Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2023)32/ANN2

Unclassified

English - Or. English

21 November 2023

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Cancels & replaces the same document of 17 August 2023

Annex II – (Q)SAR prediction reporting format (QPRF) v.2.0

Series on Testing and Assessment
No. 386

The (Q)SAR Assessment Framework document is available under the cote ENV/CBC/MONO(2023)32.

An editable format is available at https://www.oecd.org/chemicalsafety/risk-assessment/qsar-assessment-framework-annex-2-qsar-prediction-reporting-format.docx

JT03532231

OECD Environment, Health and Safety Publications
SERIES ON TESTING AND ASSESSMENT
NO. 386

Annex II – (Q)SAR prediction reporting format (QPRF) v.2.0

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2023

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

- **Also published in the Series on Testing and Assessment: link**

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

E-mail: ehscont@oecd.org

OECD Environment, Health and Safety Publications

Annex II – (Q)SAR prediction reporting format (QPRF) v.2.0

	Element	Explanation
1.	**General information**	*Information about the compilation of the current QPRF is provided in this section.*
1.1.	**Date of QPRF**	*Report the date of compilation of the QPRF. Example: 30th January 2023.*
1.2.	**QPRF author and contact details**	*Report the contact details of the author of the QPRF.*
2.	**Substance**	*Information about the substance under analysis. Some substances might be associated to more than one structure. The information on the structure(s) used as input is expected in Section 5.1.*
2.1.	**CAS number**	*Report the CAS number.*
2.2.	***EC number***	*Report the EC number.*
2.3.	**Other regulatory numerical identifiers**	*Report other numerical identifiers, e.g. METI number*
2.4.	**Chemical name**	*Report the chemical name (e.g., IUPAC and/or CAS names)*
2.5.	**Structural formula**	*Report the structural formula.*
2.6.	**Structural and composition information** ***a.*** **SMILES**	*Report the SMILES of the substance, if available.*
	b. **InChI**	*Report the InChI code of the substance, if available.*
	c. **Other structural representation**	*Indicate if another structural representation was used to generate the prediction. Indicate whether this information is included as supporting information. Example: "mol file used and included in the supporting information".*
	d. **Stereochemical features**	*Indicate whether the substance is a stereoisomer and consequently may have properties that depend on the orientation of its atoms in space.*
	e. **Composition information**	*Comment on whether there is more than one constituent, impurity or additive in the composition of the substance that should be considered for the assessment.*
	Comments on substance information	*Add any other information, as relevant*
3.	**Model and software**	*Information about the model and software used to make the prediction*
3.1	**Model** **a. Model or submodel name**	*Identify the model used to make the prediction*

	b. Model version	*Identify, where relevant, the version number and/or date of the model and submodel.*
	c. Reference to QMRF	*Provide relevant information about the QMRF that stores information about the model used to make the prediction.*
	Comments on model	*add any other information, as relevant*
3.2	**Software** **a. Software name**	*Identify the software used to make the prediction*
	b. Software version	*Identify the version number of the software.*
	c. Software reference	*Provide relevant information about the software*
	d. Software availability	*Provide information on the availability of the software (e.g. whether it is available for download, if it is commercial or free to use.)*
	Comments on software	*Add any other information, as relevant*
4.	**Prediction**	*Information about the prediction of the model*
4.1	**a. Predicted Property**	*Define the property for which the model provides predictions (this information should correspond to the information provided in the QMRF under fields 3.2 and 3.3). Information such as species, duration, should be included in the description of the property, where possible. Examples: 28 days repeated dose oral toxicity to rat. 96-hour toxicity to fish (fresh water). Provide information about the specifics of the data curation procedure, e.g. inclusion/exclusion rules, for binary endpoints the definitions of the positives and negatives etc.*
	b. Test guideline covered	*indicate if the data in the training set covers one or more specific test guidelines.*
	c. Dependent variable	*Report the dependent variable for which the model provides predictions including any transformations introduced for modelling purposes (note that this information should correspond to the information provided in the QMRF under field 3.5). Example: NOEL, LOAEL, LC_{50}*
	Comments on the predicted property	*Add any other information, as relevant*
4.2	**a. Predicted value**	*Report the predicted value (including units) obtained from the application of the model to the input structure. For alert-based expert models, report the alert triggered together with the reasoning. Example: "aromatic amine - mutagenicity, plausible".*
	b. Predicted value (comments)	*The predicted value is categorical (e.g. yes/no or low/medium/high), explain the cut-off values that were used as the basis for classification.*

