KATE2025

Ecotoxicity Prediction System—Internet Version

Operating Manual (March 27, 2025, version)

KATE20255 is a tool for the prediction of chemical ecotoxicity based on the following data:

- 50% lethal concentration (LC_{50}) in the fish acute toxicity test
- 50% effective concentration (EC_{50}) in the *Daphnia magna* immobilization test
- 50% effective concentration (EC_{50}) in the algal growth inhibition test
- No-observed-effect concentration (NOEC) in the fish early-life-stage toxicity test
- No-observed-effect concentration (NOEC) in the *Daphnia magna* reproduction test
- No-observed-effect concentration (NOEC) in the algal growth inhibition test

Values predicted by KATE2025 are for reference use only and cannot be used to satisfy the requirements for ecotoxicity test results necessary for notification under Japanese Chemical Substances Control Law (Act on the Evaluation of Chemical Substances and Regulation of Their Manufacture, etc.).

For questions, please contact the administrators at

Health and Environmental Risk Division, National Institute for Environmental Studies KATE Contact Desk

kate@nies.go.jp

Operating Manual Revision History

Version	Date of issue	Revision history
0.1	March 29, 2018	Manual for KATE2017 beta version
0.8	January 30, 2019	Provisional version
1.0	May 23, 2019	Official version
1.0.1	June 4, 2019	Minor update
1.0.2	July 30, 2019	Explanation of JSME Editor was updated
2.0	March 30, 2022	Manual for KATE2020 version 3.0
	March 30, 2022	KATE2020 version 3.0 released
3.0	March 30, 2023	Manual for KATE2020 version 4.0
	March 30, 2023	KATE2020 version 4.0 released
4.0	March 25, 2024	Manual for KATE2020 version 5.0
	March 25, 2024	KATE2020 version 5.0 released
5.0	March 27, 2025	Manual for KATE2025 version 1.0
	March 27, 2025	KATE2025 version 1.0 released

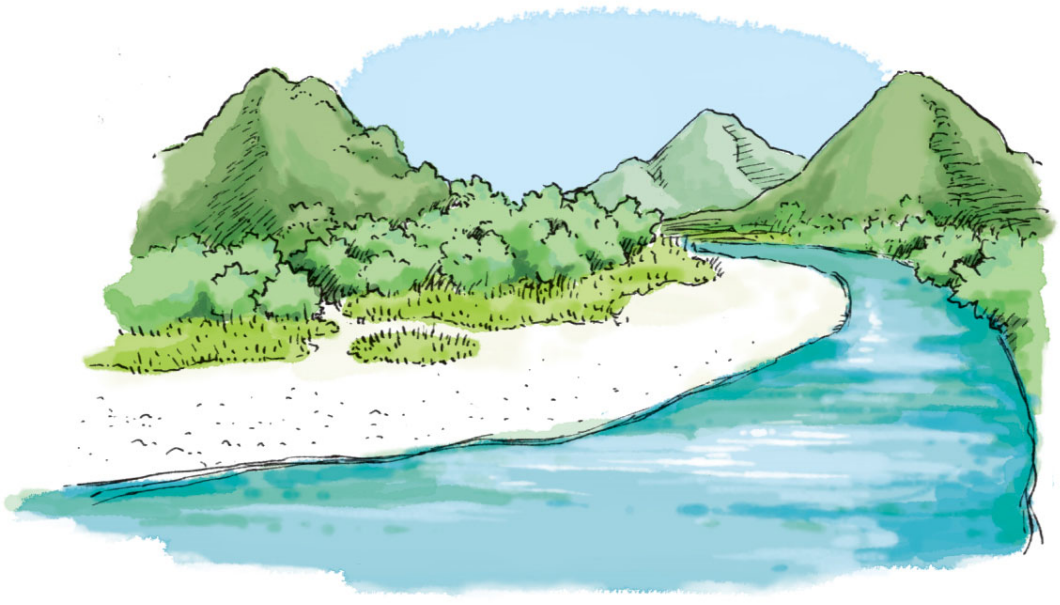


Table of Contents

List of Abbreviations

EC_{50}: 50% effective concentration

The concentration of a chemical dissolved in test water expected to produce a certain effect in 50% of test organisms.

KATE: KAshinhou Tool for Ecotoxicity

The name of the ecotoxicity QSAR system developed by the Center for Health and Environmental Risk Research of the National Institute for Environmental Studies, Japan. It is pronounced as "Kate".

KOWWIN™

A program for estimating log P values of organic compounds. The program is part of the EPI (Estimation Programs Interface) Suite™, which is a suite of estimation programs developed by the US EPA for rapid toxicity screening of chemicals.

LC_{50}: 50% lethal concentration

The concentration of a chemical dissolved in test water that causes death in 50% of test organisms.

Log P: Logarithm of the octanol/water partition coefficient

The logarithm of the ratio of the concentration of a chemical between the solvents 1-octanol and water at equilibrium. It is considered an index of the hydrophobicity of a chemical substance. Log P values ignore ionized query chemical. (http://www.eic.or.jp/ecoterm/?act=view&serial=295; description retrieved from EIC Net on March 1, 2022)

NOEC: No observed effect concentration

The concentration of a chemical causing no statistically or biologically significant increase in the frequency or intensity of any effect in the tested group compared with the control group. Concentration division just under LOEC (Least Observed Effect Concentration). (https://www.env.go.jp/content/000212303.pdf; description retrieved from the glossary of the Ministry of the Environment of Japan on March 1, 2022)

(Q)SAR: (Quantitative) structure–activity relationship

The relationship between the structural characteristics or the physicochemical constant of a chemical and its biological activities (e.g., toxicity) is called the Structure–Activity Relationship (SAR), and the quantitative relationship is called the Quantitative Structure–Activity Relationship (QSAR). When both are referred to, (Q)SAR is used. For example, SAR refers to an estimation of the toxicity level of a chemical based on the presence of a specific functional group. A model to quantitatively calculate the toxicity or other properties of a chemical based on the structure is called a QSAR model. (https://www.env.go.jp/content/000212303.pdf; description retrieved from the glossary of the Ministry of the Environment of Japan on March 1, 2022)

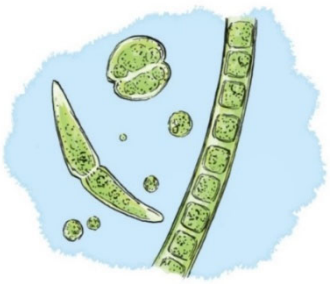
SMARTS: SMiles ARbitrary Target Specification

An extension of SMILES that is used to describe molecular substructures (see https://www.daylight.com/dayhtml_tutorials/languages/smarts/ for a tutorial on using the notation; accessed on March 1, 2025)

SMILES: Simplified Molecular-Input Line-Entry System

Line notation that uses ASCII characters to describe the molecular structure of a chemical (see https://www.daylight.com/smiles/index.html for a tutorial on using the notation; accessed on March 1, 2025)

US EPA: United States Environmental Protection Agency



1. Introduction

(1) What is KATE?

KATE (KAshinhou Tool for Ecotoxicity) is a quantitative structure–activity relationship (QSAR)–based tool for the prediction of chemical ecotoxicity. The tool was developed by the Center for Environmental Risk Research of the National Institute for Environmental Studies (currently Health and Environmental Risk Division, National Institute for Environmental Studies) under contract with the Japanese Ministry of the Environment[1].

The beta version of the tool was released as KATE2017 on NET in 2018, which was then updated in early 2019. Version 1.0 of KATE2017 on NET was officially released in mid-2019. In early 2020, version 1.0 of KATE2020 was released, followed in early 2021 by version 2.0.

Latest version, KATE 2025 version 1.0, uses ecotoxicity test data published by the Japanese MOE[3] as well as fish acute toxicity test data from the US EPA fathead minnow database[5] as chemicals dataset. QSAR equations are updated when additional test results become available.

A query chemical is entered into KATE2020 by using simplified molecular-input line-entry system (SMILES) notation, which can be obtained from within KATE2020 by entering a CAS number or by drawing a chemical structure in the provided JavaScript Molecule Editor (JSME Editor). The SMILES string is then used to perform QSAR prediction based on log P value. Currently, KATE2025 predictions are based on the following data:[2]

- 50% lethal concentration in the fish acute toxicity test (OECD TG 203)
- No-observed-effect concentration in the fish early-life-stage toxicity test (OECD TG 210)
- 50% effective concentration in the *Daphnia magna* acute immobilization test (OECD TG 202)
- No-observed-effect concentration in the *Daphnia magna* reproduction test (OECD TG 211)
- 50% effective concentration and no-observed-effect concentration in the algal growth inhibition test (OECD TG 201)

(2) What is KATE2025?

KATE2025 is an updated version of KATE2020 released since 2020 February with some improvements, which is a web application operated on web browser.

(3) Major Updates from KATE2020 version 5.0 to KATE2025 version 1.0

Improvements of user interfaces

i) Multiple chemical prediction functionality has been updated to be able to predict up to 1,000 substances at a time.
ii) In the QSAR class information detail screen, the display of substance information can now be switched using tabs.
iii) Other improvements have been made to the overall appearance and user interface.

