007. History of says; and report from working group on toxicity measures and top concentration for
 safe use as applied to the safety assessment of novel foods and foods derived from in vitro cytogenetics assays (chromosome aberrations and micronucleus). Mutat. Res.
 genetically modiﬁ ed organisms. Food Chem. Toxicol. 45, 2513–2525. https://doi. 723, 77–83. https://doi.org/10.1016/j.mrgentox.2011.01.003.
 org/10.1016/j.fct.2007.05.028. GENE-TOX, 1998. https://toxnet.nlm.nih.gov/newtoxnet/genetox.htm.
Corvi, R., Vilardell, M., Aubrecht, J., Piersma, A., 2016. Validation of transcriptomics- Gentronix, 2018. Gentronix. http://www.gentronix.co.uk/product/greenscreen/.
 based in vitro methods. Adv. Exp. Med. Biol. 856, 243–257. https://doi.org/10.1007/ Gollapudi, B.B., Lynch, A.M., Heﬂ ich, R.H., Dertinger, S.D., Dobrovolsky, V.N., Froetschl,
 978-3-319-33826-2\_10. R., Horibata, K., Kenyon, M.O., Kimoto, T., Lovell, D.P., Stankowski Jr., L.F., White,
Custer, L.L., Powley, M.W., 2015. Genotoxicity testing of API. In: Graziano, M.J., P.A., Witt, K.L., Tanir, J.Y., 2015. The in vivo pig-a assay: a report of the international
 Jacobson-Kram, D. (Eds.), Genotoxicity and Carcinogenicity Testing of workshop on genotoxicity testing (IWGT) workgroup. Mutat. Res. Genet. Toxicol.
 Pharmaceuticals. Springer International Publishing, pp. 35–54. Environ. Mutagen 783, 23–35. https://doi.org/10.1016/j.mrgentox.2014.09.007.
Dearﬁ eld, K.L., Thybaud, V., Cimino, M.C., Custer, L., Czich, A., Harvey, J.S., Hester, S., GRAS, 2018. Food Substances Generally Recognized as Safe. https://www.fda.gov/Food/
 Kim, J.H., Kirkland, D., Levy, D.D., Lorge, E., Moore, M.M., Ouedraogo-Arras, G., IngredientsPackagingLabeling/GRAS/SCOGS/default.htm.
 Schuler, M., Suter, W., Sweder, K., Tarlo, K., van Benthem, J., van Goethem, F., Witt, Hasselgren, C., Muthas, D., Ahlberg, E., Andersson, S., Carlsson, L., Noeske, T., Stålring,
 K.L., 2011. Follow-up actions from positive results of in vitro genetic toxicity testing. J., Boyer, S., 2013. CHEMOINFORMATICS and beyond. Chemoinformatics for Drug
 Environ. Mol. Mutagen. 52, 177–204. https://doi.org/10.1002/em.20617. Discovery. John Wiley & Sons, Inc, pp. 267–290.
Dearﬁ eld, K.L., Gollapudi, B.B., Bemis, J.C., Benz, R.D., Douglas, G.R., Elespuru, R.K., Hayashi, M., 2016. The micronucleus test-most widely used in vivo genotoxicity test.
 Johnson, G.E., Kirkland, D.J., LeBaron, M.J., Li, A.P., Marchetti, F., Pottenger, L.H., Genes Environ. 38, 18. https://doi.org/10.1186/s41021-016-0044-x.
 Rorije, E., Tanir, J.Y., Thybaud, V., van Benthem, J., Yauk, C.L., Zeiger, E., Luijten, Hennig, U.G., Rudd, N.L., Hoar, D.I., 1988. Kinetochore immunoﬂ uorescence in micro-
 M., 2017. Next generation testing strategy for assessment of genomic damage: a nuclei: a rapid method for the in situ detection of aneuploidy and chromosome
 conceptual framework and considerations. Environ. Mol. Mutagen. 58, 264–283. breakage in human ﬁ broblasts. Mutat. Res. 203, 405–414.
 https://doi.org/10.1002/em.22045. Honda, H., Fujita, Y., Kasamatsu, T., Fuchs, A., Fautz, R., Morita, O., 2018. Necessity for
Dertinger, S.D., Torous, D.K., Hayashi, M., MacGregor, J.T., 2011a. Flow cytometric retrospective evaluation of past-positive chemicals in in vitro chromosomal aberra-
 scoring of micronucleated erythrocytes: an eﬃ cient platform for assessing in vivo tion tests using recommended cytotoxicity indices. Genes Environ. 40, 2. https://doi.
 cytogenetic damage. Mutagenesis 26, 139–145. https://doi.org/10.1093/mutage/ org/10.1186/s41021-017-0091-y.
 geq055. Hsu, C.-W., Hewes, K.P., Stavitskaya, L., Kruhlak, N.L., 2018. Construction and applica-
Dertinger, S.D., Torous, D.K., Hayashi, M., MacGregor, J.T., 2011b. Flow cytometric tion of (Q)SAR models to predict chemical-induced in vitro chromosome aberrations.
 scoring of micronucleated erythrocytes: an eﬃ cient platform for assessing in vivo Regul. Toxicol. Pharmacol. 99, 274–288. https://doi.org/10.1016/j.yrtph.2018.09.
 cytogenetic damage. Mutagenesis 26, 139–145. https://doi.org/10.1093/mutage/ 026.
 geq055. Hughes, C., Rabinowitz, A., Tate, M., Birrell, L., Allsup, J., Billinton, N., Walmsley, R.M.,
DSSTox, 2018. Distributed Structure-Searchable Toxicity (DSSTox) Database. https:// 2012. Development of a high-throughput Gaussia luciferase reporter assay for the
 www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox- activation of the GADD45a gene by mutagens, promutagens, clastogens, and aneu-
 database. gens. J. Biomol. Screen 17, 1302–1315. https://doi.org/10.1177/
ECHA, 2008. Guidance on Information Requirements and Chemical Safety Assessment 1087057112453312.
 Chapter R.6: QSARs and Grouping of Chemicals. https://echa.europa.eu/ ICH, 2012. S2(R1) Genotoxicity Testing and Data Interpretation for Pharmaceuticals
 documents/10162/13632/information\_requirements\_r6\_en.pdf/77f49f81-b76d- Intended for Human Use.
 40ab-8513-4f3a533b6ac9. ICH, 2014. M7 Assessment and Control of DNA Reactive (Mutagenic) Impurities in
ECHA, 2011. Guidance on Information Requirements and Chemical Safety Assessment Pharmaceuticals to Limit Potential Carcinogenic Risk.
 Chapter R.4: Evaluation of Available Information. https://echa.europa.eu/ ICH, 2017. M7 Assessment and Control of DNA Reactive (Mutagenic) Impurities in
 documents/10162/13643/information\_requirements\_r4\_en.pdf/d6395ad2-1596- Pharmaceuticals to Limit Potential Carcinogenic Risk. (R1).
 4708-ba86-0136686d205e. INCHEM, 2015. http://www.inchem.org/.
ECHA, 2017. Read-Across Assessment Framework. https://echa.europa.eu/documents/ IRIS, 2015. https://www.epa.gov/iris.
 10162/13628/raaf\_en.pdf. JECDB, 2018. https://www.echemportal.org/echemportal/index.action.
ECHA, 2018. https://echa.europa.eu/. JECFA, 2018. Joint FAO/WHO Expert Committee on Food Additives (JECFA). http://
EFSA, 2011. Scientiﬁ c opinion on genotoxicity testing strategies applicable to food and apps.who.int/food-additives-contaminants-jecfa-database/search.aspx .
 feed safety assessment. EFSA J. 9, 2379. Kaina, B., 2004. Mechanisms and consequences of methylating agent-induced SCEs and
EFSA, 2016. EFSA Panel on Plant Protection products and their Residues. Guidance on the chromosomal aberrations: a long road traveled and still a far way to go. Cytogenet.
 establishment of the residue deﬁ nition for dietary risk assessment. EFSA J. 14https:// Genome Res. 104, 77–86. https://doi.org/10.1159/000077469.
 doi.org/10.2903/j.efsa.2016.4549. e04549-n/a. Kim, S., Jun, D.H., Kim, H.J., Jeong, K.C., Lee, C.H., 2011. Development of a high-content
ELSIE, 2018. http://www.elsiedata.org/elsie-database/. screening method for chemicals modulating DNA damage response. J. Biomol. Screen
EPA, 2018. Strategic Plan to Promote the Development and Implementation of 16, 259–265. https://doi.org/10.1177/1087057110392993.
 Alternative Test Methods within the TSCA Program. Kirkland, D.J., 1994. Preface. Mutat. Res. Environ. Mutagen Relat. Subj. 312, 195–199.
EPA, 2019. https://comptox.epa.gov/dashboard. https://doi.org/10.1016/0165-1161(94)00006-9.
European Commission, J.R.C.J., 2018a. EURL ECVAM Genotoxicity & Carcinogenicity Kirkland, D.J., 2003. Preface. Mutat. Res. Genet. Toxicol. Environ. Mutagen 540,
 Consolidated Database of Ames Positive Chemicals. https://eurl-ecvam.jrc.ec. 119–121. https://doi.org/10.1016/j.mrgentox.2003.07.001.
 europa.eu/databases/genotoxicity-carcinogenicity-db. Kirkland, D., Fowler, P., 2010. Further analysis of Ames-negative rodent carcinogens that
European Commission, J.R.C.J., 2018b. ToxRTool - Toxicological Data Reliability are only genotoxic in mammalian cells in vitro at concentrations exceeding 1 mM,
 Assessment Tool. https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive- including retesting of compounds of concern. Mutagenesis 25, 539–553. https://doi.
 publications/toxrtool. org/10.1093/mutage/geq041.
Farmahin, R., Williams, A., Kuo, B., Chepelev, N.L., Thomas, R.S., Barton-Maclaren, T.S., Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., Elhajouji, A.,
 Curran, I.H., Nong, A., Wade, M.G., Yauk, C.L., 2017. Recommended approaches in Glatt, H., Hastwell, P., Hayashi, M., Kasper, P., Kirchner, S., Lynch, A., Marzin, D.,
 the application of toxicogenomics to derive points of departure for chemical risk Maurici, D., Meunier, J.-R., Müller, L., Nohynek, G., Parry, J., Parry, E., Thybaud, V.,
 assessment. Arch. Toxicol. 91, 2045–2065. https://doi.org/10.1007/s00204-016- Tice, R., van Benthem, J., Vanparys, P., White, P., 2007a. How to reduce false po-
 1886-5. sitive results when undertaking in vitro genotoxicity testing and thus avoid un-
FDA, 2018. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm. necessary follow-up animal tests: report of an ECVAM Workshop. Mutat. Res. Genet.
FEMA, 2018a. FEMA GRAS Flavoring Substance List. https://www.femaﬂ avor.org/fema-Toxicol. Environ. Mutagen 628, 31–55. https://doi.org/10.1016/j.mrgentox.2006.
 gras%E2%84%A2-ﬂ avoring-substance-list. 11.008.
FEMA, 2018b. The Flavor and Extract Manufacturers Association of the United States Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., Elhajouji, A.,
 (FEMA). https://www.femaﬂ avor.org/. Glatt, H., Hastwell, P., Hayashi, M., Kasper, P., Kirchner, S., Lynch, A., Marzin, D.,
Ford, K.A., 2016. Reﬁ nement, reduction, and replacement of animal toxicity tests by Maurici, D., Meunier, J.R., Muller, L., Nohynek, G., Parry, J., Parry, E., Thybaud, V.,
 computational methods. ILAR J. 57, 226–233. https://doi.org/10.1093/ilar/ilw031. Tice, R., van Benthem, J., Vanparys, P., White, P., 2007b. How to reduce false po-
Fowler, P., Smith, K., Young, J., Jeﬀ rey, L., Kirkland, D., Pfuhler, S., Carmichael, P., sitive results when undertaking in vitro genotoxicity testing and thus avoid un-
 2012a. Reduction of misleading (“false ”) positive results in mammalian cell necessary follow-up animal tests: report of an ECVAM Workshop. Mutat. Res. 628,
 19

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 31–55. https://doi.org/10.1016/j.mrgentox.2006.11.008. silico toxicology protocols. Regul. Toxicol. Pharmacol. 96, 1–17. https://doi.org/10.
Kirkland, D.J., Hayashi, M., Jacobson-Kram, D., Kasper, P., MacGregor, J.T., Müller, L., 1016/j.yrtph.2018.04.014.
 Uno, Y., 2007c. The international workshops on genotoxicity testing (IWGT): history NIOSH, 2018. https://www.cdc.gov/niosh/index.htm.
 and achievements. Mutat. Res. Genet. Toxicol. Environ. Mutagen 627, 1–4. https:// Nishihara, K., Huang, R., Zhao, J., Shahane, S.A., Witt, K.L., Smith-Roe, S.L., Tice, R.R.,
 doi.org/10.1016/j.mrgentox.2006.08.012. Takeda, S., Xia, M., 2016. Identiﬁ cation of genotoxic compounds using isogenic DNA
Kirkland, D.J., Hayashi, M., Jacobson-Kram, D., Kasper, P., MacGregor, J.T., Müller, L., repair deﬁ cient DT40 cell lines on a quantitative high throughput screening platform.
 Uno, Y., 2007d. Summary of major conclusions from the 4th IWGT, san Francisco, Mutagenesis 31, 69–81. https://doi.org/10.1093/mutage/gev055.
 9–10 september, 2005. Mutat. Res. Genet. Toxicol. Environ. Mutagen 627, 5–9. NTP, 2018. https://tools.niehs.nih.gov/cebs3/ui/.
 https://doi.org/10.1016/j.mrgentox.2006.08.009. OECD, 1986a. Test No. 479: Genetic Toxicology: in Vitro Sister Chromatid Exchange
Kirkland, D.J., Hayashi, M., Jacobson-Kram, D., Kasper, P., Gollapudi, B., Müller, L., Uno, Assay in Mammalian Cells. Section 4. OECD Publishing, Paris. https://doi.org/10.
 Y., 2011. Summary of major conclusions from the 5th IWGT, Basel, Switzerland, 1787/9789264071384-en.
 17–19 August 2009. Mutat. Res. Genet. Toxicol. Environ. Mutagen 723, 73–76. OECD, 1986b. Test No. 481: Genetic Toxicology: Saacharomyces Cerevisiae, Miotic
 https://doi.org/10.1016/j.mrgentox.2011.03.003. Recombination Assay. Section 4. OECD Publishing, Paris. https://doi.org/10.1787/
Kirkland, D.J., 2000. Preface. Environ. Mol. Mutagen. 35, 159–161. https://doi.org/10. 9789264071421-en.
 1002/(SICI)1098-2280(2000)35:3<159::AID-EM1>3.0.CO;2-N . OECD, 1986c. Test No. 482: Genetic Toxicology: DNA Damage and Repair, Unscheduled
Klimisch, H.J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the DNA Synthesis in Mammalian Cells in Vitro. Section 4. OECD Publishing, Paris.
 quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. https://doi.org/10.1787/9789264071445-en.
 Pharmacol. 25, 1–5. https://doi.org/10.1006/rtph.1996.1076. OECD, 1986d. Test No. 484: Genetic Toxicology: Mouse Spot Test, OECD Guidelines for
Knight, A.W., Birrell, L., Walmsley, R.M., 2009. Development and validation of a higher the Testing of Chemicals. Section 4. OECD Publishing, Paris. https://doi.org/10.
 throughput screening approach to genotoxicity testing using the GADD45a-GFP 1787/9789264071483-en.
 GreenScreen HC assay. J. Biomol. Screen 14, 16–30. https://doi.org/10.1177/ OECD, 1986e. Test No. 485: Genetic Toxicology, Mouse Heritable Translocation Assay.
 1087057108327065. Section 4. OECD Publishing, Paris. https://doi.org/10.1787/9789264071506-en.
Leadscope, 2018. http://www.leadscope.com/toxicity\_databases/. OECD, 1997a. Test No. 471: Bacterial Reverse Mutation Test. Section 4. OECD Publishing,
Li, H.H., Hyduke, D.R., Chen, R., Heard, P., Yauk, C.L., Aubrecht, J., Fornace Jr., A.J., Paris. https://doi.org/10.1787/9789264071247-en.
 2015. Development of a toxicogenomics signature for genotoxicity using a dose-op- OECD, 1997b. Test No. 486: Unscheduled DNA Synthesis (UDS) Test with Mammalian
 timization and informatics strategy in human cells. Environ. Mol. Mutagen. 56, Liver Cells in Vivo. Section 4. OECD Publishing, Paris. https://doi.org/10.1787/
 505–519. https://doi.org/10.1002/em.21941. 9789264071520-en.
Li, H.H., Chen, R., Hyduke, D.R., Williams, A., Frotschl, R., Ellinger-Ziegelbauer, H., OECD, 2010. Test No. 487: in Vitro Mammalian Cell Micronucleus Test. https://doi.org/
 O'Lone, R., Yauk, C.L., Aubrecht, J., Fornace Jr., A.J., 2017. Development and vali- 10.1787/9789264091016-en.
 dation of a high-throughput transcriptomic biomarker to address 21st century genetic OECD, 2013. Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation
 toxicology needs. Proc. Natl. Acad. Sci. U. S. A. 114, E10881–e10889. https://doi. Assays. Section 4. OECD Publishing, Paris. https://doi.org/10.1787/
 org/10.1073/pnas.1714109114. 9789264203907-en.
MacGregor, J.T., Bishop, M.E., McNamee, J.P., Hayashi, M., Asano, N., Wakata, A., OECD, 2014. Guidance on Grouping of Chemicals, second ed. Series on Testing &
 Nakajima, M., Saito, J., Aidoo, A., Moore, M.M., Dertinger, S.D., 2006. Flow cyto- Assessment No. 194. http://www.oecd.org/oﬃ cialdocuments/
 metric analysis of micronuclei in peripheral blood reticulocytes: II. An eﬃ cient publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en.
 method of monitoring chromosomal damage in the rat. Toxicol. Sci. 94, 92–107. OECD, 2015. Guidance Document on Revisions to OECD Genetic Toxicology Test
 https://doi.org/10.1093/toxsci/kﬂ 076. Guidelines. https://www.oecd.org/env/ehs/testing/Draft%20Guidance
Mahadevan, B., Snyder, R.D., Waters, M.D., Benz, R.D., Kemper, R.A., Tice, R.R., Richard, %20Document%20on%20OECD%20Genetic%20Toxicology%20Test%20Guidelines.
 A.M., 2011. Genetic toxicology in the 21st century: reﬂ ections and future directions. pdf.
 Environ. Mol. Mutagen. 52, 339–354. https://doi.org/10.1002/em.20653. OECD, 2016a. Overview of the Set of OECD Genetic Toxicology Test Guidelines and
Marchant, C.A., 2012. Computational toxicology: a tool for all industries. Wiley Updates Performed in 2014-2015, Series on Testing and Assessment Nr. 238. Paris.
 Interdiscip. Rev.: Comput. Mol. Sci. 2, 424–434. https://doi.org/10.1002/wcms.100. OECD, 2016b. Test No. 473: in Vitro Mammalian Chromosomal Aberration Test.
Martus, H.J., Hayashi, M., Honma, M., Kasper, P., Gollapudi, B., Mueller, L., Schoeny, R., OECD, 2016c. Test No. 473: in Vitro Mammalian Chromosomal Aberration Test. https://
 Uno, Y., Kirkland, D.J., 2015. Summary of major conclusions from the 6th interna- doi.org/10.1787/9789264264649-en.
 tional workshop on genotoxicity testing (IWGT), Foz do Iguacu, Brazil. Mutat. Res. OECD, 2016d. Test No. 474: Mammalian Erythrocyte Micronucleus Test. Section 4. OECD
 Genet. Toxicol. Environ. Mutagen 783, 1–5. https://doi.org/10.1016/j.mrgentox. Publishing, Paris. https://doi.org/10.1787/9789264264762-en.
 2015.02.004. OECD, 2016e. Test No. 475: Mammalian Bone Marrow Chromosomal Aberration Test.
Mav, D., Shah, R.R., Howard, B.E., Auerbach, S.S., Bushel, P.R., Collins, J.B., Gerhold, Section 4. OECD Publishing, Paris. https://doi.org/10.1787/9789264264786-en.
 D.L., Judson, R.S., Karmaus, A.L., Maull, E.A., Mendrick, D.L., Merrick, B.A., Sipes, OECD, 2016f. Test No. 476: in Vitro Mammalian Cell Gene Mutation Tests Using the Hprt
 N.S., Svoboda, D., Paules, R.S., 2018. A hybrid gene selection approach to create the and Xprt Genes. Section 4. OECD Publishing, Paris. https://doi.org/10.1787/
 S1500+ targeted gene sets for use in high-throughput transcriptomics. PLoS One 9789264264809-en.
 13https://doi.org/10.1371/journal.pone.0191105. e0191105. OECD, 2016g. Test No. 478: Rodent Dominant Lethal Test. Section 4. OECD Publishing,
Miller, E.C., Miller, J.A., 1981. Searches for ultimate chemical carcinogens and their re- Paris. https://doi.org/10.1787/9789264264823-en.
 actions with cellular macromolecules. Cancer 47, 2327–2345. OECD, 2016h. Test No. 483: Mammalian Spermatogonial Chromosomal Aberration Test.
Mishima, M., 2017. Chromosomal aberrations, clastogens vs aneugens. Front Biosci Section 4. OECD Publishing, Paris. https://doi.org/10.1787/9789264264847-en.
 (Schol Ed). 9, 1–16. OECD, 2016i. Test No. 487: in Vitro Mammalian Cell Micronucleus Test. Section 4. OECD
Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Cifone, M., Delongchamp, R., Publishing, Paris. https://doi.org/10.1787/9789264264861-en.
 Fellows, M., Gollapudi, B., Jenkinson, P., Kirby, P., Kirchner, S., Muster, W., Myhr, B., OECD, 2016j. Test No. 489: in Vivo Mammalian Alkaline Comet Assay. Section 4. OECD
 O'Donovan, M., Oliver, J., Omori, T., Ouldelhkim, M.-C., Pant, K., Preston, R., Riach, Publishing, Paris. https://doi.org/10.1787/9789264264885-en.
 C., San, R., Stankowski Jr., L.F., Thakur, A., Wakuri, S., Yoshimura, I., 2003. Mouse OECD, 2016k. Test No. 490: in Vitro Mammalian Cell Gene Mutation Tests Using the
 lymphoma thymidine kinase gene mutation assay: international workshop on geno- Thymidine Kinase Gene. Section 4. OECD Publishing, Paris. https://doi.org/10.
 toxicity tests workgroup report—plymouth, UK 2002. Mutat. Res. Genet. Toxicol. 1787/9789264264908-en.
 Environ. Mutagen 540, 127–140. https://doi.org/10.1016/j.mrgentox.2003.07.003. OECD, 2018. Existing Chemicals Database. https://hpvchemicals.oecd.org/UI/Default.
Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Burlinson, B., Cifone, M., Clarke, J., aspx.
 Delongchamp, R., Durward, R., Fellows, M., Gollapudi, B., Hou, S., Jenkinson, P., OECD, 2019. OECD QSAR Toolbox. https://qsartoolbox.org/.
 Lloyd, M., Majeska, J., Myhr, B., O'Donovan, M., Omori, T., Riach, C., San, R., Olaharski, A.J., Gonzaludo, N., Bitter, H., Goldstein, D., Kirchner, S., Uppal, H., Kolaja, K.,
 Stankowski Jr., L.F., Thakur, A.K., Van Goethem, F., Wakuri, S., Yoshimura, I., 2006. 2009. Identiﬁ cation of a kinase proﬁ le that predicts chromosome damage induced by
 Mouse lymphoma thymidine kinase gene mutation assay: follow-up meeting of the small molecule kinase inhibitors. PLoS Comput. Biol. 5https://doi.org/10.1371/
 International Workshop on Genotoxicity Testing– Aberdeen, Scotland, 2003–Assay journal.pcbi.1000446. e1000446.
 acceptance criteria, positive controls, and data evaluation. Environ. Mol. Mutagen. OSHA, 2018. https://www.osha.gov/.
 47, 1–5. https://doi.org/10.1002/em.20159. PAN, 2018. http://www.pesticideinfo.org/.
Myatt, G.J., Beilke, L.D., Cross, T.K.P., 2016. In Silico Tools and Their Application. Parry, J.M., Parry, E., Phrakonkham, P., Corvi, R., 2010a. Analysis of published data for
 Reference Module in Chemistry. Molecular Sciences and Chemical Engineering 2016. top concentration considerations in mammalian cell genotoxicity testing.
Myatt, G.J., Ahlberg, E., Akahori, Y., Allen, D., Amberg, A., Anger, L.T., Aptula, A., Mutagenesis 25, 531–538. https://doi.org/10.1093/mutage/geq046.
 Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D., Parry, J.M., Parry, E., Phrakonkham, P., Corvi, R., 2010b. Analysis of published data for
 Brigo, A., Burden, N., Cammerer, Z., Cronin, M.T.D., Cross, K.P., Custer, L., Dettwiler, top concentration considerations in mammalian cell genotoxicity testing.
 M., Dobo, K., Ford, K.A., Fortin, M.C., Gad-McDonald, S.E., Gellatly, N., Gervais, V., Mutagenesis 25, 531–538. https://doi.org/10.1093/mutage/geq046.
 Glover, K.P., Glowienke, S., Van Gompel, J., Gutsell, S., Hardy, B., Harvey, J.S., PharmaPendium, 2018. http://www.pharmapendium.com/.
 Hillegass, J., Honma, M., Hsieh, J.H., Hsu, C.W., Hughes, K., Johnson, C., Jolly, R., REACH, 2006. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=
 Jones, D., Kemper, R., Kenyon, M.O., Kim, M.T., Kruhlak, N.L., Kulkarni, S.A., CELEX:02006R1907-20161011&from=EN.
 Kummerer, K., Leavitt, P., Majer, B., Masten, S., Miller, S., Moser, J., Mumtaz, M., Rosefort, C., Fauth, E., Zankl, H., 2004. Micronuclei induced by aneugens and clastogens
 Muster, W., Neilson, L., Oprea, T.I., Patlewicz, G., Paulino, A., Lo Piparo, E., Powley, in mononucleate and binucleate cells using the cytokinesis block assay. Mutagenesis
 M., Quigley, D.P., Reddy, M.V., Richarz, A.N., Ruiz, P., Schilter, B., Seraﬁ mova, R., 19, 277–284. https://doi.org/10.1093/mutage/geh028.
 Simpson, W., Stavitskaya, L., Stidl, R., Suarez-Rodriguez, D., Szabo, D.T., Teasdale, RTECS, 2018. https://www.cdc.gov/niosh/rtecs/default.html.
 A., Trejo-Martin, A., Valentin, J.P., Vuorinen, A., Wall, B.A., Watts, P., White, A.T., Saha, L.K., Kim, S., Kang, H., Akter, S., Choi, K., Sakuma, T., Yamamoto, T., Sasanuma, H.,
 Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C., Hasselgren, C., 2018. In Hirota, K., Nakamura, J., Honma, M., Takeda, S., Dertinger, S., 2018. Diﬀ erential
 20