	c. Unit	*Indicate the unit in which the predicted value is expressed*
	Comments on the predicted value	*Add any other information, as relevant*
5	**Input**	*Information about the input used to generate the prediction. It should be detailed enough to allow reproducibility of the prediction by others when using the same model and software.*
5.1	**a. Input structure**	*Specify what kind of input was used to represent the input structure (e.g., SMILES, CAS RN, mol file) and associated value*
	b. Stereochemical features	*indicate whether the substance is a stereoisomer and consequently may have properties that depend on the orientation of its atoms in space. Identify the stereochemical features that may affect the reliability of predictions for the substance, e.g. cis-trans isomerism, chiral centers. Are these features encoded in the structural representations mentioned above?*
	c. Tautomerism	*Indicate whether the substance is known to undergo tautomerism and what impact it may have on the prediction.*
	Comments on the input structure	*Add any other information, as relevant*
5.2	**Descriptors**	*Report, for each descriptor ((at least for descriptors manually input)* *a. The value and unit of the descriptor* *b. If the descriptor is measured experimentally or calculated* *c. Reference to the descriptor source* *d. A justification if different types of descriptors are used for the prediction compared to those that were used for model development and validation*
	Comments on descriptors	*Add any other information, as relevant*
5.3	**Model and/or software settings**	*For models or software that allow customization of their algorithm, indicate the custom settings used to generate the prediction*
	Comments on settings	*Add any other information, as relevant*
6	**Applicability domain (AD) and limitations**	*Information about how the substance relates to the AD as defined by the model developers and any other documented limitations. Any other reliability considerations can be reported in Section 7.*
6.1	**Applicability domain (AD) and limitations**	*In this section, the assessment of the AD is limited to the AD definition as intended by model developers. In the Reliability section, other AD considerations can be added.*

	a. AD assessment	*Indicate if the prediction is within or outside the defined AD of the model. Specify if the assessment was done manually by the user or automatically by the software. If the AD is not defined, indicate in this field "Not applicable".*
	b. AD assessment justification	*Describe why the substance is within or outside the AD assessment*
	c. Any other limitation	*Indicate if any other known limitation beyond those defined in the applicability domain are applicable to the prediction. Example: if the input structure is a perfluorinated substance and it has been documented that the model cannot predict these types of substances, the use of the model for this particular substance needs additional justification.*
	Comments on AD	*Add any other information, as relevant*
7	**Reliability assessment**	*Information about reliability of the prediction beyond the AD as defined by the model developers.*
7.1	**Reproducibility**	*Indicate if it is possible to reproduce the prediction. This requires e.g., public availability of the software implementing the model and/or a clear description of the model.*
	Comments on reproducibility	*Add any other information, as relevant*
7.2	**Overall performance of the model**	*With a reference to the predictive performance of the model reported in the QMRF, indicate why the performance is considered acceptable for the intended regulatory purpose*
7.3	**Additional reliability aspects based on the training set**	*Discuss whether the input structure is covered by the training set in terms of:*
	a. Descriptor space	*Discuss if the values of the descriptors used in the model of the chemical are within the limits (i.e. highest and lowest values) of the descriptors associated with training set chemicals.*
	b. Structural fragment space	*Discuss if the chemical contains fragments that are represented in the model training set, including stereoisomerism when relevant*
	c. Response space	*Discuss if the predicted value of the chemical falls within the response (i.e., predicted property) space of the model training set*
	d. Mechanism considerations	*Discuss if the mechanism of action of the chemical, if possible (where relevant), is covered by the model, i.e. that there are substances in the training set following (or expected to follow) the same mechanism as the input structure.*
	e. Metabolic considerations	*Discuss if the metabolism of the chemical, where relevant, is covered by model, i.e. if the training set of the model includes substances which have (or are expected to have) a similar metabolism compared to the input structure*
	Comments on additional reliability aspects	*Add any other information, as relevant*

7.4	**Analogues**	*List the structural and/or mechanistic analogues with associated experimental data that can be used to support the reliability of the prediction. For each analogue indicate:*
	a. Identifiers	*E.g. CAS number, the structural formula, the SMILES code*
	b. Source of the analogue	*From the training or test sets, or accessible from other sources (in this case it should be explained how the structural analogue was retrieved)*
	***c.* Experimental value for the property of interest**	*Self-explanatory field name*
	d. Reference for experimental value	*To verify the validity of the experimental data*
	e. Predicted value for the property of interest	*Self-explanatory field name*
	f. Accuracy of the prediction	*Comparison between experimental and predicted value*
	g. Comments on similarity	*Justification about the reasons that qualify a structural analogue as relevant*
	Considerations on analogues	*Discuss how predicted and experimental data for structural analogues support the prediction of the chemical under consideration.*
	Comments on analogues	*Add any other information, as relevant*
7.5	**Other reliable information on the property**	*Indicate if any other information other than the prediction is available for the property. Additional information can be obtained from additional models or different type of information. Provide considerations on how the other information fit with the prediction.*
7.6	**Conclusion on reliability**	*Summary of reliability assessment*
8	**Purpose of use (for regulatory applications)**	*This information aims to facilitate considerations about the adequacy of the (Q)SAR prediction or the result derived from multiple predictions for a specific regulatory use*
8.1.	**Regulatory purpose**	*Explain the regulatory purpose for which the prediction is being used.*
8.2.	**Approach for regulatory interpretation of the prediction or result derived from multiple predictions**:	*Describe how the result is going to be interpreted in light of the specific regulatory purpose (e.g. by applying an algorithm or regulatory criteria). This may involve the need to convert the units of the dependent variable (e.g. from log molar units to mg/l). It may also involve the application of another algorithm, an assessment factor, or regulatory criteria, and the use or consideration of additional information in a weight-of-evidence assessment. For a result derived from multiple predictions, explain how the individual predictions were integrated to obtain the final result.*
8.3.	**Regulatory interpretation of the result**	*Report the interpretation of the model result in relation to the defined regulatory purpose.*

