



LJMU Research Online

Myatt, GJ, Ahlberg, E, Akahori, E, Allen, D, Amberg, A, Anger, LT, Aptula, A, Auerbach, S, Beilke, L, Bellion, P, Benigni,, R, Bercu, J, Booth, ED, Bower, D, Brigo, A, Burden, N, Cammerer, Z, Cronin, MTD, Cross, KP, Custer, L, Dettwiler, M, Dobo, K, Ford, KA, Fortin, MC, Gad-McDonald, SE, Gellatly, N, Gervais, V, Glover, KP, Glowienke, S, van Gompel, J, Gutsell, S, Hardy, B, Harvey, JS, Hillegass, J, Honma, M, Hsieh, JH, Hsu, CW, Hughes, K, Johnson, C, Jolly, R, Jones, D, Kemper, R, Kenyon, MO, Kim, MT, Kruhlak, NL, Kulkarni, SA, Kümmerer, K, Leavitt, P, Majer, B, Masten, S, Miller, S, Moser, J, Mumtaz, M, Muster, W, Neilson, L, Oprea, TI, Patlewicz, G, Paulino, A, Piparo, EL, Powley, M, Quigley, DP, Reddy, MV, Richarz, AN, Ruiz, P, Schilter, B, Serafimova, R, Simpson, W, Stavitskaya, L, Stidl, R, Suarez-Rodriguez, D, Szabo, DT, Teasdale, A, Trejo-Martin, A, Valentin, JP, Vuorinen, A, Wall, BA, Watts, P, White, AT, Wichard, J, Witt, KL, Woolley, A, Woolley, D, Zwickl, C and Hasselgren, C

In silico toxicology protocols

<http://researchonline.ljmu.ac.uk/8506/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Myatt, GJ, Ahlberg, E, Akahori, E, Allen, D, Amberg, A, Anger, LT, Aptula, A, Auerbach, S, Beilke, L, Bellion, P, Benigni,, R, Bercu, J, Booth, ED, Bower, D, Brigo, A, Burden, N, Cammerer, Z, Cronin, MTD, Cross, KP, Custer, L, Dettwiler, M. Dobo. K. Ford. KA. Fortin. MC. Gad-McDonald. SE. Gellatlv. N.

LJMU has developed [LJMU Research Online](http://researchonline.ljmu.ac.uk/) for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research.

<http://researchonline.ljmu.ac.uk/>

You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

1 ***In silico* toxicology protocols★**

2 Glenn J. Myatt^{a,*}, Ernst Ahlberg^b, Yumi Akahori^c, David Allen^d, Alexander Amberg^e, Lennart T. Anger^e,
3 Aynur Aptula^f, Scott Auerbach^g, Lisa Beilke^h, Phillip Bellionⁱ, Romualdo Benigni^j, Joel Bercu^k, Ewan D.
4 Booth^l, Dave Bower^a, Alessandro Brigo^m, Natalie Burdenⁿ, Zoryana Cammerer^o, Mark T. D. Cronin^p, Kevin
5 P. Cross^a, Laura Custer^q, Magdalena Dettwiler^r, Krista Dobo^s, Kevin A. Ford^t, Marie C. Fortin^u, Samantha
6 E. Gad-McDonald^v, Nichola Gellatlyⁿ, Véronique Gervais^w, Kyle P. Glover^x, Susanne Glowienke^y, Jacky Van
7 Gompel^z, Steve Gutsell^f, Barry Hardy^{aa}, James S. Harvey^{bb}, Jedd Hillegass^q, Masamitsu Honma^{cc}, Jui-Hua
8 Hsieh^{dd}, Chia-Wen Hsu^{ee}, Kathy Hughes^{ff}, Candice Johnson^a, Robert Jolly^{gg}, David Jones^{hh}, Ray Kemperⁱⁱ,
9 Michelle O. Kenyon^s, Marlene T. Kim^{ee}, Naomi L. Kruhlak^{ee}, Sunil A. Kulkarni^{ff}, Klaus Kümmerer^{jj}, Penny
10 Leavitt^q, Bernhard Majer^{kk}, Scott Masten^g, Scott Miller^a, Janet Moser^{ll,mm}, Moiz Mumtazⁿⁿ, Wolfgang
11 Muster^m, Louise Neilson^{oo}, Tudor I. Oprea^{pp}, Grace Patlewicz^{qq}, Alexandre Paulino^{rr}, Elena Lo Piparo^{ss},
12 Mark Powley^{ee}, Donald P. Quigley^a, M. Vijayaraj Reddy^{tt}, Andrea-Nicole Richarz^{uu}, Patricia Ruizⁿⁿ, Benoit
13 Schilter^{ss}, Rositsa Serafimova^{vv}, Wendy Simpson^f, Lidiya Stavitskaya^{ee}, Reinhard Stidl^{kk}, Diana Suarez-
14 Rodriguez^f, David T. Szabo^{ww}, Andrew Teasdale^{xx}, Alejandra Trejo-Martin^k, Jean-Pierre Valentin^{yy}, Anna
15 Vuorinenⁱ, Brian A. Wall^{zz}, Pete Watts^{aaa}, Angela T. White^{bb}, Joerg Wichard^{bbb}, Kristine L. Witt^g, Adam
16 Woolley^{ccc}, David Woolley^{ccc}, Craig Zwickl^{ddd}, Catrin Hasselgren^a

- 17 a) Leadscope, Inc. 1393 Dublin Rd, Columbus, OH 43215, USA
18 b) Predictive Compound ADME & Safety, Drug Safety & Metabolism, AstraZeneca IMED Biotech
19 Unit, Mölndal, Sweden
20 c) Chemicals Evaluation and Research Institute, 1-4-25 Kouraku, Bunkyo-ku, Tokyo 112-0004 Japan
21 d) Integrated Laboratory Systems, Inc., Research Triangle Park, NC, USA
22 e) Sanofi, R&D Preclinical Safety Frankfurt, Industriepark Hoechst, D-65926 Frankfurt am Main,
23 Germany

- 24 f) Unilever, Safety and Environmental Assurance Centre, Colworth, Beds, UK
- 25 g) The National Institute of Environmental Health Sciences, Division of the National Toxicology
26 Program, Research Triangle Park, NC 27709, USA
- 27 h) Toxicology Solutions Inc., San Diego, CA, USA
- 28 i) DSM Nutritional Products, Kaiseraugst, Switzerland
- 29 j) Alpha-PreTox , via G.Pascoli 1, 00184 Roma, Italy
- 30 k) Gilead Sciences, 333 Lakeside Drive, Foster City, CA, USA
- 31 l) Syngenta, Product Safety Department, Jealott's Hill International Research Centre, Bracknell,
32 Berkshire, RG42 6EY, UK
- 33 m) Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche
34 Innovation Center Basel, Switzerland
- 35 n) National Centre for the Replacement, Refinement and Reduction of Animals in Research
36 (NC3Rs), Gibbs Building, 215 Euston Road, London NW1 2BE, UK
- 37 o) Janssen Research & Development, 1400 McKean Road, Spring House, PA, 19477, USA
- 38 p) School of Pharmacy and Chemistry, Liverpool John Moores University, Liverpool, L3 3AF, UK
- 39 q) Bristol-Myers Squibb, Drug Safety Evaluation, 1 Squibb Dr, New Brunswick, NJ 08903, USA
- 40 r) Elanco Animal Health, Schwarzwaldallee 215, 4058 Basel, Switzerland
- 41 s) Pfizer Global Research & Development, 558 Eastern Point Road, Groton, CT, 06340, USA
- 42 t) Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
- 43 u) Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers, the
44 State University of New Jersey, 170 Frelinghuysen Rd, Piscataway, NJ, 08855, USA
- 45 v) Gad Consulting Services, 4008 Barrett Drive, Suite 201, Raleigh, NC 27609, USA
- 46 w) Biologie Servier, 905 route de Saran, 45520 Gidy, France

- 47 x) Defense Threat Reduction Agency, Edgewood Chemical Biological Center, Aberdeen Proving
48 Ground, MD 21010, USA
- 49 y) Novartis Pharma AG, Pre-Clinical Safety, Werk Klybeck, CH-4057, Basel, Switzerland
- 50 z) Janssen Pharmaceutical Companies of Johnson & Johnson, 2340 Beerse, Belgium
- 51 aa) Douglas Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, CH-4057 Basel / Basel-
52 Stadt, Switzerland
- 53 bb) GlaxoSmithKline Pre-Clinical Development, Park Road, Ware, Hertfordshire, SG12 0DP, UK
- 54 cc) National Institute of Health Sciences, Tokyo, Japan
- 55 dd) Kelly Government Solutions, Research Triangle Park, NC 27709, USA
- 56 ee) FDA Center for Drug Evaluation and Research, Silver Spring, MD 20993, USA
- 57 ff) Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, ON, K1A 0K9, Canada
- 58 gg) Toxicology Division, Eli Lilly and Company, Indianapolis, IN, USA
- 59 hh) Medicines and Healthcare products Regulatory Agency, 151 Buckingham Palace Road, London,
60 SW1W 9SZ, UK
- 61 ii) Vertex Pharmaceuticals Inc., Discovery and Investigative Toxicology, 50 Northern Ave, Boston,
62 MA, USA
- 63 jj) Institute for Sustainable and Environmental Chemistry, Leuphana University Lüneburg,
64 Scharnhorststraße 1/C13.311b, 21335 Lüneburg, Germany
- 65 kk) Shire, Industriestrasse 20, 1221, Wien, Austria
- 66 ll) Chemical Security Analysis Center, Department of Homeland Security, 3401 Ricketts Point Road,
67 Aberdeen Proving Ground, MD 21010-5405, USA
- 68 mm) Battelle Memorial Institute, 505 King Avenue, Columbus, OH 43210, USA
- 69 nn) Agency for Toxic Substances and Disease Registry, US Department of Health and Human
70 Services, Atlanta, GA, USA

71 oo) British American Tobacco, Research and Development, Regents Park Road, Southampton,
72 Hampshire. SO15 8TL, UK

73 pp) Translational Informatics Division, Department of Internal Medicine, Health Sciences Center, The
74 University of New Mexico, NM. USA

75 qq) U.S. Environmental Protection Agency, National Center for Computational Toxicology, Research
76 Triangle Park, NC 27711, USA

77 rr) SAPEC Agro, S.A., Avenida do Rio Tejo, Herdade das Praias, 2910-440 Setúbal, Portugal

78 ss) Chemical Food Safety Group, Nestlé Research Center, Lausanne, Switzerland

79 tt) Merck Research Laboratories, West Point, PA 19486, USA

80 uu) European Commission, Joint Research Centre, Directorate for Health, Consumers and Reference
81 Materials, Chemical Safety and Alternative Methods Unit, Via Enrico Fermi 2749, 21027 Ispra
82 (VA), Italy

83 vv) European Food Safety Authority, Via Carlo Magno 1A, 43126 Parma, Italy

84 ww) RAI Services Company, 950 Reynolds Blvd., Winston-Salem, NC 27105, USA

85 xx) AstraZeneca, Macclesfield, Cheshire, UK

86 yy) UCB Biopharma SPRL, Chemin du Foriest – B-1420 Braine-l'Alleud, Belgium

87 zz) Colgate-Palmolive Company, Piscataway, New Jersey 08854, USA

88 aaa) Bibra, Cantium House, Railway Approach, Wallington, Surrey, SM6 0DZ, UK

89 bbb) Bayer Pharma AG, Investigational Toxicology, Muellerstr. 178, D-13353 Berlin, Germany

90 ccc) ForthTox Limited, PO Box 13550, Linlithgow, EH49 7YU, UK

91 ddd) Transendix LLC, 1407 Moores Manor, Indianapolis, IN 46229, USA

92

93

94 ✧ FDA CDER Disclaimer: The findings and conclusions in this manuscript have not been formally
95 disseminated by the FDA and should not be construed to represent any agency determination or
96 policy. The mention of commercial products, their sources, or their use in connection with
97 material reported herein is not to be construed as either an actual or implied endorsement of
98 such products by the Department of Health and Human Services.

99 MHRA Disclaimer: Any opinions expressed in this document are the author's and are not
100 necessarily shared by other assessors at the Medicines and Healthcare products Regulatory
101 Agency (MHRA). As such, they cannot be considered to be UK policy. The mention of
102 commercial products, their sources, or their use in connection with material reported herein is
103 not to be construed as either an actual or implied endorsement of such products by the UK's
104 MHRA.

105 CDC /ATSDR Disclaimer: The findings and conclusions in this report are those of the author(s)
106 and do not necessarily represent the official position of the Centers for Disease Control and
107 Prevention or the Agency for Toxic Substances and Disease Registry. Mention of trade names is
108 not an endorsement of any commercial product.

109 EPA Disclaimer: The views expressed in this article are those of the author(s) and do not
110 necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of
111 trade names or commercial products does not constitute endorsement or recommendation for
112 use.

113 EFSA disclaimer: This paper reflects Dr. Serafimova's personal view and is not endorsed by the
114 European Food Safety Authority

115 Health Canada Disclaimer: The findings and conclusions in this report are those of the author(s)
116 and do not necessarily represent the official position of Health Canada. The mention of
117 commercial products, their sources, or their use should not be construed as either an actual or
118 implied endorsement of such products by Health Canada.

119 DHS Disclaimer: The findings and conclusions in this report are those of the author(s) and do not
120 necessarily represent the official position of the U.S. Government, Department of Homeland
121 Security (DHS), DHS Science and Technology (S&T) Homeland Security Advanced Research
122 Projects Agency (HSARPA), or the Chemical Security Analysis Center (CSAC). In no event shall
123 either the U.S. Government, DHS, HSARPA, CSAC, or the author(s) have any responsibility or
124 liability for any consequences of any use, misuse, inability to use, or reliance upon the
125 information contained herein, nor do any of the parties set forth in this disclaimer warrant or
126 otherwise represent in any way the accuracy, adequacy, efficacy, or applicability of the contents
127 hereof. The use of trade, firm, company, or corporation names including product descriptions
128 does not constitute an official DHS endorsement of any service or product.

129 European Commission: The views expressed are solely those of the authors and the content
130 does not necessarily represent the views or position of the European Commission.