C. Hasselgren, et al. Regulatory Toxicology and Pharmacology 107 (2019) 104403
 micronucleus frequency in isogenic human cells deﬁ cient in DNA repair pathways is a for discriminating aneugen- vs. clastogen-induced micronuclei. Environ. Mol.
 valuable indicator for evaluating genotoxic agents and their genotoxic mechanisms. Mutagen. 31, 340–344.
 Environ. Mol. Mutagen. 59, 529–538. https://doi.org/10.1002/em.22201. TOXNET, 2018. https://toxnet.nlm.nih.gov/.
Sakai, R., Kondo, C., Oka, H., Miyajima, H., Kubo, K., Uehara, T., 2014. Utilization of TSCA, 2016. Toxic Substances Control Act (TSCA). https://www.congress.gov/bill/
 CDKN1A/p21 gene for class discrimination of DNA damage-induced clastogenicity. 114th-congress/senate-bill/697/all-info.
 Toxicology 315, 8–16. https://doi.org/10.1016/j.tox.2013.10.009. VITIC, 2018. https://www.lhasalimited.org/products/vitic-nexus.htm.
Snyder, R.D., 2000. Use of catalytic topoisomerase II inhibitors to probe mechanisms of Wexler, P., 2001. TOXNET: an evolving web resource for toxicology and environmental
 chemical-induced clastogenicity in Chinese hamster V79 cells. Environ. Mol. health information. Toxicology 157, 3–10. https://doi.org/10.1016/S0300-483X(00)
 Mutagen. 35, 13–21. 00337-1.
Snyder, R.D., 2010. Possible structural and functional determinants contributing to the Wilde, S., Dambowsky, M., Hempt, C., Sutter, A., Queisser, N., 2017. Classiﬁ cation of in
 clastogenicity of pharmaceuticals. Environ. Mol. Mutagen. 51, 800–814. https://doi. vitro genotoxicants using a novel multiplexed biomarker assay compared to the ﬂ ow
 org/10.1002/em.20626. cytometric micronucleus test. Environ. Mol. Mutagen. 58, 662–677. https://doi.org/
Sofuni, T., Matsuoka, A., Sawada, M., Ishidate Jr., M., Zeiger, E., Shelby, M.D., 1990. A 10.1002/em.22130.
 comparison of chromosome aberration induction by 25 compounds tested by two Witt, K.L., Hsieh, J.H., Smith-Roe, S.L., Xia, M., Huang, R., Zhao, J., Auerbach, S.S., Hur,
 Chinese hamster cell (CHL and CHO) systems in culture. Mutat. Res. 241, 175–213. J., Tice, R.R., 2017. Assessment of the DNA damaging potential of environmental
Stanton, K., Kruszewski, F.H., 2016. Quantifying the beneﬁ ts of using read-across and in chemicals using a quantitative high-throughput screening approach to measure p53
 silico techniques to fulﬁ ll hazard data requirements for chemical categories. Regul. activation. Environ. Mol. Mutagen. 58, 494–507. https://doi.org/10.1002/em.
 Toxicol. Pharmacol. 81, 250–259. https://doi.org/10.1016/j.yrtph.2016.09.004. 22112.
Stavitskaya, L., Minnier, B.L., Benz, R.D., Kruhlak, N.L., 2013. Development of Improved Yamamoto, K.N., Hirota, K., Kono, K., Takeda, S., Sakamuru, S., Xia, M., Huang, R.,
 QSAR Models for Predicting A-T Base Pair Mutations. Poster presented at the Genetic Austin, C.P., Witt, K.L., Tice, R.R., 2011. Characterization of environmental chemi-
 Toxicity Association, Newark, Delaware. cals with potential for DNA damage using isogenic DNA repair-deﬁ cient chicken
Sweet, D.V., Anderson, V.P., Fang, J.C.F., 1999. An overview of the registry of toxic ef- DT40 cell lines. Environ. Mol. Mutagen. 52, 547–561. https://doi.org/10.1002/em.
 fects of chemical substances (RTECS): critical information on chemical hazards. 20656.
 Chem. Health Saf. 6, 12–16. https://doi.org/10.1016/S1074-9098(99)00058-1. Zeiger, E., Gollapudi, B., Aardema, M.J., Auerbach, S., Boverhof, D., Custer, L., Dedon, P.,
Torous, D.K., Dertinger, S.D., Hall, N.E., Tometsko, C.R., 1998a. An automated method Honma, M., Ishida, S., Kasinski, A.L., Kim, J.H., Manjanatha, M.G., Marlowe, J.,
 for discriminating aneugen- vs. clastogen-induced micronuclei. Environ. Mol. Pfuhler, S., Pogribny, I., Slikker, W., Stankowski Jr., L.F., Tanir, J.Y., Tice, R., van
 Mutagen. 31, 340–344. https://doi.org/10.1002/(SICI)1098-2280(1998) Benthem, J., White, P., Witt, K.L., Thybaud, V., 2015. Opportunities to integrate new
 31:4<340::AID-EM6>3.0.CO;2-M. approaches in genetic toxicology: an ILSI-HESI workshop report. Environ. Mol.
Torous, D.K., Dertinger, S.D., Hall, N.E., Tometsko, C.R., 1998b. An automated method Mutagen. 56, 277–285. https://doi.org/10.1002/em.21923.
 21

Source PDF: /instem.com/www.instem.com/docs/sot-2021/out-of-domain-final.pdf

 Regulatory Toxicology and Pharmacology 102 (2019) 53–64
 Contents lists available at ScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepage: www.elsevier.com/locate/yrtph
Principles and procedures for handling out-of-domain and indeterminate T
results as part of ICH M7 recommended (Q)SAR analyses☆
Alexander Amberga, Roxanne V. Andayab, Lennart T. Angera, Chris Barberc, Lisa Beilked,
Joel Bercue, Dave Bowerf, Alessandro Brigog, Zoryanna Cammererh, Kevin P. Crossf,
Laura Custeri, Krista Doboj, Helga Geretsk, Véronique Gervaisl, Susanne Glowienkem,
Stephen Gomezn, Jacky Van Gompelo, James Harveyp, Catrin Hasselgrenb, Masamitsu Honmaq,
Candice Johnsonf, Robert Jollyr, Raymond Kempers, Michelle Kenyonj, Naomi Kruhlakt,
Penny Leavitti, Scott Millerf, Wolfgang Musterg, Russell Navenu, John Nicolettev, Alexis Parentym,
Mark Powleyw, Donald P. Quigleyf, M. Vijayaraj Reddyw, Jennifer C. Sasakib,1 ,
Lidiya Stavitskayat, Andrew Teasdalex, Alejandra Trejo-Martine, Sandy Weinerh,
Dennie S. Welchv, Angela Whitep, Joerg Wichardy, David Woolleyz, Glenn J. Myattf,∗
a Sanoﬁ , R&D Preclinical Safety Frankfurt, Industriepark Hoechst, D-65926, Frankfurt am Main, Germany
b Genentech, Inc., 1 DNA Way, South San Francisco, CA, 94080, USA
c Lhasa Limited, Leeds, UK
d Toxicology Solutions Inc., San Diego, CA, USA
e Gilead Sciences, 333 Lakeside Drive, Foster City, CA, USA
f Leadscope, Inc., 1393 Dublin Rd, Columbus, OH, 43215, USA
g Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Switzerland
h Janssen Research & Development, 1400 McKean Road, Spring House, PA, 19477, USA
i Bristol-Myers Squibb, Drug Safety Evaluation, 1 Squibb Dr, New Brunswick, NJ, 08903, USA
j Pﬁ zer Global Research & Development, 558 Eastern Point Road, Groton, CT, 06340, USA
k UCB Biopharma SPRL, Chemin du Foriest, B-1420, Braine-l’Alleud, Belgium
l Servier Group, Gidy, France
m Novartis Pharma AG, Pre-Clinical Safety, Werk Klybeck, CH-4057, Basel, Switzerland
n Consultant to Theravance Biopharma US, Inc., 901 Gateway Blvd, South San Francisco, CA, 94080, USA
o Janssen Pharmaceutical Companies of Johnson & Johnson, 2340, Beerse, Belgium
p GlaxoSmithKline, Park Road, Ware, Hertfordshire, SG12 0DP, UK
q National Institute of Health Sciences, Tokyo, Japan
r Toxicology Division, Eli Lilly and Company, Indianapolis, IN, USA
s Vertex Pharmaceuticals Inc., Discovery and Investigative Toxicology, 50 Northern Ave, Boston, MA, USA
t FDA Center for Drug Evaluation and Research, Silver Spring, MD, USA
u Takeda, Cambridge, MA, USA
v AbbVie Inc., North Chicago, IL, USA
w Merck Research Laboratories, West Point, PA, 19486, USA
x AstraZeneca, Macclesﬁ eld, Cheshire, UK
y Bayer Pharma AG, Investigational Toxicology, Muellerstr. 178, D-13353, Berlin, Germany
z ForthTox Limited, PO Box 13550, Linlithgow, EH49 7YU, UK
ABSTRACT
The International Council for Harmonization (ICH) M7 guideline describes a hazard assessment process for impurities that have the potential to be present in a drug
substance or drug product. In the absence of adequate experimental bacterial mutagenicity data, (Q)SAR analysis may be used as a test to predict impurities’ DNA
reactive (mutagenic) potential. However, in certain situations, (Q)SAR software is unable to generate a positive or negative prediction either because of conﬂ icting
 ☆ FDA CDER Disclaimer: This article reﬂ ects the views of the authors and should not be construed to represent FDA's views or policies. The mention of commercial
products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by
the Department of Health and Human Services.
 ∗ Corresponding author. Present address: Leadscope, Inc., 1393 Dublin Road, Columbus, OH, 43215, USA.
 E-mail address: gmyatt@leadscope.com (G.J. Myatt).
 1 Current aﬃliation: Denali Therapeutics, South San Francisco, CA 94080.
https://doi.org/10.1016/j.yrtph.2018.12.007
Received 14 September 2018; Received in revised form 10 December 2018; Accepted 14 December 2018
Available online 15 December 2018
0273-2300/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license
(http://creativecommons.org/licenses/BY/4.0/).

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
information or because the impurity is outside the applicability domain of the model. Such results present challenges in generating an overall mutagenicity prediction
and highlight the importance of performing a thorough expert review. The following paper reviews pharmaceutical and regulatory experiences handling such
situations. The paper also presents an analysis of proprietary data to help understand the likelihood of misclassifying a mutagenic impurity as non-mutagenic based
on diﬀ erent combinations of (Q)SAR results. This information may be taken into consideration when supporting the (Q)SAR results with an expert review, especially
when out-of-domain results are generated during a (Q)SAR evaluation.
1. Introduction It is often deﬁ ned using structural features and/or properties of the
 training or reference set chemicals. Diﬀ erent modeling algorithms use
 In 2014, the International Council for Harmonization (ICH) issued distinct approaches to compute this applicability domain and therefore
their M7 guideline (“Assessment and control of DNA reactive (muta- diﬀ er in coverage (Ellison et al., 2010; Chakravarti, 2012; Hanser et al.,
genic) impurities in pharmaceuticals to limit potential carcinogenic 2016; Myatt et al., 2016; Williams et al., 2016). A (Q)SAR model de-
risk”), which was revised in 2017 (ICH M7(R1), 2017). The guideline termines whether the impurity is outside the applicability domain of
describes a hazard assessment process for impurities that reside or are the model and such a result will be referred to as out-of-domain
reasonably likely to be present in a drug substance or product. In the throughout this publication; however, diﬀ erent systems may use other
absence of adequate experimental mutagenicity and/or carcinogenicity terms such as not-in-domain).
results, a structure-based computational toxicology or (Q)SAR2 analysis The ICH M7 guideline describes that the (Q)SAR analysis may be
may be used as a test to predict DNA reactive (mutagenic) potential. (Q) supported by an expert review, especially in situations where the results
SAR is a commonly used and relatively mature approach for predicting are inconclusive (i.e., indeterminate or out-of-domain) as well as where
mutagenicity (Myatt et al., 2016). Based on the high predictive con- there are valid reasons to overturn or refute a prediction. The stan-
ﬁ dence levels (Dobo et al., 2012; Greene et al., 2015) and the cost of dardized use of expert review has been detailed in several publications
running such analysis relative to in vitro or in vivo studies, (Q)SAR as- (Powley, 2015; Barber et al., 2015; Amberg et al., 2016) including as an
sessments balance the need for a fast and eﬃ cient analysis while en- in silico toxicology workﬂ ow that can be utilized under ICH M7 to
suring patient safety (Amberg et al., 2016). The (Q)SAR results, in turn, generate predictions with improved accuracy that are consistent be-
support the assignment of each impurity to one of ﬁ ve classes (shown in tween diﬀ erent experts (Myatt et al., 2018).
Table 1; Müller et al., 2006). This class assignment determines whether The principles of the ICH M7 guideline are now routinely followed
the impurity (1) requires no additional action, (2) requires additional by the pharmaceutical industry and international regulatory agencies.
laboratory testing, or (3) needs to be controlled below thresholds de- Although the (Q)SAR assessment and expert review of the results has
ﬁ ned in the guideline. been discussed in a number of publications, there are some speciﬁ c
 The guideline recommends such (Q)SAR assessments to be based on challenges associated with managing out-of-domain and indeterminate
the results from two complementary (Q)SAR methodologies: expert results, which have not been fully addressed. The following paper
rule-based and statistical-based (ICH M7(R1), 2017, 2017; Myatt et al., outlines current regulatory and industry approaches for handling out-
2016). The results from these models are combined, based on the “the of-domain and indeterminate results based on an industry survey. A
absence of structural alerts” (ICH M7(R1), 2017), to generate an overall series of case studies from regulatory submissions are provided to il-
prediction to support the class assignment. That is, if both systems are lustrate how out-of-domain or indeterminate results can be put into
non-alerting the class assignment is non-mutagenic. context. The paper also includes an assessment of the likelihood of
 (Q)SAR models use datasets of historical data as well as general misclassifying a mutagenic impurity as not mutagenic based on dif-
scientiﬁ c knowledge (such as structural alerts) from the literature based ferent combinations of (Q)SAR results (e.g., a negative expert rule-
on known mechanisms of DNA reactive mutagenicity to generate a based result and an out-of-domain statistical-based result). How this
prediction. Since the models are based on what is known, they may not information can be taken into consideration as part of an overall as-
be able to predict with suﬃ cient conﬁ dence a clear positive or negative sessment is discussed.
outcome for novel chemicals. This may be due to conﬂ icting evidence
such that the inﬂ uence of substituents on the reactivity of an alerting 2. Methodology
chemical moiety is not fully understood. These results will be referred
to as indeterminate predictions in this publication (individual systems A general request was made to the pharmaceutical industry and
may refer to them as equivocal or other similar terms). regulatory authorities to outline current practices for handling out-of-
 Another area where (Q)SAR models can present a challenge to users domain and indeterminate results. This information was collated and
is when the structure being assessed falls outside the training or re- summarized in Section 4 (Discussion).
ference set used to generate the model. Such domain analysis is re- To help understand the likelihood of misclassifying a mutagenic
quired as part of the (Q)SAR assessment since the ICH M7 guideline impurity as non-mutagenic based on diﬀ erent combinations of (Q)SAR
states that both methodologies should follow the general validation results, a request was made to run the (Q)SAR models generally used for
principles set forth by the Organization for Economic Co-operation and ICH M7 assessment over proprietary chemicals for which bacterial re-
Development (OECD) (OECD, 2007). The third OECD validation prin- verse mutation assay (Ames) data were available and provide a table
ciple requires the (Q)SAR model to assess whether each impurity is containing the ﬁ elds shown in Table 2. This included running diﬀ erent
within the applicability domain of the model. (Netzeva et al., 2005; systems as detailed in the supplementary material.
OECD, 2007; Carrió et al., 2014; Powley, 2015; Patlewicz et al., 2016).The results were compiled into a single consolidated table for ana-
The applicability domain is generally deﬁ ned as a region of chemical lysis. This involved a step to harmonize the results from diﬀ erent sys-
space within which a model makes predictions with a given reliability. tems (including expert rule-based and statistical-based methodologies
 from Leadscope Inc. and Lhasa Limited) into the following calls for each
 2 The term “(Q)SAR” refers to (Quantitative) Structure-Activity Relationship methodology:
and is used as an acronym for computational models that predict a biological  Positive: A positive call (i.e., predicted to be mutagenic).
response (such as mutagenicity) based on the chemical structure of the test
molecule. The term collectively refers to both quantitative and non-quantitative  Negative: A negative call that is within the applicability domain of
(e.g., expert rule-based) structure-activity relationships by placing the “Q” in the model (i.e., predicted to be non-mutagenic).
parentheses.  Indeterminate: An indeterminate or equivocal call that is within
 54