8.4	**Uncertainty**	*Estimate and comment on the uncertainty of the prediction for the input structure. The methodology proposed in the OECD (Q)SAR Assessment Framework can be used for this purpose.*
8.5.	**Conclusion**	*Provide an assessment of whether the final result is considered adequate for a regulatory conclusion, or whether additional information is required (if this is the case the content of additional information should be specified).*

1. The MS word version is available at XXXXXX

2. Further, a mapping with between field numbers in QPRF v.2.0 and QPRF format v.1.1 in ECHA Guidance on information requirements and chemical safety assessment Chapter R6[1] is provided.

QPRF mapping	QPRF v2.0	QPRF v.1.1
Administrative information	1	2
Date of QPRF	1.1	2.1
QPRF author and contact details	1.2	2.2
Substance	2	1
CAS number	2.1	1.1
EC number	2.2	1.2
Other numerical identifiers	2.3	**NA**
Chemical name	2.4	1.3
Structural formula	2.5	1.4
Structural and composition information	2.6	1.5
SMILES	2.6.a	1.6.a
InChI	2.6.b	1.6.b
Other structural representation	2.6.c	1.6.c
Stereochemical features	2.6.d	1.6.d
Composition information	2.6.e	**NA**
Comments on substance information	-	**NA**
Model and software	**3**	
Model name	3.1.a	3.2.a
Version	3.1.b	3.2.b
Reference to QMRF	3.1.c	3.2.c
Comments on model	-	**NA**
Software implementation	3.2.a	**NA**
Software name	3.2.b	**NA**
Software version	3.2.c	**NA**
Software reference	3.2.d	**NA**
Software availability	3.2.e	**NA**
Comments on model and software	-	**NA**
Prediction	4	
Property	4.1.a	3.1a
Test guideline(s) covered	4.1.b	**NA**
Dependent variable	4.1.c	3.1b
Comments on predicted property	-	NA
Predicted value	4.2.a	3.2.d
Predicted value (comments)	4.2.b	3.2.e
Unit	4.2.c	**NA**

[1] https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf/77f49f81-b76d-40ab-8513-4f3a533b6ac9

Comments on predicted value	-	**NA**
Input	5	
Input structure	5.1	
Input structure	5.1.a	3.2.f
Stereochemical features	5.1.b	1.5.d
Tautomerism	5.1.c	**NA**
Comments on the input structure	-	**NA**
Descriptor value and unit	5.2.a	3.2.g
Descriptor measured or calculated	5.2.b	3.2.g
Descriptor reference	5.2.c	3.2.g
Comments on descriptors	-	**NA**
User defined settings	5.3	**NA**
Comments on settings	-	**NA**
Applicability domain and limitations	6	
AD assessment	6.1.a	3.3a
AD assessment methodology	6.1.b	3.3a
Any other limitations	6.1.c	3.3a
Comments on AD	-	3.3a
Reliability assessment	7	
Reproducibility	7.1	**NA**
Comments on reproducibility	-	**NA**
Overall performance of the model	7.2	**NA**
Descriptor space	7.3.a	3.3.a.i
Structural space	7.3.b	3.3.a.ii
Response space	7.3.c	**NA**
Mechanistic considerations	7.3.d	3.3.a.iii - 3.5
Metabolic considerations	7.3.e	3.3.a.iv
Comments on additional reliability	-	**NA**
Analogues: identifiers	7.4.a	3.3.b - c
Analogues: source	7.4.b	3.3.b - c
Analogues: Experimental value for the property of interest	7.4.c	3.3.b - c
Analogues: Reference for experimental value	7.4.d	3.3.b - c
Analogues: predicted value for the property of interest	7.4.e	3.3.b - c
Analogues: Accuracy of the prediction	7.4.f	3.3.b - c
Analogues: comments on similarity	7.4.g	3.3.b - c
Considerations on structural analogues	7.4	3.3.b - c
Comments on analogues	-	3.3.b - c
Other reliable information on the property	7.5	**NA**
Conclusion on reliability	7.6	**NA**
Purpose of use (for regulatory applications)	8	

Regulatory purpose	8.1	4.1
Approach for regulatory interpretation of the prediction or result derived from multiple predictions	8.2	4.2
Regulatory outcome	8.3	4.3
Uncertainty	8.4	3.4
Conclusion	8.5	4.4