Updates to QSAR models

i) Reviewing toxicity test results for acute and chronic tests for fish and daphnid, recalculated toxicity of some chemicals based on the latest test guideline, some chemicals (e.g., mixture of isomers) were altered from training set to support chemicals.
ii) Two sulfur-containing compounds, whose toxicity testing was done in FY2023, were added to the training set.
iii) Some of the regression equations of QSAR classes were recalculated.

(4) Major Updates from KATE2020 version 4.0 to version 5.0

Improvements of user interfaces

iv) Input screen and multiple chemicals prediction results screen were improved
v) Download functionality of multiple chemicals prediction results was added
vi) Display of structure class definition was improved

Updates to QSAR models

iv) Reviewing toxicity test results for alga acute and chronic, recalculated toxicity of some chemicals based on the latest test guideline, some chemicals (e.g., mixture of isomers) were altered from training set to support chemicals.
v) Two aliphatic primary amines and one epoxide, whose toxicity testing was done in FY2022, were added to the training set.
vi) Recalculated statistics of regression equations of QSAR classes

(5) Major Updates from KATE2020 version 3.0 to version 4.0

Improvements of user interfaces

i) Log in system was revised so that users can use KATE2020 without user ID and password.
ii) Some interfaces (prediction results screen, verify QSAR screen etc.) were modified.

Updates to QSAR models

i) Reviewing toxicity test results for fish acute, recalculated toxicity of some chemicals based on the latest test guideline, some chemicals were altered from training set to support chemicals.
ii) Two thiols (fish and Daphnid acute) and one imide (Daphnid acute) compounds, whose toxicity testing was done in FY2021, were added to training set.
iii) Recalculated statistics of regression equations of QSAR classes

Updates to QSAR class names

i) Scrutinizing the names of all QSAR classes, some of them were fixed.

(6) Major Updates from KATE2020 version 2.0 to version 3.0

Updates to QSAR models

i) Reviewing toxicity test results for daphnid acute, recalculated toxicity of some chemicals based on the latest test guideline, some chemicals were altered from training set to support chemicals.
ii) Recalculated statistics of regression equations of QSAR classes

Updates to QSAR class names

i) Scrutinizing the names of all QSAR classes that satisfy statistic criteria ($R^2 \geq 0.7$, $Q^2 \geq 0.5$, and $n \geq 5$), some of them were fixed.

(7) Major Updates from KATE2020 version 1.1 to version 2.0

Updates to QSAR models

i) One of the criteria for QSAR classes displayed on the prediction results screen by default was changed from $Q^2 \geq 0.6$ to $Q^2 \geq 0.5$.
ii) The algal chronic toxicity QSAR equation "CNOS_X basic aromatic n unreactive" was changed in accordance with the correction of toxicity data.

Updates to displays and user interfaces

i) An individual structure judgement result for each substructure was added.
ii) The ability to display a format for printing was added.
iii) An issue in batch mode where prediction would stop before completion when an error was encountered was fixed.
iv) Some expressions were corrected.
vii) An issue in batch mode where prediction would stop before completion when an error was encountered was fixed.

Updates to structure class names

i) Prefixes such as "CN_X" were removed or added.
ii) Typos were fixed.
iii) Unnecessary notations were removed.
iv) Prefix notations were unified.
v) The "reactive/unreactive" notations were corrected.
vi) Predicted toxicity types were corrected.

(8) Major Updates from KATE2020 version 1.0 to version 1.1

i) The predicted toxicity value was updated to be displayed in exponential notation (ex. 2.3e-7) when the value is greater than 10^{-5}, equal to 10^{6}, or less than 10.
ii) The predicted toxicity value was updated to be displayed to two significant figures.
iii) An issue where prediction was not always executed, even when the input information was correct, was fixed.
iv) Other slight changes in some expressions were made.

(9) Major Updates from KATE2017 on NET to KATE2020 version 1.0

Updates to QSAR models

i) The tool for estimating log P was changed from ClogP to KOWWIN™. Some QSAR models were modified in accordance with this change.
ii) KOWWIN™ measured value was removed from the priority list for log P value to leave only two options: 1. user input value and 2. KOWWIN™ estimated value.
iii) Chemical substances with log P > 6.0 were excluded from calculation of the QSAR equation.
iv) Additional training set data and QSAR classes were added, and some were removed.
v) Chemicals that were not used for QSAR calculation (support chemicals with data with log P > 6.0, data with an inequality sign, and outliers) were updated to be displayed as information.

Updates to displays and user interfaces

i) Data lists of training set data and support chemicals included in the QSAR class were added.
ii) Definition lists of structure class corresponding to QSAR class were added.
iii) A sorting function was added to the chemical list on the "Verify QSAR" screen.
iv) The prediction and confidence intervals were linked to the additional regression line calculated by excluding training set data.
v) A checkbox to skip KOWWIN™ calculation was added.

(10) Development History up to the Release of KATE2017 on NET

The beta version of the KATE system was released in January 2008 and KATE2009 on the Internet was released in March 2009. In March 2011, KATE2009 was updated to KATE2011 with the addition of chemicals data, updated rules for structure classification, updated structure judgement, and addition of substructures related to skin sensitization.

In March 2018, the beta version of KATE2017 on NET was released. This release was followed by the official release of version 1.0 in January 2019. The main changes between KATE2011 and KATE2017 on NET were updates to the substructure language and search program from FITS (a combined substructure language and search program) to a combination of SMARTS notation and the CDK search program, and the addition of predictions for algal and chronic toxicity, toxicity data with inequality signs, and structure classes. In addition, the QSAR models received a significant update, the log P calculation module was changed from ClogP (Daylight Chemical Information Systems, Inc.) to KOWWIN™ (US EPA), displays, user interfaces were updated, and the display language was changed to English.

(11) About Support Chemicals

In KATE2025, the following data are not used for QSAR model construction and are

displayed as “Support Chemicals” for information purposes only:

i) data with estimated log P > 6.0
ii) data with an inequality sign (such as limit test)
iii) mixture of isomers, etc.

Data with an inequality sign that falls within the applicability domain of log P are also used for the judgement for applicability domain regarding structure.

(12) About log P

KATE2025 utilizes the log P prediction model KOWWIN™ (with permission from the US EPA) to obtain log P values used for the toxicity prediction[6]. Users must acknowledge that they agree to the KOWWIN™ Licensing Policy each time at login to KATE2025. The policy is provided here for reference:

KOWWIN v1.69 (April 2015)

c 2000-2015 U.S. Environmental Protection Agency

KOWWIN is owned by the U.S. Environmental Protection Agency and is protected by copyright throughout the world.

Permission is granted for individuals to download and use the software on their personal and business computers.

Users may not alter, modify, merge, adapt or prepare derivative works from the software. Users may not remove or obscure copyright, tradename, or proprietary notices on the program or related documentation.

KOWWIN contained therein is a tradename owned by the U.S. Environmental Protection Agency.

(13) Disclaimer

KATE2025 is provided as a tool for obtaining information on the potential degree of ecotoxicological effects of chemicals. The Japanese MOE and the National Institute for Environmental Studies give no guarantee about the accuracy of ecotoxicity values provided by KATE2025 and assume no responsibility whatsoever for any damages resulting from the use of ecotoxicity values predicted by KATE2025.

In addition, values predicted by KATE2025 cannot be used to satisfy the requirements for ecotoxicity test results necessary for notification under the Japanese Chemical Substance Control Law (Act on the Evaluation of Chemical Substances and Regulation of Their Manufacture, etc.).

For copyright information and instructions for linking to KATE2025, please visit the policy page of the KATE website: https://kate.nies.go.jp/spolicy-en.html.

(14) Acknowledgments

KATE2025 uses data obtained from the following software applications and libraries. We express our sincere appreciation to each of the development teams.

☐ Open Babel (https://openbabel.org/docs/index.html)

☐ JSME Molecular Editor (https://jsme-editor.github.io/)

- B Bienfait and P Ertl, JSME: A free molecule editor in JavaScript, J. Cheminform. 5:24 (2013). doi:10.1186/1758-2946-5-24.

☐ CDK (Chemistry Development Kit) (https://cdk.github.io/)

- E Willighagen et al., The Chemistry Development Kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching, J. Cheminform. 9:33 (2017). doi:10.1186/s13321-017-0220-4.
- JW May and C Steinbeck, Efficient ring perception for the Chemistry Development Kit, J. Cheminform. 6:3 (2014). doi:10.1186/1758-2946-6-3.
- C Steinbeck et al., Recent developments of the Chemistry Development Kit (CDK) - an open-source Java library for chemo- and bioinformatics, Curr. Pharm. Des 12:2111-2120 (2006). doi:10.2174/138161206777585274.
- C Steinbeck et al., The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics, J. Chem. Inf. Comput. Sci. 43:493-500 (2003). doi:10.1021/ci025584y.

☐ KOWWIN™ (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface)

(all URLs accessed March 1, 2025)

(15) References

1) https://kate.nies.go.jp (accessed March 1, 2025)
2) https://www.env.go.jp/chemi/sesaku/01.html (accessed March 1, 2025)
3) https://www.env.go.jp/chemi/sesaku/seitai.html (accessed March 1, 2025)
4) https://archive.epa.gov/med/med_archive_03/web/html/fathead_minnow.html (accessed March 1, 2025)
5) A. Furuhama, T. Toida, N. Nishikawa, Y. Aoki, Y. Yoshioka, and H. Shiraishi: Development of an ecotoxicity QSAR model for the KAshinhou Tool for Ecotoxicity (KATE) system, March 2009 version, SAR QSAR Environ. Res., 21 (5), 403 (2010).
6) https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface/ (accessed March 1, 2025)

2. Overview of KATE2025

(1) QSAR prediction Procedures

Figure 2-1 shows a summary of how KATE2025 performs toxicity prediction for a query chemical.