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
Table 1
The ICH M7 hazard classiﬁ cations.
 Class Deﬁ nition
 1 Known mutagenic carcinogens
 2 Known mutagens with unknown carcinogenic potential (bacterial mutagenicity positive,a no rodent carcinogenicity data)
 3 Alerting structure unrelated to the structure of the drug substance, no mutagenicity data
 4 Alerting structure, same alert in drug substance or compounds related to the drug substance (e.g. process intermediates) which have been tested and are non-mutagenic
 5 No structural alerts, or alerting structure with suﬃ cient data to demonstrate lack of mutagenicity or carcinogenicity
 a Or other relevant positive mutagenicity data indicative of DNA-reactivity-related induction of gene mutations (e.g., positive ﬁ ndings in in vitro gene mutation
studies).
 Table 2
 Fields to analyze as part of the request to pharmaceutical companies.
 Field Description
 Compound A unique number to reference the compound but nothing to identify the chemical structurea.
 Experimental Ames result Include 1 for clear mutagens, 0 for clear non-mutagens. Do not include equivocal results.
 Ames test description A short description of how the Ames test was performed (e.g. GLP OECD 471/ICH S2 Ames).
 Statistical-based model result The prediction results and/or indication of whether it is out-of-domain or indeterminate.
 Probability or other score Any additional output, such as the probability of a positive outcome.
 Rule-based (structural alert) result The prediction result and/or indication of whether it is out-of-domain or indeterminate.
 Precision or other score Any additional output, such as the precision of the alert.
 a This column may be blank.
 the applicability domain. negative prediction and these predictions are inside the applicability
  Out-of-domain: The (Q)SAR model considered the chemical outside domain of the models. The models used in this analysis are outlined in
 the applicability domain. the supplementary material. The table includes the number of chemi-
 cals (“Count”) and the proportion of experimentally determined mu-
 In order to assess if additional model output that reﬂ ects the tagenic compounds in each category. The total number of chemicals
probability of a positive response can be used to support an overall included in Table 3 is 10,083 out of the 15,886 chemicals analyzed
assessment where the statistical-based model is out-of-domain, the (63.5%). When both predictions are positive approximately 60% of the
following two situations are also considered when the system generates chemicals are mutagenic, whereas when both are negative approxi-
a probability score (such as the Leadscope Genetox Statistical Suite), as mately 8% are mutagenic. When the results are not in agreement, the
outlined in the supplemental material: proportion of mutagenic compounds is between these two values.
 Table 4 illustrates diﬀ erent situations when there is at least one out-
  Out-of-domain with probability of being positive < 0.2: The of-domain result, which represents approximate 25% of cases in this
 compound is considered outside the applicability domain of the study. The highest proportion of mutagenic compounds is when the
 statistical-based model; however, a probability score of less than 0.2 expert rule-based model generates a positive or indeterminate result.
 is generated When the statistical-based model is out-of-domain and the expert rule-
  Out-of-domain with probability of being positive 0.2–0.4: The based model is negative, there is a reduction in the proportion of mu-
 compound is considered outside the applicability domain of the tagenic compounds identiﬁ ed. There is also a reduction when both
 statistical-based model; however, a probability score of between 0.2 models are out-of-domain. It should be noted that there were no ex-
 and 0.4 is generated amples in this study where the expert rule-based model result is out-of-
 domain and the statistical-based model is positive, negative or in-
 The rules used to harmonize the results across the diﬀ erent systems determinate.
are included in the supplemental information. Table 5 shows a more detailed analysis where the statistical-based
 model result is out-of-domain and the expert rule-based model result is
3. Results negative. The table shows that the proportion of mutagenic compounds
 is lower when the statistical-based model generates a low probability
 score (less than 0.2) and no alerts are identiﬁ ed, even though the sta-
 The following analysis of proprietary data was performed to help tistical-based model result is out-of-domain.
understand the likelihood of misclassifying a mutagenic impurity as not Table 6 summarizes diﬀ erent scenarios where there is an in-
mutagenic (i.e., a false negative prediction) using diﬀ erent combina- determinate call in one or both of the (Q)SAR methodologies which
tions of (Q)SAR results. The analysis is based on historical data from
proprietary collections that include similar chemicals to those in a ty- Table 3
pical assessment of impurities such as low molecular weight chemicals Summary of in domain predictions generated for the two (Q)SAR methodolo-
used as starting materials and API (Active Pharmaceutical Ingredient)- gies.
like chemicals similar to the synthetic intermediates. The results were
generated based on the methodology outlined in Section 2. The total Statistic-based Expert rule- Counta Percentage of results that were
number of chemicals considered was 15,886, which generally re- result based result experimentally identiﬁ ed Ames
 mutagens
presents chemicals that were not used in building the models since (Q)
SAR models for regulatory use are usually built using data from the Positive Positive 1253 59.7%
public domain. The proportion of mutagenic compounds across the Negative Positive 499 37.5%
entire proprietary collection was 17.25%. It should be noted that no Positive Negative 353 24.7%
 Negative Negative 7978 8.1%
proprietary information was transferred as part of this process.
 Table 3 shows the results where each model generates a positive or a Out of 15,886 compounds tested.
 55

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
Table 4 Table 6
Summary of analysis where at least one of the methods generates an out-of- Diﬀ erent scenarios that include an indeterminate call from one or both of the
domain result. methodologies.
 Statistic-based Expert rule-based Count Percentage of results that were Statistic-based Expert rule-based Count Percentage of results that were
 result result experimentally identiﬁ ed Ames result result experimentally identiﬁ ed
 mutagens Ames mutagens
 Out-of-domain Positive 296 36.2%Indeterminate Positive 516 50.6%
 Out-of-domain Indeterminate 78 28.2% Out-of-domain Indeterminate 78 28.2%
 Out-of-domain Out-of-domain 1558 11.8% Positive Indeterminate 155 27.7%
 Out-of-domain Negative 2027 11.8%Indeterminate Negative 668 23.2%
 Indeterminate Indeterminate 93 20.4%
 Negative Indeterminate 314 11.8%
represents approximately 12% of examples in this study. The highest
proportion of mutagens is shown when the expert rule-based model is 4.2. (Q)SAR expert review
positive and the statistical-based model is indeterminate, whereas the
lowest proportion of mutagens is shown when the statistical-based 4.2.1. Expert review deﬁ nition and resulting actions
model reports a negative and the rule-based model output is in- The application of expert knowledge has been shown to improve (Q)
determinate. The percentage of mutagens for other scenarios is between SAR predictive performance, particularly in resolving ambiguous out-
these two values. comes such as out-of-domain results or indeterminate predictions,
 which is consistent with other published accounts (Dobo et al., 2012;
4. Discussion Sutter et al., 2013). To support such an expert review, available com-
 putational methods generally provide information on the certainty of
4.1. Overview the prediction, such as a probability of a positive outcome. In addition,
 these methods often describe how the model generated the result. In the
 Out-of-domain and/or indeterminate results are often encountered case of statistical-based methodologies, it is often possible to examine
as part of an ICH M7 impurity assessment. This has been quantiﬁ ed, in the training set and/or database analogs with detailed experimental
part, through an analysis of new drugs approved in 2016 and 2017 that data to understand how structural features or physicochemical prop-
showed 18% of the impurities had an out-of-domain result (Powley, erties inﬂ uence the model's prediction. For rule-based methodologies, it
2017). These out-of-domain and/or indeterminate results are often is often possible to inspect the structural features responsible for acti-
challenging for both pharmaceutical companies and regulatory agen- vation or deactivation of the alert along with an examination of plau-
cies to generate an overall ICH M7 classiﬁ cation. Although a con- sible mechanisms, examples, and associated references for any acti-
servative approach would be to assume that indeterminate or out-of- vated alerts. Any expert review can make use of this information
domain (Q)SAR results are positive, this would compromise the desired alongside the knowledge of the reviewer concerning DNA reactive
utility of the computational analysis and could result in unnecessary mutagenicity, the quality of the experimental data, metabolism and
additional drug development costs and delay the approval of new knowledge derived from proprietary data (e.g., unpublished proprietary
medicines. alerting chemicals).
 A variety of approaches for handling out-of-domain and/or in- To quantify the actions resulting from such a review, the US Food
determinate results are being used across pharmaceutical companies as and Drug Administration's (FDA) Center for Drug Evaluation and
well as regulatory agencies to support an overall prediction (deﬁ ned as Research (CDER) between May 2016 and April 2017 analyzed 519
“the absence of structural alerts” in the ICH M7 guideline). This is re- impurities for bacterial mutation using software from Leadscope,
ﬂ ected in a further breakdown of drugs approved in 2016 and 2017 (2017), Lhasa Limited (Lhasa, 2017), and MultiCASE ( MultiCASE,
which quantiﬁ ed the diﬀ erent approaches for handling such situations 2017). (Kruhlak et al., 2017) The expert-reviewed predictions were
including applying expert knowledge (for 70% of the submitted im- concordant with the consensus (Q)SAR results 87% of the time with:
purities), applying an additional model (for 6% of the submitted im-
purities), test/control (for 21% of the submitted impurities), and no  2.1% of the negative consensus predictions changed to positive after
follow-up (for 3% of the submitted impurities) (Powley, 2017). The the expert review
following sections are based on the responses from the pharmaceutical  4.2% of the positive consensus predictions changed to negative after
industry and regulatory agencies as to how they handle out-of-domain the expert review
and indeterminate results. This will cover diﬀ erent expert review  61% of the indeterminate consensus predictions changed to negativeafter the expert review
strategies in addition to using another (Q)SAR model. The discussion
will also review the results from the analysis of the likelihood of mis-  11% of the indeterminate consensus predictions changed to positive
classifying a mutagenic impurity as non-mutagenic based on the dif- after the expert review
ferent combinations of (Q)SAR results from the diﬀ erent methodolo-  28% of the indeterminate consensus predictions were not changed
gies. These results (i.e., 4.2% of the positive consensus predictions
 changed to negative after the expert review) support the observation
 that the ICH M7 classiﬁ cation paradigm (that any positive prediction
 yields a positive overall call) risks generating a number of false positive
Table 5
Summary showing the eﬀ ects of the conﬁ dence scores.
 Statistic-based result Expert rule-based result Count Percentage of results that were experimentally identiﬁ ed Ames mutagens
 Out-of-domain with probability of being positive 0.2–0.4 Negative 339 17.1%
 Out-of-domain with probability of being positive < 0.2 Negative 1415 8.8%
 56

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
predictions that can be corrected through the application of expert (highlighted in blue). Based on the structural similarity to triphenyl-
knowledge. Furthermore, it supports the conclusion that the majority of phosphine (CAS 603-35-0) and triphenylphosphine oxide (CAS 791-28-
indeterminate predictions are not meaningful signals and can be readily 6), which are both non-mutagenic in an Ames test (OECD SIDS, 2012),
downgraded to negative through a review of the model output and impurities B and C were concluded to be non-mutagenic (class 5).
supporting analogs from the public domain.
 4.2.3. Assessment of non-reactive groups
4.2.2. Use of analogs in an expert review For out-of-domain or indeterminate (Q)SAR results, additional
 Structurally and/or toxicologically meaningful non-mutagenic supporting analysis to conﬁ rm that the impurity lacks any DNA-reactive
analog(s) from public or proprietary databases or chemicals related to potential has also been used (Powley, 2015). This includes a visual
the impurity (such as the API or synthetic intermediate) are often used assessment of the compound to assure the lack of valid DNA-reactive
to support an overall prediction for an impurity when additional evi- alerts with plausible mechanisms, taking into consideration any unique
dence is needed. This approach is sometimes referred to as read-across alerts from proprietary information (Amberg et al., 2016) or knowledge
(Ball et al., 2016). As part of conducting a read-across assessment, the of metabolic activation. This assessment often takes into consideration
adequacy of the experimental design and results for the analog(s) are the strength of other model result(s) since models engineered under
evaluated, as reviewed in Amberg et al. (2016). Speciﬁ c analogs are statistical-based or expert rule-based algorithms predict mutagenicity
often selected that cover any structural features that the model(s) and consider applicability domain in diﬀ erent ways, and when one
identiﬁ ed as being potentially reactive (ICH M7(R1), 2017). It is valid model is deﬁ cient, the other may be reliably used to make up for the
and usual practice to discount mutagenic analogs when they contain deﬁ ciencies. Such an assessment may be supported by an inspection of
one or more additional structural moieties that are more likely re- the (Q)SAR model output when such reports visually illustrate a lack of
sponsible for the mutagenic result. potentially reactive features.
 The following examples illustrate expert reviews of out-of-domain In addition, speciﬁ c searches for signiﬁ cant functional groups and
results based on structural analogs. other substructures present in the impurity (performed manually or
 Case study 1 (Impurity A) - Expert review based on analogs: An automated by a software application) against a database containing
impurity (shown in Fig. 1) was predicted to be negative by an expert mutagenicity data (including proprietary data) is often performed. This
rule-based model. The impurity was out-of-domain in a statistical-based may indicate that the impurity contains a potentially reactive group
model as the compound contained a fragment not present in any che- when the results contain signiﬁ cantly more positive examples linked to
micals in the training set and no nearest neighbors were identiﬁ ed in that particular substructure than would be expected by chance (i.e.,
the model. The impurity is structurally similar to the API, which was enrichment of positives over background rate). Additionally, the (Q)
experimentally determined to be non-mutagenic. The prediction results SAR model may be applied to any substructure of the impurity to help
for both the impurity and the API were identical, and the API was also determine the reactive potential for the components, when the whole
predicted to be out-of-domain by the statistical model for the same chemical is out-of-domain.
reasons as the impurity. Since the impurity is structurally similar to the There are several substructures, such as protection groups (e.g., tert-
API and the only diﬀ erence is the addition of a non-reactive group (a butyloxycarbonyl, or BOC), where their presence within an impurity
hydroxyl group), the overall prediction is non-mutagenic and the im- may change the prediction from negative to out-of-domain. In these
purity is assigned to class 5. A class 4 assignment was not used in this cases, such substructures are speciﬁ cally known to block that portion of
situation since neither the API nor the impurity share an alert asso- the molecule from chemical reactivity (Amberg et al., 2016). Therefore,
ciated with mutagenicity (i.e., the out-of-domain fragment was not the (Q)SAR model could be run on the substructure without the BOC
considered an alert for mutagenicity). Some sponsors may consider this group instead (i.e., on the free amine in the case of a BOC-protected-
a class 4 compound to highlight that structural comparison with a amine) and this resulting prediction used as part of an expert review.
known non-mutagenic analog has been performed. The US FDA inter- Case study 3 (Impurity D) - Expert review based on an analysis
prets this situation as a class 5. of potentially reactive features: As shown in Fig. 3, the rule-based
 Case study 2 (Impurities B and C) - Expert review based on expert system identiﬁ ed no alert but determined one or more features
analogs: Cyclohexyldiphenylphosphine oxide (Impurity B; CAS 13689- were present in the impurity that were not found in the reference set.
20-8) and cyclohexyldiphenylphosphine (Impurity C; CAS 6372-42-5) Therefore, it is assigned as negative; however, there is uncertainty since
are impurities (shown in Fig. 2) occurring in the synthesis of a drug it contains unclassiﬁ ed features. The statistical-based model determined
substance. Neither showed structural concern for mutagenicity using the impurity is out-of-domain. It is known that the impurity reacts with
the expert rule-based model and were considered within the applic- water to form diphenyl phosphoric acid (838-85-7) and hydrazoic acid
ability domain. The statistical-based model predicted them as negative (7782-79-8). Since hydrazoic acid shows evidence of mutagenicity, a
but both molecules were out-of-domain due to the phosphine moiety conservative action would be to assigned it to class 3; however, since
 Fig. 1. (Q)SAR assessment of impurity A.
 57

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
 Fig. 2. (Q)SAR assessment of impurities B and C.
 58

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
 Fig. 3. (Q)SAR assessment of impurity D.
(Sodium) Azide is negative in a 2-year cancer bioassay (NTP, 1991)it domain results are similar. One of the biggest diﬀ erences is that there is
could be assigned as class 5 with appropriate justiﬁ cation. a supporting dataset for the indeterminate prediction but it falls in the
 Case study 4 (Impurity E) - Expert review based on features not middle of the positive and negative predictive space. One eﬀ ective
covered: As shown in Fig. 4, no alerts were identi ﬁ ed using the expert strategy can be to review the training set for secondary features not
rule-based system; however, it was determined to contain structural contained in the impurity that could skew the prediction towards in-
features not present in any of the reference set chemicals. Therefore, it determinate from either positive or negative. Also, a lack of similarity
was assigned as negative; however, there is uncertainty since it contains of the impurity to the underlying training set chemicals can be used to
unclassiﬁ ed features. The statistical-based system also generated an overrule such a call. For example, a statistical-based model prediction
out-of-domain call. Since impurity E will rapidly hydrolyze in an aqu- of a high molecular weight impurity containing a hindered epoxide was
eous environment to the aniline, which is experimentally Ames nega- indeterminate; however, an inspection of the training set indicated that
tive (aniline is a publicly known non-mutagenic compound in Salmo- the majority of the training set compounds responsible for the in-
nella and E. coli (NTP, 1980).), the impurity is assigned to class 5. determinate call were unhindered and hence the prediction may be
 overruled.
4.2.4. Situations when (Q)SAR methodology uses sub-models, i.e., GC
versus AT primary reversion site 4.3. Using an additional model
 Some (Q)SAR systems include a battery of models including those
for the traditional Ames (i.e., four strains to detect GC base pairs at the Although multiple available models may be built from the same or
primary site of reversion) and an additional model for the AT base pair similar public databases, diﬀ erent modeling techniques, as well as
reversion site (i.e., E. coli WP2 or E. coli WP2 uvrA, or E. coli WP2 uvrA methods for assessing the applicability domain, may give diﬀ erent re-
(pKM101), or S. typhimurium TA102). The database of compounds used sults. For example, a new model may generate a similar result that is
to build the model for the GC base pair mutations is typically larger within the applicability domain, whereas the initial model's result was
than that used for the AT reversion site. Therefore, it is more likely that out-of-domain.
a compound will be out-of-domain in the model for the AT reversion In addition to using another public model directly, an alternative is
site. A prediction just at the GC primary reversion site may be suﬃ cient to enhance an existing model through inclusion of proprietary structure
to support a valid prediction in many cases. However, where the im- (s) to increase the domain of the original model without substantially
purities contain speciﬁ c AT alerting fragment(s), such as oxidizing changing the original model. This has been particularly useful when
mutagens, cross-linking agents, and hydrazines, the four strains in the many related compounds are out-of-domain and the expansion of the
traditional Ames would not be able to detect the mutagenicity of an model includes one or more chemicals (e.g., API or key Ames tested
impurity. In this case, further interrogation of the impurity in strains intermediate) that are structurally related to the impurities. The addi-
that detect the AT base pair reversion site may be warranted. tion of these structures is often suﬃ cient to bring the impurity within
 the applicability domain but might change the probability score (or
4.2.5. Speciﬁ c consideration for expert review of indeterminate (Q)SAR equivalent conﬁ dence score) and, in limited situations, the prediction of
 Strategies for handling both indeterminate predictions and out-of- mutagenic potential for the impurity. Another approach is to create a
 Fig. 4. (Q)SAR assessment of impurity E.
 59