131 * Corresponding author. Present address: Leadscope, Inc., 1393 Dublin Road, Columbus, OH 43215, USA.
132 E-mail address: gmyatt@leadscope.com (G.J. Myatt).

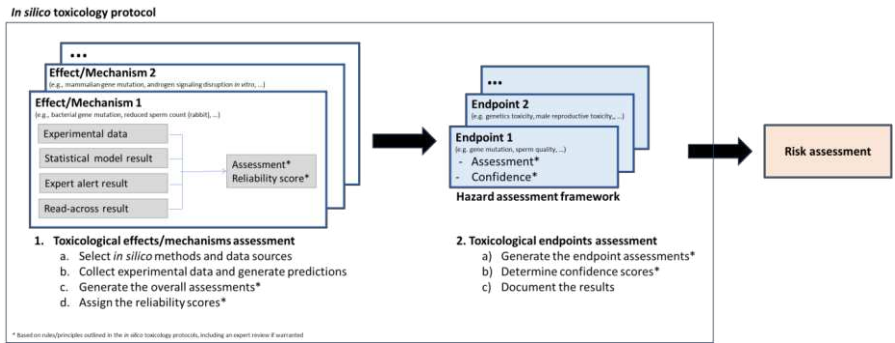
133

134 **Abstract**

135 The present publication surveys several applications of *in silico* (i.e., computational) toxicology
136 approaches across different industries and institutions. It highlights the need to develop standardized
137 protocols when conducting toxicity-related predictions. This contribution articulates the information
138 needed for protocols to support *in silico* predictions for major toxicological endpoints of concern (e.g.,
139 genetic toxicity, carcinogenicity, acute toxicity, reproductive toxicity, developmental toxicity) across
140 several industries and regulatory bodies. Such novel *in silico* toxicology (IST) protocols, when fully
141 developed and implemented, will ensure *in silico* toxicological assessments are performed and
142 evaluated in a consistent, reproducible, and well-documented manner across industries and regulatory
143 bodies to support wider uptake and acceptance of the approaches. The development of IST protocols is
144 an initiative developed through a collaboration among an international consortium to reflect the state-
145 of-the-art in *in silico* toxicology for hazard identification and characterization. A general outline for
146 describing the development of such protocols is included and it is based on *in silico* predictions and/or
147 available experimental data for a defined series of relevant toxicological effects or mechanisms. The
148 publication presents a novel approach for determining the reliability of *in silico* predictions alongside
149 experimental data. In addition, we discuss how to determine the level of confidence in the assessment
150 based on the relevance and reliability of the information.

151

152 **Graphical abstract**



153

154

155 **Keywords:** *In silico*, *in silico* toxicology, computational toxicology, predictive toxicology, QSAR, expert

156 alert, expert review.

157

158 **Highlights**

- 159 • General outline of *in silico* toxicology protocols is described
- 160 • A reliability score for predictions alongside experimental data is discussed
- 161 • A checklist for performing an expert review of the *in silico* results is outlined
- 162 • A hazard assessment framework is proposed that includes *in silico* results

163

164 **Word count:**

165 Abstract = 211

166 Text = 8,258

167 References = 2,688

168 **1. Introduction**

169 *In silico* toxicology (IST) methods are computational approaches that analyze, simulate, visualize, or
170 predict the toxicity of chemicals. IST encompasses all methodologies for analyzing chemical and
171 biological properties generally based upon a chemical structure that represents either an actual or a
172 proposed (i.e., virtual) chemical. Today, *in silico* approaches are often used in combination with other
173 toxicity tests; however, the approaches are starting to be used to generate toxicity assessments
174 information with less need to perform any *in vitro* or *in vivo* studies depending on the decision context.
175 IST uses models which can be encoded within software tools to predict the potential toxicity of a
176 chemical and in some situations to quantitatively predict the toxic dose or potency. These models are
177 based on experimental data, structure-activity relationships, and scientific knowledge (such as structural
178 alerts reported in the literature).

179 There are a number of different situations where *in silico* methods serve an important role in the hazard
180 assessment of existing chemicals or new substances under development that would benefit from the
181 development of *in silico* toxicology protocols. These include:

- 182 • emergency situations where rapid understanding of potential toxicological consequences from
183 exposure is needed in the absence of existing toxicological testing data;
- 184 • cases where there is only a limited supply of a test material available;
- 185 • scenarios where there are challenges to conduct laboratory studies;
- 186 • instances where synthesis of a complex test material is not feasible; and
- 187 • situations where a less time-consuming and less expensive high-throughput approach than an
188 experimental test is needed.

189 IST methods are one approach to generating additional information for complementing and ultimately
190 enhancing the reliability or supporting a risk assessment, including an understanding of the structural
191 and/or mechanistic basis that may contribute ideas for the rational design of new chemicals,
192 development of a testing strategy or an overall weight-of-evidence evaluation. IST inherently supports
193 the principle of the 3Rs (replacement, refinement and reduction) relating to the use of animals in
194 research (Russell and Burch, 1959; Ford 2016). Table 1 outlines fifteen specific uses of IST to illustrate
195 the diversity of applications that currently can benefit from *in silico* methods. Stanton and Kruszewski
196 (2016) recently quantified the benefits of using *in silico* and read-across methods where they
197 determined that the approach used across two voluntary high-production-volume (HPV) chemical
198 programs for 261 chemicals obviated the use of 100,000 – 150,000 test animals and saved 50,000,000
199 US\$ to 70,000,000 US\$.

200 The increased interest and acceptance of *in silico* methods for regulatory data submission and chemicals
201 evaluation is driving the adoption of its use for regulatory purposes. Several guidance documents have
202 been drafted to improve standardization, harmonization, and uptake of *in silico* methods by regulatory
203 authorities including the International Council for Harmonization (ICH) M7 guideline (assessment and
204 control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk)
205 (ICH M7, 2017(R1)), the European Union's Registration, Evaluation, Authorization, and restriction of
206 Chemicals (REACH) regulation (EU 2006; ECHA 2008; ECHA 2015), European Food Safety Authority
207 (EFSA) residue guidance (EFSA 2016), Canada's chemicals management plan assessments for new and
208 existing substances under the Canadian Environmental Protection Act (CEPA) (Canada 2016), and the
209 Toxic Substances Control Act (TSCA) (TSCA 2016). A number of national and international initiatives have
210 focused on developing specific documents supporting the use of *in silico* tools. The OECD has published
211 a series of (Quantitative) Structure-Activity Relationship (Q)SAR validation principles that are discussed
212 in detail in Section 2.3.2. (OECD 2004, OECD 2007) Other initiatives include the North American Free

213 Trade Agreement pesticides Quantitative Structure-Activity Relationship (QSAR) guidance (NAFTA 2012),
214 considerations on the use of *in silico* approaches for assessing cosmetics ingredients (Amaral et al.,
215 2014), European Food Safety Agency report (EFSA 2014), European Chemicals Agency REACH supporting
216 documentation (ECHA 2008; ECHA 2016, 2017), Organization for Economic Co-operation and
217 Development (OECD) documentation (OECD 2007; OECD 2014; OECD 2015), and the ICH M7 guideline
218 for prediction of mutagenicity (ICH M7, 2017(R1)), along with complementary peer reviewed
219 publications outlining the process for implementation of such computational assessments (e.g., Amberg
220 et al., 2016; Barber et al., 2015; Powley et al., 2015; Schilter et al., 2014). Certain projects have provided
221 substantial guidance on the documentation of the models and prediction results (JRC 2014; Patlewicz et
222 al., 2016) as well as principles and workflows to support safety assessments (Bassan and Worth, 2008;
223 ECHA 2015; Worth et al., 2014; Berggren et al., 2017; Amaral et al., 2017).

224 These prior initiatives provide a robust foundation for the current project to establish the IST protocols
225 described here; however, several issues have hindered the general acceptance and use of *in silico*
226 methods on a larger scale. In particular, there remains a lack of generally accepted procedures for
227 performing *in silico* assessments for the toxicological endpoints. The lack of such procedures or
228 protocols has led to inconsistency in the application and use of *in silico* tools across different
229 organizations, industries, and regulatory agencies (e.g., searching databases, applying predictive models
230 and alerts, performing an expert review/assessment, documenting and communicating the results and
231 associated uncertainties). The use of traditional experimental evidence coupled with *in silico*
232 information to support hazard identification and risk assessment also varies both across, and often
233 within, organizations. Although not always, such *ad hoc* approaches may be time-consuming and the
234 results poorly accepted. Standardization of protocols will enhance the acceptability of the methods and
235 their results by end users. Additionally, there are misconceptions about when *in silico* predictions are
236 appropriate to use as well as a lack of defined consensus processes for interpreting the result(s) of such

237 predictions (Bower et al., 2017; SCCS 2016). Some scientists view *in silico* methods as a “black box” that
238 inhibits their ability to critically assess the predictions and their reliability. (Alves et al., 2016) Others lack
239 expertise to interpret the results of *in silico* predictions, and some have an unrealistic expectation that
240 an *in silico* prediction can always provide an unerring definitive assessment.

241 Standardization of *in silico* tool use and interpretation of results would greatly reduce the burden on
242 both industry and regulators to provide confidence in or justification for the use of these approaches.
243 The objective of developing IST protocols is to define *in silico* assessment principles so the results can
244 be generated, recorded, communicated, archived and then evaluated in a uniform, consistent and
245 reproducible manner. Incorporating these principles routinely into the use of *in silico* methods will
246 support a more transparent analysis of the results and serves to mitigate “black box” concerns¹. This
247 approach is similar to guideline studies that provide a framework for the proper conduct of
248 toxicological studies and assurance in the validity of the results (such as OECD Guidelines for the
249 Testing of Chemicals) (OECD 2017). The development of these protocols is driven by consensus
250 amongst leading scientists representing industry, private sector and governmental agencies.
251 Consequently, this project provides an important step towards a quality-driven science for IST or good
252 *in silico* practice .

253 Herein, we provide a framework to develop a series of procedures for performing an *in silico* assessment
254 to foster greater acceptance. These IST protocols are being created for a number of toxicological
255 endpoints (e.g., genetic toxicity, carcinogenicity, acute toxicity, reproductive toxicity, developmental
256 toxicity) as well as other related properties (e.g., biodegradation and bioaccumulation) that could
257 impact the chemical hazard classification. Throughout this publication, these toxicological and related

¹It should be noted that black box models may be acceptable in certain situations, such as compound filtering and virtual screening, as long as they show acceptable performance in validation studies; however, for most applications the acceptance of this class of models is low.

258 endpoints are referred to as “major endpoints” and the protocols are referred to as IST protocols. These
259 protocols will support the assessment of hazards and in some cases the prediction of quantitative
260 values, such as a No Observed Adverse Effect Levels (NOAELs); however, these protocols do not define
261 how a risk assessment will be performed. This publication outlines the components of an IST protocol,
262 including schematics to describe how a prediction could be performed, approaches to assess the
263 reliability and confidence of the results, and items that may be considered as part of an expert review.
264 This publication also outlines the process for creating the IST protocols through an international
265 consortium comprising representatives across regulatory agencies, government research agencies,
266 different industrial sectors, academia and other stakeholders. Specific endpoint-dependent
267 considerations will be described in future separate publications and IST protocols (developed as a result
268 of this process) will also be published for widespread use and for incorporation into different technology
269 platforms.

270 **2. *In silico* toxicology protocols**

271 **2.1 Overview**

272 Each IST protocol describes the prediction process in a consistent, transparent, and well-documented
273 manner. This includes recommendations on how to:

- 274 1) plan the *in silico* analyses including identifying what toxicological effects or mechanisms to
275 predict (Section 2.2), what *in silico* methodologies to use (Section 2.3.1), and other selection
276 criteria for the *in silico* methods (Section 2.3.2),
- 277 2) conduct the appropriate individual software predictions (Section 2.3.3) and further database
278 searches (Section 2.5),

279 3) perform and document the *in silico* analysis (Sections 2.6 and 2.7) including expert review
280 (Section 2.4), and

281 4) report and share the information and assessment results, including information about
282 uncertainties (Section 2.9).

283 Section 2.8 provides a template for the individual IST protocols for major toxicological endpoints. IST
284 protocols could be applicable for use with several *in silico* programs, including different *in silico* models
285 and databases.

286 **2.2 Toxicological effects and mechanisms**

287 In an experimental approach, hazard is evaluated based on specific observations (toxicological effects)
288 during toxicity studies. Often, toxicity of a chemical involves a biological event: a non-specific or specific
289 interaction with a vital biological structure, which causes sequential perturbation of a physiological
290 pathway at a cellular, tissue, organ and/or system level, leading to a toxicological effect observed at the
291 organism level. Experiments evaluating the potential of a chemical to cause such a biological event (e.g.,
292 *in vitro* analysis of specific interaction with a cellular receptor or inhibition of an enzyme or non-specific
293 cytotoxicity), may support hazard assessment and provide information about the mechanism of toxicity.
294 Such an approach is utilized in the Adverse Outcome Pathway (AOP), where identification of a molecular
295 initiating event supports assessment of the related adverse outcome at the organism level (Bell et al.,
296 2016; OECD 2016a; OECD 2016b). A computational approach to hazard assessment may address the
297 two complementary levels of hazard identification in a similar way (i.e., predicting the resulting
298 manifestation (effect) or the molecular perturbation (mechanism) that led to the toxicological effect).

299 Each IST protocol defines a series of known toxicological effects and mechanisms relevant to the
300 assessment of the major toxicological endpoint. For example, in the reproductive toxicity IST protocol,

301 the list of toxicological effects/mechanisms may include reduced sperm count, androgen signaling
302 disruption *in vitro*, and so on. Within each IST protocol, these effects/mechanisms may be species
303 and/or route of administration specific.

304 Figure 1 outlines a general approach to performing an *in silico* assessment. For each toxicological
305 effect/mechanism, relevant information (as defined in the IST protocol) is collected, including any
306 available experimental data as well as *in silico* predictions. The experimental data and/or *in silico* results
307 are then analyzed and an overall assessment of the toxicological effect or mechanism is generated
308 alongside a reliability score (defined in Section 2.6.2) that reflects the quality of the results. The
309 assessment results and reliability scores for a range of relevant toxicological effects/mechanisms are
310 then used to support a hazard assessment within the hazard assessment framework.