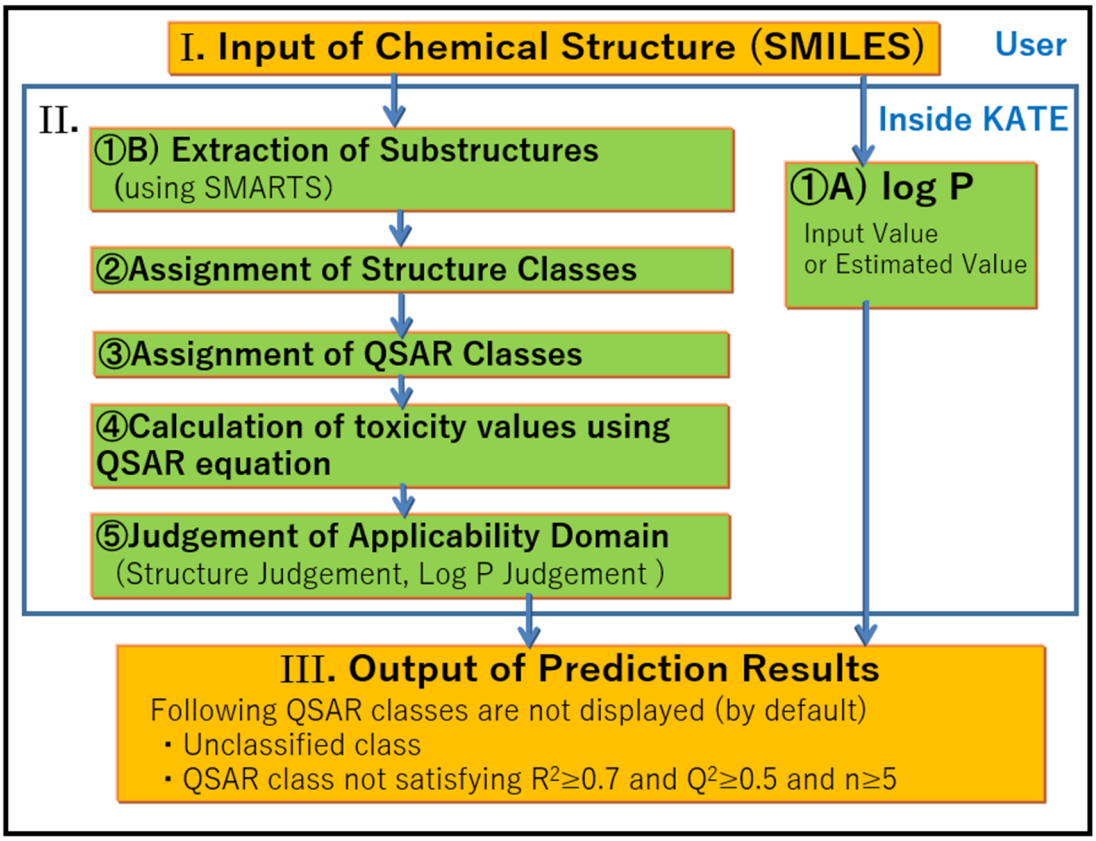


Figure 2-1. Summary of QSAR prediction in KATE2020

I. The structure of the query chemical is input using SMILES notation.

II. Based on the SMILES string input by the user, KATE2025 calculates QSAR equations, predicts toxicity values, and judges the applicability domain. This is accomplished in five steps:

① For the query chemical (entered chemical substance)

A) log P value is estimated (or the input value is used).

B) substructure/s is/are extracted.

② Structure class[*1] is assigned based on the types and numbers of the extracted substructures.

③ For each predicted toxicity type, a QSAR class[*2] corresponding to the structure class is assigned (multiple classes may be assigned).

④ For each assigned QSAR class, the toxicity value is calculated using the QSAR equation[*3].

⑤ The applicability domain is judged (log P judgement and structure judgement).

*1 Refer to “Section 3.5 Structure class extraction”.

*2 Classification defined based on the structure of the chemical for each predicted toxicity type.

*3 Model built based on training set data included in the QSAR class. Here, a simple regression equation with log P as a descriptor.

III. Output of prediction results

Unclassified class*4 and QSAR classes not satisfying $R^2 \geq 0.7$, $Q^2 \geq 0.5$ and $n \geq 5$ are not displayed*5 by default.

*4 QSAR class that is not assigned to any QSAR classes of KATE2025.

*5 R^2, Q^2, and n are the coefficient of determination, internal validation index (leave-one-out method), and the number of training set data respectively, which have been calculated for each QSAR class in advance.

Figure 2-2 shows a representative prediction flow using 1-pyridin-3-ylethanone as the query chemical.

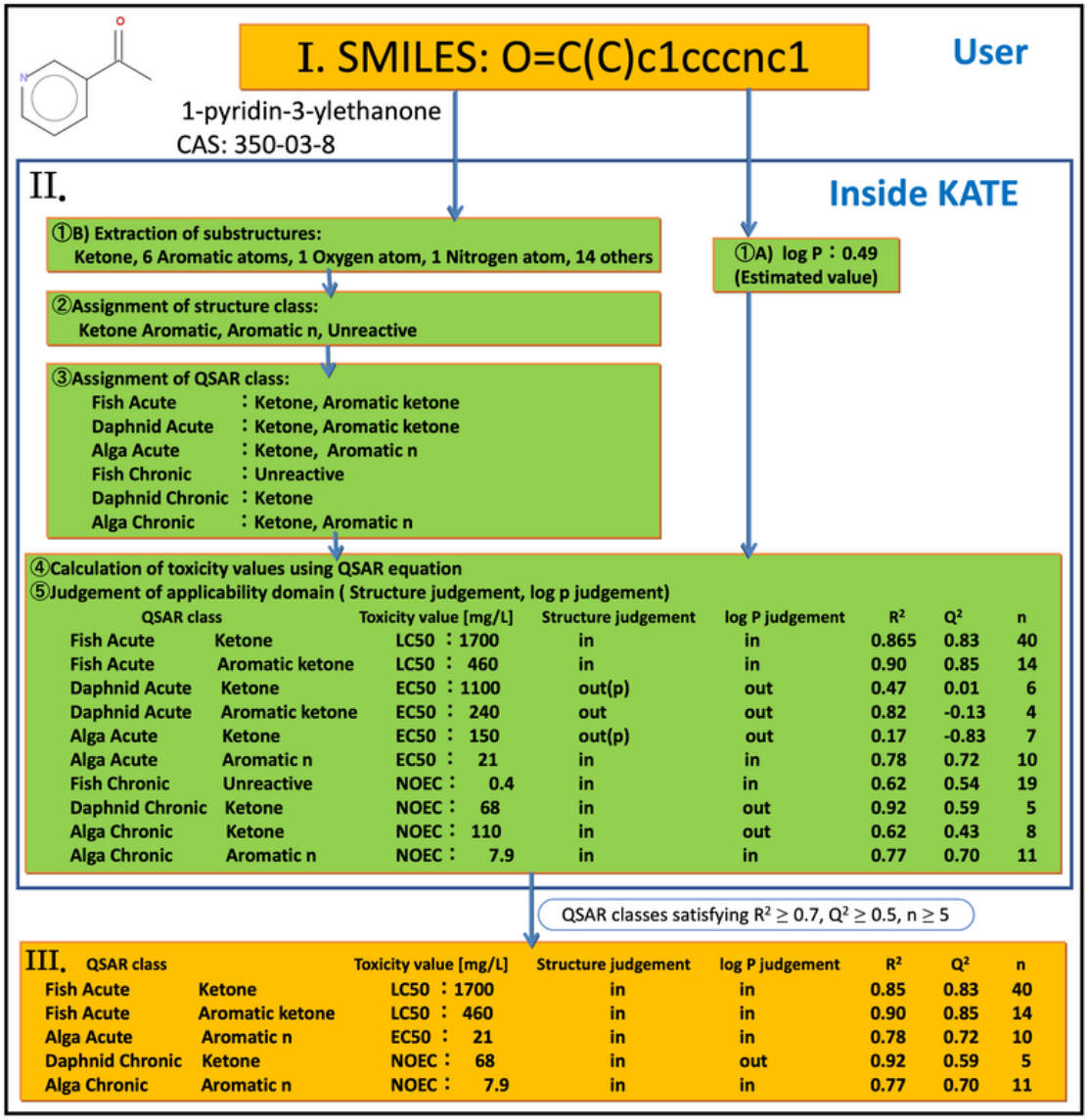


Figure 2-2. Representative prediction flow using 1-pyridin-3-ylethanone as the query chemical

*Actual names of Ketone, Aromatic Ketone, Aromatic n and Unreactive in Figure 2-2 are as follows.