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
new model using a training set built either exclusively from proprietary 4.4. Class assignments - test or control to TTC?
data or proprietary data combined with publicly available data (Jolly
et al., 2015). Such modiﬁ ed models may need additional documenta- Any situation where one or more of the results are out-of-domain or
tion describing the speciﬁ c modiﬁ cations (such as the chemicalsindeterminate, in general, requires an expert review to provide support
added), as well as evidence that the revised model is consistent with the for an overall negative prediction (as outlined previously). In the ab-
OECD's (Q)SAR validation principles (OECD, 2007). sence of any supportive evidence for out-of-domain or indeterminate
 Case study 5 (Impurity F) - Assessment based on running an call(s) (e.g., there are no adequate non-mutagenic analogs or it is not
addition model and an expert review of analogs: For the impurity possible to verify that certain structural features of the impurity are not
shown in Fig. 5, an expert rule-based model did not identify any alerts reactive or another model generates an out-of-domain call), the pre-
but ﬂ agged it as containing a structural feature shared with an ex- diction would be considered uncertain and an Ames test may be pru-
perimentally determined mutagenic analog. A statistical-based model dent to make a ﬁ nal conclusion. Alternatively, it can be treated as
indicated the compound was out-of-domain. A second statistical-based mutagenic (class 3) and controlled to the Threshold of Toxicological
model predicted the compound as indeterminate, highlighting the Concern (TTC) deﬁ ned in the ICH M7 guideline. In situations where
oxime group as a potentially reactive fragment. An examination of the there is suﬃ cient supportive evidence and any positive signals from the
structural analogs supporting the oxime group as a potentially reactive models are refuted as part of an expert review, the impurity is generally
fragment showed that the examples most closely-related to Impurity F assigned to class 5 or class 4 if any present alert is shared with an
were mutagenic. In addition, two analogs were identiﬁ ed based on a empirically non-mutagenic chemical such as an API.
substructure search of supplemental databases for the oxime group and
both were experimentally determined mutagens. As a result of the po- 4.5. Regulatory review – US FDA experience
tential reactivity of the oxime group, the impurity was assigned to class
3. ICH M7 submissions are handled by the individual review divisions
 Case study 6 (Impurity G) - An assessment based on running a at US FDA/CDER. The reviewers assess the information provided by the
third model and expert review of analogs: An impurity (shown in pharmaceutical applicant, including information on the software and
Fig. 6) was predicted to be negative by an expert rule-based model; models used, the results from the software, the overall conclusions and
however, it contained a fragment spanning part of the “R1 to N to R1” any associated expert review documentation for consistency with the
(highlighted in blue) that was not present in any chemical in the re- ICH M7 guideline and to ensure the results and expert review are valid.
ference set of the expert rule-based system. The statistical-based model In cases where the reviewer has questions or concerns, the (Q)SAR
determined the compound to be out-of-domain. A second statistical- submission is provided to FDA/CDER's internal Computational
based system predicted the impurity to be negative. The impurity is Toxicology Consulting Service (CTCS) for evaluation (Rouse et al.,
structurally related to the API (API 2), which was predicted negative 2017). It should be noted that the reviewer will not re-run the predic-
(and in domain) for all three models run, and known to be experi- tions, but the Computational Toxicology Consultation Service staﬀ may.
mentally negative. A further search for analogs identiﬁ ed a compound Examples of such situations are:
that contained a similar fragment to the “R1 to N to R1” fragment
(highlighted in blue) that was negative in strain TA100. Based on the 1. poorly documented evaluations, unfamiliar software, software that
weight-of-the-evidence, the impurity was determined to be non-muta- does not allow for prediction interpretation consistent with ICH M7,
genic and assigned to class 5. or models that are not compliant with the OECD validation princi-
 Case study 7 (Impurity H) - Assessment based on running an ples
additional model and expert review of an analog (API): This im- 2. situations when only a single methodology was used or only read-
purity (shown in Fig. 7) was negative in an expert rule-based model and across (Ball et al., 2016) was used
out-of-domain in a statistical-based model. A second statistical-based 3. when the overall conclusions conﬂ ict with the individual model
model was run and it was determined to be negative and within the predictions, without an explanation
applicability domain of the model. The API was negative in the Ames 4. when the most recent version of the software was not run and a
test, predicted negative in the expert rule-based model and out-of-do- change in the prediction is anticipated
main in a statistical model with a second statistical model providing an
indeterminate call due to a low conﬁ dence negative prediction. Further In 2016, 217 consultation requests (for a total of 473 chemicals)
assessment was made that the substituent on the impurity that was not were related to ICH M7 submissions, with 90% for generic drug ap-
contained in the parent was qualiﬁ ed by a negative Ames test on the plications and 10% for new drug applications (Kruhlak et al., 2017). In
same substructure (analog) with a similar environment. Hence, the cases where the US FDA/CDER performs an independent (Q)SAR as-
impurity is assigned to class 5. sessment of the impurities, this includes:
 1 a review of the (Q)SAR data submitted
 2. a further structure-based search (using exact, substructure and
 Fig. 5. (Q)SAR assessment of impurity F.
 60

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
 Fig. 6. (Q)SAR assessment of impurity G.
 similarity-based searches) for additional experimental data on the negative and false positive predictions as much as practically possible.
 impurities or any analogs The information in Figs. 8 and 9 can be used to support such an analysis
 3. an independent (Q)SAR assessment using models from Leadscope of the likelihood of misclassifying a mutagenic impurity as non-muta-
 Inc. (Leadscope, 2017), Lhasa Limited (Lhasa, 2017), and Multi- genic based on diﬀ erent combinations of (Q)SAR results. This is im-
 CASE Inc. (MultiCASE, 2017) portant since all follow-on activities may compromise the desired high
 4. an expert review of the results and related literature throughput goals if they are not tied to an assessment of the overall risk.
 A detailed expert review of out-of-domain and indeterminate results has
 Section 4.2.1. provides some additional information on performing time and cost implications, since it may require the gathering of a group
an expert review at the US FDA. of cross-discipline experts, performing literature searches, and/or in-
 stigating an additional analysis (such as a legal review) in order to
4.6. Analysis of the likelihood of misclassifying a mutagenic impurity as reveal analogs that were previously designated as proprietary. On the
non-mutagenic other hand, assuming all out-of-domain or indeterminate results are
 potentially mutagenic has almost certainly greater time and cost im-
 A computational assessment of impurities should ideally balance the plications, such as the need to perform additional laboratory test(s) (as
need for a rapid analysis of multiple compounds while limiting false well as the possible synthesis of the impurity) and/or implementation of
 Fig. 7. (Q)SAR assessment of impurity H.
 61

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
 Fig. 8. Illustration of the number of times diﬀ erent (Q)SAR results are encountered.
 Fig. 9. Summary of the likelihood of misclassifying a mutagenic impurity as non-mutagenic for diﬀ erent combinations of results.
control strategies. Hence, the level of any additional analysis, such as similar to the pre-expert review performance (94%) reported in Dobo
the extent of an expert review, should ideally take into consideration et al. (2012) and also in the same range as the reproducibility of the
the likelihood of misclassifying a mutagenic impurity as non-mutagenic assay (McCann et al., 1984; Jolly et al., 2015). The analysis also shows
(i.e., a false negative prediction). that when the statistical-based model is negative (within the applic-
 As discussed previously, there are a variety of approaches to resolve ability domain of the model) and the expert rule-based model is in-
out-of-domain results. For example, as part of an expert review addi- determinate, then the likelihood of misclassifying a mutagenic impurity
tional supportive evidence may be provided, including suitable analogs, is also similar to two clear negatives (i.e., 11.8% vs. 8.1%). In addition,
an analysis of the lack of reactive potential as well as running another one of the most common scenarios is when a statistical-based model is
model (as discussed earlier). The likelihood of misclassifying a muta- out-of-domain and an expert rule-based model is negative (in domain).
genic impurity as non-mutagenic when one of the methodologies gen- A subset of these examples is shown where the calculated probability of
erates an out-of-domain result is in large part dependent on the result being positive from the statistical-based model is less than 0.2. This
from the other model. Based on the analysis shown in Fig. 9, if a sta- subset represents 1415 cases where the percentage of experimental
tistical-based model is out-of-domain and an expert rule-based model is mutagens is close to the case where both methodologies are clear ne-
positive, 36.2% of compounds are shown to be positive, whereas if the gative. An expert review based on the low conﬁ dence or probability
statistical-based model is out-of-domain and the expert rule-based score alongside an assessment consistent with a clear negative, as dis-
model is negative then only 11.3% of compounds are positive. When cussed in Amberg et al. (2016), may be appropriate (i.e., “a rapid visual
the (Q)SAR model presents a result that is indeterminate, it may be inspection of the results by the expert can be used to verify that no valid
prudent to examine the basis for the indeterminate call and determine alerts for mutagenicity with a plausible mechanism were overlooked by
through an expert review whether it can be refuted for valid reason, as the two (Q)SAR methodologies”).
discussed in Amberg et al. (2016) (e.g. a shared alert with known ne- It is also interesting to note that when the statistical-based model is
gative [ICH M7 class 4], an explanation based on the mechanism, an out-of-domain and the expert rule-based model is positive then the
assessment of the relevance of features or underlying data from statis- percentage of Ames mutagens is 36.1%, which is similar to the situation
tical-based methodologies, expert reviews based on chemical analogs when the statistical-based model is negative and the expert rule-based
from public or in-house sources, a visual inspection by an expert or an model is positive (37.5%). This may indicate that many of the chemicals
assessment of the strength of the single prediction). predicted to be out-of-domain by the statistical-based model are related
 Fig. 8 summarizes the frequency for the diﬀ erent (Q)SAR combi- to novel APIs lacking reactive features. It is also worth pointing out that
nations and Fig. 9 the percentage of results that were experimentally when an expert rule-based result is indeterminate, then the results from
identiﬁ ed Ames mutagens. The charts show the most common scenario the statistical-based model are correlated with the percentage of Ames
is when the two systems predict the chemicals as a clear negative. The positive. When the statistical-based result is negative and the expert
proportion of experimentally determined mutagens in this situation is rule-based result it indeterminate, the percentage of Ames positives is
 62

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
11.8% whereas when the statistical-based result is positive the per- References
centage of Ames positives is 27.7% which illustrates the value of using
the two complementary methodologies. Amberg, A., Beilke, L., Bercu, J., Bower, D., Brigo, A., Cross, K.P., Custer, L., Dobo, K.,
 There are a number of limitations with this type of cross-pharma- Dowdy, E., Ford, K.A., Glowienke, S., Gompel, J.V., Harvey, J., Hasselgren, C.,
ceutical company analysis of proprietary data. It is not possible to see Honma, M., Jolly, R., Kemper, R., Kenyon, M., Kruhlak, N., Leavitt, P., Miller, S.,
 Muster, W., Nicolette, J., Plaper, A., Powley, M., Quigley, D.P., Reddy, M.V., Spirkl,
the individual chemicals to verify that there are no duplicate chemicals; H.-P., Stavitskaya, L., Teasdale, A., Weiner, S., Welch, D.S., White, A., Wichard, J.,
however, this number will be low since diﬀ erent proprietary collections Myatt, G.J., 2016. Principles and procedures for implementation of ICH M7 re-
generally cover separate areas of the chemical space. The analysis is commended (Q)SAR analyses. Regul. Toxicol. Pharmacol. 77, 13–24. https://doi.org/
 10.1016//j.yrtph.2016.02.004.
also based on a harmonized evaluation of diﬀ erent models and algo- Ball, N., Cronin, M.T., Shen, J., Blackburn, K., Booth, E.D., Bouhifd, M., Donley, E.,
rithms. In addition, the overall proportion of positive compounds dif- Egnash, L., Hastings, C., Juberg, D.R., Kleensang, A., Kleinstreuer, N., Kroese, E.D.,
fers across the individual proprietary collections; however, the general Lee, A.C., Luechtefeld, T., Maertens, A., Marty, S., Naciﬀ , J.M., Palmer, J., Pamies, D.,
analysis is based on chemicals similar to APIs as well as chemicals ty- Penman, M., Richarz, A.N., Russo, D.P., Stuard, S.B., Patlewicz, G., van Ravenzwaay,
 B., Wu, S., Zhu, H., Hartung, T., 2016. Toward Good Read-Across Practice (GRAP)
pically used in chemical synthesis. Finally, there may be bias in-guidance. ALTEX 33, 149–166. https://doi.org/10.14573/altex.1601251.
troduced since historical testing of impurities is generally performed on Barber, C., Amberg, A., Custer, L., Dobo, K.L., Glowienke, S., Gompel, J.V., Gutsell, S.,
only class 3. Harvey, J., Honma, M., Kenyon, M.O., Kruhlak, N., Muster, W., Stavitskaya, L.,
 Teasdale, A., Vessey, J., Wichard, J., 2015. Establishing best practise in the appli-
 cation of expert review of mutagenicity under ICH M7. Regul. Toxicol. Pharmacol.
5. Summary and conclusion 73, 367–377. https://doi.org/10.1016/j.yrtph.2015.07.018.
 Carrió, P., Pinto, M., Ecker, G., Sanz, F., Pastor, M., 2014. Applicability Domain Analysis
 (ADAN): A robust method for assessing the reliability of drug property predictions. J.
 As part of any ICH M7 analysis, an essential step is to combine the Chem. Inf. Model. 54, 1500–1511. https://doi.org/10.1021/ci500172z.
results from the two (Q)SAR methodologies consistent with the lan- Chakravarti, S.K., Saiakhov, R.D., Klopman, G., 2012. Optimizing predictive performance
 of CASE Ultra expert system models using the applicability domains of individual
guage in the ICH M7 guideline (i.e., “absence of structural alerts”). Out- toxicity alerts. J. Chem. Inf. Model. 52 (10), 2609–2618. https://doi.org/10.1021/
of-domain results present particular challenges as to how this overall ci300111r. 2012 Oct 22.
prediction should be generated. The underlying basis for this applic- Dobo, K.L., Greene, N., Fred, C., Glowienke, S., Harvey, J.S., Hasselgren, C., Jolly, R.,
 Kenyon, M.O., Munzner, J.B., Muster, W., Neft, R., Reddy, M.V., White, A.T., Weiner,
ability domain analysis is often based upon a structural assessment, S., 2012. In silico methods combined with expert knowledge rule out mutagenic po-
using similarity measures or an analysis of unknown fragments, po- tential of pharmaceutical impurities: an industry survey. Regul. Toxicol. Pharmacol.
tentially in combination with other information. The results from such 62, 449–455. https://doi.org/10.1016/j.yrtph.2012.01.007.
 Ellison, C.M., Madden, J.C., Judson, P., Cronin, M.T., 2010. Using in silico tools in a
applicability domain analysis may be presented as either inside or weight of evidence approach to aid toxicological assessment. Mol. Inform. 29 (1–2),
outside the applicability domain. However, additional information may 97–110. https://doi.org/10.1002/minf.200900006.
be taken into consideration as part of an expert review that includes a Greene, N., Dobo, K.L., Kenyon, M.O., Cheung, J., Munzner, J., Sobol, Z., Sluggett, G.,
weight-of-evidence re-assessment of the applicability domain, including Zelesky, T., Sutter, A., Wichard, J., 2015. A practical application of two in silico
 systems for identiﬁ cation of potentially mutagenic impurities. 2015 Jul. Regul.
relevant non-mutagenic analogs or additional model output (e.g., the Toxicol. Pharmacol. 72 (2), 335–349. https://doi.org/10.1016/j.yrtph.2015.05.008.
conﬁ dence score or a calculated probability score of less than 0.2). The Epub 2015 May 15.
likelihood of misclassifying a mutagenic impurity as non-mutagenic in Hanser, T., Barber, C., Marchaland, J.F.F., Werner, S., 2016. Applicability domain: to-
 wards a more formal deﬁ nition. SAR QSAR Environ. Res. 27, 865–881. https://doi.
situations with out-of-domain calls when no alerts or other positive/ org/10.1080/1062936X.2016.1250229.
indeterminate signals are identiﬁ ed is similar to when there are two ICH M7, 2017. (R1). Assessment and control of DNA reactive (mutagenic) impurities in
clear negative results based on the analysis in this paper. pharmaceuticals to limit potential carcinogenic risk. http://www.ich.org/ﬁ leadmin/
 Public\_Web\_Site/ICH\_Products/Guidelines/Multidisciplinary/M7/M7\_R1\_
 The current paper has reviewed diﬀ erent ways of handling out-of- Addendum\_Step\_4\_31Mar2017.pdf.
domain and indeterminate results, including how to generate an overall Jolly, R., Ahmed, K.B., Zwickl, C., Watson, I., Gombar, V., 2015. An evaluation of in-
prediction and any follow-up expert review or additional analysis (in- house and oﬀ -the-shelf in silico models: implications on guidance for mutagenicity
 assessment. Regul. Toxicol. Pharmacol. 71 (3), 388–397. https://doi.org/10.1016/j.
cluding an expert review based on analogs or an assessment of reactive yrtph.2015.01.010.
features as well as the option of running another model) that is con- Kruhlak, N., 2017. A Regulatory Perspective on the Use of (Q)SAR to Evaluate Potential
sidered adequate based on the likelihood of misclassifying a mutagenic Genotoxic Impurities. In: Presentation at American College of Toxicology (ACT)
 Annual Meeting, November 2017, Palm Springs, CA.
impurity. Over time, models that are routinely updated with the latest Leadscope, 2017. Leadscope, Inc. http://www.leadscope.com/.
information, including the incorporation of knowledge from proprie- Lhasa, 2017. Lhasa Limited. https://www.lhasalimited.org/.
tary databases whenever possible, will reduce the number of false po- McCann, J., Horn, L., Kaldor, J., 1984. An evaluation of Salmonella (Ames) test data in
sitives and false negatives, as well as out-of-domain and indeterminate the published literature: application of statistical procedures and analysis of muta-
 genic potency. Mutat. Res. 134 (1), 1–47. https://doi.org/10.1016/0165-1110(84)
results, further streamlining the ICH M7 (Q)SAR assessment process. 90013-7.
 Müller, L., Mauthe, R.J., Riley, C.M., Andino, M.M., Antonis, D.D., Beels, C., Degeorge, J.,
 Knaep, A.G.D., Ellison, D., Fagerland, J.A., Frank, R., Fritschel, B., Galloway, S.,
Acknowledgement Harpur, E., Humfrey, C.D., Jacks, A.S., Jagota, N., Mackinnon, J., Mohan, G., Ness,
 D.K., O'Donovan, M.R., Smith, M.D., Vudathala, G., Yotti, L., 2006. A rationale for
 Research reported in this publication was supported by the National determining, testing, and controlling speciﬁ c impurities in pharmaceuticals that
 possess potential for genotoxicity. Regul. Toxicol. Pharmacol. 44, 198e211.
Institute of Environmental Health Sciences of the National Institutes of MultiCASE, 2017. MultiCASE Inc. http://www.multicase.com/.
Health under Award Number R44ES026909. The content is solely the Myatt, G.J., Beilke, L.D., Cross, K.P., 2016. In Silico Tools and their Application. In:
responsibility of the authors and does not necessarily represent the Reedijk, J. (Ed.), Reference Module in Chemistry, Molecular Sciences and Chemical
 Engineering. Elsevier. https://doi.org/10.1016/B978-0-12-409547-2.12379-0 .
oﬃ cial views of the National Institutes of Health. Myatt, G.J., Ahlberg, E., Akahori, Y., Allen, D., Amberg, A., Anger, L.T., Aptula, A.,
 Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D.,
 Brigo, A., Burden, N., Cammerer, Z., Cronin, M.T.D., Cross, K.P., Custer, L.P.,
Appendix A. Supplementary data Dettwile, M., Dobo, K., Ford, K.A., Fortin, M.C., Gad-McDonald, S.E., Gellatly, N.,
 Gervais, V., Glover, K.P., Glowienke, S., Van Gompel, J., Gutsell, S., Hardy, B.,
 Supplementary data to this article can be found online at https:// Harvey, J.S., Hillegass, J., Honma, M., Hsieh, J.-H., Hsu, C.-W., Hughes, K., Johnson,
doi.org/10.1016/j.yrtph.2018.12.007. C., Jolly, R., Jones, D., Kemper, R., Kenyon, M.O., Kim, M.T., Kruhlak, N.L., Kulkarni,
 S.A., Kümmerer, K., Leavitt, P., Majer, B., Masten, S., Miller, S., Moser, J., Mumtaz,
 M., Muster, W., Neilson, L., Oprea, T., Patlewicz, G., Paulino, A., Piparo, E.L., Powley,
 M., Quigley, D.P., Reddy, M.V., Richarz, A.-N., Ruiz, P., Schilter, B., Seraﬁ mova, R.,
Transparency document Simpson, W., Stavitskaya, L., Stidl, R., Suarez-Rodrigue, D., Szabo, D.T., Teasdale, A.,
 Trejo-Martin, A., Valentin, J.-P., Vuorinen, A., Wall, B.A., Watts, P., White, A.T.,
 Transparency document related to this article can be found online at Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C., Hasselgren, C., 2018. In
https://doi.org/10.1016/j.yrtph.2018.12.007 . silico toxicology protocols. Regul. Toxicol. Pharmacol. 98, 1–17.
 63

A. Amberg et al. Regulatory Toxicology and Pharmacology 102 (2019) 53–64
Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Advances in Experimental Medicine and Biology Validation of Alternative Methods
 Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova- for Toxicity Testing. pp. 165–187. https://doi.org/10.1007/978-3-319-33826-2\_6.
 Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, Powley, M.W., 2015. (Q)SAR assessments of potentially mutagenic impurities: a reg-
 D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C., 2005. Current Status of ulatory perspective on the utility of expert knowledge and data submission. Regul.
 Methods for Deﬁ ning the Applicability Domain of (Quantitative) Structure-Activity Toxicol. Pharmacol. 71, 295–300. https://doi.org/10.1016/j.yrtph.2014.12.012.
 Relationships. The Report and Recommendations of ECVAM Workshop 52. ATLA 33, Powley, M., 2017. Presentation made at the 2017 Genetic Toxicity Association annual
 155–173. meeting “(Q)SAR Evaluation of Potentially Mutagenic Impurities: Regulatory
NTP, 1980. Bacterial Mutagenicity Study for Aniline (712474). http://tools.niehs.nih. Experience with Out of Domain Results” by Mark Powley.
 gov/cebs3/ntpViews/?activeTab=detail&studyNumber=712474. Rouse, R., Kruhlak, N., Weaver, J., Burkhart, K., Patel, V., Strauss, D.G., 2017. Translating
NTP, 1991. NTP Technical Report on the Toxicology and Carcinogenesis Studies of new science into the drug review process: The US FDA's division of applied regulatory
 Sodium Azide (CAS NO. 26628-22-8) in F344/N Rate (Gavage Studies). September science. Therapeut. Innovat. Regul. Sci. 1–12. https://doi.org/10.1177/
 1991. NTP TR 389. https://ntp.niehs.nih.gov/ntp/htdocs/lt\_rpts/tr389.pdf. 2168479017720249.
OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure-activity Sutter, A., Amberg, A., Boyer, S., Brigo, A., Contrera, J.F., Custer, L.L., Dobo, K.L.,
 Relationships [(Q)SAR] Models, OECD Environment Health and Safety Publications Gervais, V., Glowienke, S., Gompel, J.V., Greene, N., Muster, W., Nicolette, J., Reddy,
 Series on Testing and Assessment No. 69. (ENV/JM/MONO(2007)2). http://www. M.V., Thybaud, V., Vock, E., White, A.T., Müller, L., 2013. Use of in silico systems and
 oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure- expert knowledge for structure-based assessment of potentially mutagenic impurities.
 activity-relationship-q-sar-models-9789264085442-en.htm . Regul. Toxicol. Pharmacol. 67, 39–52. https://doi.org/10.1016/j.yrtph.2013.05.001.
OECD SIDS, 2012. Triphenylphosphine. Organisation for Economic Co-operation and Williams, R.V., Amberg, A., Brigo, A., Coquin, L., Giddings, A., Glowienke, S., Greene, N.,
 Development. SIDS Initial Assessment Proﬁ les agreed in the course of the OECD HPV Jolly, R., Kemper, R., O'Leary-Steele, C., Parenty, A., Spirkl, H.-P., Stalford, S.A.,
 Chemicals Programme from 1993 to 2011. Series on Testing & Assessment No. 166. Weiner, S.K., Wichard, J., 2016. It's diﬃ cult, but important, to make negative pre-
 ENV/JM/MONO(2012)4/PART5. 27-Feb-2012. dictions. Regul. Toxicol. Pharmacol. 76, 79–86. https://doi.org/10.1016/j.yrtph.
Patlewicz, G., Worth, A.P., Ball, N., 2016. Validation of Computational Methods. 2016.01.008.
 64