311 **2.3 *In silico* predictions**

312 **2.3.1 *In silico* methodologies**

313 Several organizations develop and make available computer software packages for predicting toxicity or
314 physicochemical properties of query chemical(s). These systems generally contain one or more models,
315 where each model predicts the compound's putative toxicological effect or mechanism of action. For
316 example, a model may predict the results for bacterial gene mutation using data generated from the
317 bacterial reverse mutation test or Ames test. These models may be revised over time as more data
318 become available, structure-activity relationships are better characterized, and any data set used is
319 updated. Each new or updated model is given a different version number because the results from
320 different model versions may vary and it is important to track the source of the results. (Amberg et al.,
321 2016)

322 All IST protocols will identify the toxicological effects or mechanisms to be predicted as discussed in
323 Section 2.2. These predictions may be dichotomous (e.g., predict mutagenic or non-mutagenic
324 compounds), quantal (e.g., Globally Harmonized System [GHS] Classification and Labeling² scheme) or
325 quantitative/continuous (e.g., prediction of median toxic dose [TD₅₀] values). The specific IST protocols
326 will detail the type of prediction(s) ideally generated.

327 The major *in silico* prediction methodologies include the following:

328 • **Statistical-based (or QSAR).** This methodology uses a mathematical model that was derived
329 from a training set of example chemicals. The training set includes the chemicals that were
330 found to be positive and negative in a given toxicological study (e.g., the bacterial reverse
331 mutation assay) or to induce a continuous response (e.g., NOAEL in teratogenicity) that the
332 model will predict. As part of the process to generate the model, physicochemical property-
333 based descriptors (e.g., molecular weight, octanol water partition coefficient [log P]), electronic
334 and topological descriptors (e.g., quantum mechanics calculations), or chemical structure-based
335 descriptors (e.g., the presence or absence of different functional groups) are generated and
336 used to describe the training set compounds. The model encodes the relationship between
337 these descriptors and the (toxicological) response. After the model is built and validated (OECD
338 2007; Myatt et al., 2016), it can be used to make a prediction. The (physico)chemical descriptors
339 incorporated into the model are then generated for the test compound and are used by the
340 model to generate a prediction. This prediction is only accepted when the test compound is
341 sufficiently similar to the training set compounds (i.e., it is considered within the applicability
342 domain of the QSAR model, often considering the significance of descriptors). (Netzeva et al.,
343 2005; Carrió et al., 2014; Patlewicz et al., 2016) This applicability domain analysis may be

² A chemical is assigned to a category (e.g., 1, 2, 3, 4, or 5) based on distinct ranges of quantitative values (e.g., LD₅₀). Examples of such ranges include LD₅₀ <5mg/kg (i.e., category 1) or 50-300mg/kg (i.e., category 3).

344 performed automatically by some software to determine whether the training set compounds
345 share similar chemical and/or biological properties with the test chemical.

346 • **Expert rule-based (or expert/structural alerts).** This methodology uses structural rules or alerts
347 to make predictions for specific toxicological effects or mechanisms of toxicity. These rules are
348 derived from the literature or from an analysis of data sets generated by scientists. Structural
349 alerts are defined as molecular substructures that can activate the toxicological effect or
350 mechanism. The rules may also encode situations where the alert is deactivated. Expert rule-
351 based models often include a description of the toxic mechanism and examples from the
352 literature or other reference sources to justify the structural alert. A positive prediction is
353 generally made when a structural alert is present (without deactivating structural features or
354 properties) in the test compound. When no alerts are triggered for a test chemical, a negative
355 prediction may be generated for well investigated endpoints; however, additional analysis is
356 generally required to make this assessment as discussed further in Section 2.4.3.

357 • **Read-across:** Read-across uses data on one or more analogs (the “source”) to make a prediction
358 about a query compound or compounds (the “target”). Source compounds are identified that
359 have a structurally or toxicologically meaningful relationship to the target compound, often
360 underpinned by an understanding of a plausible biological mechanism shared between the
361 source and target compounds. The toxicological experimental data from these source
362 compounds can then be used to “read-across” to the specific target compound(s). Read-across is
363 an intellectually-derived endpoint-specific method that provides justification for why a chemical
364 is similar to another chemical (with respect to chemical reactivity, toxicokinetics,
365 mechanism/mode of action, structure, physicochemical properties, and metabolic profile). (Wu
366 et al., 2010; ECETOC 2012; Patlewicz et al., 2013a; Patlewicz et al., 2013b; OECD 2014; Blackburn

367 and Stuard, 2014; Patlewicz (2014); Patlewicz et al., 2015; Schultz et al., 2015; Ball et al., 2016;
368 ECHA 2017b)

369 • **Other approaches:** In certain cases, other *in silico* methodologies may be appropriate. Examples
370 include the use of molecular dynamics (e.g., simulating interactions of a query chemical with a
371 metabolic enzyme) and receptor binding as an indication of a possible Molecular Initiating Event
372 (e.g., estrogen receptor-ligand docking).

373 Each IST protocol will include an assessment of key computational aspects and specific issues to
374 consider. For example, when performing read-across, issues such as the data quality of the source
375 compound(s), how to perform an assessment of non-reactive chemical features and selection of
376 grouping approaches used to form categories will be discussed to ensure source compound(s) are
377 sufficiently similar, both chemically and biologically, for the endpoint being considered.

378 Each methodology has its strengths and weaknesses, which often depend on the type of toxicological
379 effect or mechanism being predicted. This will be discussed in the individual IST protocols. In addition,
380 there may be cases of unique or novel compounds for which it is not possible to make a prediction or for
381 which confidence in the predictions is so low as to render it meaningless or unhelpful.

382 **2.3.2 *In silico* methods selection criteria**

383 *In silico* methods selection may include the following five considerations:

384 1. **Relevant toxicological effects or mechanisms.** As discussed in Section 2.2, each IST protocol will
385 define a series of toxicological effects or mechanisms relevant to a specific endpoint and
386 appropriate *in silico* models need to be selected that predict these specific effects or
387 mechanisms.

388 2. **Model validity.** Best practices for validation of (Q)SAR *in silico* models have been documented in
389 a number of publications (Cherkasov et al.; 2014, Raies and Bajic, 2016; Myatt et al., 2016), and
390 models built using these best practices may be preferred. The OECD has published a series of
391 validation principles for *in silico* models (OECD 2004; OECD 2007) and valid statistical-based or
392 expert rule-based *in silico* methods. Such (Q)SAR methods have: 1) a defined endpoint; 2) an
393 unambiguous algorithm; 3) a defined domain of applicability; 4) appropriate measures of
394 goodness-of-fit, robustness and predictivity; and 5) a mechanistic interpretation, if possible. Any
395 *in silico* model must include documentation that supports an assessment of the model's
396 scientific validity, including the toxicological effect or mechanism being predicted, version
397 number, type of methodology, training set size and content, as well as any predictive
398 performance information. Validation performance is documented in report formats such as the
399 QSAR Model Reporting Format (QMRF) (JRC 2014). The level of adherence to the OECD
400 principles and the performance statistics need to be appropriate for the purpose of the
401 assessment.

402 3. **Chemical space.** Often, *in silico* models will only make predictions for specific classes of
403 chemicals, the so called "applicability domain". The chosen *in silico* model(s) may report the
404 applicability domain assessment to demonstrate its proficiency for this class of compounds. Vice
405 versa, only models are ideally chosen where the query compound is in the applicability domain.
406 (Netzeva et al., 2005; Carrió et al., 2014; Patlewicz et al., 2016)

407 4. **Model combinations.** Complementary or independent *in silico* models may be selected, as
408 concurring results increase the reliability of the prediction (as discussed in Section 2.6.2).

409 5. **Supporting an expert review.** For QSAR models, tools to help the expert review (see Section 2.4)
410 include the ability to allow examination of the descriptors and weightings used in the model,
411 underlying training set data, and how the applicability domain assessment was defined. For

412 expert rule-based systems, this could include how the alert was defined (including any factors
413 that activate or deactivate the alert), any mechanistic understanding associated with the alert,
414 citations, and any relevant known examples of alerting chemicals.

415 Read across may be used when there are experimental data from high quality databases for one or more
416 substances which are similar enough to the target chemical of interest. The Read-Across Assessment
417 Framework (RAAF), or similar published and established frameworks, may be used to document the
418 read-across assessment and to support its scientific plausibility (ECHA 2017b; Patlewicz et al., 2013b;
419 Blackburn & Stuard 2014; Schultz et al., 2015; Patlewicz et al., 2015). The OECD has also produced
420 guidance on the process of grouping chemicals and other considerations as part of a read-across
421 assessment (OECD 2014), and ECHA has generated guidelines on the process of performing a valid read-
422 across assessment (ECHA 2008).

423 **2.3.3 Running the *in silico* models**

424 All *in silico* systems require an electronic representation of the chemical structure and any errors in this
425 representation will result in invalid predictions. Therefore, it is important to ensure that the chemical
426 structure is properly curated and entered following conventions set out by the model's developer,
427 including appropriate representations for tautomers, aromaticity, salt forms, stereochemistry, charges,
428 and specific functional groups (e.g., nitro or carboxylic acid groups). It is possible that different formats
429 (i.e., SMILES vs. MOL files) may be processed differently. It is also important to verify that the software
430 correctly interprets the structural representation during processing, particularly for complex molecules.
431 For some types of chemicals, *in silico* models may not be applicable due to the structural representation
432 or the unsuitability of the experiment assay for the specific chemical class. Examples include non-
433 discrete chemical substances, UVCBs (unknown/variable composition, complex reaction products and

434 biologicals), metals, inorganics, polymers, mixtures, organometallics and nano-materials. (Mansouri et
435 al., 2016)

436 Some models, such as statistical-based models, allow for prediction settings to be adjusted or turned off
437 (e.g., they report “positive” when a value is greater than a predetermined threshold). The settings are
438 ideally selected in a way that does not compromise the model’s validity (such as changing the validation
439 statistics of the model) and appropriately reported.

440 A thorough documentation of all selected models and computer software packages including, version
441 numbers, and any parameters set, is needed as part of the materials and methods in sufficient detail to
442 assess and potentially repeat the analysis (discussed in Section 2.9). In addition, the results need to be
443 presented in enough detail to fully understand how they were generated and to critically assess the
444 findings.

445 **2.4 *In silico* expert review**

446 **2.4.1 Overview**

447 As with *in vitro* or *in vivo* study data, *in silico* predictions may be critically assessed and an expert review
448 of the output is often prudent (Dobo et al., 2012; Sutter et al., 2013). Frameworks for conducting an
449 expert review ensure that it is performed in a consistent and transparent manner. Examples of such a
450 review framework include the Office of Health Assessment and Translation (OHAT) systematic review
451 and evidence integration (Rooney et al., 2014), weight-of-evidence assessments (ECHA 2017a), and
452 Integrated Approaches to Testing and Assessment (IATA) (OECD 2016a; OECD 2016b).

453 The purpose of an *in silico* expert review is to evaluate the reliability of the prediction. The outcome of
454 the review provides information to include in the assessment of the toxicological effect or mechanism.

455 As part of this review, the expert might agree with, or refute, individual *in silico* predictions. In addition,

456 these reviews might support cases when a chemical is out of the applicability domain of the model,
457 support the use of an equivocal prediction (i.e., there is evidence both for and against the supposition),
458 or support cases where multiple predictions do not agree. A checklist of items to consider and report
459 will help to ensure such reviews are performed in a consistent manner (as illustrated in Tables 2 and 3).
460 This review may include knowledge from proprietary information available within an organization from
461 the testing of related chemicals.

462 When an expert review assesses multiple predictions from different *in silico* systems, it is important to
463 justify how they complement each other with regard to the training set (i.e., the use of relevant
464 guideline studies plus relevant chemical classes), methodology (e.g., expert rule-based vs. statistical-
465 based vs. read-across), or QSAR descriptor sets.

466 It is essential to document the reasoning and decisions of the expert review steps so they can be
467 retraced at any time, including the information used as the basis for the review.

468 **2.4.2 Expert review of statistical models**

469 An expert review of a statistical-based model involves a critical assessment of how the model generated
470 the prediction. This includes examining the weightings of the model descriptors (e.g., structural features
471 or physicochemical properties related to toxicity), underlying data, chemical space of the training set of
472 the model, and the experimental results for analog compounds and model performance for these
473 analogs (e.g., nearest-neighbor list of compounds) (Amberg et al., 2016). This may also incorporate an
474 understanding of the mechanism of toxicity or knowledge of factors that activate or deactivate the
475 toxicity. The items described in Table 2 provide a checklist of elements to consider as part of any QSAR
476 expert review to ensure such a review is as objective as possible, transparent and based on a consistent
477 set of considerations. An expert review may increase the reliability of statistical model results based on
478 one or more elements defined in Table 2.

479 Individual IST protocols will outline specific points to consider when performing an expert review, such
480 as how the similarity of analogs could be assessed.

481 **2.4.3 Expert review of expert rule-based (structural) alert systems**

482 An expert review of the results from an expert rule-based alert system may involve inspection of the
483 underlying information as well as external knowledge. Special emphasis needs to be placed on the
484 assessment of chemicals where no alerts are identified in the expert alert system. When no alert is fired
485 (i.e., it is not predicted active), it is often not reported if the prediction is negative, equivocal, or out of
486 the applicability domain of the model and often no prediction is generated. An expert review may
487 increase the reliability of the results based on one or more elements defined in Table 3.

488 **2.4.4 Read-across expert review**

489 Read-across contains an expert assessment by its nature: it requires expert judgment of the analogs,
490 their data and extrapolation to the query chemical. For example, read-across assessments performed
491 and documented according to the RAAF (i.e., following the detailed RAAF Assessment Elements), or
492 similar frameworks, as discussed earlier, incorporate an expert review as part of the assessment. This
493 type of assessment includes a strong justification for biological plausibility of any analogs selected
494 (including an assessment of the structural differences and similarities to the target structure, and an
495 analysis of potential metabolism). It also includes an expert assessment when a read-across prediction
496 concludes there is an absence of effects. In addition, an assessment of supporting evidence (including
497 the reliability of the source data), any weight-of-evidence considerations, and an assessment of any
498 possible bias in the selection of source chemicals is required.