Ketone	: COS_X ketone unreactive
Aromatic Ketone	: COS_X ketone unreactive aromatic
Aromatic n	: aromatic n unreactive Alga
Unreactive	: CNO_X unreactive Fish chronic, w/ N,O

(2) Assignment of QSAR Equations, Prediction of Toxicity Values, and Judgement of Applicability Domain

KATE2025 performs the following process to construct the QSAR equations and predict toxicity values based on the structure of the query chemical.

(i) Extraction of substructures

Based on the list of substructure definitions (SMARTS), the number of each substructure contained in the query chemical is determined. The CDK library is utilized for the calculation of the number of substructures by SMARTS.

(ii) Assignment of structure classes

Based on the list of structure class definitions ("AND/OR" combination of substructures), all the structure classes that match the structure of the query chemical are extracted.

(iii) Assignment of QSAR class

The system assigns QSAR classes that correspond to the structure class of the query chemical for each type of predicted toxicity based on the list of QSAR class definitions, where each QSAR class is defined by a type of predicted toxicity and a structure class. KATE2025 may assign more than one QSAR class for each type of predicted toxicity. If the query chemical is not classified to any QSAR class, it is assigned to "Unclassified class".

(iv) Calculation of toxicity values by using QSAR equations

There is a QSAR equation corresponding to each QSAR class, and so the log P value of the query chemical is substituted into each QSAR equation to calculate the log (1/toxicity value((in mmol/L); the system then converts that value into a toxicity value (in mg/L) using the molecular weight of the query chemical.

(v) Judgement of applicability domain

KATE20255 judges whether a predicted toxicity value of the query chemical is within the applicability domain. It performs two types of judgment: A) structure judgement, and B) log P judgement. When the results of both judgement types fall within the applicability domain, the predicted toxicity value of KATE2025 is considered appropriate.

A) Structure Judgement

KATE2025 judges whether the structure of the query chemical falls within the applicability domain of the QSAR class classified in (ii) by comparing the "substructures for structure judgement" *1 (Figure 2-3). There are three possible judgements: in, in (conditionally), and out of. A QSAR class judged to be "in" or "in (conditionally)" is considered to be within the applicability domain in terms of structure.

in: Within the applicability domain

All the "substructures for structure judgement" of the query chemical are found in the "substructures for structure judgement" extracted from the training set data*2 in the QSAR class (pink and orange area in Figure 2-3), or the query chemical contains no "substructures for structure judgement".

in(p):Provisionally within the applicability domain

The query chemical does not meet the condition of "in", but all the "substructures for structure judgement" of the query chemical are found in "substructures for structure judgement" extracted from the training set data in the QSAR class or those from the Narcotic Group class *3 (pink, orange and yellow areas in Figure 2-3).

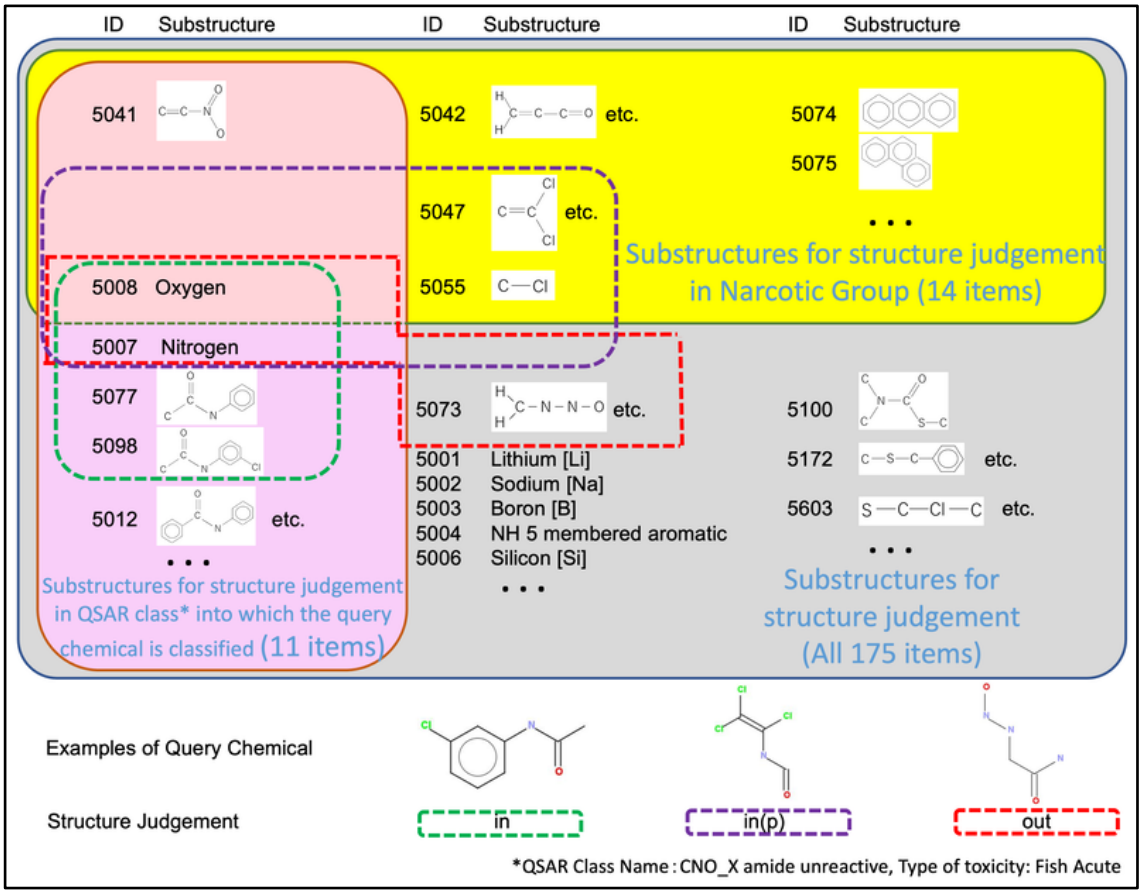
out: Out of the applicability domain

The query chemical does not meet the conditions of "in" or "in (conditionally)"; that is, in the query chemical, there is at least one "substructures for structure judgement" that is in neither the "substructures for structure judgement" extracted from the training set data of the QSAR class nor those from the Narcotic Group class (grey area in Figure 2-3).

1 Substructures introduced for the structure judgement in KATE, which contain toxicologically characteristic structures, are also used for structural classification (for details see the KATE2025 technical document.
* https://kate.nies.go.jp/doc/KATE2025_tech_doc-en.pdf

*2 The "substructures for structure judgement" extracted from the training set data also includes substructures of chemicals with an inequality sign whose log P judgement is "in" (within the applicability domain).

*3 Baseline toxicity not based on specific physiological activity (narcotic effect). In KATE2025, QSAR classes whose toxicity is explained only by a simple narcotic effect are defined for each type of predicted toxicity. Examples include aliphatic hydrocarbons, sulfoxide, aliphatic ethers, aromatic ethers, aliphatic ketones, aromatic ketones, and alcohols. These QSAR classes are grouped and defined as the Narcotic Group for each type of predicted toxicity.



* The green frame encloses substructures extracted when predicting with the leftmost chemical in "Examples of Query Chemical". The same applies to the purple and red frames.
* "etc." indicates multiple substructures have the same ID, so only one example is shown.

Figure 2-3. Example of Structure Judgement

B) Log P Judgement

KATE2025 judges whether the log P value falls within the applicability domain, based on whether the log P value of the query chemical is between the minimum and maximum log P values of all training set data in the QSAR class concerned.

In KATE2025, all chemical substances with log P > 6, which are highly hydrophobic and have low prediction accuracy, are judged as being out of the applicability domain.

in: Within the applicability domain (Figure 2-4).

out: Out of the applicability domain (Figure 2-5).

out(p): Out of the applicability domain, but the log P value of the query chemical takes a value between the minimum and maximum log P values of all chemicals (both the training set data and the support chemicals) in the QSAR class concerned (Figure 2-6).

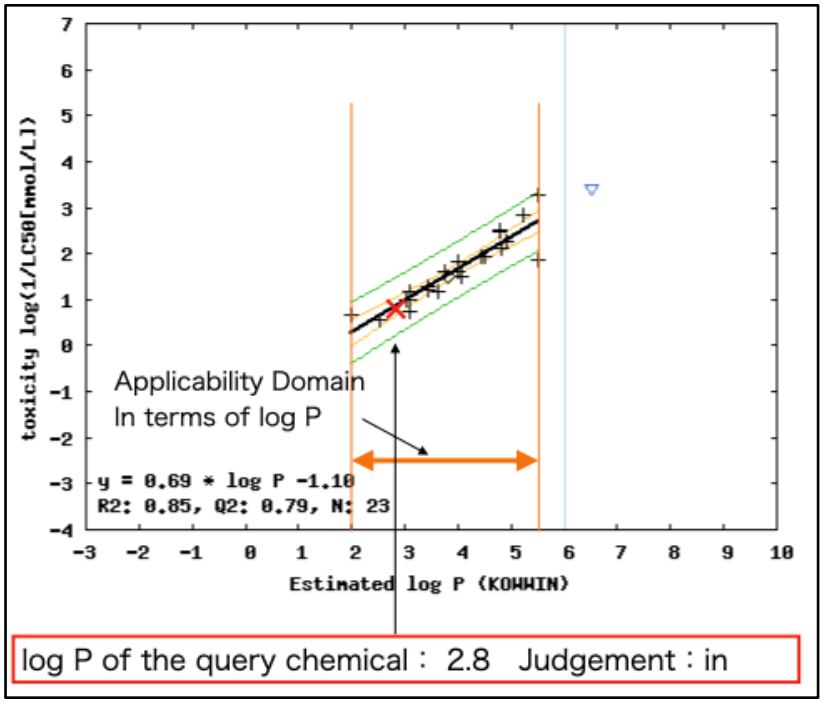


Figure 2-4. Example of log P judgement (in)