Source PDF: /instem.com/www.instem.com/docs/sot-2021/ICH-M7-Principles-and-Procedures-Paper.pdf

 Regulatory Toxicology and Pharmacology 77 (2016) 13e24
 Contents lists available at ScienceDirect
 Regulatory Toxicology and Pharmacology
 journal homepage: www.elsevier.com/locate/yrtph
Principles and procedures for implementation of ICH M7
recommended (Q)SAR analyses\*
Alexander Amberg a, Lisa Beilke b, Joel Bercu c, Dave Bower d, Alessandro Brigo e,
Kevin P. Cross d, Laura Custer f, Krista Dobo g, Eric Dowdy c, Kevin A. Ford h,
Susanne Glowienke i, Jacky Van Gompel j, James Harvey k, Catrin Hasselgren d,
Masamitsu Honma l, Robert Jolly m, Raymond Kemper n, Michelle Kenyon g,
Naomi Kruhlak o, Penny Leavitt f, Scott Miller d, Wolfgang Muster e, John Nicolette p,
Andreja Plaper q, Mark Powley o, Donald P. Quigley d, M. Vijayaraj Reddy r,
Hans-Peter Spirkl a, Lidiya Stavitskaya o, Andrew Teasdale s, Sandy Weiner t,
Dennie S. Welch p, Angela White k, Joerg Wichard u, Glenn J. Myatt d, \*
a Sanoﬁ -Aventis Deutschland GmbH, Frankfurt, Germany
b Toxicology Solutions, San Diego, CA, USA
c Gilead, Foster City, CA, USA
d Leadscope, Columbus, OH, USA
e Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center, Basel, Switzerland
f Bristol-Myers Squibb, New Brunswick, NJ, USA
g Pﬁ zer, Groton, CT, USA
h Genentech, South San Francisco, USA
i Novartis Institutes for Biomedical Research, Basel, Switzerland
j Janssen, Beerse, Belgium
k GlaxoSmithKline, Ware, Herts, UK
l National Institute of Health Sciences, Tokyo, Japan
m Eli Lilly and Company, Indianapolis, IN, USA
n Vertex, Boston, MA, USA
o FDA Center for Drug Evaluation and Research, Silver Spring, MD, USA
p AbbVie Inc., North Chicago, IL, USA
q KRKA, Novo Mesto, Slovenia
r Merck Research Laboratories, West Point, PA, USA
s AstraZeneca, Macclesﬁ eld, Cheshire, UK
t Janssen, Spring House, PA, USA
u Bayer HealthCare, Berlin, Germany
article info abstract
Article history: The ICH M7 guideline describes a consistent approach to identify, categorize, and control DNA reactive,
Received 3 February 2016 mutagenic, impurities in pharmaceutical products to limit the potential carcinogenic risk related to such
Accepted 5 February 2016 impurities. This paper outlines a series of principles and procedures to consider when generating (Q)SAR
Available online 11 February 2016 assessments aligned with the ICH M7 guideline to be included in a regulatory submission. In the absence
 of adequate experimental data, the results from two complementary (Q)SAR methodologies may be
Keywords: combined to support an initial hazard classiﬁ cation. This may be followed by an assessment of additional
ICH M7 information that serves as the basis for an expert review to support or refute the predictions. This paper
 \* FDA CDER Disclaimer: The ﬁ ndings and conclusions in this manuscript have not
been formally disseminated by the FDA and should not be construed to represent
any agency determination or policy. The mention of commercial products, their
sources, or their use in connection with material reported herein is not to be
construed as either an actual or implied endorsement of such products by the
Department of Health and Human Services.
 \* Corresponding author. Present address: Leadscope, Inc., 1393 Dublin Road,
Columbus, OH 43215, USA.
 E-mail address: gmyatt@leadscope.com (G.J. Myatt).
http://dx.doi.org/10.1016/j.yrtph.2016.02.004
0273-2300/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

14 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24
Impurities elucidates scenarios where additional expert knowledge may be beneﬁ cial, what such an expert review
(Q)SAR may contain, and how the results and accompanying considerations may be documented. Furthermore,
Mutagenic impurities the use of these principles and procedures to yield a consistent and robust (Q)SAR-based argument to
Ames test support impurity qualiﬁ cation for regulatory purposes is described in this manuscript.
Toxicity databases © 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license
Expert review (http://creativecommons.org/licenses/by/4.0/).
1. Introduction details concerning the use of expert knowledge in the context of an
 ICH M7 (Q)SAR analysis and Powley (2015) provided general rec-
 The ICH M7 guideline (“Assessment and control of DNA reactive ommendations concerning the format and content of a (Q)SAR
(mutagenic) impurities in pharmaceuticals to limit potential analysis report to support regulatory submission.
carcinogenic risk”) provides a framework for assessing and con- The ICH M7 guideline is currently being implemented
trolling DNA reactive impurities in a pharmaceutical product (ICH throughout the pharmaceutical industry and international regula-
M7, 2015a). The guideline describes the process whereby actual tory agencies. A number of speciﬁ c difﬁ culties are being encoun-
and potential impurities or degradation products likely to be pre- tered that are not fully addressed in existing publications. These
sent in the drug substance and drug product are identiﬁ ed and include: (1) the process of assessing the adequacy of sufﬁ cient
outlines how a hazard assessment should be performed. When no in vivo and/or in vitro data; (2) the generation of an overall
adequate experimental mutagenicity and/or carcinogenicity assessment from the two (Q)SAR methodologies which individually
results are available, a structure-based computational toxicology or generate positive, negative, or inconclusive predictions as well as
(Q)SAR1 analysis may be able to predict the mutagenic potential of out-of-domain classiﬁ cations; (3) when to apply expert knowledge
an impurity. The hazard assessment process leads to the assign- that could potentially refute a (Q)SAR prediction; (4) what rationale
ment of each impurity to one of ﬁ ve classes described in Table 1. may be considered for use in such an expert review; and (5) an
Brieﬂ y, class 1 impurities are to be controlled “… at or belowoutline for a standardized report to ensure the results are consis-
compound-speciﬁ c acceptable limit” (ICH M7, 2015b), class 2 or 3 tently documented, transparent and complete.
impurities are to be controlled at or below acceptable limits Fig. 1 summarizes the process of implementing a (Q)SAR anal-
(appropriate Threshold of Toxicological Concern or TTC) and classes ysis of potential mutagenic impurities. The ﬁ rst step is to collect any
4 and 5 are to be treated as non-mutagenic impurities (ICH M7, relevant data from public sources (such as from the literature) for
2015a; Kasper and Müller, 2015). each impurity. This information can be supplemented with relevant
 Prior to the publication of ICH M7, many regional guidance in-house test results. In general, adequate negative bacterial
documents and scientiﬁ c papers were published, each contributing mutagenicity and/or carcinogenicity laboratory data are sufﬁ cient
to the thought process followed in a mutagenic impurity risk to assign the impurity to class 5, whereas adequate positive data
assessment (EMA, 2006, 2010; FDA, 2008; Müller et al., 2006). This would result in assigning the impurity to classes 1 or 2. The ade-
included regulatory guidance documents from the European quacy of the data used in these classiﬁ cations should be critically
Medicines Agency (EMA, 2006, 2010) and a draft guidance from the reviewed. In the absence of adequate data, a (Q)SAR analysis may be
US Food and Drug Administration (FDA, 2008) that outlined a used for this class assignment. The (Q)SAR results are used to assign
methodology for assessing DNA-reactive compounds based on the impurity to ICH M7 classes 3e5. This may include the genera-
available data as well as mutagenicity predictions from (Q)SAR tion of an expert review to accept or refute any predictions. Positive
models. Sutter et al. (2013) outlined the different (Q)SAR method- overall assessments are assigned to class 3, with negative overall
ologies available and highlighted the importance of applying expert assessments generally assigned to class 5; however, where a spe-
knowledge to predictions, a concept also discussed by Dobo et al. ciﬁ c argument based on shared alerts with a compound known to
(2012), Kruhlak et al. (2012), Naven et al. (2012), Barber et al.be non-mutagenic is made, these compounds may be assigned to
(2015) and Stavitskaya et al. (2015). Dobo et al. (2012) demon- class 4.
strated improved accuracy with expert input on negative pre- This paper outlines a number of practical principles and pro-
dictions. Powley (2015), Greene et al. (2015), Stavitskaya et al. cedures that can be used in generating a (Q)SAR assessment aligned
(2015) and Barber et al. (2015) recently provided additionalwith ICH M7 as part of a regulatory submission, including accom-
Table 1
Deﬁ nition of the ICH M7 hazard classiﬁ cations.
 Class Deﬁ nition
 1 Known mutagenic carcinogens
 2 Known mutagens with unknown carcinogenic potential (bacterial mutagenicity positive,a no rodent carcinogenicity data)
 3 Alerting structure, unrelated to the structure of the drug substance; no mutagenicity data
 4 Alerting structure, same alert in drug substance or compounds related to the drug substance (e.g., process intermediates) which have been tested and are non-
 mutagenic
 5 No structural alerts, or alerting structure with sufﬁ cient data to demonstrate lack of mutagenicity or carcinogenicity
 a Or other relevant positive mutagenicity data indicative of DNA-reactivity-related induction of gene mutations (e.g., positive ﬁ ndings in in vivo gene mutation studies).
 1 The term “(Q)SAR” refers to (Quantitative) Structure-Activity Relationship and panying expert analysis. The paper provides a brief overview of the
is used as an acronym for computational models that predict a biological response process of identifying and reviewing available data from public and
(such as mutagenicity) based on the chemical structure of the test molecule. The in-house databases as well as the literature. In the absence of
term collectively refers to both quantitative and non-quantitative structure-activity adequate data, the principles for combining the (Q)SAR results from
relationships by placing the “Q” in parentheses.

 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24 15
 Fig. 1. Flow chart depicting an ICH M7 (Q)SAR assessment.
complementary methodologies will be described. This paper will TA100) as well as Escherichia coli WP2 strains or Salmonella typhi-
discuss when and how to generate a supplemental expert review murium TA102 (which are similar in mutation detection), exposed
that may concur with or refute any prediction. A series of case to the test substance both in the presence and absence of an
studies are presented to illustrate the different principles and appropriate metabolic activation system, with concentrations for
procedures described. Many of these case studies are from phar- soluble non-cytotoxic substances up to 5 mg/plate or 5 ml/plate.
maceutical projects but have not been reviewed or accepted by a Studies pre-dating the publication of the OECD guideline are
regulatory authority unless stated otherwise. This paper will also generally acceptable when they were performed in a manner
provide suggestions detailing the contents of an expert analysis and consistent with the OECD guideline (OECD, 1997).
delineate its inclusion in a regulatory submission. Pending sufﬁ cient justiﬁ cation (e.g., difﬁ cult to synthesize im-
 purities), data from other study designs, using fewer test strains or
2. Assessing available data lower drug concentrations, may be used when the quality of the
 data and study design is considered appropriate. Decisions to
 According to the ICH M7 guideline (ICH M7, 2015a), the ﬁ rst step accept suboptimal assays may be inﬂ uenced by an analysis of the
in the hazard assessment is “… database and literature searches for risk versus the beneﬁ t. Deviations from the standard test protocols
carcinogenicity and bacterial mutagenicity data …” Since data may are acceptable in certain situations, for example, where a limited
have been generated within a pharmaceutical manufacturer's or- number of strains have been tested yet it has been shown that those
ganization, a search of proprietary in-house data may be performed strains are sensitive to any identiﬁ ed structural alert, as outlined in
alongside open access or commercial database searches. Table 2 Note 2 of the ICH M7 guideline (ICH M7, 2015a). The assessment of
lists a number of open access and commercial databases contain- data may also take into account structural classes that result in false
ing mutagenicity and/or carcinogenicity data. Since it is unrealistic positives under certain experimental conditions, such as an inter-
to search all possible databases individually, utilizing a database action between a test material containing an acid halide or sulfonyl
containing up-to-date information from many of these sources halide and DMSO in the Ames test (Amberg et al., 2015). It should
provides a useful alternative. A number of such services arebe noted that Ames data tested on a limited number of strains may
described in Table 2. be considered as part of the weight of evidence in any accompa-
 In addition ICH recently published a draft addendum to ICH M7. nying expert analysis. Validation statistics of limited strain models
Included within this addendum are a series of permissible limits for can be used to support the expert analysis (Diehl et al., 200 0; Zeiger
a range of commonly used reagents (ICH M7, 2015b). et al., 1985). Other reported genetic toxicity testing battery results
 The focus of ICH M7 is on DNA-reactive impurities, which are are not generally relevant in this context, but may be considered on
generally identiﬁ ed using the Bacterial Reverse Mutation Assay, a case-by-case basis when no or inadequate Ames data are avail-
commonly referred to as the Ames assay (OECD, 1997). An Ames able, such as, positive mouse lymphoma studies with increases in
assay may have been performed on the speciﬁ c impurity, either by large colonies, when the assay and data meet up to date criteria for
the pharmaceutical manufacturer or identiﬁ ed from a search of positive results (OECD 490, 2015).
open access or commercial databases. Any results from a database The ICH M7 Addendum (Step 2) discusses what factors consti-
search should return information necessary to understand the tute an adequate rodent carcinogenicity study (ICH M7, 2015b). An
adequacy of the study. An adequately performed negative bacterial adequate negative rodent carcinogenicity study is sufﬁ cient to
mutagenicity study is generally sufﬁ cient to assign the impurity to categorize the impurity as class 5. A positive result with evidence of
class 5, which is treated as a non-mutagenic impurity. Positive re- a mutagenic mechanism from an adequately performed study may
sults may be used to assign the impurity to class 2 (known muta- be used to categorize the compound as class 1 (known mutagenic
gens with unknown carcinogenic potential). The adequacy of any carcinogen). There may also be situations where a compound is
Ames data used in both the class 2 or class 5 assignments should be positive in the rodent carcinogenicity study and negative in the
critically reviewed as discussed in Greene et al. (2015), in line with bacterial mutagenicity study. For example, carcinogens that are
the principles of Klimisch (Klimisch et al., 1997)aswellasbe negative in the bacterial mutation study may act through a non-
generally consistent with the discussion in Note 2 of the ICH M7 mutagenic mechanism such as by causing hormonal imbalance or
guideline (ICH M7, 2015a). These publications indicate that the proliferative changes leading to cancer. When mechanisms are
Ames test data should be available for inspection and should clearly demonstrated, these cases are considered outside the scope
include at least ﬁ ve strains of bacteria, including four strains of of ICH M7. When a genotoxic threshold is demonstrated per ICH M7
S. typhimurium (TA1535; TA1537 or TA97a or TA97; TA98; and in an in vivo follow-up test e.g. rat micronucleus, a Permissible Daily

16 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24
Table 2
Databases containing information on carcinogenicity and mutagenicity data.
 Database Description Reference
 ATSDR Open access database from the Agency for Toxic Substances and Disease Registry (ATSDR) includes toxicological ATSDR, 2015
 proﬁ les for the hazardous substances including genotoxicity
 CCRIS Open access database covering chemical carcinogens, including structures and experimental data, covering the period Young, 2002; CCRIS, 2011
 1985e2011
 CPDB Open access Carcinogenicity Potency DataBase covering the period 1980e2011 Gold, 1997, 2001, 2005; CPDB
 2011
 DSSTox Open access Distributed Structure-Searchable Toxicity (DSSTox) Database Network including content from other DSSTox-Archive, 2012
 sources (e.g. CPDB, ISSCAN)
 ECHA Open access European Chemicals Agency (ECHA) database containing actual data and read across results for chemicals ECHA, 2015
 manufactured and imported in Europe
 ExPub Commercial application that includes access to the GENE-TOX and CCRIS databases ExPub, 2015
 GENE-TOX GENE-TOX provides genetic toxicology (mutagenicity) test data from expert peer review of open scientiﬁ c literature for GENE-TOX, 1998
 more than 3000 chemicals from the United States Environmental Protection Agency (EPA)
 IARC Open access International Agency for Research on Cancer (IARC) monographs including carcinogenicity classiﬁ cation IARC, 2015
 IPS INCHEM Open access International Program on Chemical Safety search for variety of summary documents INCHEM, 2015
 IRIS Open access data from the EPA in support of human health risk assessment, focusing on hazard identiﬁ cation and dose IRIS, 2015
 eresponse assessment
 ISSCAN Open access database on chemical carcinogens, including structures and experimental data from Istituto Superiore di Benigni et al., 2008
 Sanit/C18a
 JECDB Open access Japanese Existing Chemical Data Base (JECDB) containing high production volume chemicals JECDB, 2015
 Leadscope Commercial genetic toxicity and rodent carcinogenicity databases from numerous sources (including US FDA CDER Leadscope, 2015
 product approval reviews, FDA CFSAN, NTP, CCRIS, and so on) as well as ongoing data harvesting from the literature.
 Currently includes genetic toxicity data for 11,028 compounds and 179,732 test results and rodent carcinogenicity data
 for 3598 compounds and 11,538 test results.
 MultiCASE QSAR model training sets containing mutagenicity and rodent carcinogenicity data from public and proprietary sources MultiCASE, 2015
 including the FDA, GENETOX, NTP, CCRIS and IARC.
 NTP Open access database of National Toxicology Program results Tennant, 1991; NTP, 2015
 PAN Open access Pesticide Action Network (PAN) Pesticide Database PAN, 2014
 Pharma Pendium Commercial toxicity data from FDA and EMA approval documents Pharmapendium, 2015
 RTECS Commercial database available through third parties (e.g. Leadscope) currently containing 10,517 Tumorigenic studies Sweet, 1999; RTECS, 2015
 for 3724 compounds and 46,385 Mutation studies for 13,343 compounds
 ToxNet/ Open access on-line toxicity search system from the US National Library of Medicine with access to archived versions of Wexler, 2001; ToxNet, 2015
 ChemIDPlus CCRIS, GENE-TOX, CPDB
 TRACE from Commercial service for TRACE includes information from peer-reviewed toxicology and nutrition journals as well as Anderson, 2000; BIBRA, 2015;
 BIBRA secondary sources and websites. In addition to the primary literature on the health effects of chemicals, TRACE covers Robinson, 2000
 ofﬁ cial publications and evaluations issued by authoritative groups.
 VITIC from Lhasa Commercial data from published and unpublished sources (15,000 records for carcinogenicity and nearly 95,000 VITIC, 2015
 Limited records with mutagenicity Ames data) from a number of sources including IARC Monographs, European Chemicals
 Bureau (IUCLID) and NTP.
Exposure (PDE) approach may be considered (ICH M7, 2015a). 3. Generating (Q)SAR predictions
2.1. Case study 1: identifying a compound with historical data In the absence of sufﬁ cient experimental mutagenicity and/or
 carcinogenicity data for a speciﬁ c impurity, the ICH M7 guideline
 In case study 1, a public database search identiﬁ ed a historical recommends the use of (Q)SAR models for evaluating the muta-
bacterial mutagenicity study with a negative result for the impurity, genic potential. This (Q)SAR assessment should utilize models that
as shown in Fig. 2. This search identiﬁ ed a 5-strain Ames study by focus on “… bacterial mutagenicity predictions …” and the guideline
which the compound may be assigned to class 5 due to sufﬁ cient suggests the use of the two complementary methodologies: “expert
evidence for absence of mutagenicity in an adequately performed rule-based” and “statistical-based.” The guideline goes on to state
in vitro reverse mutation assay. that the “… (Q)SAR models … should follow the general validation
 Fig. 2. Example 1 showing the results of a database search.