499 **2.5 Assessment of available experimental data**

500 Experimental data may have been previously generated and reported for a chemical being assessed, for
501 example, in the literature or through a public or proprietary database. To support the identification of

502 experimental data, each IST protocol will identify a series of relevant study types and specific result(s)
503 corresponding to the identified toxicological effects or mechanisms, as discussed in Section 2.2. To
504 illustrate, in the assessment of the toxicological effect/mechanism bacterial gene mutation (part of the
505 genetic toxicity IST protocol), the overall mutagenic or non-mutagenic results from a bacterial reverse
506 mutation assay may be used. A more complex example is in the assessment of the toxicological
507 effect/mechanism of sperm morphology (part of the reproductive IST protocol). Here, specific results
508 from potentially different study types, such as one- or two- generation reproductive studies, repeated
509 dose toxicity studies or segment I (fertility) studies, and possibly also from different species (rat, mouse,
510 rabbit) will be applicable.

511 The selection of experimental study types need focus on those that have general value based on
512 scientific justification. This includes study types that have widespread use in risk assessments, regulatory
513 acceptance and that follow internationally recognized test guidelines. In addition, other types of data
514 may be considered relevant on a case-by-case basis. Numerous guidance documents discuss acceptable
515 studies, their relevancy, and their use in hazard identification, hazard characterization and risk
516 assessment. These include guidance documents from the ICH (ICH 2017), OECD (OECD 2017), European
517 Food Safety Authority (EFSA) (EFSA 2017a), Scientific Committee on Consumer Safety (SCCS) (SCCS
518 2017), REACH /ECHA (ECHA 2008; ECHA 2015), United States Environmental Protection Agency (EPA)
519 Office of Chemical Safety and Pollution Prevention (OCSPP 2015), and National Institute of
520 Environmental Health Sciences (NIEHS) (NIEHS 2017) guidance documents. Such guidance documents
521 provide a useful basis for test considerations but may not always be harmonized across legislation,
522 industrial sector or geographical regions, as requirements may differ across guidance documents.

523 The IST protocols will discuss how to assess and document the experimental data and uncertainties to
524 ensure the proper justification of the experimental results' reliability, including defining what specific

525 elements or fields are important to document. With older studies pre-dating existing guidelines, it will
526 often still be possible to perform an expert review to determine the adequacy of the data, but it will be
527 important to document specifically why the study results were considered acceptable or dismissed as
528 unacceptable. The IST protocols will also provide recommendations on how to select a result when
529 multiple studies (with potentially conflicting results) for the same effect or mechanism are reported.

530 Klimisch scores are a widely used approach adopted to support an assessment of experimental data
531 reliability (Table 4; Klimisch et al., 1997). The Klimisch score (1 to 4) is based on factors including
532 whether the test was compliant with the OECD principles of Good Laboratory Practices (GLP) or Good *In*
533 *Vitro* Methods Practices (GIVIMP) standards (OECD 2016c), whether the data were generated using
534 accepted test guidelines, whether the data are available for independent inspection, and the quality of
535 the report. ECHA uses this score, for example, as part of its data submission process (ECHA 2011), and
536 there are tools to support the assignment of Klimisch scores (ECVAM 2017; Schneider et al., 2009).
537 Another approach to the assessment of the reliability of the experimental data is the Science in Risk
538 Assessment and Policy (SciRAP) application, a web-based reporting and evaluation resource created to
539 help understand how academic toxicity-related studies can be used as part of any regulatory assessment
540 (Molander et al., 2014). An approach proposed by EFSA is a detailed analysis of different parameters of
541 the study (e.g. statistical power; verification of measurement methods and data; control of experimental
542 variables that could affect measurements; universality of the effects in validated test systems using
543 relevant animal strains and appropriate routes of exposure, etc.) with detailed documentation of the
544 process (EFSA, 2011).

545 2.6 Combined assessment of experimental data and *in silico* predictions

546 2.6.1 Toxicological effect or mechanism assessment

547 Reliable data, generally defined by Klimisch scores 1 or 2 reviewed by an expert (see Table 4), is ideally
548 used for the toxicological effect or mechanism (shown in Figure 1) whenever available³. In the absence
549 of adequate experimental data, results from one or more *in silico* models can be used to support
550 assessment of the toxicological effect or mechanism. When multiple *in silico* model results, from
551 potentially different methodologies, or QSAR models using different descriptors and/or training sets, are
552 generated per toxicological effect or mechanism, the individual results need to be compiled to provide
553 one overall assessment, as shown in Figure 1. This assessment may take into consideration information
554 from any expert review of the *in silico* results, as certain results may need to be refuted. Similarly, when
555 there are data assigned Klimisch 3 or 4 and/or there are *in silico* results, this information needs to be
556 compiled into an overall assessment. Individual IST protocols will document such procedures.

557 There are multiple approaches to compile results. A cautious approach is to use the most conservative
558 data or prediction for this assessment. For example, when predicting the results of the bacterial reverse
559 mutation test using two models, if either model's prediction result is mutagenic then the overall
560 assessment is mutagenic. Other options include a weight-of-evidence or consensus approach or
561 selection of the prediction with the highest confidence (e.g., predictive probability score and relevance
562 of analogous structures). Specific considerations per endpoint may be addressed in the individual IST
563 protocols and may be dependent on the problem formulation.

³ As mentioned in Section 2.5, where high quality experimental data are available (as shown in Figure 1), it may not be necessary to run *in silico* models. However, generating *in silico* predictions for chemicals with known values is sometimes performed to verify experimental 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 additional weight-of-evidence or for other reasons.

564 **2.6.2 Reliability scores**

565 Reliability, in this context, is defined as the inherent quality of the experimental study (Klimisch, 1997)
566 and/or *in silico* analysis. It is used to support any hazard assessment, in combination with other
567 information. A reliability score (RS) is associated with the toxicological effect or mechanism assessment
568 (as shown in Figure 1). As noted earlier, when data from the literature or other sources are considered,
569 Klimisch scores can be used to assess the reliability of the results. However, the Klimisch framework was
570 never intended to assess the reliability of *in silico* predictions. It is also important to note that regardless
571 of the approach taken, reliability assessments will contain subjective decisions.

572 A number of general factors can affect the reliability of *in silico* results:

- 573 • **Multiple *in silico* results:** Combining results from multiple complementary or independent *in*
574 *silico* tools which use different methodologies or QSAR descriptors and/or training sets, has
575 been shown to improve overall sensitivity, but it can lower specificity by increasing false positive
576 rates (Myatt et al., 2016). In the case of quantitative predictions, such process are overly
577 conservative estimates. Hence, consistency across several different models can increase the
578 reliability of the results.
- 579 • **Expert review:** A plausible and well-documented read-across (consistent with the RAAF or
580 similar frameworks) may be acceptable as part of a REACH regulatory submission as an
581 alternative to experimental data. A structured expert review is implicit in any read-across
582 assessment (as discussed in Section 2.4.4). Similarly, an explicit expert review (following the
583 elements described in Sections 2.4.2 and 2.4.3) of the *in silico* predictions can improve the
584 reliability of the final results, especially for negative predictions. (Dobo et al., 2012)

585 To generate an overall reliability score for assessments based on experimental data and/or *in silico*
586 predictions, the Klimisch score has been adapted (as shown in Figure 2) to include an assessment of *in*
587 *silico* prediction results.

588 Experimental data assigned a Klimisch score of 1 or 2 is assigned a score of RS1 and RS2, respectively, in
589 this revised scheme. *In silico* results are not assigned a score of RS1 or RS2 since adequate experimental
590 data is preferred over *in silico* predictions. Since *in silico* results may be used directly as part of certain
591 regulatory submissions, whereas experimental data with a Klimisch score of 3 or 4 would not (or only as
592 supporting data under REACH, for example), the next two categories (RS3 and RS4) represent, in part, *in*
593 *silico* predictions. The following may be acceptable as part of a regulatory submission: (1) an adequately
594 performed read-across prediction (EU 2006), or (2) an expert review of *in silico* and/or other
595 experimental data (ICH M7, 2017(R1); EU 2006); they are assigned a reliability score of RS3. A score of
596 RS4 would be assigned when two or more predictive models are available that are complementary, with
597 concurring results (with no expert review), and no supporting literature data are available. Examples
598 include those predictive models that use either substantially different QSAR descriptors and/or QSAR
599 training sets or different *in silico* methodologies. If two or more *in silico* model results do not agree, then
600 an expert review would be required to assess the results. This review might increase the confidence in
601 the assessment, resulting in an increased reliability score of RS3. A single acceptable (as discussed in
602 Section 2.3.2) *in silico* model result, without further expert review, is afforded the same reliability score
603 of RS5 as an actual test result of lowest reliability (Klimisch 3 or 4). The *in silico* result is placed in the
604 same category as low reliability data because such models inform decisions based on a series of
605 compounds or trends. However, this reliability score may be increased following expert review. This
606 reliability score closely follows the ICH M7 guideline, where submissions corresponding to reliability
607 scores RS1-RS4 would be accepted according to the guideline. In addition to this score, it may be helpful

608 to document any additional considerations that may be important to the overall assessment. Individual
609 IST protocols may deviate from this scheme with appropriate justification.

610 **2.6.3 Worked examples**

611 Three examples from Amberg et al. (2016) illustrate how the framework described in this publication
612 can be used for determining a toxicological effect or mechanism assessment and reliability score, based
613 on experimental data and/or *in silico* predictions. Assessing reliability is an initial step in the overall
614 assessment of hazard, where it will be combined with other information, including an evaluation of the
615 relevance of the information, to support decision making.

616 In the example in Figure 3, no experimental data were identified. Two *in silico* models were run; the
617 statistical-based model prediction was negative and the expert rule-based alert prediction was negative.
618 The initial score would be RS4 based on multiple concurring prediction results; however, an expert
619 review was performed on the results from both methodologies and the negative result was confirmed
620 with increased reliability. The review concluded there were no potentially reactive features in the
621 chemical. This resulted in a negative overall assessment and a reliability score of RS3 (as a result of the
622 expert review increasing the reliability).

623 In the example in Figure 4, no experimental data were identified. Two *in silico* models were run; the
624 statistical model prediction was positive and the expert alert prediction was positive. No expert review
625 of the results was performed. The overall assessment was therefore positive and a reliability score of
626 RS4 was assigned as a result of two concurring positive predictions using complementary *in silico*
627 methodologies but without expert review.

628 In the example in Figure 5, no experimental data were identified. Two *in silico* models were run; the
629 statistical model prediction was positive and the expert alert prediction was negative. An expert review

630 was performed on the results from both methodologies, refuting the statistical model's positive
631 prediction. This review was based on an analysis of the test chemical's potential to react with DNA and
632 the highlighted structural feature was determined to be irrelevant for the mechanism of interaction with
633 DNA. This resulted in a negative overall assessment and a reliability score of RS3 (as a result of the
634 expert review increasing the reliability).

635 **2.7 Hazard assessment framework**

636 **2.7.1 Toxicological endpoints**

637 Figure 6 illustrates a general scheme for the prediction of a major toxicological endpoint. In this scheme,
638 the specific toxicological effects or mechanisms are used to support the assessment of a series of
639 toxicological endpoints. These toxicological endpoint assessments are, in turn, used in the overall
640 assessment of the major toxicological endpoint. In Figure 6, effect/mechanism 1 is identified as being
641 relevant to an assessment of a specific toxicological endpoint (Endpoint 1). For example, bacterial gene
642 mutation (effect/mechanism 1) is relevant to the assessment of gene mutation (endpoint 1). Endpoint 1
643 is, in turn, one of the endpoints that are relevant to the major toxicological endpoint (e.g., genetic
644 toxicity). Other identified toxicological effects or mechanisms are associated with toxicological
645 endpoints as shown in Figure 6. For example, the mammalian gene mutation (effect/mechanism 2) is
646 also relevant to the assessment of gene mutations (endpoint 1) and clastogenicity (endpoint 2) is
647 another endpoint to be used in the assessment of genetic toxicity (a major toxicological endpoint).
648 Figure 6 also includes another example to illustrate how this scheme might be used to assess male
649 reproductive toxicity.

650 The hazard assessment framework scheme for each IST protocol will contain different numbers of
651 toxicological endpoints as needed to support the assessment of each major toxicological endpoint in a
652 complete and transparent manner.

653 It is noteworthy that only the toxicological endpoints required to support a particular problem
654 formulation need to be assessed. For example, in certain applications only an assessment of gene
655 mutation may be needed (i.e., it may not be necessary to compute clastogenicity or the genetic toxicity
656 major toxicological endpoint).

657 **2.7.2 Relevance**

658 Relevance, in this context, is defined as the scientific predictivity of the each toxicological effect or
659 mechanism for the purpose of assessing a specific toxicological endpoint. As shown in Figure 6, the
660 assessment of toxicological endpoints may be based on the associated toxicological effects or
661 mechanisms. To support a transparent overall analysis, the relevance of the toxicological
662 effect/mechanism information in support of the assessment of the associated toxicological endpoint will
663 be defined in the IST protocols. This relevance will be based on the collective experience of the
664 consortium and available validation information.

665 **2.7.3 Toxicological endpoint assessment**

666 The assessment of each toxicological endpoint (as shown in Figure 6) is a function of all associated
667 toxicological effects or mechanisms and, in some cases, other toxicological endpoints. For example, in
668 Figure 6, bacterial gene mutation and mammalian gene mutation (toxicological effects or mechanisms)
669 are associated with gene mutation, whereas gene mutation and clastogenicity (both toxicological
670 endpoints) are associated with genetic toxicity. Rules or general principles for combining all associated
671 results for each endpoint will be defined in the upcoming IST protocols. For example, a rule may state
672 that if one of the associated effects/mechanisms is positive then the endpoint assessment is positive.
673 These rules or principles will take into consideration how combinations of different toxicological
674 effects/mechanisms are evaluated to generate an assessment for any toxicological endpoint which may
675 include a sequence of steps and incorporate Boolean logic.