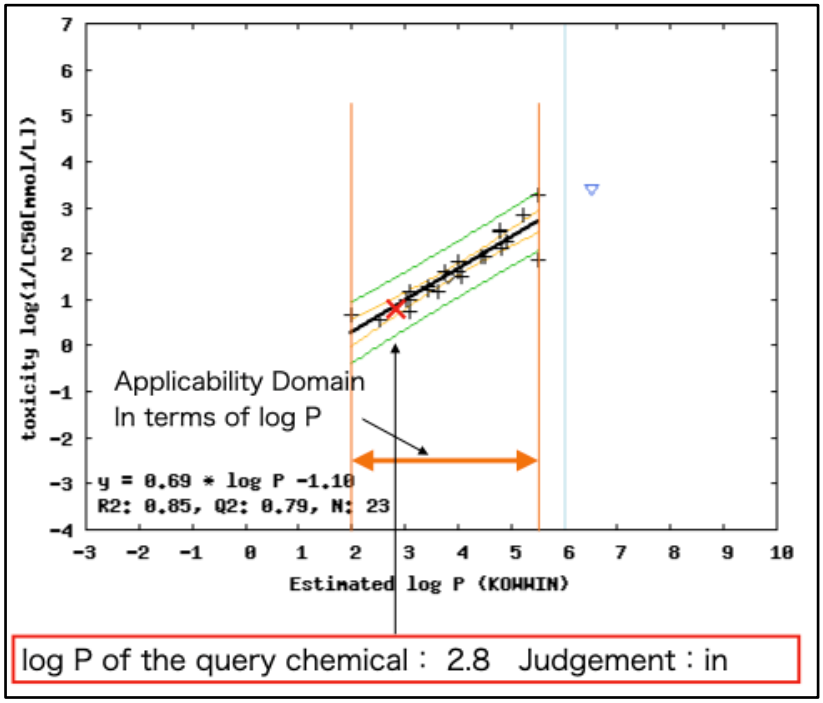


Figure 2-5. Example of log P judgement (out)

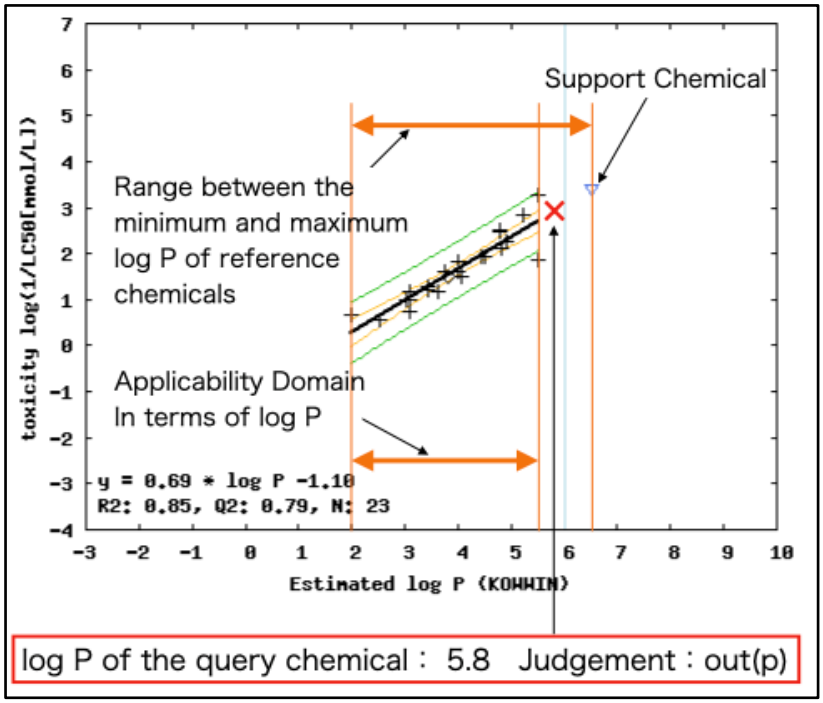


Figure 2-6. Example of log P judgement (out(p))

3. Log In

Since the release of KATE2020 ver. 4.0 in March 2023, users no longer need user ID and password to login. Accessing KATE2025 login screen (https://kate.nies.go.jp/nies/index.php) in figure 3, users only have to (1) agree with the term of agreement and then (2) click "Start the session" button. Data entered by the user, such as SMILES, and output results (e.g., predicted toxicity values and QSAR classes) are stored only in the session, not on the KATE server, and they are automatically deleted when the session expires. The session will expire when web browser is closed or user stopped using KATE2025 for one hour. Furthermore, the previous information is overwritten and deleted every time a prediction is made.

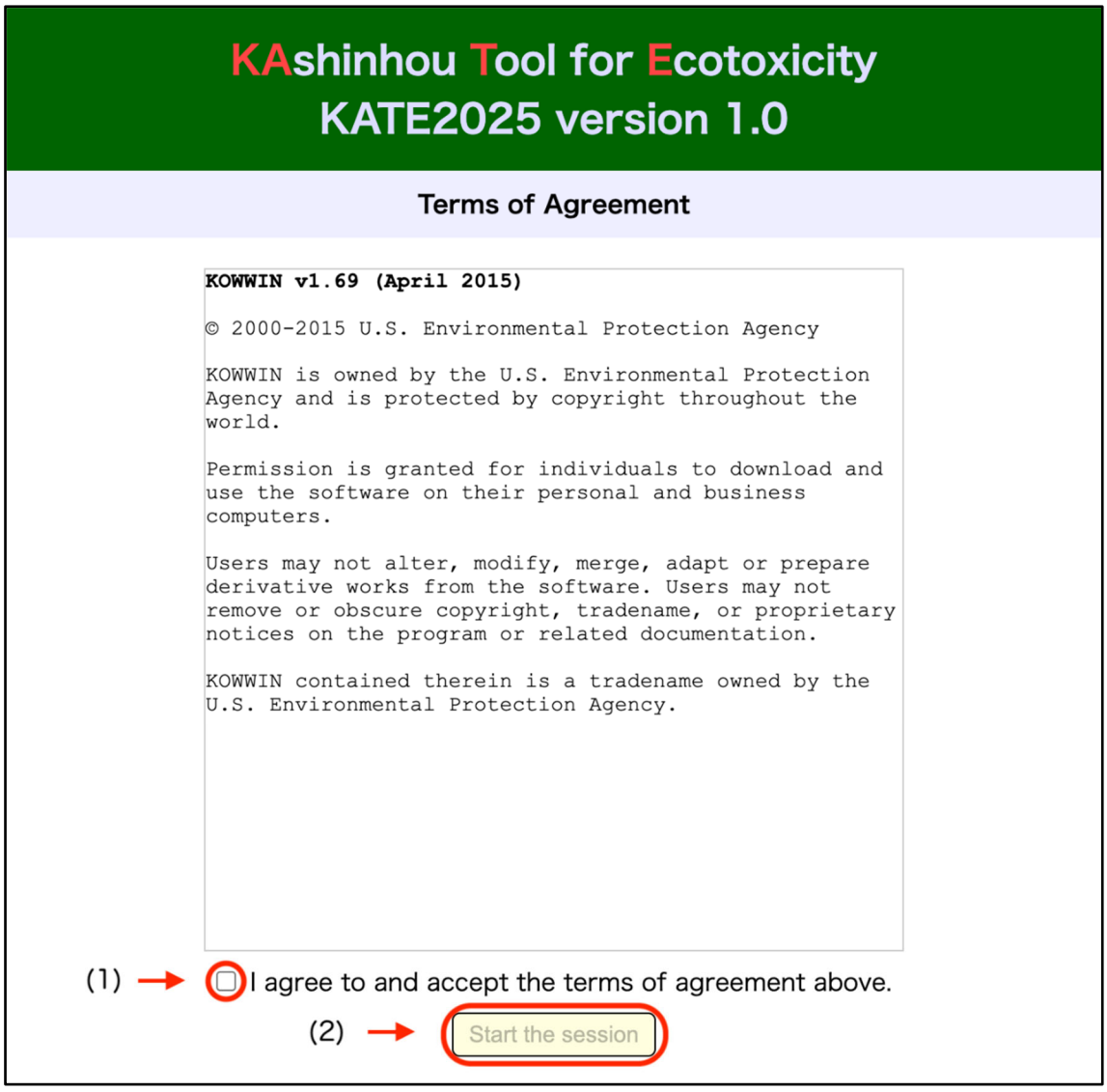


Figure 3. Login screen

4. Input of Query Chemical

After login, the input screen is displayed (Figure 4-1). KATE2025 performs predictions based on the SMILES string of a query chemical. In the center of the screen is a field for the input of a SMILES string. Above that are fields for entering a CAS number or chemical name for conversion to a SMILES string. SMILES strings can also be generated from a drawing of a chemical structure created in the JSME Editor tool within KATE2025.