 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24 17
principles set forth by the Organisation for Economic Co-operation and 4.2. Negative assessments and expert reviews
Development (OECD).” (OECD, 2007a)
 Commonly used statistical-based models include the Leadscope A regulatory evaluation of potentially mutagenic impurities
Genetox Statistical QSAR, CASE Ultra from MultiCASE, Inc., and should allow for the analysis of many compounds while main-
Sarah Nexus from Lhasa Limited and commonly used expert rule- taining a high degree of sensitivity. This can reasonably be achieved
based methodologies include the Leadscope Genetox Expert using negative predictions from two recommended (Q)SAR meth-
Alerts and Derek Nexus from Lhasa Limited. The most recent odologies for each compound, without the need for a detailed
version of each model is preferred for the (Q)SAR analysis; how- expert analysis, as long as the methodologies use an automated
ever, it is generally accepted that there are limited changes between domain assessment. If additional veriﬁ cation is desired, a rapid
different versions and that in practice there are few if any reported visual inspection of the results by the expert can be used to verify
changes in overall predictions, in particular of negative predictions that no valid alerts for mutagenicity with a plausible mechanism
being reversed. Recommendations for setting computational model were overlooked by the two (Q)SAR methodologies (Powley, 2015;
parameters have been provided by Stavitskaya et al., 2013. For Barber et al., 2015).
example (at the time of publication), with the Leadscope expert-
rule based methodology (Leadscope Model Applier: Genetox 4.2.1. Case study 2: clear negative prediction from two
Expert Alerts Suite), the domain assessment should be turned on, methodologies
and with the Leadscope statistical-based methodology (Leadscope In Fig. 3, the depicted impurity was automatically determined to
Model Applier: Genetox Statistical (Q)SAR Suite) probabilities /C21 0.6 be within the applicability domain of both the expert rule-based
set to positive, probabilities < 0.4 set to negative and the domain and the statistical-based models and negative predictions were
assessment turned on. generated by both methodologies. The statistical-based model
 (Q)SAR models adhering to OECD principles would ideally considered all atoms and bonds in the analysis (i.e., in this
generate the following prediction results that can be used directly modelling system, no atoms or bonds appear in black) as shown in
to assess the individual impurities: positive (predicted to be Fig. 3. A quick review of this information may be sufﬁ cient to
mutagenic) and negative (predicted to be non-mutagenic). How- conclude that the overall prediction for this impurity is non-
ever, there are a number of reasons why a (Q)SAR model does not mutagenic and it can be assigned to class 5.
always generate such a classiﬁ cation. The ﬁ rst reason is that the
system may determine that the impurity is out-of-domain, that is, it 4.2.2. Case study 3: refuted negative prediction from two
is incapable of making a prediction since the system does not methodologies
adequately cover the structural features of the impurity (OECD O-(2-Hydroxyethyl)hydroxylamine is shown in Fig. 4 and had a
validation principle #3). The second reason is that the prediction negative prediction for bacterial mutagenicity using both the
results may be categorized as equivocal or indeterminate due to expert rule-based and the statistical-based models. However, there
weak or conﬂ icting evidence, such that a deﬁ nitive prediction is conﬂ icting evidence for the mutagenic response of different hy-
cannot be made with adequate conﬁ dence. The third is where a droxylamine salts in the public domain. It was therefore concluded
prediction system is technically unable to process certain types of that a potential mutagenic response on the basis of the hydroxyl-
chemicals, such as for coordination compounds.2 amine moiety should be further evaluated. O-(2-Hydroxyethyl)
 hydroxylamine was submitted for Ames assay testing where it
 induced mutations in strain TA1535 in the absence of S9.
4. Considerations for an overall assessment and expert review
 When a negative prediction is made in only a single method-
4.1. Overview ology and an inconclusive prediction or an out-of-domain assign-
 ment made in the second methodology, it may be necessary to
 The ICH M7 guideline states that the “… absence of structural inspect the results in more detail before generating an overall
alerts …” from the two suggested (Q)SAR methodologies is sufﬁ - conclusion. Both situations are discussed in Sections 4.4 and 4.5.
cient to assign the impurity to class 5. Since any individual meth-
odology may generate results such as a positive prediction, a 4.3. Positive prediction and expert reviews
negative prediction, an inconclusive prediction, or an out-of-
domain assignment, it is important to consider how these indi- A positive prediction from either of the methodologies may lead
vidual results may be used to derive an overall mutagenic or non- to an overall positive prediction. Positive predictions may be
mutagenic assessment consistent with the language in the guide- refuted through an expert analysis, if appropriate. There are several
line. The ICH M7 guideline goes on to state that the results from the issues to consider when writing an expert review refuting positive
(Q)SAR methodologies may, if warranted, be examined further. This (Q)SAR results including the relevance of any alerting features or
expert review may provide “… additional supportive evidence on corresponding training set compounds, the ability of the chemical
relevance of any positive, negative, conﬂ icting or inconclusive pre- environment proximate to the alerting feature to mitigate the
diction and provide a rationale to support the ﬁ nal conclusion.” (ICH mutagenicity and information from chemical analogs (Powley,
M7, 2015a) This review has been shown to improve performance 2015; Stavitskaya et al., 2015; Barber et al., 2015). A positive
(Stavitskaya et al., 2015; Sutter et al., 2013) and provide a basis for assessment may be based on results from a single or multiple
refuting the (Q)SAR results (Powley, 2015; Stavitskaya et al., 2015;
Barber et al., 2015). The following sections outline a series of gen-
eral principles that describe (1) how an overall assessment may be
performed, (2) when an expert review may be provided, and (3)
what such an expert analysis may contain.
 2 A coordination complex or metal complex consists of a central atom or ion Fig. 3. Example 2. (For interpretation of the references to color in this ﬁ gure legend,
(generally metallic) and a surrounding array of bound molecules or ions. the reader is referred to the web version of this article.)

18 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24
 that the impurities would also be non-mutagenic and should be
 assigned to class 4.
 4.3.1.2. Case study 5: refuting a positive prediction based on an ICH
 Fig. 4. Example 3 (O-(2-Hydroxyethyl)hydroxylamine). M7 class 4 analysis. Example 5 (shown in Fig. 6) was predicted to be
 positive by both the expert rule-based and the statistical-based
 models as a result of the primary aromatic amine (highlighted in
models and each positive result should be individually evaluated as red). The most relevant and structurally similar analog in the rule-
the underlying reasons for the positive result may be different. The based alert system, and the only analog containing a similarly
following sections outline different points to be considered when substituted aniline as is present in the test structure, was experi-
refuting a positive prediction. mentally negative for bacterial mutation (shown in Fig. 6)(NTP,
 1980). It has been reported that the triﬂ uoromethyl groups in the
4.3.1. Shared alert with known negative (ICH M7 class 4) meta position to the amine are strongly deactivating for mutage-
 The ICH M7 guideline includes the following statement: “An nicity (Ahlberg et al., 2016). The most analogous structure from the
impurity with a structural alert that is shared (e.g., same structural alert in the statistical model was run in a different rule-based
alert in the same position and chemical environment) with the drug system and was predicted negative, since it contains a strong
substance or related compounds can be considered as non-mutagenic deactivating group. Example 5 is also fully contained within the
… if the testing of such material in the bacterial mutagenicity assay drug substance, for which the GLP Ames assay was negative. The
was negative.” (ICH M7, 2015a) The ﬁ rst step is to identify theweight-of-evidence suggests that it is unlikely to be mutagenic and
structural basis for the impurity's (Q)SAR result (from the matched was therefore assigned to class 4. This expert review has been
expert rule and/or the statistical-based model(s)). Next, a related reviewed and accepted by a regulatory authority.
compound with negative Ames data (such as the Active Pharma-
ceutical Ingredient or API, or another related impurity) is identiﬁ ed 4.3.2. An explanation of the mechanism
that also contains the same highlighted structural features (“known A positive prediction is triggered by an alert or a signiﬁ cant
negative”). The following questions may then be asked: statistical-based model feature that is present in the impurity. This
 fragment's associated mutagenic potential may be based on a
 /C15 Are there any additional structural alerts present in the impurity reasonable mechanistic rationale and/or there may be sufﬁ cient
 that are not present in the known negative comparator com- positive examples matching the fragment; however, the environ-
 pound? If so, it may not be possible to completely refute the ment around the alerting moiety within this speciﬁ c impurity may
 positive (Q)SAR result and apply the class 4 argument. preclude reaction at this site. It is possible to construct an expert
 /C15 Is the alert in the same chemical environment in the impurity as review to refute the prediction (Powley, 2015; Stavitskaya et al.,
 in the comparator compound? Chemical reactivity of an alerting 2015; Barber et al., 2015). In situations where a compound is pre-
 moiety may be mitigated by the presence of another feature in dicted negative by an expert rule-based methodology, yet predicted
 both molecules. Factors to consider in this comparison include positive by a statistical-based methodology, it may be helpful to
 (1) differences in the electron charge density (i.e. electron rich understand why the compound containing any highlighted group is
 or electron deﬁ cient) around the speciﬁ c alerting structure, (2) not positive in the expert rule-based system. Does the alert deﬁ -
 the steric environment proximal to the alerting structure, (3) the nition contain any exceptions to the rule?
 solubility or (4) the size or shape of the impurity.
 4.3.2.1. Case study 6: refuting a positive prediction based on a
 mechanism analysis. In Example 6, the potential impurity was
4.3.1.1. Case study 4: refuting a positive prediction based on an ICH predicted to be positive by the statistical-based model but negative
M7 class 4 analysis. Fig. 5 represents a series of similar impurities by the expert rule-based model. As shown in Fig. 7, the main
that were predicted to be positive in the statistical-based model. contribution to the positive prediction by the statistical-based
The common features responsible for the positive prediction are model was the feature highlighted in red. In reviewing the com-
summarized and highlighted in red in Fig. 5. 5-strain GLP Ames data pounds supporting the alerting fragment, it was found that the
conducted according to OEDC 471 and ICH S2(R1) guidelines were alerting fragment was highly inﬂ uenced by the mutagenicity data
generated for one structure (known negative) and were applied to on alkyl sulfonate esters, dialkyl sulfates, or sultones (see Fig. 8),
other impurities where R, R1, R2 or R3 varied but without addi- which are known alerts for mutagenicity (Ashby and Tennant,
tional alerting functionality (shown in Fig. 5). The impurities in case 1988; Benigni and Bossa, 2008). Example 6 is a mono-alkyl sul-
study 4 are considered analogs of the known negative compound fate esters; these are consistently negative in the Ames assay
and all share the same highlighted positive structural features. The (OECD, 2007b) and are not alkylating agents. Mono-alkyl sulfate
known negative comparator in combination with negative pre- esters are negatively charged at physiological pH and therefore are
dictions in the expert rule-based model was sufﬁ cient to predict less electrophilic than their alkyl sulfonate counterparts. The
 mono-alkyl sulfate esters in the training set were also non-
Fig. 5. A series of chemicals all predicted to be positive in a statistical-based model
based on the feature highlighted in red. (For interpretation of the references to color in Fig. 6. Example 5 and analog. (For interpretation of the references to color in this
this ﬁ gure legend, the reader is referred to the web version of this article.) ﬁ gure legend, the reader is referred to the web version of this article.)

 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24 19
 interaction between a test material containing an acid halide or
 sulfonyl halide and DMSO in the Ames test (Amberg et al., 2015).
 4.3.3.1. Case study 7: refuting a positive prediction based on a
 mechanism and coincidental features. In this example the potential
 impurity was predicted to be positive by the statistical-based
Fig. 7. Example 6 predicted to be positive in the statistical-based methodology, pri- model but negative by the expert rule-based model. Example 7 is
marily based on the feature highlighted in red. (For interpretation of the references to an N-oxide of a non-aromatic amine bearing a phenyl/aryl group.
color in this ﬁ gure legend, the reader is referred to the web version of this article.) The major contributing features are highlighted in red in Fig. 9.
 Firstly, the training sets inadequately represent N-oxide of a non-
 aromatic amine bearing a phenyl/aryl group, whose predicted
 mutagenic activity was inﬂ uenced by other co-occurring alerting
 features. Secondly, the literature indicates that the tertiary alkyl
 amine N-oxides are non-mutagenic. Finally, a structural analysis
 was performed for mutagenicity on nitrogenous aryl compounds
 and their corresponding N-oxides using TRPþ reversion in E. coli
 (Pai et al., 1978). This structural analysis included 10 tertiary aryl
 Fig. 8. Structural deﬁ nitions for alkyl sulfonate esters, dialkyl sulfates, and sultones. amines and their corresponding N-oxides. As part of the weight of
mutagenic except for the mono-alkyl sulfate esters of known the evidence, it was concluded that primary aromatic amines and
mutagenic poly aromatic hydrocarbons such as benz(a)anthracene their corresponding hydroxylamines, and N-hydroxycarbamates
and chrysene. Therefore, Example 6 is predicted to be non-were mutagenic, but not the tertiary aryl amines or their corre-
mutagenic. sponding tertiary N-oxides, as shown in Fig. 10 (Pai et al., 1978).
 Hypothetically, dealkylation of the amine to yield a primary aro-
 matic amine is a potential mechanism of mutagenicity; however, in
4.3.3. The relevance of features from statistical-based case study 7, this would yield aniline, which is known to lack
methodologies mutagenic potential. The lack of mutagenicity following deal-
 A positive prediction from a statistical-based model may be kylation is further supported by the observation that the parent
refuted if the structural features that are the basis for the “alert” in drug substance (API) structure contains the corresponding primary
the model (“positive contributing features”) are not relevant, aromatic amine and was negative in the bacterial reverse mutation
illustrated as follows: assay. Based on analysis of the training sets, a negative expert rule-
 based prediction, literature analysis for tertiary amine N-oxides,
 /C15 Coincidental features: Structural features are identiﬁ edand its structural similarity to the drug substance, Example 7 is
 through machine-learning when building statistical-based predicted to be non-mutagenic.
 models. Positive features are identiﬁ ed when present in a
 group of predominantly mutagenic training set compounds. 4.3.3.2. Case study 8: refuting a positive prediction based on coinci-
 However, these mutagenic compounds could also contain other dental features. Example 8 is shown in Fig. 11 and was predicted to
 structural features that better represent the actual moiety be positive by the statistical-based model and negative by the
 responsible for the observed mutagenicity. In these cases, the expert rule-based model. An expert review of Example 8 described
 statistical model has identiﬁ ed coincidental features. If the below concluded that the probability of mutagenicity is low based
 positive prediction was based primarily upon these coincidental on a review of the training set and a comparison with the drug
 features then an expert analysis refuting the prediction may be substance, which was negative in the bacterial reverse mutation
 made (Powley, 2015; Barber et al., 2015). One example of such a assay. The most relevant model features were evaluated and found
 situation is where an amine oxide is ﬂ agged in a set of aromatic to contain examples of another alert more likely to be responsible
 nitro compounds. for the positive prediction (see supplemental material for more
 /C15 Mitigating features: A positive prediction may be refuted if the details). These features included a planar anthracene-like tricyclic
 positive model features are mitigated by negative features aromatic core; however, the polycyclic core of Example 8 is puck-
 present at or proximal to the same reaction center. ered, due to the presence of sp3 carbon atoms, with CeH bonds
 /C15 Limited training set examples: It is possible that a positive almost orthogonal to either plane deﬁ ned by any two fused rings,
 model feature was derived from a small number of examples. An hence making the structure non-planar. Example 8 was therefore
 expert analysis may refute a positive prediction made primarily predicted to be non-mutagenic.
 using such features.
 /C15 No signiﬁ cant positive model features: The positive prediction 4.3.3.3. Expert reviews based on chemical analogs from public or in-
 may result from very small contributions from many unrelated house sources. Experimental Ames data for structural analogs can
 or unconnected positive model features.
 /C15 Irrelevant training set examples: It is possible that a positive
 model feature was derived from a set of compounds covering
 multiple structural classes. It is also possible that some of these
 structural classes do not apply to the speciﬁ c impurity (they are
 part of a different chemical series) and an expert review to
 refute the positive prediction may be an option if the impurity is
 within one of the non-mutagenic chemical classes.
 /C15 Underlying data are incorrect or not adequate: It may be Fig. 9. Example 7 with features contributing to the positive prediction highlighted in
 possible to identify model features based on data that are not red. (For interpretation of the references to color in this ﬁ gure legend, the reader is
 correct as a result of certain experimental conditions, such as an referred to the web version of this article.)

20 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24
Fig. 10. Structural deﬁ nitions for primary aromatic amines and their corresponding hydroxylamines, and N-hydroxycarbamates as well as tertiary aryl amines and corresponding
tertiary N-oxides.
 /C15 Visual inspection by an expert: One approach to assess
 inconclusive predictions is for a chemist or toxicologist to
 visually inspect the results to verify there are no valid alerts for
 mutagenicity with a plausible mechanism. For example, the
 chemist or toxicologist who is visually inspecting the results
 may have knowledge of mutagenicity alerts and/or mechanisms
 Fig. 11. Example 8. derived from proprietary data not built into the (Q)SAR models.
 It may be important to consider portions of the molecule (e.g.
also be used when the training sets do not contain suitablefunctional groups) not represented in the (Q)SAR models. Sys-
numbers of related structures (Powley, 2015; Barber et al., 2015; tematic substructural searching of functional groups not
Stavitskaya et al., 2015). Sufﬁ ciently similar analogs from the liter- considered by the models may also support the identiﬁ cation of
ature, public databases or in-house information may be used to features that are positively associated with bacterial mutage-
provide justiﬁ cation for refuting a positive or overruling annicity data (i.e. there is a statistically signiﬁ cantly greater
inconclusive prediction. The number of analogs and the degree of number of positive examples than would be expected by
structural similarity needs to be assessed on a case-by-case basis chance). The expert may also consider whether the structural
(Powley, 2015). features highlighted by the statistical-based models show sig-
 niﬁ cant association with bacterial mutagenicity.
 /C15 Strength of a single prediction: Where only a single method-
4.3.3.4. Case study 9: refuting a positive prediction using data from ology has generated a prediction, an assessment of the strength
chemical analogs. Example 9 was predicted to be positive by the of this prediction may be made to determine whether it is suf-
statistical-based methodology and negative by the expert rule- ﬁ cient as the basis of an overall conclusion.
based methodology. A database search identiﬁ ed a number of
close analogs as shown in Fig. 12. All analogs were experimentally
non-mutagenic in the Ames assay. The extension of the carbon side 4.4.1. Case study 10: assessing an inconclusive prediction using the
chain of Example 9 should not increase its reactivity compared to literature
the analogs. Example 9 is therefore predicted to be non-mutagenic. Example 10 (Fig. 13) was predicted to be negative by the expert
This compound, in fact, has been shown to be experimentally rule-based methodology and inconclusive by the statistical-based
negative for bacterial mutagenicity (Carmellino, 1993). This methodology; in the latter the most signiﬁ cant contribution was
example is used to illustrate the concept of an analog search and, as from the primary aromatic amine. As discussed in Ahlberg et al.
part of this analysis, it is necessary to assess the adequacy of the (2016), primary aromatic amines are mutagenic only in the pres-
underlying Ames data. ence of an activating functional group. Both functional groups (the
 bromo group in the para position and the carboxylate in the ortho
4.4. Expert reviews for inconclusive (Q)SAR results position) are not activating according to Ahlberg et al. (2016) (based
 on an analysis of primary aromatic amine data from public and
 Inconclusive predictions are generated when there is not proprietary databases) and therefore Example 10 was predicted to
enough evidence to make a mutagenic or non-mutagenic predic- be non-mutagenic. This compound has been tested in a standard
tion with adequate conﬁ dence. In general, all approaches discussed Ames assay using 5 strains and is non-mutagenic (Greene et al.,
earlier to refute a positive or negative prediction can reasonably be 2015).
applied to an inconclusive prediction for a covered (i.e., within the
applicability domain) compound in an attempt to resolve the pre- 4.4.2. Case study 11: assessing an inconclusive prediction using
diction and generate a negative or positive overall conclusion. The analogs
following outlines several potential approaches to assessing the Example 11 (shown in Fig. 14) was predicted to be negative by
results as part of an expert review to reach a conclusion that the the expert rule-based model and inconclusive by the statistical-
impurity is likely mutagenic or non-mutagenic. based model. Since Example 11 contains a hydrazine substructure
 and speciﬁ c classes of hydrazines are known to be mutagenic, an
 analysis based on the evaluation of published Ames assay data for
 Fig. 12. Example 9 with analogs. Fig. 13. Example 10.