676 **2.7.4 Toxicological endpoint confidence**

677 Confidence, in this context, is defined as a score that combines the reliability and relevance of the
678 associated toxicological effects or mechanisms. This is an additional score associated with toxicological
679 endpoints. The score may, in some cases, use other toxicological endpoint confidence scores (as shown
680 in Figure 6). This score will also take into consideration the completeness of the information available;
681 for example, the confidence score may be lowered when information on an effect or mechanism is
682 missing. It will also include complementary effects or mechanisms that need to be considered. This
683 score will be generated based on a series of general principles and/or rules defined in each IST protocol.
684 Each protocol will outline the different confidence values to generate, such as high, medium or low.

685 A confidence score is one of the most important items to generate. Different decision contexts tolerate
686 a different level of confidence in the assessment result as exemplified in the following two scenarios.

687 1) *Scenario 1*. The decision is to prioritize a large number of chemicals to screen as part of
688 product development. In this scenario, selecting a small subset of compounds using *in silico*
689 methods supports strategic resource utilization with the eventual goal of reducing overall
690 costs.

691 2) *Scenario 2*. A regulatory submission for a new cosmetic ingredient is being prepared based
692 on results from *in silico* methods.

693 Although in both scenarios, toxicological endpoint assessments generated at the highest level of
694 confidence would be preferable, Scenario 1 could still make beneficial use of lower confidence
695 predictions because the safety consequences of a false negative is lower than in Scenario 2. Therefore, a
696 risk assessment which takes into account the acceptable tolerance for a wrong prediction can be used to
697 evaluate the necessity for high confidence.

698 The assignment of the confidence score for each toxicological endpoint has to support the decision
699 context(s), regulatory framework and the type of product being assessed. Minimum confidence scores
700 for regulatory purposes may need to be set; however for other applications, the use of these scores may
701 be based on the individual organization's risk tolerance or based on the context, a decision on the
702 maximum permitted effort to be expended (since higher confidence score may be generated with
703 additional resources), or an organization's internal policy for using the confidence scores for specific
704 tasks.

705 **2.7.5 Expert review of toxicological endpoints**

706 In certain situations, an expert review of the toxicological endpoint assessment and/or confidence may
707 be warranted, and specific points to consider as part of such an expert review will be detailed in the
708 individual IST protocols. This review may take into consideration the context of the assessment, that is,
709 the type of product being assessed and any potential regulatory framework. It may be helpful to
710 document any additional considerations concerning the assessment and confidence to support an
711 overall assessment.

712 **2.8 *In silico* toxicology protocol components**

713 Ongoing efforts are concentrated on the development of individual IST protocols for major endpoints
714 including genetic toxicity, carcinogenicity, acute toxicity, repeated dose toxicity, reproductive toxicity,
715 and developmental toxicity. Table 5 outlines proposed common components for these IST protocols.

716 **2.9 Reporting formats**

717 Standardized reporting of the results and expert review is good scientific practice and assures that when
718 such information is communicated to regulatory authorities, it is complete, consistent and transparent;

719 this may avoid requests for additional information and maintain a consistent, expedient, and streamline
720 regulatory review process. Table 6 outlines a proposed structure for the report format.

721 The proposed report format is more comprehensive than existing data formats by including information
722 on overall assessment and expert reviews. For example, the “QSAR prediction reporting format” (QPRF;
723 JRC 2014) could be used to report the individual model results (as shown in Section D of Table 6), or
724 “QSAR model reporting format” (QMRF) can be used to report the QSAR model’s details (as shown in
725 Section H of Table 6).

726 The new proposed report format collects enough details on how the predictions were generated to
727 enable another expert to repeat the process. It is also important that the reasoning and decisions of the
728 expert review steps are transparently documented and can be retraced at any time, including the
729 information used as their basis for conclusions.

730 **3. Summary and outlook**

731 IST is poised to play an increasingly significant role in the assessment of chemicals in a range of chemical
732 exposure scenarios that have the potential to impact public health. Thus, this is an opportune time for
733 the development of IST protocols. As expected, the quality and quantity of experimental data will vary
734 as will the available *in silico* methods. For example, experimental data could be from a variety of
735 sources, studies, protocols and laboratories using or not using GLP standards. Similarly, several *in silico*
736 methods and approaches are available for assessment of toxicity. Thus, accepted selection criteria have
737 to be defined for experimental data and *in silico* methods, for consistent and uniform use. The
738 development of IST protocols will support the use and adoption of *in silico* methods in the same manner
739 in which *in vitro* and *in vivo* test guidelines support the use and adoption of those assays.

740 Figure 7 summarizes the steps to perform an *in silico* assessment consistent with the framework defined
741 in this publication. The key elements needed for the development of IST protocols are outlined in this
742 publication, including: 1) how to select, assess and integrate *in silico* predictions alongside experimental
743 data for defined toxicological effects or mechanisms, including a new methodology for establishing the
744 reliability of this assessment, 2) a hazard assessment framework for systematic assessment of these
745 toxicological effects or mechanisms to predict specific endpoints and assess the confidence in the
746 results. Wherever possible, this is based on mechanistic knowledge on different biological levels of
747 organization. (Bell et al., 2016; OECD 2016a; OECD 2016b) Overall, the IST protocols will contain
748 information to ensure predictions are performed in a consistent, repeatable, transparent and ultimately
749 accepted manner and will include a checklist (as defined in Section 2.4) to guide an expert review of the
750 information. Each individual IST protocol will address how predictions will be performed in alignment
751 with the framework discussed in this publication. These new protocols will provide specific guidance for
752 each toxicological endpoint, including situations where no AOP or IATA is currently available. These
753 protocols build on and fully incorporate wherever possible the considerable work previously reported,
754 such as the OECD validation principles (see Sections 2.3.2), IATAs (see Sections 2.2), AOPs (see Sections
755 2.2), read-across frameworks (see Sections 2.3.2, 2.6.2), the Klimisch score (see Sections 2.5, 2.6.1,
756 2.6.2) and the QMRF/QPRF (see Sections 2.3.2, 2.9).

757 The IST protocols do not define how a risk assessment will be performed; they solely define the process
758 which will lead to the prediction of the potential toxicity (hazard) of a chemical. Risk analysis depends on
759 the exposure scenario, industry, regulatory framework and decision context based on the level of
760 tolerated uncertainty and is performed in the hands of an expert.

761 The process of developing IST protocols requires an understanding of the best practices and science
762 across various organizations, different industries and regulatory authorities. To develop such protocols,

763 an international consortium was established comprising regulators, government agencies, industry,
764 academics, model developers, and consultants across many different sectors. This consortium initially
765 developed the overall strategy outlined in this publication. Working subgroups will develop individual
766 IST protocols for major endpoints including genetic toxicity, carcinogenicity, acute toxicity, reproductive
767 toxicity, and developmental toxicity. As each IST protocol is established, it will be reviewed internally
768 within each organization and published. This process will evolve over time, as computational technology
769 progresses, as will the assays and other information relevant to assessing these major endpoints
770 emerges. Hence, similar to other test guidelines, the IST protocols will need to be periodically reviewed
771 and updated. The implementation of IST protocols will also require user-friendly tools for performing
772 such analyses and reporting the results, education, as well as further collaboration with organizations to
773 support global adoption.

774 **Acknowledgements**

775 Research reported in this publication was supported by the National Institute of Environmental Health
776 Sciences of the National Institutes of Health under Award Number R43ES026909. The content is solely
777 the responsibility of the authors and does not necessarily represent the official views of the National
778 Institutes of Health. The authors would like to thank Sachin Bhusari and George Pugh from the Coca-
779 Cola Company for their input into the manuscript.

780

781 **References**

- 782 Alves, V., Muratov, E., Capuzzi, S., Politi, R., Low, Y., Braga, R., Zakharov, A.V., Sedykh, A., Mokshyna, E.,
783 Farag, S., Andrade, C., Kuz'min, D., Fourches, D., Tropsha, A. Alarms about structural alerts. *Green Chem.*
784 2016 Aug 21; 18(16): 4348–4360. Doi: 10.1039/C6GC01492E
- 785 Amaral, R.T.D., Ansell, J., Aptula, N., Ashikaga, T.; Chaudhry, Q.; Hirose, A., Jaworska, J., Kojima, H.,
786 Lafranconi, M., Matthews, E., Milstein, S., Roesler, C., Vaillancourt, E., Verma, R., Worth, A., Yourick J.,
787 2014. Report for the International Cooperation on Cosmetics Regulation. In *Silico Approaches for Safety*
788 *Assessment of Cosmetic Ingredients*.
789 <https://www.pharmamedtechbi.com/~media/Supporting%20Documents/The%20Rose%20Sheet/36/IC>
790 [CR%20In%20Silico%20Report.pdf](https://www.pharmamedtechbi.com/~media/Supporting%20Documents/The%20Rose%20Sheet/36/IC)
- 791 Amaral, R., Amores Da Silva, P., Ansell, J., Boislevé, F., Dent, M., Hatao, M., Hirose, A., Kasai, Y., Kojima,
792 H., Kern, P., Kreiling, R., Milstein, S., Oliveira, J., Richarz, A., Taalman, R., Vaillancourt, E., Verma, R.,
793 Vieira, N.C., Weiss, C. Report for the International Cooperation on Cosmetics Regulation, Joint
794 Regulators-Industry Working Group: Integrated Strategies for Safety Assessments of Cosmetic
795 Ingredients - Part I. <http://www.iccr-cosmetics.org/files/4715/0824/0761/ICCR->
796 [11_JWG_Integrated_Strategies_Integrated_Strategies_for_Safety_Assessments_of_Cosmetic_Ingredien](http://www.iccr-cosmetics.org/files/4715/0824/0761/ICCR-)
797 [ts_-_Part_I.pdf](http://www.iccr-cosmetics.org/files/4715/0824/0761/ICCR-)
- 798 Amberg, A., Harvey, J.S., Czich, A., SpirkI, H.-P., Robinson, S., White, A., Elder, D.P. 2015. Do
799 Carboxylic/Sulfonic acid halides really present a mutagenic and carcinogenic risk as impurities in final
800 drug products? *Org. Process Res. Dev.* 19, 1495 - 1506. doi: 10.1021/acs.oprd.5b00106
- 801 Amberg, A., Beilke, L., Bercu, J., Bower, D., Brigo, A., Cross, K.P., Custer, L., Dobo, K., Dowdy, E., Ford,
802 K.A., Glowienke, S., Gompel, J.V., Harvey, J., Hasselgren, C., Honma, M., Jolly, R., Kemper, R., Kenyon, M.,

803 Kruhlak, N., Leavitt, P., Miller, S., Muster, W., Nicolette, J., Plaper, A., Powley, M., Quigley, D.P., Reddy,
804 M.V., SpirkI, H.-P., Stavitskaya, L., Teasdale, A., Weiner, S., Welch, D.S., White, A., Wichard, J., Myatt,
805 G.J., 2016. Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses.
806 *Regulatory Toxicology and Pharmacology* 77, 13–24. doi:10.1016/j.yrtph.2016.02.004

807 Barber, C., Amberg, A., Custer, L., Dobo, K.L., Glowienke, S., Gompel, J.V., Gutsell, S., Harvey, J., Honma,
808 M., Kenyon, M.O., Kruhlak, N., Muster, W., Stavitskaya, L., Teasdale, A., Vessey, J., Wichard, J., 2015.
809 Establishing best practise in the application of expert review of mutagenicity under ICH M7. *Regulatory*
810 *Toxicology and Pharmacology* 73, 367–377. doi:10.1016/j.yrtph.2015.07.018

811 Ball, N., Cronin, M.T., Shen, J., Blackburn, K., Booth, E.D., Bouhifd, M., Donley, E., Egnash, L., Hastings, C.,
812 Juberg, D.R., Kleensang, A., Kleinstreuer, N., Kroese, E.D., Lee, A.C., Luechtefeld, T., Maertens, A., Marty,
813 S., Naciff, J.M., Palmer, J., Pamies, D., Penman, M., Richarz, A.N., Russo, D.P., Stuard, S.B., Patlewicz, G.,
814 van Ravenzwaay, B., Wu, S., Zhu, H., Hartung, T., 2016. Toward Good Read-Across Practice (GRAP)
815 guidance. *ALTEX* 33, 149-66. <http://dx.doi.org/10.14573/altex.1601251>

816 Ball, D. and Norwood, D., 2012 *Leachables and Extractables Handbook*. Wiley. Pages 58-79.

817 Bassan, A., Worth, A.P., 2008. The Integrated Use of Models for the Properties and Effects of Chemicals
818 by means of a Structured Workflow. *QSAR & Combinatorial Science* 27, 6–20.
819 doi:10.1002/qsar.200710119

820 Bell, S.M., Angrish, M.M., Wood, C.E., Edwards, S.W., 2016. Integrating Publicly Available Data to
821 Generate Computationally Predicted Adverse Outcome Pathways for Fatty Liver. *Toxicological Sciences*
822 150, 510–520. doi:10.1093/toxsci/kfw017

823 Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A.-N., Bois, F.Y., Exner, T., Leite, S., Grunsvén,
824 L.A.V., Worth, A., Mahony, C., 2017. Ab initio chemical safety assessment: A workflow based on
825 exposure considerations and non-animal methods. *Computational Toxicology* 4, 31–44.
826 doi:10.1016/j.comtox.2017.10.001

827 Blackburn, K., Stuard, S.B., 2014. A framework to facilitate consistent characterization of read across
828 uncertainty. *Regulatory Toxicology and Pharmacology* 68, 353–362. doi:10.1016/j.yrtph.2014.01.004

829 Bossuyt, M.V., Hoeck, E.V., Raitano, G., Manganelli, S., Braeken, E., Ates, G., Vanhaecke, T., Miert, S.V.,
830 Benfenati, E., Mertens, B., Rogiers, V., 2017. (Q)SAR tools for priority setting: A case study with printed
831 paper and board food contact material substances. *Food and Chemical Toxicology* 102, 109–119.
832 doi:10.1016/j.fct.2017.02.002