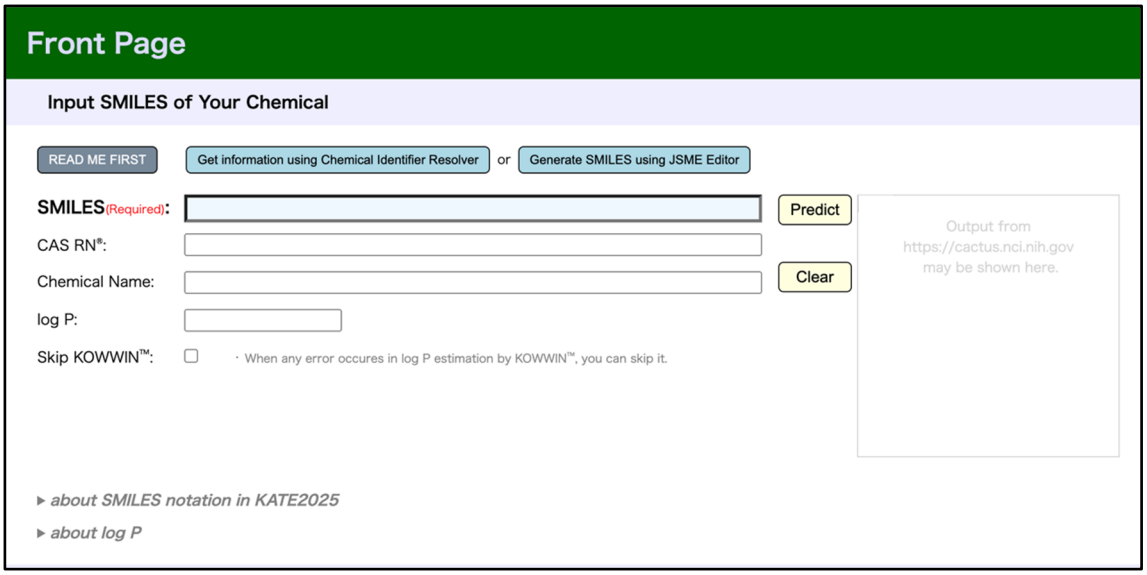


Figure 4-1. Input screen

(1) Acceptable and Unacceptable SMILES Strings

SMILES strings for almost all organic compounds, as well as some inorganic nitrogen compounds (e.g., hydrazine), can be entered into KATE2025. However, the string must

- contain C or N
- not include elements other than H, C, N, O, F, Si, P, S, Cl, As, Br, Sn, and I
- not include ions (although, ammonium [N+] and [n+] can be entered)
- not express a mixture (i.e., a SMILE string that includes ".")
- be converted to the protonated form if it includes [Na], [K], [Li], [Na+], [K+], or [Li+] (e.g., "c1ccccc1O[Na]" should be entered as "c1ccccc1O").

Unacceptable SMILES strings will prompt the system to return an error message.

(2) Direct Input of a SMILES String

Enter a SMILES string in the SMILES input field (Figure 4-1), click the "Predict" button (Figure 4-2), and the QSAR Prediction Results screen will be displayed. In this example, the SMILES string for pyridine-3-ylmethanamine has been entered.

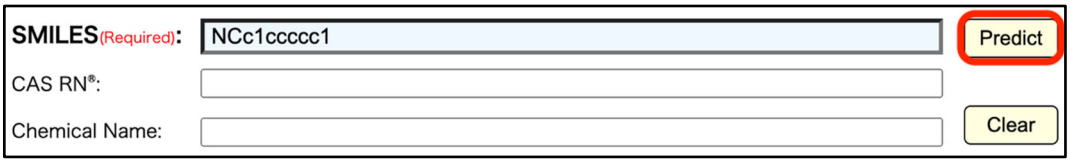


Figure 4-2. “Predict” button next to the SMILES input box

(3) Convert a Drawing of a Chemical Structure into a SMILES String

You can draw a chemical structure using the JavaScript Molecule Editor (JSME Editor) within KATE2025 and convert the drawing to a SMILES string.

Step 1: Click “JSME Editor” under the SMILES string input box (Figure 4-3) to open the JSME Editor window (Figure 4-4).

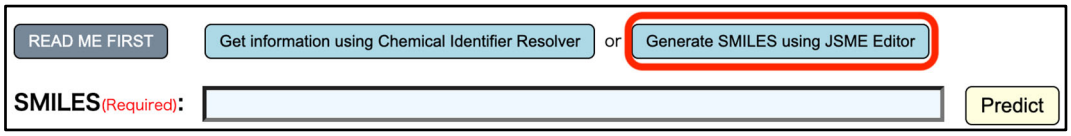


Figure 4-3. Link to the “JSME Editor”

Step 2: In the editor window, draw the structure of the query chemical and click the “Submit smiles to KATE” button. The chemical structure will be converted into the corresponding SMILES string and the string will be automatically entered in the SMILES input field (Figure 4-5). Click the “Predict” button to display the QSAR Prediction Results screen. Detailed instructions on how to use the JSME Editor can be found at the developers’ webpage: https://jsme-editor.github.io/ (accessed March 1, 2024)

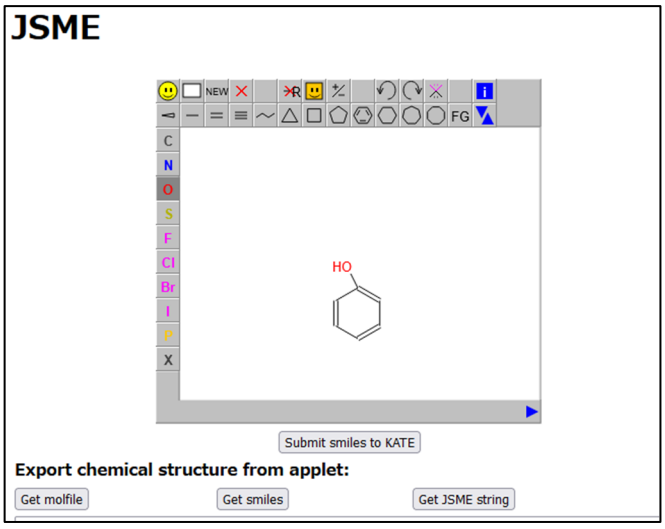


Figure 4-4. JSME Editor

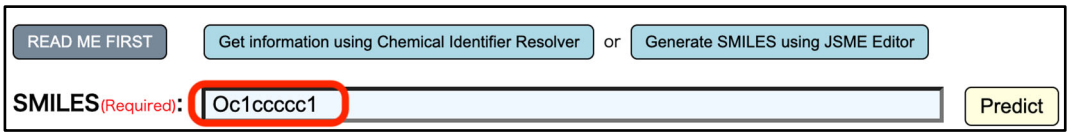


Figure 4-5. Conversion of a structure drawn in JSME Editor into a SMILES string

(4) Convert a CAS Number or Chemical Name into a SMILES String

KATE2025 includes tools to convert a CAS number or chemical name into a SMILES string and vice versa.

Enter the CAS number or chemical name and click the "Get information using Chemical Identifier Resolver" button (Figure 4-6). The structure associated with the CAS number or chemical name will be converted to a SMILES string, and the string will be automatically entered in the SMILES field. In addition, the IUPAC name, SMILES string, and chemical structure of the query chemical will be presented above the CAS number entry box, and the IUPAC name will be entered into the Name field (Figure 4-7). Click the "Predict" button to display the QSAR Prediction Results screen.

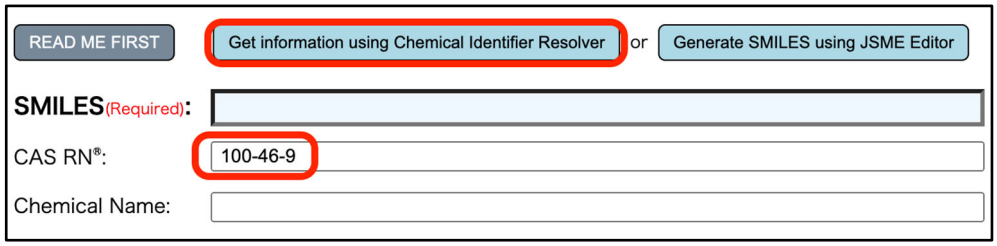


Figure 4-6. CAS input box and "CAS to SMILES, IUPAC Name" button

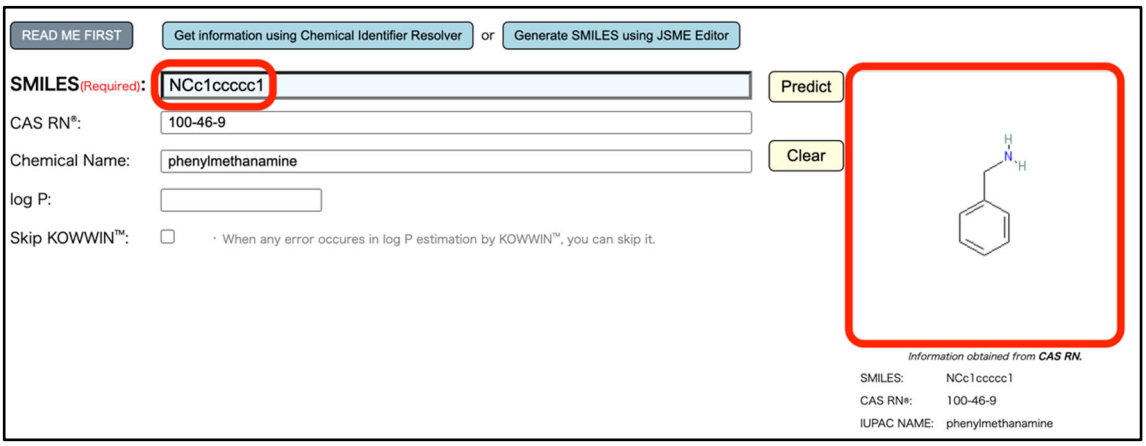


Figure 4-7. Conversion of a CAS number to a SMILES string

(5) User-input Log P Values

If the log P value of the query chemical is known, the value can be entered in the field below the SMILES string input box (Figure 4-8). The entered log P value will be preferentially used in the toxicity prediction.