 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24 21
 The mechanism of mutagenicity associated with aromatic amines
 requires oxidation by cytochrome P450 to hydroxylamines and
 then further activation by O-acylation. The O-acylated N-arylhy-
 droxylamine is converted to a highly electrophilic nitrenium ion,
 which then reacts with DNA (Benigni and Bossa, 2011). Aromatic
 amines found within pharmaceutical intermediates are more likely
 to be negative in the bacterial mutation assay than those that have
 data available in the public literature, according to an analysis of in-
 house databases (McCarren et al., 2011). This has been attributed to
 Fig. 14. Example 11 with analog. the bias towards larger molecular weight compounds in drug
 development with increased steric hindrance to formation of the
 reactive mutagenic metabolite or decreased ability of the metabo-
structural analogs was performed. This assessment led to the lite to cross bacterial cell walls (Glende et al., 2002; Hatch et al.,
identiﬁ cation of numerous structural analogs tested in the Ames 2001; Benigni, 2005). For example, it has been reported that the
assay including the analog shown in Fig. 14 that was reported to be addition of bulky alkyl groups away from the amino group changes
mutagenic. Hence, Example 11 was predicted to be mutagenic. a mutagenic aromatic amine to a non-mutagenic species (Glende
 et al., 2002). Hydrolysis or metabolism to generate a small aro-
4.5. Expert reviews for “out of domain” statements matic amine that may be mutagenic is not possible in Example 12.
 Therefore, Example 12 is predicted to be non-mutagenic due to the
 When an impurity is presented to a model that is sufﬁ ciently size of the compound which results in a potential lower bioavail-
different from the types of chemicals used in the reference/training ability, and inhibited formation of the putative reactive nitrenium
set, the model should not make a prediction, in accordance with metabolite.
OECD validation principle #3. These out-of-domain results, how-
ever, may also be assessed as part of an expert review. As back- 4.5.2. Case study 13: assessing an out-of-domain prediction using a
ground to the analysis, it may be helpful to understand why the similar analog
model was unable to make a prediction for this speciﬁ c impurity. In Example 13 was out-of-domain by the statistical-based models
a similar manner to inconclusive results that were discussed earlier, and predicted to be negative by the expert rule-based model.
it may be possible to generate an expert review for an out-of- Example 13 (shown in Fig. 16) is very similar to the drug substance,
domain result based on: (1) a visual inspection by an expertwhich was also out-of-domain for the statistical based models. The
chemist or toxicologist, (2) an assessment of the strength of a change in position was concluded not to change the potential for
prediction by a single methodology, (3) an understanding of rele- mutagenic reactivity, since there were no alerting features on the
vant mutagenic mechanisms, and (4) data for structural analogs. drug substance or the impurity (based on the expert rule-based
Another approach that may be helpful in assessing this type of model). Therefore, based on its structural similarity to the drug
result is to investigate whether the out-of-domain result is attrib- substance (which was negative for bacterial mutagenicity in the
utable to the addition of a non-reactive group. The ﬁ rst step as part Ames assay), Example 13 was predicted to be non-mutagenic.
of this assessment is to determine if there are any similar chemicals
that were predicted negative or where there is a negative experi- Case study 13 illustrates an expert analysis based on a change of
mental result. If the only difference from the out-of-domain a substituent position. Changes in the position of heteroatoms
structure is the addition of a non-reactive group (e.g. an amine within the ring can also be important to consider. For example, 3-
protected by two tert-butoxycarbonyl (Boc) groups or other non- Aminoisoxazole is non-mutagenic and 5-amino-4-chloro-3-
alerting fragment) and as long as this group could not cause an methylisoxazole is mutagenic, as shown in Fig. 17. Both are exam-
additional functional group to become an activated alert, then this ples of primary aromatic amines, where the aromatic system is a 5-
scenario may be used to address an out-of-domain situation. membered heterocycle and both rings contain a single nitrogen and
 Running another model is also an option to address an out-of- oxygen; however, the position of these heteroatoms is different in
domain or indeterminate (Q)SAR prediction; however, it should the two compounds relative to the primary aromatic amine. These
be noted that running a third model is not required by ICH M7. compounds, along with an analysis of the structure-activity rela-
Similar to the ﬁ rst two models, the third model should also follow tionship, are discussed in Ahlberg et al. (2016).
the OECD (20 07a) (Q)SAR validation principles to ensure that one is
simply not running models until one with a less stringent appli- 4.5.3. Case study 14: assessing an out-of-domain prediction using
cability domain calculation is found. public analogs
 Aminoacetonitrile (Example 14) was out-of-domain for the
4.5.1. Case study 12: assessing an out-of-domain response based on statistical-based models and predicted to be negative by the expert
the mechanism rule-based model. No standardized Ames testing has been per-
 Example 12 is a large compound containing greater than 30 formed with aminoacetonitrile. However, data from structurally
non-hydrogen atoms (Fig. 15). Example 12 was determined to be
out-of-domain by the statistical-based model. The example also
contains an aromatic amine moiety which is structurally alerting.
 Fig. 15. Example 12 (>30 non-hydrogen atoms). Fig. 16. Example 13 alongside the drug substance which is negative in the Ames assay.

22 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24
 pragmatic approach would be to either perform an Ames test or
 assign the impurity to class 3. However, in situations when no
 experimental data are generated, expert knowledge could be used
 to supersede even these predicted outcomes, with the caveat that it
 should include justiﬁ able scientiﬁ c evidence for regulatory
 acceptance.
 Fig. 17. Examples of how the position of heteroatoms may inﬂ uence mutagenicity. 4.5.5. Case study 16: assessing an out-of-domain result from two
 methodologies
 Example 16 (shown in Fig. 20) was concluded to be out-of-
 domain by both the expert rule-based and the statistical-based
 models as a result of the novelty of the R-group. Example 16 is
 similar to the drug substance; the only difference is that the pri-
 mary amine group of the drug substance has been has been con-
 verted to the bis-boc imide, shown in Fig. 20, through Boc
 protection of the primary amine. The drug substance is also
 concluded to be out-of-domain by both (Q)SAR methodologies;
 however, it has been tested and is non-mutagenic in the standard
 5-strain Ames test. Since there is no expected reactivity from the
 bis-boc functionality, Example 16 is predicted to be non-mutagenic
 (which was conﬁ rmed experimentally in a standard 5 strain Ames
 assay). As in Case Study 14, given the bis-boc protection serves to
 reduce reactivity, it could be reasonable to classify this as a non-
 mutagenic compound despite the lack of predictions in both
 methodologies.
 5. Reporting
 Fig. 18. Example 14 with analogs (including Ames results). The ﬁ nal report may include a description of the methodologies
similar compounds suggest that it is non-mutagenic (see Fig. 18). used, a summary of the results along with any expert reviews that
This included 3-aminopropionitrile (Analog 2) that was tested should be transparent and “include supporting information to arrive
negative in TA98, TA100, TA1535, TA1537, and TA1538 with and at the overall conclusion for Class 4 and Class 5 impurities” (ICH M7,
without metabolic activation. (CCRIS 3-aminopropionitrile) The 2015a). The selection of the impurities to be reported is depen-
single analog that was mutagenic (3-chloropropionitrile) contains dent on the stage of development, as shown in Table 3, which
an additional alerting structure (monofunctional alkyl chloride) not presents a summary from the ICH M7 guideline.
shared with aminoacetonitrile. In addition to these nearest neigh- The following elements may be included in the report of a
bors, aminoacetonitrile is also structurally similar to cyanamide (Q)SAR assessment consistent with ICH M7 with the level of detail
which is also non-mutagenic in a 5-strain Ames assay with E. coli dependent on the stage of development:
(FIOSH, 2014). Therefore, Example 14 was predicted to be non-
mutagenic. 1. Materials and methods
 /C15 Software, models and databases used, along with version
 numbers and parameters set
4.5.4. Case study 15: assessing an out-of-domain based on the 2. Summary of the results and conclusions
addition of a non-reactive group /C15 Chemical structure of the impurity that may include high-
 Example 15 is the Boc protected form of Compound Y (shown in lighting to illustrate what the software has identiﬁ ed as
Fig. 19). Example 15 was predicted to be negative by the expert rule- structural features associated with or not associated with
based methodology but out-of-domain for the statistical-based positive bacterial mutagenicity data (when this highlighting
methodology. Compound Y was predicted to be negative in the can be generated automatically by the system)
statistical-based methodology. Boc protection is used to prevent /C15 Experimental data and/or (Q)SAR results from both method-
chemical reactivity of the secondary amine and can be cleaved under ologies (the experimental and (Q)SAR results may be in
acidic conditions (Schelhass and Waldmann, 1996). Therefore, different tables or sections)
Example 15 is also not predicted to be mutagenic given its similarity /C15 Overall conclusion based on the prediction results and any
and reduced chemical reactivity compared to Compound Y. expert review (i.e., mutagenic or non-mutagenic) along with
 class 1e5 assignment
 Situations can arise where it is not possible to generate a (Q)SAR
prediction with either methodology due to the impurity being out-
of-domain, or both methodologies returning inconclusive pre-
dictions. When no model is able to generate a prediction, a
 Fig. 19. Example 15 with analog Y (predicted negative). Fig. 20. Example 16 alongside the non-mutagenic drug substance.

 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24 23
Table 3
Reporting requirements at each development phase.
 Development phase Reporting requirements
 Phase 1 clinical trials of 14 days or less Class 1 and class 2 impurities; cohorts of concern
 Phase 1 clinical greater than 14 days or Phase Class 1, class 2 and class 3 impurities; cohorts of concern
 2a clinical trials
 Phase 2b clinical trials or Phase 3 clinical List of actual/potential impurities assessed by (Q)SAR, Class 1, class 2 and class 3 impurities; plan for control, bacterial
 trials mutagenicity test results.
 Common Technical Document (Marketing List of actual/potential impurities assessed by (Q)SAR, Class 1, class 2, class 3, class 4, and class 5 impurities; supporting
 Application) information, plan for control, bacterial mutagenicity study reports.
 /C15 Summary of any supporting expert reviews or remarks Appendix A. Supplementary data
 3. Supporting information
 /C15 Expert review(s) supporting or refuting the (Q)SAR result, Supplementary data related to this article can be found at http://
 along with examples and references to illustrate dx.doi.org/10.1016/j.yrtph.2016.02.004.
 4. References, especially those used to support an expert review, if
 applicable
 5. Appendices References
 /C15 Complete bacterial mutagenicity study reports at the time of Ahlberg, E., Amberg, A., Beilke, L.D., Bower, D., Cross, K.P., Custer, L., Dobo, K.,
 marketing application may be included in the appendices or Ford, K.A., VanGompel, J., Harvey, J., Honma, M., Jolly, R., Joossens, E., Kemper, R.,
 cross-referenced or hyperlinked from another section Kenyon, M., Kruhlak, N., Kuhnke, L., Leavitt, ., Neilan, C., Naven, R., Quigley, D.P.,
 Shuey, D., Spirkl, H.P., Stavitskaya, L., Teasdale, A., White, A., Wichard, J.,
 When models are used that are not familiar to regulatory Zwickl, C., Myatt, G.J., 2016. Extending (Q)SARs to incorporate proprietary
 knowledge for regulatory purposes: a case study using aromatic amine muta-
agencies, it will be necessary to provide additional documentation genicity. Regul. Toxicol. Pharmacol. 77, 1e12.
showing how these models are consistent with the OECD (Q)SAR Amberg, A., Harvey, J.S., Czich, A., Spirkl, H.-P., Robinson, S., White, A., Elder, D.P.,
validation principles (OECD, 2007a). The supplemental information 2015. Do Carboxylic/Sulfonic acid halides really present a mutagenic and
 carcinogenic risk as impurities in ﬁ nal drug products? Org. Process Res. Dev. 19,
contains an example of a regulatory submission using examples 1495e15 0 6.
described above. Anderson, C., Copestake, P., Robinson, L., 2000. A specialist toxicity database
 (TRACE) is more effective than its larger, commercially available counterparts.
 Toxicology 151, 37e43.
6. Conclusions Ashby, J., Tennant, R.W., 1988. Chemical structure, Salmonella mutagenicity and
 extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222
 The ICH M7 guideline provides a framework for assessing DNA chemicals tested in rodents by the U.S. NCI/NTP. Mutat. Research Genetic Tox-
 icol. 204, 17e11 5.
reactive impurities and describes how these impurities may be ATSDR, 2015. http://www.atsdr.cdc.gov/.
controlled. This framework is currently being implemented across Barber, C., Amberg, A., Custer, L., Dobo, K.L., Glowienke, S., Gompel, J.V., Gutsell, S.,
the pharmaceutical industry and international regulatory agencies. Harvey, J., Honma, M., Kenyon, M.O., Kruhlak, N., Muster, W., Stavitskaya, L.,
An important component of this guideline is the use of (Q)SAR as an Teasdale, A., Vessey, J., Wichard, J., 2015. Establishing best practise in the
 application of expert review of mutagenicity under ICH M7. Regul. Toxicol.
alternative to conventional testing for the assessment of the Pharmacol. 73, 367e377.
mutagenic potential of drug substance impurities. (Q)SAR models Benigni, R., 2005. Structure/C0 activity relationship studies of chemical mutagens and
represent a state-of-the art approach to predicting mutagenicity carcinogens: mechanistic investigations and prediction approaches. Chem. Rev.
 105, 1767e18 0 0.
that balances the need for high-throughput while maximizing pa- Benigni, R., Bossa, C., Richard, A.M., Yang, C., 2008. A novel approach: chemical
tient safety. relational databases, and the role of the ISSCAN database on assessing chemical
 This paper has outlined a number of practical principles and carcinogenicity. Ann. Ist. Super. Sanita 44, 48e56.
 Benigni, R., Bossa, C., 2008. Structure alerts for carcinogenicity, and the Salmonella
procedures to be considered when conducting a (Q)SAR analysis assay system: a novel insight through the chemical relational databases tech-
consistent with the ICH M7 guideline. This includes, in the absence nology. Mutat. Res. 659, 248e261.
of adequate experimental data, how to combine the results from Benigni, R., Bossa, C., 2011. Mechanisms of chemical carcinogenicity and mutage-
 nicity: a review with implications for predictive toxicology. Chem. Rev. 111,
the recommended (Q)SAR models, when to consider generation of 2507e2536.
a detailed expert review, and what such a review may contain. The BIBRA, 2015. http://www.bibra-information.co.uk/trace-unique-chemical-database/.
contents of a full report for inclusion as part of a regulatory sub- Carmellino, M.L., Massolini, G., Pagani, G., Zanif, F., Leri, G., 1993. Fungicidal and
mission have been outlined. Through adoption of common princi- genotoxic activity of new substituted pyridazinones. Framaca 48, 1427e1438.
 CCRIS, 2011. http://toxnet.nlm.nih.gov/newtoxnet/ccris.htm.
ples and procedures, the practical implementation of a (Q)SAR CPDB, 2011. http://toxnet.nlm.nih.gov/newtoxnet/cpdb.htm.
analysis consistent with the ICH M7 guideline will become more CCRIS, 3-aminopropionitrile. http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?
 dbsþccris:@termþ@rnþ151-18-8.
standardized, consistent, and transparent. Additionally, the gener- Diehl, M.S., Willaby, S.L., Snyder, R.D., 2000. Comparison of the results of a modiﬁ ed
ation and review of these reports should become more streamlined miniscreen and the standard bacterial reverse mutation assays. Environ. Mol.
over time for both pharmaceutical manufacturers and regulatory Mutagen. 36, 72e77.
agencies. Dobo, K.L., Greene, N., Fred, C., Glowienke, S., Harvey, J.S., Hasselgren, C., Jolly, R.,
 Kenyon, M.O., Munzner, J.B., Muster, W., Neft, R., Reddy, M.V., White, A.T.,
 Weiner, S., 2012. In silico methods combined with expert knowledge rule out
Acknowledgments mutagenic potential of pharmaceutical impurities: an industry survey. Regul.
 Toxicol. Pharmacol. 62, 449e455.
 DSSTox-Archive, 2012. ftp://ftp.epa.gov/dsstoxftp/DSSTox\_Archive\_20120411/.
 We acknowledge Maurizio Franzini for helping with some of the ECHA, 2015. http://echa.europa.eu/.
expert review examples. EMA, 2006. Guideline on the Limits of Genotoxic Impurities. CPMP/SWP/5199/02
 EMEA/CHMP/QWP/251344/2006. http://www.ema.europa.eu/ema/pages/
 includes/document/open\_document.jsp?webContentId¼WC500002903.
Transparency document EMA, 2010. Questions and Answers on the ‘Guideline on the Limits of Genotoxic
 Impurities’. http://www.ema.europa.eu/ema/pages/includes/document/open\_
 Supplementary data related to this article can be found at http:// document.jsp?webContentId¼WC500002907.
 ExPub, 2015. http://www.expub.com/index.html.
dx.doi.org/10.1016/j.yrtph.2016.02.004. FDA, 2008. Draft guidance for industry e genotoxic and carcinogenic impurities in

24 A. Amberg et al. / Regulatory Toxicology and Pharmacology 77 (2016) 13e24
 drug substances and products: recommended approaches. MultiCASE, 2015. http://www.multicase.com/case-ultra-models#bacterial\_muta\_
FIOSH, 2014 Federal Institute for Occupational Safety and Health, 2014. CLH Report: bundle.
 Proposal for Harmonised Classiﬁ cation and Labelling. Substance Name: Cyan- Naven, R.T., Greene, N., Williams, R.V., 2012. Latest advances in computational
 amide. Federal Ofﬁ ce for Chemicals, Dortmund, Germany. genotoxicity prediction. Expert Opin. Drug Metab. Toxicol. 8, 1579e1587.
GENE-TOX, 1998. http://toxnet.nlm.nih.gov/newtoxnet/genetox.htm. NTP, 1980. http://tools.niehs.nih.gov/cebs3/ntpViews/?activeTab¼summary&
Glende, C., Klein, M., Schmitt, H., Erdinger, L., Boche, G., 2002. Transformation of studyNumber¼758178.
 mutagenic aromatic amines into non-mutagenic species by alkyl substituents. NTP, 2015. http://tools.niehs.nih.gov/cebs3/ui/.
 Part II: alkylation far away from the amino function. Mutat. Res. 515, 15e38 OECD, 1997. Test 471. Bacterial Reverse Mutation Test OECD Guideline for Testing of
 [Research Support, Non-U.S. Gov't]. Chemicals Section 4 1997 July. http://www.oecd-ilibrary.org/environment/test-
Gold, L.S., Slone, T.H., Ames, B.N., Manley, N.B., Garﬁ nkel, G.B., Rohrbach, L., 1997. no-471-bacterial-reverse-mutation-test\_9789264071247-en.
 Chapter 1: carcinogenic potency database. In: Gold, L.S., Zeiger, E. (Eds.),OECD, Organisation for Economic Co-operation and Development, 2007a. Guidance
 Handbook of Carcinogenic Potency and Genotoxicity Databases. CRC Press, Boca Document on the Validation of (Quantitative) Structure-activity Relationships
 Raton, FL, pp. 1e605. [(Q)SAR] Models, OECD Environment Health and Safety Publications Series on
Gold, L.S., Manley, N.B., Slone, T.H., Ward, J.M., 2001. Compendium of chemical Testing and Assessment No. 69 (ENV/JM/MONO(2007)2).
 carcinogens by target organ: results of chronic bioassays in rats, mice, hamsters, OECD, Organisation for Economic Co-operation and Development, 2007b. SIDS
 dogs and monkeys. Toxicol. Pathol. 29, 639e652. Initial Assessment Report for SIAM 25. Category of Alkyl Sulfates, Alkane Sul-
Gold, L.S., Manley, N.B., Slone, T.H., Rohrbach, L., Garﬁ nkel, G.B., 2005. Supplement fonates and a-Oleﬁ n Sulfonates. 16e19 October 2007.
 to the carcinogenic potency database (CPDB): results of animal bioassays OECD 490, 2015. http://www.oecd-ilibrary.org/environment/test-no-490-in-vitro-
 published in the general literature through 1997 and by the national toxicology mammalian-cell-gene-mutation-tests-using-the-thymidine-kinase-gene\_
 program in 1997 and 1998. Toxicol. Sci. 85, 747e800. 9789264242241-en.
Greene, N., Dobo, K.L., Kenyon, M.O., Cheung, J., Munzner, J., Sobol, Z., Sluggett, G., Pai, V., Bloomﬁ eld, S.F., Jones, J., Gorrod, J.W., 1978. Mutagenicity testing of
 Zelesky, T., Sutter, A., Wichard, J., 2015. A practical application of two in silico nitrogenous compounds and their N-oxidised products using TRPþ reversion in
 systems for identiﬁ cation of potentially mutagenic impurities. Regul. Toxicol. E. Coli. In: Gorrod, J.W. (Ed.), Biological Oxidation of Nitrogen. Elsevier/North-
 Pharmacol. 72, 335e349. Holland Biomedical Press, pp. 375e383.
Hatch, F.T., Knize, M.G., Colvin, M.E., 2001. Extended quantitative structure-activity PAN, 2014. http://www.pesticideinfo.org/.
 relationships for 80 aromatic and heterocyclic amines: structural, electronic, Pharmapendium, 2015. http://www.pharmapendium.com/.
 and hydropathic factors affecting mutagenic potency. Environ. Mol. Mutagen. Powley, M.W., 2015. (Q)SAR assessments of potentially mutagenic impurities: a
 38, 268e291. regulatory perspective on the utility of expert knowledge and data submission.
IARC, 2015. http://Monographs.iarc.fr/ENG/Classiﬁ cation/index.php. Regul. Toxicol. Pharmacol. 71, 295e300.
ICH M7, 2015a. ICH M7 e assessment and control of DNA reactive (mutagenic) Robinson, L., Mcilwaine, I., Copestake, P., Anderson, C., 2000. Comparative evalua-
 impurities in pharmaceuticals to limit potential carcinogenic risk. http://www. tion of the performance of online databases in answering toxicology queries.
 fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ Int. J. Inf. Manag. 20, 79e87.
 UCM347725. RTECS, 2015. http://www.cdc.gov/niosh/rtecs/default.html.
ICH M7, 2015b. Addendum to ICH M7: assessment and control of DNA reactive Schelhaas, M., Waldmann, H., 1996. Protecting group strategies in organic synthesis.
 (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk: Angew. Chem. Int. Ed. Engl. 35, 2056e2083.
 application of the principles of the ICH M7 guideline to calculation ofStavitskaya, L., Aubrecht, J., Kruhlak, N.L., 2015. Chemical structure-based and tox-
 compound-speciﬁ c acceptable intakes M7 (R1). http://www.ich.org/ﬁ leadmin/ icogenomics models. In: Graziano, M.J., Jacobson-Kram, D. (Eds.), Genotoxicity
 Public\_Web\_Site/ICH\_Products/Guidelines/Multidisciplinary/M7/M7\_ and Carcinogenicity Testing in Pharmaceuticals. Springer, pp. 13e34.
 Addendum\_Step\_2.pdf. Stavitskaya, L., Minnier, B.L., Benz, R.D., Kruhlak, N.L., 2013. FDA Center for drug
INCHEM, 2015. http://www.inchem.org. evaluation and research,“Development of improved Salmonella mutagenicity
IRIS, 2015. http://www.epa.gov/iris/. QSAR models using structural ﬁ ngerprints of known toxicophores”. In: Society
JECDB, 2015. http://www.echemportal.org/echemportal/. of Toxicology, 52nd Annual Meeting. Retrieved from. https://www.leadscope.
Kasper, P., Müller, L., 2015. Genotoxic impurities in pharmaceuticals. In:com/media/SOT%202013%20Stavitskaya%20CDER%20poster.pdf.
 Graziano, M.J., Jacobson-Kram, D. (Eds.), Genotoxicity and Carcinogenicity Sutter, A., Amberg, A., Boyer, S., Brigo, A., Contrera, J.F., Custer, L.L., Dobo, K.L.,
 Testing in Pharmaceuticals. Springer, pp. 55e74. Gervais, V., Glowienke, S., Gompel, J.V., Greene, N., Muster, W., Nicolette, J.,
Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating Reddy, M.V., Thybaud, V., Vock, E., White, A.T., Müller, L., 2013. Use of in silico
 the quality of experimental toxicological and ecotoxicological data. Regul. systems and expert knowledge for structure-based assessment of potentially
 Toxicol. Pharmacol. 25, 1e5. mutagenic impurities. Regul. Toxicol. Pharmacol. 67, 39e52.
Kruhlak, N.L., Benz, R.D., Zhou, H., Colatsky, T.J., 2012. (Q)SAR modeling and safety Sweet, D.V., Anderson, V.P., Fang, J., 1999. An overview of the registry of toxic effects
 assessment in regulatory review. Clin Pharmacol Ther. 91, 529e534. of chemical substances (RTECS): critical information on chemical hazards.
Leadscope, 2015. http://www.leadscope.com/toxicity\_databases/. Chem. Health Saf. 6, 12e16.
McCarren, P., Springer, C., Whitehead, L., 2011. An investigation into pharmaceuti- Tennant, R.W., 1991. The genetic toxicity database of the national toxicology pro-
 cally relevant mutagenicity data and the inﬂ uence on Ames predictive poten- gram: evaluation of the relationships between genetic toxicity and carcinoge-
 tial. J. Cheminformatics 3, 51. nicity. Environ. Health Perspect. 96, 47e51.
Müller, L., Mauthe, R.J., Riley, C.M., Andino, M.M., Antonis, D.D., Beels, C.,ToxNet, 2015. http://toxnet.nlm.nih.gov/.
 Degeorge, J., Knaep, A.G.D., Ellison, D., Fagerland, J.A., Frank, R., Fritschel, B., VITIC, 2015. http://www.lhasalimited.org/products/vitic-nexus.htm.
 Galloway, S., Harpur, E., Humfrey, C.D., Jacks, A.S., Jagota, N., Mackinnon, J., Wexler, P., 2001. TOXNET: an evolving web resource for toxicology and environ-
 Mohan, G., Ness, D.K., O'Donovan, M.R., Smith, M.D., Vudathala, G., Yotti, L., mental health information. Toxicology 157, 3e10.
 2006. A rationale for determining, testing, and controlling speciﬁ c impurities in Young, R.R., 2002. Genetic toxicology: web resources. Toxicology 173, 103e121.
 pharmaceuticals that possess potential for genotoxicity. Regul. Toxicol. Phar- Zeiger, E., Risko, K.J., Margolin, B.H., 1985. Strategies to reduce the cost of muta-
 macol. 44, 198e211. genicity screening with the Salmonella assay. Environ. Mutagen. 7, 901e911.