833 Bower, D., Cross, K.P., Esche, S., Myatt, G.J., Quigley, D.P., 2017. *In silico* Toxicology: An Overview of
834 Toxicity Databases, Prediction Methodologies, and Expert Review, in: Richardson R.J., Johnson, D.E.,
835 (Eds.) *Computational Systems Pharmacology and Toxicology*. Royal Society of Chemistry.
836 DOI:10.1039/9781782623731-00209

837 Canada 2016. Chemicals Management Plan. Government of Canada. [https://www.canada.ca/en/health-](https://www.canada.ca/en/health-canada/services/chemical-substances/chemicals-management-plan.html)
838 [canada/services/chemical-substances/chemicals-management-plan.html](https://www.canada.ca/en/health-canada/services/chemical-substances/chemicals-management-plan.html)

839 Carrió, P., Pinto, M., Ecker, G., Sanz, F., Pastor, M., 2014. Applicability Domain Analysis (ADAN): A Robust
840 Method for Assessing the Reliability of Drug Property Predictions. *Journal of Chemical Information and*
841 *Modeling* 54, 1500–1511. doi:10.1021/ci500172z

842 CDRH 2016. Use of International Standard ISO 10993-1. Biological evaluation of medical devices - Part 1:
843 Evaluation and testing within a risk management process. Guidance for Industry and Food and Drug

Formatted: Spanish (Spain)

844 Administration Staff. June 16, 2016
845 [https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/uc](https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm348890.pdf)
846 [m348890.pdf](https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm348890.pdf)

847 Chen, B., Zhang, T., Bond, T., Gan, Y., 2015. Development of quantitative structure activity relationship
848 (QSAR) model for disinfection byproduct research: A review of methods and resources, J. Hazard Mater.
849 299, 260-279. Doi: 10.1016/j.jhazmat.2015.06.054

850 Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P.,
851 Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J.,
852 Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., 2014 QSAR modeling: where have you been? Where
853 are you going to? J Med Chem. 57, 4977-5010. doi: 10.1021/jm4004285

854 Dobo, K.L., Greene, N., Fred, C., Glowienke, S., Harvey, J.S., Hasselgren, C., Jolly, R., Kenyon, M.O.,
855 Munzner, J.B., Muster, W., Neft, R., Reddy, M.V., White, A.T., Weiner, S., 2012. *In silico* methods
856 combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry
857 survey. Regul. Toxicol. Pharmacol. 62, 449-455. doi: 10.1016/j.yrtph.2012.01.007

858 ECETOC 2012: ECETOC Technical Report No. 116: Category approaches, Read-across, (Q)SAR.
859 [http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-116-Category-approaches-Read-](http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-116-Category-approaches-Read-across-QSAR.pdf)
860 [across-QSAR.pdf](http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-116-Category-approaches-Read-across-QSAR.pdf)

861 ECHA 2008. Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs
862 and grouping of chemicals.
863 [https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf/77f49f81-](https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf/77f49f81-b76d-40ab-8513-4f3a533b6ac9)
864 [b76d-40ab-8513-4f3a533b6ac9](https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf/77f49f81-b76d-40ab-8513-4f3a533b6ac9)

865 ECHA 2011. Guidance on information requirements and chemical safety assessment Chapter R.4:
866 Evaluation of available Information.
867 [https://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf/d6395ad2-](https://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf/d6395ad2-1596-4708-ba86-0136686d205e)
868 1596-4708-ba86-0136686d205e

869 ECHA 2015. Guidance on the Application of the CLP Criteria Guidance to Regulation (EC) No 1272/2008
870 on classification, labelling and packaging (CLP) of substances and mixtures.
871 [https://echa.europa.eu/documents/10162/23036412/clp_en.pdf/58b5dc6d-ac2a-4910-9702-](https://echa.europa.eu/documents/10162/23036412/clp_en.pdf/58b5dc6d-ac2a-4910-9702-e9e1f5051cc5)
872 e9e1f5051cc5

873 ECHA 2016. Practical guide: How to use and report (Q)SARs.
874 https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

875 ECHA 2017a. Weight of evidence. [https://echa.europa.eu/support/registration/how-to-avoid-](https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/weight-of-evidence)
876 unnecessary-testing-on-animals/weight-of-evidence

877 ECHA 2017b. Read-Across Assessment Framework (RAAF).
878 https://echa.europa.eu/documents/10162/13628/raaf_en.pdf

879 ECVAM 2017. ToxRTool - Toxicological data Reliability Assessment Tool. [https://eurl-](https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool/toxrtool-toxicological-data-reliability-assessment-tool)
880 [ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool/toxrtool-toxicological-data-](https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/toxrtool/toxrtool-toxicological-data-reliability-assessment-tool)
881 reliability-assessment-tool

882 EFSA 2011. European Food Safety Authority; Submission of scientific peer-reviewed open literature for
883 the approval of pesticide active substances under Regulation (EC) No 1107/2009. EFSA Journal 9, 1-49.
884 doi:10.2903/j.efsa.2011.2092. <https://www.efsa.europa.eu/en/efsajournal/pub/2092>

885 EFSA 2014. Modern methodologies and tools for human hazard assessment of chemicals. EFSA Journal
886 2014;12 1-87.doi:10.2903/j.efsa.2014.3638. <https://www.efsa.europa.eu/de/efsajournal/pub/3638>

887 EFSA 2016. Guidance on the establishment of the residue definition for dietary risk assessment: EFSA
888 Panel on Plant Protection Products and their Residues (PPR), EFSA Journal 14, 1-12.
889 <http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2016.4549/epdf>

890 EFSA 2017a. European Food Safety Authority. <http://www.efsa.europa.eu/>

891 EFSA 2017b. Public consultation on Guidance on The Use of the Weight of Evidence Approach in
892 Scientific Assessments. <https://www.efsa.europa.eu/en/consultations/call/170306-0>

893 Ellison, C.M., Piechota, P., Madden, J.C., Enoch, S.J., Cronin, M.T.D., 2016. Adverse Outcome Pathway
894 (AOP) Informed Modeling of Aquatic Toxicology: QSARs, Read-Across, and Interspecies Verification of
895 Modes of Action. Environmental Science & Technology 50, 3995–4007. doi:10.1021/acs.est.5b05918

896 EU 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the council of 18 December
897 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).
898 <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02006R1907-20161011&from=EN>

899 EU 2009a. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November
900 2009 on cosmetic products (recast) (Text with EEA relevance). [http://eur-](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:342:0059:0209:en:PDF)
901 [lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:342:0059:0209:en:PDF](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:342:0059:0209:en:PDF)

902 EU 2009b. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October
903 2009 concerning the placing of plant protection products on the market and repealing Council Directives
904 79/117/EEC and 91/414/EEC. [http://eur-lex.europa.eu/legal-](http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32009R1107)
905 [content/en/TXT/?uri=CELEX%3A32009R1107](http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32009R1107)

Formatted: French (France)

906 EU 2012. Guidance Document on the Assessment of the Equivalence of Technical Materials of
907 Substances Regulated UNDER Regulation (EC) No 1107/2009 SANCO/10597/2003 –rev. 10.1
908 [13https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_guidance_equivalence-chem-](https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_guidance_equivalence-chem-)
909 [substances_en.pdf](https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_guidance_equivalence-chem-substances_en.pdf)

910 Ford, K.A., 2016. Refinement, Reduction, and Replacement of Animal Toxicity Tests by Computational
911 Methods. *ILAR J.* 57 (2), 226-233. doi: 10.1093/ilar/ilw031

912 Freidig, A., Dekkers, S., Verwei, M., Zvinavashe, E., Bessems, J., Sandt, J.V.D., 2007. Development of a
913 QSAR for worst case estimates of acute toxicity of chemically reactive compounds. *Toxicology Letters*
914 170, 214–222. doi:10.1016/j.toxlet.2007.03.008

915 Harvey, J., Fleetwood, A., Ogilvie, R., Teasdale, A., Wilcox, P., Spanhaak, S., 2017. Management of
916 organic impurities in small molecule medicinal products: Deriving safe limits for use in early
917 development. *Regulatory Toxicology and Pharmacology* 84, 116–123. doi:10.1016/j.yrtph.2016.12.011

918 Hillisch, A., Heinrich, N., Wild, H., 2015. Computational Chemistry in the Pharmaceutical Industry: From
919 Childhood to Adolescence. *ChemMedChem* 10, 1958–1962. doi:10.1002/cmdc.201500346

920 Hochstein, C., Arnesen, S., Goshorn, J., Szczur, M., 2008. Selected Resources for Emergency and Disaster
921 Preparedness and Response from the United States National Library of Medicine. *Medical Reference*
922 *Services Quarterly* 27, 1–20. doi:10.1300/j115v27n01_01

923 ICH M7, 2017 (R1). Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals
924 to limit potential carcinogenic risk.
925 http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1
926 [_Addendum_Step_4_31Mar2017.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1_Addendum_Step_4_31Mar2017.pdf)

927 ICH 2017. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for
928 Human Use. <http://www.ich.org/products/guidelines.html>

929 JRC 2014. JRC QSAR Model Database and QSAR Model Reporting Formats. [https://eurl-](https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/qrf)
930 [ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/qrf](https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/qrf) and
931 <http://qsar.db.jrc.it/qmrf/>

932 Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A Systematic Approach for Evaluating the Quality of
933 Experimental Toxicological and Ecotoxicological Data. *Regulatory Toxicology and Pharmacology* 25, 1–5.
934 doi:10.1006/rtph.1996.1076

935 Kruhlak, N.L., Benz, R.D., Zhou, H., Colatsky, T.J., 2012. (Q)SAR modeling and safety assessment in
936 regulatory review. *Clin Pharmacol Ther.* 91, 529-34. doi: 10.1038/clpt.2011.300

937 Mansouri, K., Grulke, C.M., Richard, A.M., Judson, R.S., Williams, A.J., 2016. An automated curation
938 procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling.
939 *SAR and QSAR in Environmental Research* 27, 911–937. doi:10.1080/1062936x.2016.1253611

940 Martin, T., Young, D., Lilavois, C., Barron, M., 2015. Comparison of global and mode of action-based
941 models for aquatic toxicity. *SAR and QSAR in Environmental Research* 26, 245–262.
942 doi:10.1080/1062936x.2015.1018939

943 Molander, L., Ågerstrand, M., Beronius, A., Hanberg, A., Rudén, C., 2014. Science in Risk Assessment and
944 Policy (SciRAP): An Online Resource for Evaluating and Reporting In Vivo(Eco)Toxicity Studies. *Human
945 and Ecological Risk Assessment: An International Journal* 21, 753–762.
946 doi:10.1080/10807039.2014.928104

947 Mumtaz, M.M., Suk, W.A., Yang, R.S.H., 2010. Introduction to Mixtures Toxicology and Risk Assessment.
948 in: Mumtaz, M. (Ed.) Principles and Practice of Mixtures Toxicology. Mumtaz, M (Ed.) Wiley-VCH,
949 Weinheim, Germany, p1-25. doi: 10.1002/9783527630196.ch1

950 Mumtaz, M.M., Durkin, P.R., 1992. A weight of evidence scheme for assessing interactions in chemical
951 mixtures. *Toxicol. Indus. Health.* 8, 377-406.

952 Myatt, G.J., Beilke, L.D., Cross, K.P., 2016. *In Silico* Tools and their Application in: Reedijk, J. (Ed.)
953 Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Elsevier. doi:
954 <https://doi.org/10.1016/B978-0-12-409547-2.12379-0>

955 NAFTA 2012, TWG Quantitative Structure Activity Relationships [(Q)SAR] Guidance Document.
956 <https://archive.epa.gov/pesticides/news/web/pdf/qsar-guidance.pdf>

957 Netzeva, T.I., Worth, A.P. Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn,
958 S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts,
959 D.W., Schultz, T.W., Stanton, D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C., 2005. Current
960 Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity
961 Relationships. The Report and Recommendations of ECVAM Workshop 52. *ATLA* 33 155-173.

962 NIEHS 2017. National Institute of Environmental Health Sciences. <https://www.niehs.nih.gov/index.cfm>

963 NTP 2016. West Virginia Chemical Spill: NTP Studies.
964 <https://ntp.niehs.nih.gov/results/areas/wvspill/studies/>

965 OCSPP 2017. Office of Chemical Safety and Pollution Prevention (OCSPP).
966 <https://www.epa.gov/aboutepa/about-office-chemical-safety-and-pollution-prevention-ocspp>

967 OECD 2004. The report from the expert group on (Quantitative) Structure-Activity Relationships
968 [(Q)SARs] on the principles for the validation of (Q)SARs, No. 49 (ENV/JM/MONO(2004)24).
969 [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2004\)24](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2004)24)
970 &doclanguage=en

971 OECD 2007. Guidance Document on the Validation of (Quantitative) Structure-activity Relationships
972 [(Q)SAR] Models, OECD Environment Health and Safety Publications Series on Testing and Assessment
973 No. 69 (ENV/JM/MONO(2007)2). [http://www.oecd.org/env/guidance-document-on-the-validation-of-](http://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm)
974 [quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm](http://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm)

975 OECD 2014. Guidance on grouping of chemicals, second edition. OECD Environment Health and Safety
976 Publications Series on Testing & Assessment. No. 194 ENV/JM/MONO(2014)4.
977 [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2014\)4&docla](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en)
978 [nguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en)

979 OECD 2015. Fundamental And Guiding Principles For (Q)SAR Analysis Of Chemical Carcinogens with
980 Mechanistic Considerations, Monograph 229 (ENV/JM/MONO(2015)46) , Series on Testing and
981 Assessment No. 229.
982 [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2015\)46&docl](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)46&doclanguage=en)
983 [anguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2015)46&doclanguage=en)

984 OECD 2016a. Guidance document on the reporting of defined approaches to be used within integrated
985 approaches to testing and assessment, Series on Testing & Assessment No. 255,
986 ENV/JM/MONO(2016)28.
987 [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2016\)28&docl](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)28&doclanguage=en)
988 [anguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)28&doclanguage=en)

Formatted: Spanish (Spain)