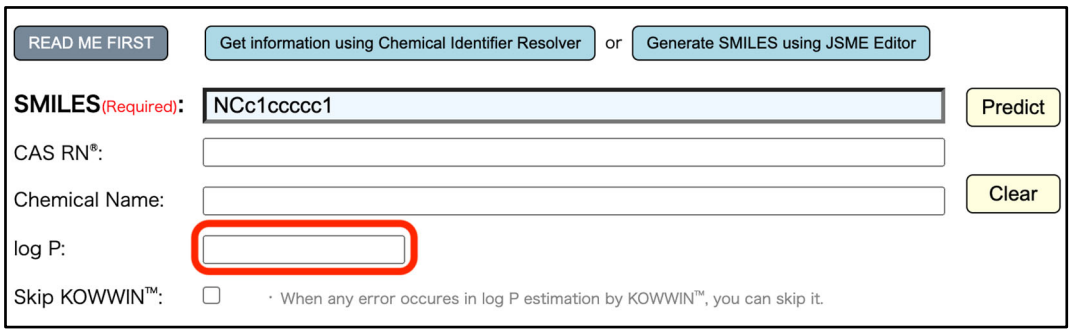


Figure 4-8. Input of a known log P value

(6) User-input CAS Number and IUPAC Name

If a CAS number or IUPAC name is entered in the CAS or Name fields, the information will displayed on the QSAR Prediction Results screen exactly as it was entered (Figure 4-9).

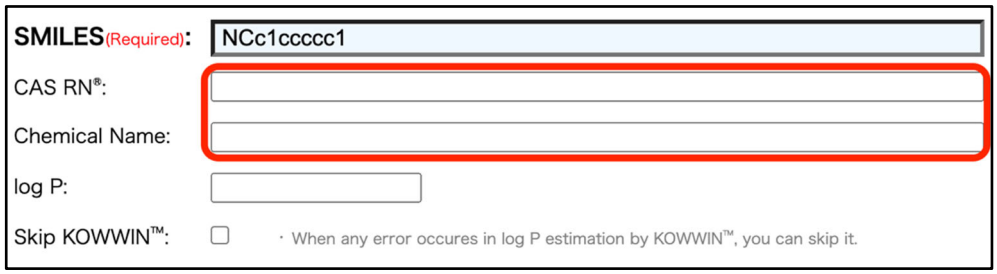


Figure 4-9. Input of CAS number and IUPAC name for inclusion at the top of the QSAR prediction results

(7) Skip KOWWIN™ Calculation

If a SMILES string that cannot be used by KOWWIN™ for the estimation of log P is entered, an error will be reported during the prediction process. To continue with the prediction, the "Skip KOWWIN Calculation" box should be clicked, and a user-define log P value should be entered (Figure 4-10). If a log P value is not entered, QSAR classes will still be assigned, but no predicted toxicity values will be calculated. A user-defined log P value can also be entered on the QSAR Prediction Results screen.

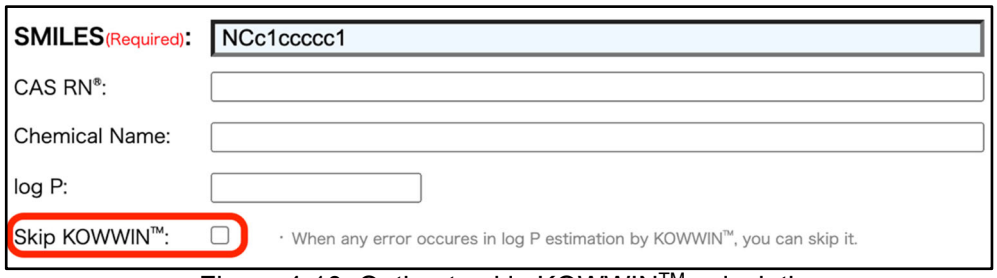


Figure 4-10. Option to skip KOWWIN™ calculation

5. QSAR Prediction Results

After entering the necessary information about the query chemical in the input page and clicking the "Predict" button, the QSAR Prediction Results screen is displayed (Figure 5-1). This screen provides a summary of the query chemical, the QSAR classes to which the query chemical was assigned for each type of predicted toxicity, the ecotoxicity prediction results for each QSAR class, and the statistics associated with each QSAR class.

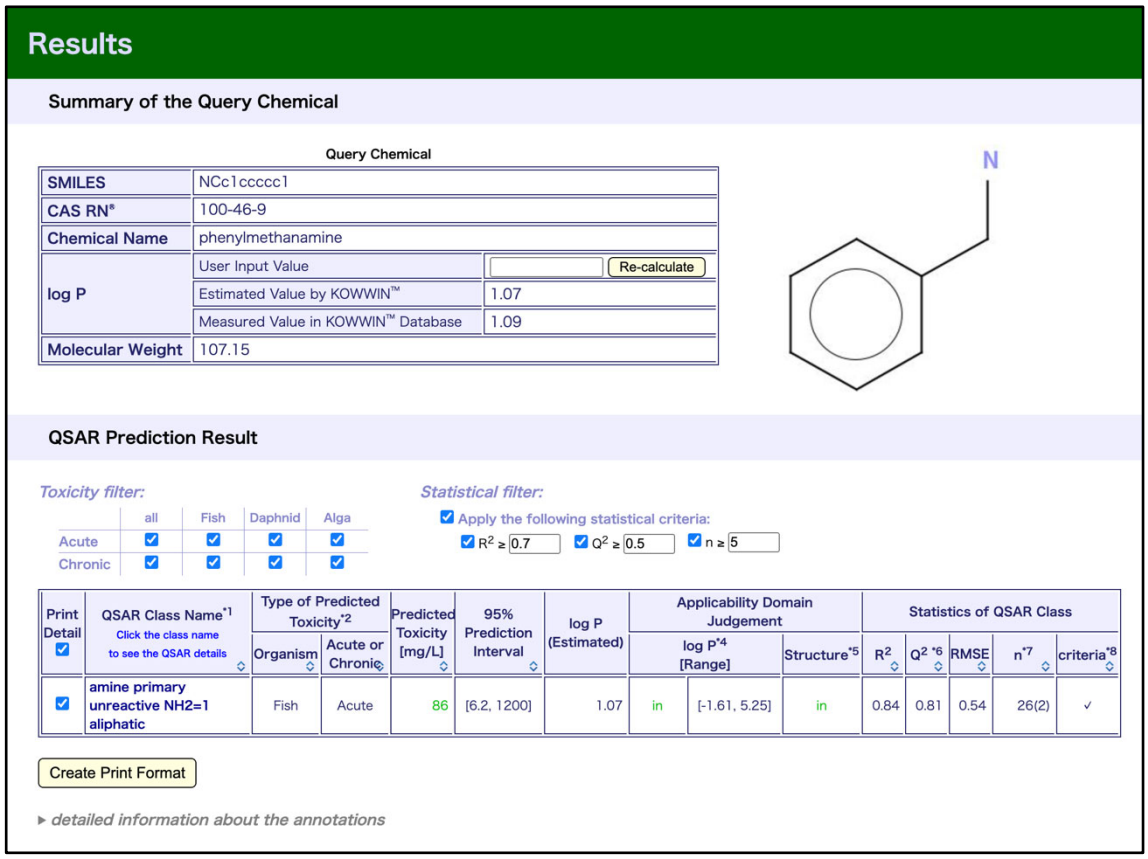


Figure 5-1. QSAR Prediction Results screen

(1) Summary of the Query Chemical

The upper part of the QSAR Prediction Results screen shows a summary of the information used for the prediction (Figure 5-2).