Source PDF: /instem.com/www.instem.com/docs/sot-2021/Developing-structure-activity-relationships-for-N-nitrosamine-activity.pdf

 Contents lists available at ScienceDirect
 Computational Toxicology
 journal homepag e: www.sci encedirect.co m/journal/ computatio nal-toxicol ogy
Developing structure-activity relationships for N-nitrosamine activity
Kevin P. Cross a,\*, David J. Ponting b
aInstem, 1393 Dublin Road, Columbus, OH, USA
bLhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, UK
ARTICLE INFO ABSTRACT
Keywords: The detection of N-nitrosodimethylamine (NDMA) in several marketed drugs led regulatory agencies to require
Nitrosamine risk assessment that N-nitrosamine risk assessments be performed on all marketed medical products [EMA/351053/2019 rev 1
Nitrosamine structure-activity relationships (2019)]. Regulation of N-nitrosamine impurity levels in pharmaceutical drug substances and products is
Structural features supporting read-across for described in the ICH M7(R1) guideline where they are referred to as “cohort-of-concern” compounds as several
N-nitrosamines are potent rodent carcinogens [Kroes et. al. 2004]. EMA, U.S. FDA and other regulatory agencies have set pro-
Nitrosamine mutagenicity and carcinogenicity visional acceptable daily intake limits for N-nitrosamines calculated from rodent carcinogenicity TD values for
Computational toxicology 50
Regulation of nitrosamine impurities experimentally measured N-nitrosamines or the measured TD50 values of close analogs. The class-specific limit
 can be adjusted based upon a structure activity relationship analysis (SAR) and comparison with analogs having
 established carcinogenicity data [EMA/369136/2020, (2020)]. To investigate whether improvements in SARs
 can more accurately predict N-nitrosamine carcinogenic potency, an ad hoc workgroup of 23 companies and
 universities was established with the goals of addressing several scientific and regulatory issues including:
 reporting and review of N-nitrosamine mutagenicity and carcinogenicity reaction mechanisms, collection and
 review of available, public relevant experimental data, development of structure–activity relationships consistent
 with mechanisms for prediction of N-nitrosamine carcinogenic potency categories, and improved methods for
 calculating acceptable intake limits for N-nitrosamines based upon mechanistic analogs. Here we describe this
 collaboration and review our progress to date towards development of mechanistically based structure–activity
 relationships. We propose improving risk assessment of N-nitrosamines by first establishing the dominant re-
 action mechanism prior to retrieving an appropriate set of close analogs for use in read-across exercises.
1. Introduction limits for N-nitrosamines are calculated from compound-specific carci-
 nogenicity data by extrapolation of rodent TD50 values. For N-nitrosa-
 Recently N-nitrosodimethylamine (NDMA) has been detected in mines without carcinogenicity data, regulatory agencies established
several pharmaceutical marketed drugs. These events have led regula- provisional AI limits for several N-nitrosamine impurities based on
tory agencies to require that N-nitrosamine risk assessments be per- structure activity relationships (SARs) with “close” analogs [3–6].
formed on all marketed medical products [1]. The need for these Currently, these regulatory limits are based on the AIs for the highly
assessments is driven by the high carcinogenic potency of several N- potent animal carcinogens NDMA and N-nitrosodiethylamine (NDEA).
nitrosamines in rodents, thus making these substances a significant However, not all N-nitrosamines are highly potent (as measured by ro-
regulatory concern [2]. Management of N-nitrosamine impurity levels in dent TD50 values), and their carcinogenic potency have been shown to
pharmaceutical drug substances and products has previously been span over 4 log units of TD50 values [7,8]. Fortunately, the class-specific
guided by ICH M7 where they are referred to as “cohort-of-concern” limit can be adjusted based upon a SAR analysis as part of a comparison
(COC) compounds. Consequently, class-specific Acceptable Intake (AI) with other similar N-nitrosamines that have established carcinogenicity
 Abbreviations: AI, acceptable intake; ADME, absorption, distribution, metabolism, and elimination; CPDB, Carcinogenicity Potency Database; COC, cohort of
concern; EMA, European Medicines Agency; EWG, electron-withdrawing group; FDA, U.S. Food and Drug Administration; LCDB, Lhasa Carcinogenicity Database;
NDEA, N-nitrosodiethylamine; NDIPA, N-nitrosodiisopropylamine; NDMA, N-nitrosodimethylamine; NDSBA, N-nitrosdisecbutylamine; NMEA, N-nitro-
somethylethylamine; NMNA, N-nitrosomethylneopentylamine; NMIPA, N-nitrosomethylisopropylamine; NMTBA, N-nitrosomethyltertbutylamine; NPDA, N-nitro-
sodiphenylamine; SAR, structure activity relationship; TD50, dose that results in a 50% excess in tumor incidence.
 \*Corresponding author.
 E-mail addresses: kevin.cross@instem.com (K.P. Cross), david.ponting@lhasalimited.org (D.J. Ponting).
https://doi.org/10.1016/j.comtox.2021.100186
Received 30 June 2021; Received in revised form 12 August 2021; Accepted 6 September 2021

K.P. Cross and D.J. Ponting
data. The EMA Assessment Report on the subject [4] states, “ It is potency dialkyl N-nitrosamines (i.e., those with the lowest TD50 values
therefore prudent to consider all N-nitrosamines containing an available in the carcinogenicity database) is that of α-carbon hydroxyl-
α-hydrogen that can be metabolically activated as potentially mutagenic ation via metabolic activation, as indicated in Fig. 1 [9– 12]. It has been
and carcinogenic to humans, however with different potencies reported that multiple stages of this, including that marked as hetero-
depending on nature of the functional group, specifics of metabolic lysis, may be catalysed by the same P450 enzyme without relaxation of
activation and repair efficiency and capacity.” conformation – resulting in the loss of the R1-bearing side as a carbox-
 To investigate whether improvements in SARs can more effectively ylic acid as opposed to an aldehyde [13,14]; however, in other cases
predict N-nitrosamine carcinogenic potency, an ad hoc workgroup of 23 such as nitrosomorpholine, the reactive aldehyde intermediate is sig-
companies and universities was established to address several scientific nificant and trapped intramolecularly [15].
and regulatory issues. These include: For small dialkyl nitrosamines, the predominant enzyme responsible
 for the activation of the nitrosamine to intermediate I is reported to be
1) reporting and review of N-nitrosamine mutagenicity and carcino- Cytochrome P450 2E1 (Cyp 2E1) [12]; however, the active site of this
 genicity reaction mechanisms, specific isoform is particularly small, and a number of other P450 iso-
2) collection and review of available, public, relevant experimental forms may become involved for larger nitrosamines. Examples of
 carcinogenicity and mutagenicity data, particular relevance are: Cyp 2A6 – also relevant for small nitrosamines
3) development of SARs consistent with mechanisms for predicting N- [11,12,14]; Cyp 2C9 – substrates with an anionic site, and of specific
 nitrosamine carcinogenic potency categories, and orientation requirements [16,17]; 2C19 – Zwitterionic compounds [17];
4) improved methods for calculating AI limits for N-nitrosamines based 2D6 – cationic site [17] and Cyp 3A4 – which is able to metabolise
 upon mechanistic analogs. particularly large substrates [17].
 Many factors can contribute to nitrosamine carcinogenicity potency,
 Herein we describe the progress made towards development of including:
mechanistically based SARs, identifying the structural features that most
affect carcinogenic potency. Specifically: 1) α -carbon substitution, and a) the relevant P450 enzymes summarised above and their levels in
2) electron-withdrawing groups on nitrosamine carcinogenicity potency various target organs – which can vary between species and between
and mutagenicity prevalence. The features that impact a SAR of a individuals [11]
complex biological process such as carcinogenesis may include a num- b) compound solubility, size, and shape [18],
ber of different events. The key events driving DNA mutagenicity from c) potential phase II conjugation (such as carboxylic acid-containing
dialkyl N-nitrosamines include: metabolic activation, DNA alkylation compounds being substrates for e.g., glucuronidation directly),
and the repair of potential DNA adducts. While these events could d) the stability of intermediates such as carbocation and diazonium ion
potentially result in different SARs, the metabolic activation mechanism stability,
is understood [9,10] to be of principal concern for the overall SAR – e) DNA adduct profiles and the level of mutagenic adducts, and
since if a nitrosamine is not metabolically activated, the SAR for binding f) DNA repair mechanisms and their capacity levels.
and repair is relevant. A three-stage consideration of the SAR, however,
may be necessary in some cases to fully explain the potency of some There can also be competing metabolic activation mechanisms, such
dialkyl N-nitrosamines. as β -carbon [9,19], γ -carbon [19], and ω -carbon hydroxylation [9,19],
 as well as mechanisms such as denitrosation [20], and trans-nitrosation
 [21], which may be either metabolically mediated (in the case of deni-
1.1. Metabolic activation mechanisms for dialkyl N-nitrosamine trosation, potentially via the same radical intermediate as α -hydroxyl-
mutagenicity ation [22]) or not.
 This investigation will focus on identifying the structural charac-
 Given the significance of the metabolic activation in understanding teristics that affect dialkyl N-nitrosamines potency and how they may be
the overall SAR, current understanding is briefly summarized here. It used to determine the relative potency of these different nitrosamines.
has been reported [9,10] that several different competing metabolism
mechanisms primarily drive the potency for dialkyl N-nitrosamines,
with uninhibited metabolic activation via α -carbon hydroxylation pro-
ducing the most potent carcinogens. The mechanism for the highest
 Fig. 1. α-carbon hydroxylation of dialkyl N-nitrosamines.

K.P. Cross and D.J. Ponting
2. Material and methods Many of these, however, have no examples in the dataset and are thus
 unable to be considered.
2.1. Dataset curation
 3. Results
 Data was extracted from historic rodent carcinogenicity and muta-
genicity sources and curated according to the respective standard pro- The curation of carcinogenicity and Ames study data described
tocols by a number of separate data-gathering exercises – Lhasa resulted in a consensus dataset of 362 dialkyl N-nitrosamines. Of these,
Limited’ s Vitic (2020) [7,23], Instem’ s Leadscope Genetox and Carci- 208 have carcinogenicity data (including TD50[27,28] values for 74 of
nogenicity Databases (2020) [24] and the now-retired Carcinogenicity these) and 281 have Ames study data. Analysis of the concordance be-
Potency Database (CPDB) [25,26] available as the Lhasa Carcinogenic- tween these endpoints has been performed elsewhere [Trejo-Martin et
ity Database [LCDB, carcdb.lhasalimited.org]. Data extraction from al, manuscript in preparation, [7,29]], and is reported to be excellent.
CPDB/LCDB was performed in-house at Lhasa Limited from the source The reasons for the lack of a TD50 for many of the carcinogenicity re-
data, extracting all data for structures that match NN(III) = O and cords include principally that for 120 compounds a study exists in the
filtering via substructure patterns in Knime (www.rdkit.org, as imple- Lhasa and/or Instem dataset that was not incorporated in the CPDB and
mented in KNIME version 4.1.0, www.knime.org) to remove those 14 compounds for which at least one record exists in the CPDB, but no
structures that match the non-dialkyl compounds shown in Fig. 2. These TD50 was able to be determined by Gold et al (typically due to a negative
compound classes, such as nitrosoureas, nitrosamides and similar com- result in the study).
pounds are known to exert mutagenic and carcinogenic potential via
different mechanisms and have therefore been excluded from this 3.1. Categorizing nitrosamine potency by structural features
analysis. A similar approach was taken to the Vitic data, using the same
substructural features, and extracting all data from the ‘Carcinogenicity’ The analysis focused on extracting and developing chemistry-based
and ‘Genetic Toxicology – in Vitro’ tables; data from the latter was then knowledge by uncovering trends in the chemical feature-activity space
filtered to Ames test or synonyms only. Data extraction from the Lead- that are represented in the database. The objective is ultimately to
scope Genetox and Carcinogenicity Databases was similarly performed encode the expert, intellectual knowledge into alerts for identification of
in-house at Instem from the source data, extracting all data for structures carcinogenicity potency categories for compounds (based on rodent
that match NN(III) = O and filtering using Leadscope substructure TD values). As it is not the intent to develop statistical (Q)SAR models
search functionality. The latter was filtered to include compounds con- 50
taining Ames test data and carcinogenicity calls. Data from these three using these features, the number of observations is not as important as is
sources were curated together manually, creating a combined dataset the relevance of chemical features to known organic chemistry reactivity
with consensus calls for carcinogenicity and Ames test data. and functional group properties.
 A closer examination of the many structural features that can affect
 dialkyl N-nitrosamines is presented in Fig. 5. This figure shows a sum-
2.2. Choice of structural features mary of all the structural features investigated thus far. Many potential
 features had few observations and the presence of multiple substituents
 Exploratory investigations were performed using a subjective anal- per compound can sometimes complicate the analysis when carbon
ysis of TD50 [27,28] potency data from the LCDB previously described hydroxylation can potentially occur on either substituent. Since the
[7,29] using substructure patterns for features previously identified relative amount of 2-year rodent carcinogenicity bioassay data is low
[9,10]. Several distinct substructural categories were identified (see and there is little expectation of new data being generated, the potency
Fig. 5) and two were chosen to investigate in more depth: trends established from analysing the carcinogenicity data were
 corroborated by comparing the Ames mutagenicity data for prevalence
1) the degree of α -branching of the nitrosamine (Fig. 3) of positive and negative results with carcinogenicity potency trends.
2) the presence or absence of electron-withdrawing groups (Fig. 4). This comparison is supported by the high sensitivity of Ames study re-
 sults in predicting rodent carcinogenicity [7,29] and the fact that
2.3. Data analysis nitrosamine mutagenicity is observed to occur via alkylation at specific
 DNA base sites (e.g., O6-guanine [32]) in a mutagenic mechanism
 The structural categories described in Figs. 3– 5 were manually [9,10].
encoded into substructure patterns using the SMARTS notation [31], Based upon these considerations, the current investigation chose to
and pattern-matching was performed against the dataset described using initially analyse and report the steric effects of α -carbon substitution and
RDKit (www.rdkit.org, as implemented in KNIME version 4.1.0, www. electronic effects of β -carbon electron-withdrawing groups on nitrosa-
knime.org). Data analysis and visualisation was performed in python mine carcinogenicity potency and mutagenicity prevalence.
(www.python.org, version 3.7.6).
 The two alkyl substituents of the molecule were considered both 3.2. The effects of degree of α-carbon substitution on nitrosamine
separately and in combination (i.e., with R1 in Fig. 3– 5 either kept as “ C carcinogenicity potency and mutagenicity prevalence
except C = O, C = N” or explicitly defined, respectively), and thus an
exponentially large number of potential feature combinations exist. The first category investigated is the degree of α -branching of the
 Fig. 2. Definitions of nitros(o)amide, nitrosourea and similar compounds.

K.P. Cross and D.J. Ponting
 Fig. 3. Visualisations of substructure patterns considered for identification of the degree of α -carbon branching.
 Fig. 4. Definitions of electron-withdrawing group patterns categorised by strength (as defined by strength of the withdrawing group [30].
nitrosamine, which has historically been reported [9,10] to have a sig- increasing chain length and ring size (though there are some notable
nificant impact on potency – indeed, dialkyl nitrosamines lacking any exceptions to this trend). Lastly, the “ No a-CH2” plot is of particular
α -carbon hydrogens are indicated by the European Medicines Agency interest. There are two compounds in this category with TD50 values;
(EMA) to be of lower concern [4]. Fig. 3 gives the structural definitions firstly, 2,6-dimethyl-N,N’ -dinitrosopiperazine contains both a
used to identify these classes. substituted and unsubstituted nitrosamine, and thus matches the sub-
 While much of the literature on nitrosamines has concentrated on structure pattern for having two isopropyl groups. However, it also has a
experiments measuring NDMA and NDEA potency, Fig. 6 shows that reactive, unsubstituted nitrosamine that is the probable source of
these small nitrosamines constitute a very potent but limited nitrosa- mutagenesis and carcinogenesis – and hence this compound is worthy of
mine set with a tight TD50 value range. Larger nitrosamines, such as inclusion in the cohort-of-concern and matches both the “ No a-CH2” (at
those for drug-like compounds, have TD50 ranges spanning 4 orders of one nitrosamine substitution site) and “ Cyclic a-CH2 (at the other).
magnitude and containing examples of compounds with much lower Secondly, nitrosodiphenylamine, which is the weakest carcinogen in the

Source PDF: /instem.com/www.instem.com/legal/IISMS.pdf

 Instem Information Security Management System
 INFORMATION SECURITY POLICY
 Intended Outcome of the system
 The intended outcome of our Information Security Management System (ISMS) is to protect information
 assets to an appropriate level of the following: -
 Confidentiality: Ensuring the information is accessible only to those who
 are authorised to have access
 Integrity: Safeguarding the accuracy and completeness of
 information and processing methods
 Availability: Ensuring access when required
 Our Promise
 Instem undertakes to keep safe the information that it receives and holds for its customers, staff and other
 stakeholders. We will only make such information available to those that need to see it and we will strive to
 ensure that all the information that we keep is necessary, complete and accurate.
 Objectives
 It is the objective of this policy and the supporting system to minimise undesired effects by identifying,
 reducing or preventing the impact of internal and external threats and vulnerabilities and to ensure:
  that business, regulatory, legislative and other information security requirements are understood
 and met;
  that we identify measurable Information Security objectives that we use to monitor and drive
 improvement;
  that the integrity of our ISMS is maintained when changes are planned and implemented, and we
 remain vigilant in an environment of constantly evolving threats;
  that all our people are aware, trained and competent in fulfilling their contribution to protect our
 information;
 The ISMS is implemented in a manner that ensures that our widely geographically dispersed business can
 still operate with world-leading efficiency and effectiveness.
 Continuous Improvement
 As an organisation we are committed to the on-going review and improvement of our ISMS.
 This policy is reviewed yearly as part of the Management review of the system.
 Approvals
 Meaning of Signature: Approval by Management
 Approver Role E-Signature Date
 Phil Reason CEO
 Changes in this issue
 General description of change Reason for change
 None Annual review
 Version: 2.3 Page 1 of 1 Ref: 99\_01\_001
 Copyright (c) 2021 Instem Group of Companies. IMPORTANT NOTICE: This document is uncontrolled in paper form.
 Refer to Instem SharePoint QMS for current version prior to use