989 OECD 2016b. Guidance document on the reporting of defined approaches and individual information
990 sources to be used within integrated approaches to testing and assessment (IATA) for skin sensitisation.
991 ENV/JM/MONO(2016)29. Series on Testing & Assessment No. 256.
992 [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2016\)29&docl](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29&docl)
993 [language=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29&doclanguage=en)

994 OECD 2016c. Draft guidance document on good in vitro methods Practices (GIVIMP) for the
995 development and implementation of in vitro methods for regulatory use in human safety assessment.
996 http://www.oecd.org/env/ehs/testing/OECD_Draft_GIVIMP_in_Human_Safety_Assessment.pdf

997 OECD 2017. OECD Guidelines for the Testing of Chemicals.
998 <http://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm>

999 Patlewicz, G., Ball, N., Booth, E.D., Hulzebos, E., Zvinavashe, E., Hennes, C., 2013. Use of category
1000 approaches, read-across and (Q)SAR: General considerations. *Regulatory Toxicology and Pharmacology*
1001 67, 1–12. doi:10.1016/j.yrtph.2013.06.002

1002 Patlewicz, G., Roberts, D.W., Aptula, A., Blackburn, K., Hubesch, B., 2013. Workshop: Use of “read-
1003 across” for chemical safety assessment under REACH. *Regulatory Toxicology and Pharmacology* 65, 226–
1004 228. doi:10.1016/j.yrtph.2012.12.004

1005 Patlewicz, G., 2014. Read-across approaches - misconceptions, promises and challenges ahead. *Altex* 31,
1006 387–396. doi:10.14573/altex.1410071

1007 Patlewicz, G., Ball, N., Boogaard, P., Becker, R., Hubesch, B., 2015. Building scientific confidence in the
1008 development and evaluation of read-across. *Regulatory Toxicology and Pharmacology* 72, 117–133.
1009 doi:10.1016/j.yrtph.2015.03.015

1010 Patlewicz, G., Worth, A.P., Ball, N., 2016. Validation of Computational Methods. *Advances in*
1011 *Experimental Medicine and Biology* Validation of Alternative Methods for Toxicity Testing 165–187.
1012 doi:10.1007/978-3-319-33826-2_6

1013 PMTA/FDA 2016. Premarket Tobacco Product Applications for Electronic Nicotine Delivery Systems.
1014 Guidance for Industry. U.S. Department of Health and Human Services, Food and Drug Administration,
1015 Center for Tobacco Products. May 2016.
1016 <https://www.fda.gov/downloads/tobaccoproducts/labeling/rulesregulationsguidance/ucm499352.pdf>

1017 Powley, M.W., 2015. (Q)SAR assessments of potentially mutagenic impurities: a regulatory perspective
1018 on the utility of expert knowledge and data submission. *Regul. Toxicol. Pharmacol.* 71, 295-300. doi:
1019 10.1016/j.yrtph.2014.12.012

1020 Raies, A.B., Bajic, V.B., 2016. *In silico* toxicology: computational methods for the prediction of chemical
1021 toxicity. *Wiley Interdiscip Rev Comput Mol Sci.* 6, 147-172. doi: 10.1002/wcms.1240

1022 Rastogi, T., Leder, C., Kümmerer, K., 2014. Designing green derivatives of β -blocker Metoprolol: A tiered
1023 approach for green and sustainable pharmacy and chemistry. *Chemosphere* 111, 493–499.
1024 doi:10.1016/j.chemosphere.2014.03.119

1025 Rooney, A.A., Boyles, A.L., Wolfe, M.S., Bucher, J.R., Thayer, K.A., 2014. Systematic review and evidence
1026 integration for literature-based environmental health science assessments. *Environ Health Perspect.*
1027 122, 711-8. doi:10.1289/ehp.1307972

1028 Russell, W.M.S., Burch, R.L., 1959. *The principles of humane experimental technique.* Methuen, London.

1029 SCCS 2016. Memorandum on the Use of *In Silico* Methods for Assessment of Chemical Hazard.
1030 SCCS/1578/16.
1031 https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_200.pdf
1032 SCCS 2017. Scientific Committee on Consumer Safety.
1033 https://ec.europa.eu/health/scientific_committees/consumer_safety_en
1034 Schilter, B., Benigni, R., Boobis, A., Chiodini, A., Cockburn, A., Cronin, M.T., Piparo, E.L., Modi, S., Thiel,
1035 A., Worth, A., 2014. Establishing the level of safety concern for chemicals in food without the need for
1036 toxicity testing. *Regulatory Toxicology and Pharmacology* 68, 275–296. doi:10.1016/j.yrtph.2013.08.018
1037 Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., Hartung,
1038 T., Hoffmann, S., 2009. “ToxRTool”, a new tool to assess the reliability of toxicological data. *Toxicology*
1039 *Letters* 189, 138–144. doi:10.1016/j.toxlet.2009.05.013
1040 Schultz, T., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D., Mahony, C., Schwarz, M., White,
1041 A., Cronin, M., 2015. A strategy for structuring and reporting a read-across prediction of toxicity.
1042 *Regulatory Toxicology and Pharmacology* 72, 586–601. doi:10.1016/j.yrtph.2015.05.016
1043 Schwetz, B., 1995. Use of mechanistic and pharmacokinetic data for risk assessment at the National
1044 Institute of Environmental Health Sciences (NIEHS). *Toxicology Letters* 79, 29–32. doi:10.1016/0378-
1045 4274(95)03354-n
1046 Seed, M.J., Agius, R.M., 2017. Progress with Structure-Activity Relationship modelling of occupational
1047 chemical respiratory sensitizers. *Curr Opin Allergy Clin Immunol.* 17, 64-71 doi:
1048 10.1097/ACI.0000000000000355

1049 Stanton, K., Kruszewski, F.H., 2016. Quantifying the benefits of using read-across and *in silico* techniques
1050 to fulfill hazard data requirements for chemical categories. *Regulatory Toxicology and Pharmacology* 81,
1051 250–259. doi:10.1016/j.yrtph.2016.09.004

1052 Sutter, A., Amberg, A., Boyer, S., Brigo, A., Contrera, J.F., Custer, L.L., Dobo, K.L., Gervais, V., Glowienke,
1053 S., Gompel, J.V., Greene, N., Muster, W., Nicolette, J., Reddy, M.V., Thybaud, V., Vock, E., White, A.T.,
1054 Müller, L., 2013. Use of *in silico* systems and expert knowledge for structure-based assessment of
1055 potentially mutagenic impurities. *Regul. Toxicol. Pharmacol.* 67, 39-52. doi: 10.1016/j.yrtph.2013.05.001

1056 TSCA 2016. Toxic Substances Control Act (TSCA). [https://www.congress.gov/bill/114th-congress/senate-](https://www.congress.gov/bill/114th-congress/senate-bill/697/all-info)
1057 [bill/697/all-info](https://www.congress.gov/bill/114th-congress/senate-bill/697/all-info)

1058 Worth, A., Barroso, J., Bremer, S., Burton, J., Casati, S., Coecke, S., Corvi, R., Desprez, B., Dumont, C.,
1059 Gouliarmou, V., Goumenou, M., Gräpel, R., Griesinger, C., Halder, M., Janusch Roi, A., Kienzler, A.,
1060 Madia, F., Munn, S., Nepelska, M., Paini, A., Price, A., Prieto, P., Rolaki, A., Schäffer, M., Triebe, J.,
1061 Whelan, M., Wittwehr, C., Zuang, V., 2014. Alternative methods for regulatory toxicology – a state-of-
1062 the-art review. JRC report EUR 26797 EN. Publications Office of the European Union.
1063 [http://publications.jrc.ec.europa.eu/repository/bitstream/JRC91361/echa_jrc_sla_report_public_05-09-](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC91361/echa_jrc_sla_report_public_05-09-14_withcover%20ipo.pdf)
1064 [14_withcover%20ipo.pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC91361/echa_jrc_sla_report_public_05-09-14_withcover%20ipo.pdf)

1065 Wu, S., Blackburn, K., Amburgey, J., Jaworska, J., Federle, T., 2010. A framework for using structural,
1066 reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based
1067 toxicological assessments. *Regulatory Toxicology and Pharmacology* 56, 67–81.
1068 doi:10.1016/j.yrtph.2009.09.006

Formatted: English (United Kingdom)

1069 **Table Legends**

1070 Table 1: Applications of *in silico* toxicology

1071 Table 2: Checklist of elements to consider as part of an expert review of a QSAR model result

1072 Table 3: Checklist of elements to consider as part of an expert review of results from expert rule-based

1073 Table 4: Summary of Klimisch scores for data reliability (adapted from Klimisch et al., 1997) (Note

1074 “restriction”, as part of scores 1 and 2, indicates restricted quality)

1075 Table 5: Common components of an IST protocol (IATA = Integrated Approaches to Testing and

1076 Assessment; AOP = Adverse Outcome Pathways)

1077 Table 6: Elements of an *in silico* toxicology report (QMRF = QSAR Model Reporting Format)

1078

1079 **Figure Legends**

1080 Figure 1: Overview of the IST protocol framework, showing how experimental data or *in silico* model(s)
1081 for each defined toxicological effect/mechanism are assessed and used to support a hazard assessment.
1082 (Note Effect/Mechanism N is used to illustrate that there can be any number of effects/mechanisms in
1083 each protocol)

1084 Figure 2: Reliability of toxicity assessments based computational models and experimental data

1085 Figure 3: Determining the bacterial gene mutation assessment and reliability score for two concurring *in*
1086 *silico* results with expert review

1087 Figure 4: Determining the bacterial gene mutation assessment and reliability score for two concurring *in*
1088 *silico* results with no expert review

1089 Figure 5: Determining the bacterial gene mutation assessment and reliability score where there is no
1090 experimental data available and conflicting *in silico* results

1091 Figure 6: Hazard assessment framework

1092 Figure 7: Summary of the IST protocol process

1093

1094

1095 **Tables**

1096

1097 **Table 1: Applications of *in silico* toxicology**

<i>In silico</i> toxicology application	Discussion
1. <i>Alternative to test data.</i>	The use of non-animal alternative methods including <i>in silico</i> approaches, may substitute for other types of tests in regulatory submissions in certain cases. Acceptable alternative methods for filling 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 21 st 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 classified and labeled for shipping or other purposes, <i>in silico</i> toxicology provides an alternative method for quickly filling the data gaps in the toxicity/safety information, such as predictions of acute toxicity to support assignment to the Globally Harmonized System of Classification and Labelling category (Freidig et al., 2007; ECHA 2015).
2. <i>As part of the weight-of-evidence in regulatory submissions.</i>	There are currently several regulatory frameworks where only specific laboratory tests for an endpoint of concern may be submitted (such as for drugs or food additives). However, in such cases, <i>in silico</i> predictions can be submitted alongside standard toxicological data to complement the assessment. This may include <i>in silico</i> assessments provided as supporting data or adjuncts to the primary <i>in vivo</i> or <i>in vitro</i> studies to give a mechanistic understanding of the observed results and/or allow a better definition of experimental needs. Additionally, <i>in silico</i> methods may be used to guide or prioritize <i>in vitro</i> 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 of alternative methods, such as IST, in the assessment of new cosmetics ingredients. In a recent memorandum, the European Commission's Scientific Committee for Consumer Safety (SCCS), which is responsible for the risk assessment of cosmetic ingredients, acknowledged the importance and limitations of <i>in silico</i> methods; the SCCS recommended that <i>in silico</i> 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. <i>Mixtures assessment.</i>	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, botanical drugs or natural products. In certain situations, it may be possible to use <i>in silico</i> methods to assess individual components since today's <i>in silico</i> analysis can only be performed on discrete identifiable 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 <i>in silico</i> 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 are multiple chemicals for interactions, such as synergistic or additive effects that may have the same, similar or different mechanisms of action (MOA).
4. <i>Assessment of impurities and degradation products.</i>	Chemicals, such as pharmaceuticals or plant protection products, may contain low levels of impurities produced during manufacturing and degradation. Many such substances, when present at levels above accepted thresholds, need to be assessed. In most cases, mutagenicity evaluation of the impurity under question is required as a first step of the risk assessment. (Harvey et al., 2017) The ICH M7 guideline provides specific recommendations for assessing drug impurities (ICH M7, 2017(R1)), including the use of two complementary computational

	toxicology methodologies (i.e., statistical and expert based models) to predict bacterial mutagenicity.
5. <i>Residues of plant protection products.</i>	Residues of plant protection products may be evaluated as a part of residue definition for dietary risk assessment of plant protection products (EU 2009b). In this context, <i>in silico</i> methods provide a useful alternative approach. (EFSA 2016)
6. <i>Assessment of extractables and leachables.</i>	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, identification, and quantification of any leachables. An <i>in silico</i> toxicological assessment, in certain situations, can provide sufficient data for the risk assessment.
7. <i>Workers' safety and occupational health.</i>	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. <i>In silico</i> 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). <i>In silico</i> approaches have been utilized to assess these major toxicological endpoints in the occupational safety setting. <i>In silico</i> methods to predict respiratory sensitization potential of industrial chemicals have recently been reviewed by Seed and Agius (2017).
8. <i>Metabolite analysis.</i>	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 identification, separation as well as possible synthesis for testing purposes may be challenging. <i>In silico</i> methods provide a practical alternative approach to understanding the safety profiles of this potentially large number of chemicals as well as to support the prediction of metabolites.
9. <i>Ecotoxicology.</i>	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 effects 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 <i>in silico</i> assessment of potentially hazardous contaminants present in water.
10. <i>Green chemistry and safer alternatives.</i>	<i>In silico</i> methods can play an important role when identifying alternative chemicals that may have a safer profile 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 specific additives. <i>In silico</i> methods can provide insights about structural features responsible for the toxicity of different groups of chemicals and thereby allow for the rational design of intrinsically safer chemicals.
11. <i>Selection of product development candidates.</i>	In early product discovery or development, many thousands of compounds may be evaluated. <i>In silico</i> methods may provide a helpful approach to selecting candidates, since <i>in silico</i> methods are inexpensive, rapid to perform, and high throughput. In addition, <i>in silico</i> 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; Myatt et al., 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 profile(s) (Myatt et al., 2016).
12. <i>Emergency response situations.</i>	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 effects on humans, wildlife, and the environment. In such emergency situations the toxicological profile 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

1098
1099
1100

	absence of previously generated data, <i>in silico</i> approaches may be a practical option for rapid hazard identification.
13. <i>Prioritizing testing of chemicals.</i>	<i>In silico</i> approaches can help prioritize <i>in vitro</i> and <i>in vivo</i> 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). <i>In silico</i> 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. <i>Rationalization of in vivo or in vitro study results.</i>	As mentioned previously in the description of the <i>in silico</i> application titled " <i>As part of the weight-of-evidence in regulatory studies</i> ", 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 <i>in silico</i> methods can also be used to tailor an <i>in vivo</i> 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, <i>in silico</i> 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).

1101 **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	<ul style="list-style-type: none"> • 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 positive outcome). The prediction reliability may be increased where a score indicating a high likelihood can be justified through an expert review of the available information.
B. Analysis of structural descriptors and corresponding training set data (see Note A)	<ul style="list-style-type: none"> • 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 descriptors might not be biologically meaningful for the predicted toxicological effect/mechanism, as discussed in Powley (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 identified were coincidental and in fact irrelevant) (Amberg et al., 2016). • Another scenario is when the structural descriptors map to experimental data that is incorrect and attributable to known problems with an assay. Again, these features may be discounted if they are not relevant to the toxicological effect 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 identified as significant by the model that are also present in the query compound may 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 and references 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 significant descriptor may be reviewed 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 effects or mechanisms associated with the descriptor could be extrapolated across different 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	<ul style="list-style-type: none"> • Is there any supporting information from the literature or elsewhere to support any correlation between the physicochemical properties

Note B)	<p>identified as significant by the model and the toxicological effect/mechanism?</p> <ul style="list-style-type: none"> • An evaluation of the quality of the experimental data of the training set chemicals used for building 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	<ul style="list-style-type: none"> • 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 affect the evaluation of the reliability of the prediction.

1102 (Note A: items to consider when the QSAR model includes structure-based descriptions; Note B: items to consider when the
1103 QSAR model includes physicochemical descriptors)

1104

1105 **Table 3: Checklist of elements to consider as part of an expert review of results from expert rule-**
 1106 **based**

Expert 4 review elements	Considerations
A. Alert score or qualitative output	<ul style="list-style-type: none"> 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 justified through an expert review of the information provided.
B. Justification of negative prediction	<ul style="list-style-type: none"> Additional considerations may be important where no alerts are identified in the test chemical. Such analysis may focus on similar analogs as well as other chemicals containing the different 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 compound was not included to elucidate the underlying cause for firing 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	<ul style="list-style-type: none"> Although the presence of a structural alert increases the potential of the chemical to exert a toxicological effect or mechanism, this effect 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	<ul style="list-style-type: none"> The reliability of the prediction can be assessed by the quality of the experimental data of the reference set substances 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 diverse structures 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 scientific literature to support the alert to understand the strengths and limitations of the experimental data supporting it.

1107

1108 **Table 4: Summary of Klimisch scores for data reliability (adapted from Klimisch et al., 1997) (Note**
1109 **“restriction”, as part of scores 1 and 2, implies restricted quality)**

Score	Description	Summary
1	Reliable without restriction	<ul style="list-style-type: none">• Well documented and accepted study or data from the literature• Performed according to valid and/or accepted test guidelines (e.g., OECD)• Preferably performed according to good laboratory practices (GLP)
2	Reliable with restriction	<ul style="list-style-type: none">• Well documented and sufficient• Primarily not performed according to GLP• Partially complies with test guideline
3	Not reliable	<ul style="list-style-type: none">• Inferences between the measuring system and test substance• Test system not relevant to exposure• Method not acceptable for the endpoint• Not sufficiently documented for an expert review
4	Not assignable	<ul style="list-style-type: none">• Lack of experimental details• Referenced from short abstract or secondary literature

1110

1111

1112 **Table 5: Common components of an IST protocol (IATA = Integrated Approaches to Testing and**
 1113 **Assessment; AOP = Adverse Outcome Pathways)**

Introduction	<ul style="list-style-type: none"> Describe the major toxicological endpoint being assessed Outline the general hazard assessment framework, including how a series of toxicological effects or mechanisms are related to one or more endpoints Provide citations to any applicable AOPs or IATAs used
<i>In silico</i> methodologies and models	<ul style="list-style-type: none"> Identify toxicological effects or mechanisms that might realistically be predicted Define what <i>in silico</i> methodologies are appropriate to use Specify additional considerations as to what constitutes an acceptable model Discuss issues to be considered as part of any read-across analysis
Experimental data	<ul style="list-style-type: none"> Define specific study types and result(s) relevant to each toxicological effect or mechanism Define and justify the relevance of the information to the assessment of the toxicological endpoint (defined in the hazard assessment framework) Define specific factors to consider when assessing the results and documenting the reliability of any available data or reference specific test guideline(s) Identify sources of data that may be considered
Toxicological effects or mechanisms assessment and reliability scores	<ul style="list-style-type: none"> Describe how each toxicological effect or mechanism assessment may be generated from available experimental data and/or <i>in silico</i> prediction(s) Define additional items to consider as part of an expert review Discuss any endpoint specific issues to consider as part of the reliability score
Toxicological endpoint assessment and confidence	<ul style="list-style-type: none"> Describe the toxicological endpoints that will be used as part of the hazard assessment framework Describe the rules or principles for determining each endpoint assessment, based on the associated effect/mechanisms or other endpoints Define the rules or principles for determining each toxicological endpoint confidence, based on the relevance and reliability (from associated effects/mechanisms) or confidence (from associated endpoints) Identify points to consider as part of any expert review
Reporting	<ul style="list-style-type: none"> Define a format for a report of the results, expert review and conclusions
Other considerations	<ul style="list-style-type: none"> Case studies

1114

1115 **Table 6: Elements of an *in silico* toxicology report (QMRF = QSAR Model Reporting Format)**

Section	Content
Title page	<ul style="list-style-type: none"> - Title (including information on the decision context) - Who generated the report and from which organization - Who performed the <i>in silico</i> analysis and/or expert review, including their organization - Date when this analysis was performed - Who the analysis was conducted for
Executive summary	<ul style="list-style-type: none"> - Provide a summary of the study - Describe the toxicity or properties being predicted - Include a table or summary showing the following: <ul style="list-style-type: none"> o The chemical(s) analyzed o Summary of <i>in silico</i> results, reviewed experimental data and overall assessment for each toxicological effect or mechanism o Summary of toxicological endpoint assessment and confidence o Summary of supporting information
Purpose	<ul style="list-style-type: none"> - Specification of the problem formulation
Materials and methods	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - Details of the results and expert review of the <i>in silico</i> models and any experimental data, including results of the applicability domain analysis - Report of any read-across analysis, including source analogs and read-across justifications
Conclusion	<ul style="list-style-type: none"> - Summarize the overall analysis including experimental data, <i>in silico</i> methods and expert review - Final prediction that is based on expert judgment
References	<ul style="list-style-type: none"> - Complete bibliographic information or links to this information, including test guidelines referred to in the experimental data, etc.
Appendices (optional)	<ul style="list-style-type: none"> - Full (or summary) study reports used or links to the report, detailed (or summary) <i>in silico</i> reports, reports on the models used (e.g., QMRF reports)

1116
1117

1118 **Figures**

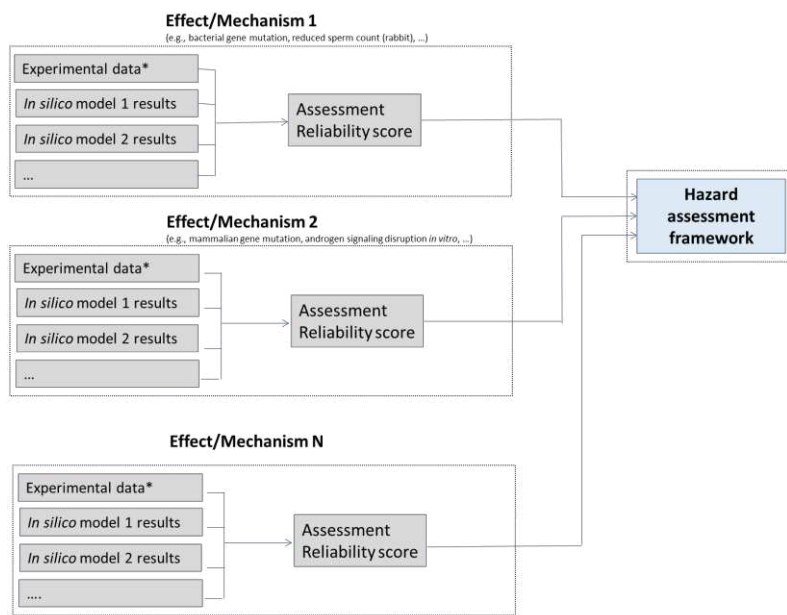
1119

1120 **Figure 1: Overview of the IST protocol framework, showing how experimental data or *in silico***

1121 **model(s) for each defined toxicological effect/mechanism are assessed and used to support a hazard**

1122 **assessment. (Note Effect/Mechanism N is used to illustrate that there can be any number of**

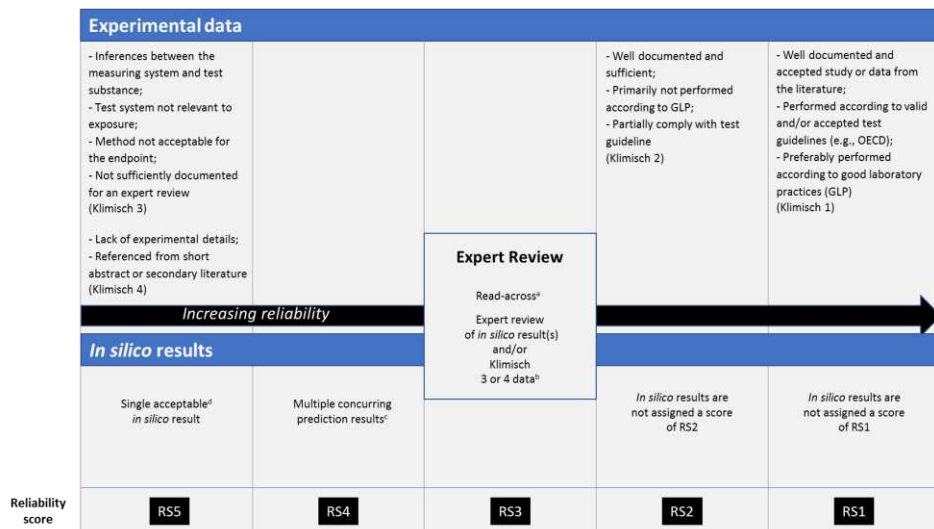
1123 **effects/mechanisms in each protocol)**



1124

* From the literature, database or study report

1125 **Figure 2: Reliability of toxicity assessments based on computational models and experimental data**

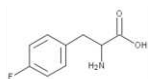


a. Read-across performed according to the RAAF or similar
 b. Expert review resulting in increased confidence
 c. 2+ concurring results from different methodologies or QSAR descriptor sets or training sets
 d. Based on the selection criteria provided in Section 3.3.2

1126

1127 **Figure 3:** Determining the bacterial gene mutation assessment and reliability score for two concurring *in*
1128 *silico* results with expert review

1129



Bacterial gene mutation

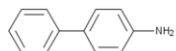


1130

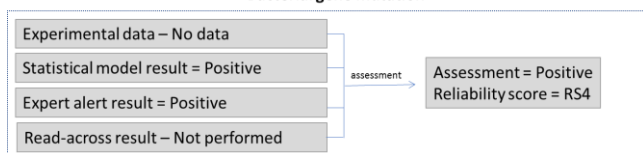
1131

1132

1133 **Figure 4:** Determining the bacterial gene mutation assessment and reliability score for two concurring *in*
1134 *silico* results with no expert review



Bacterial gene mutation



1135

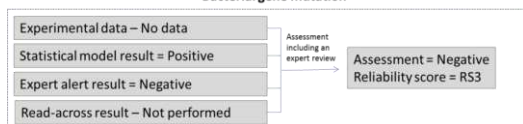
1136

1137 **Figure 5:** Determining the bacterial gene mutation assessment and reliability score where there is no
1138 experimental data available and conflicting *in silico* results



Example 6

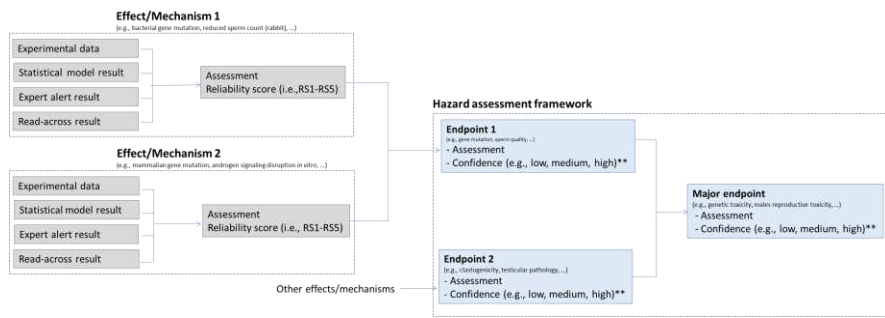
Bacterial gene mutation



1139

1140

1141 **Figure 6: Hazard assessment framework**

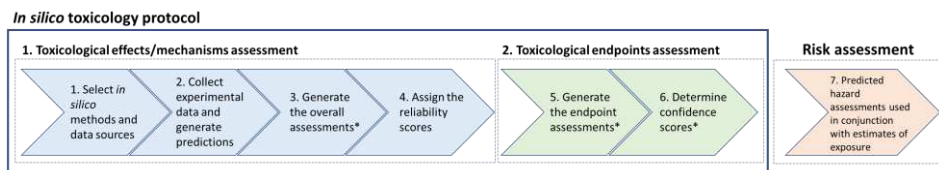


* From the literature, database or study report
 ** Function of the associated reliability, relevance and completeness

1142

1143

1144 **Figure 7: Summary of the IST protocol process**



1145 * Based on rules/principles outlined in the IST protocols, including an expert review if warranted

1146