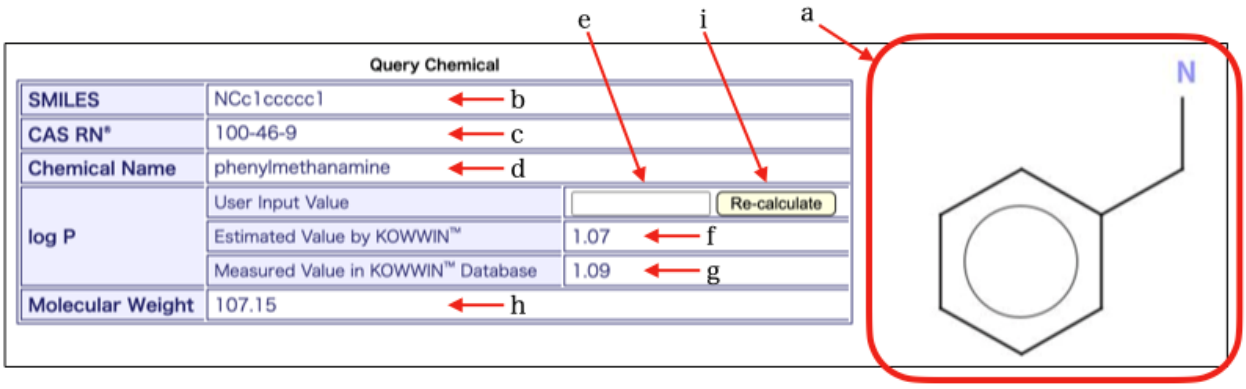


Figure 5-2. Summary of the query chemical

a Chemical structure of the query chemical.

b CAS number (displayed only when entered by the user).

Note: The CAS number is verified by means of the check digit only (the final digit in the CAS number), not the whole CAS number, and the number is tagged with "(incorrect)" if the check digit is incorrect . This may identify if an incorrect number has been input, but not whether the SMILES string matches the CAS number.

c Name of chemical substance (displayed only when entered by the user).

d SMILES string of the query chemical.

e Molecular weight of the query chemical calculated by Open Babel

f Log P value entered by the user.

g Log P value estimated by KOWWIN™.

h Measured log P value in the KOWWIN™ database.

Note: If there is more than one log P value for the query chemical in the KOWWIN™ database, all values in the database will be displayed.

i "Re-calculate" button: Clicking updates the QSAR prediction results with the log P value entered in the "User Input Value" field.

(2) QSAR Prediction Result

The middle section of the QSAR Prediction Results screen provides the QSAR class names of the query chemical, type of predicted toxicity, predicted toxicity results (in green), 95% prediction interval, log P value used for the query chemical, applicability domain judgement, and statistics related to the QSAR classes (Figure 5-3). The results are presented as a table with a series of checkboxes that can be used to filter the results.

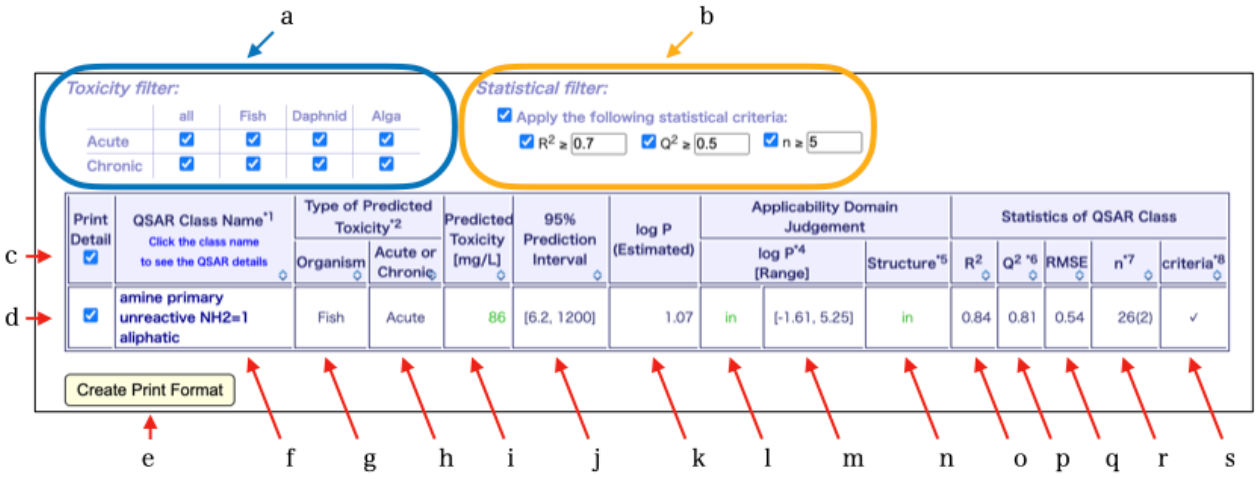


Figure 5-3. QSAR Prediction Results screen

Checkboxes

a "Include" checkboxes indicate the type of predicted toxicity results to show. All the boxes are checked by default.

b "Exclude" checkboxes: If a QSAR class meets any of the conditions stipulated for R^2 (coefficient of determination), Q^2 (internal validation index), or n (number of training set data), the results are not displayed. Which QSAR classes are shown can be specified by unchecking the boxes or changing the values. By default, all the boxes are checked, and the following values are entered: $R^2 < 0.7$, $Q^2 < 0.5$, or $n < 5$. QSAR classes with either value less than these limits are not displayed.

※When the leftmost checkbox is checked, all three checkboxes on the right side are checked. For example, checking the checkbox in the parentheses to the right of "Exclude" will check R2, Q2, and n checkboxes at the same time.

When no QSAR class is displayed for all types of predicted toxicity, "No applicable results" is displayed (Figure 5-4).

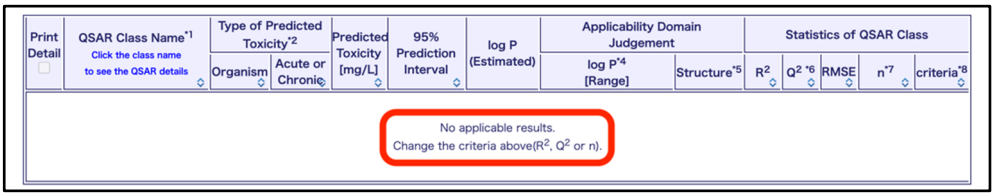


Figure 5-4. QSAR Prediction Results (No QSAR Class Shown)

c "Update" button: Clicking updates the QSAR results in line with any changes that have been made to the checkboxes.

Batch Printing

d Check boxes for inclusion in the final print format. By default, checkboxes are selected when the QSAR class meets the criteria $R^2 \geq 0.7$ and $Q^2 \geq 0.5$ and $n \geq 5$, the log P judgement is "in", and the structure judgement is "in" or "in (conditionally)".

e Button for displaying the final print format for review.

QSAR class names and links

f QSAR class names. Click the name to go to the Verify QSAR screen.

Type of predicted toxicity

g Organism (fish, daphnid, or alga)

h Acute or chronic

The following combinations are available for e and f:

Type of predicted toxicity		Testing method	Test duration	Indicator
Organism	Acute/ chronic			
Fish	Acute	Fish acute toxicity test (OECD TG 203)	96 h	LC_{50}
Daphnid	Acute	Daphnia magna immobilization test (OECD TG 202)	48 h	EC_{50}
Alga	Acute	Algal growth inhibition test (OECD TG 201)	72 h	EC_{50}
Fish	Chronic	Fish early-life-stage toxicity test (OECD TG 210)	Embryonic stage and 30 d after hatching*	NOEC
Daphnid	Chronic	Daphnia magna reproduction test (OECD TG 211)	21 d	NOEC
Alga	Chronic	Algal growth inhibition test (OECD TG 201)	72 h	NOEC

*Although the test duration of the fish early-life-stage test differs by species and the number of days before hatching, it is set as "embryonic stage and 30 days after hatching" for *Oryzias latipes* used in the ecotoxicity tests conducted by the Japanese Ministry of the Environment.

Predicted values

i Predicted toxicity value

j 95% prediction interval of the predicted toxicity value

log P

k Type of log P used for the query chemical, which is determined in the following priority order:

1. User Input: log P value entered by the user
2. Estimated: log P value estimated by KOWWIN™

l Log P value of the query chemical

Judgement of applicability domain

m Log P judgement result.

n [Minimum value, maximum value] of the descriptor log P of the training set data in the QSAR class (applicability domain in terms of log P).

o Structure judgement result.

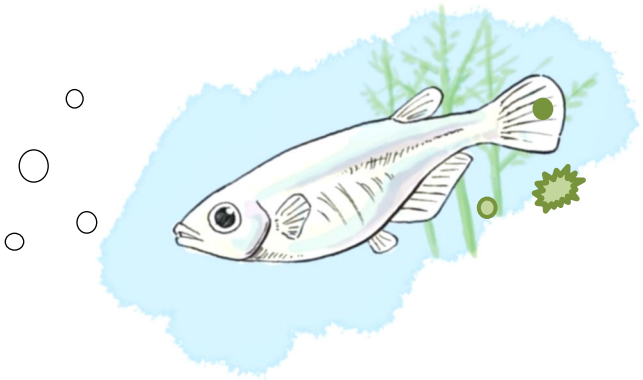
Statistics

p R^2 (coefficient of determination) of the QSAR equation

q Q^2 (internal validation index by the leave-one-out method) of the QSAR equation (for details, see the KATE2025 technical document).

r RMSE (root mean square error) of the QSAR equation

s Number of training set data used for the QSAR equation (support chemicals are not included). The value in parenthesis is the number of support chemical data points.



6. Details of QSAR Class Information

In the QSAR prediction results, clicking on a link in the "QSAR Class Name" column will open a new window showing detailed information of the QSAR class (Figures 6-1 to 6-3). On this screen, a graph of the regression equation for the QSAR class, detailed information on the training set data and support chemicals, definition of the structure class, and substructures of the query chemical are shown. The page can be considered to consist of nine sections, each of which is described below.

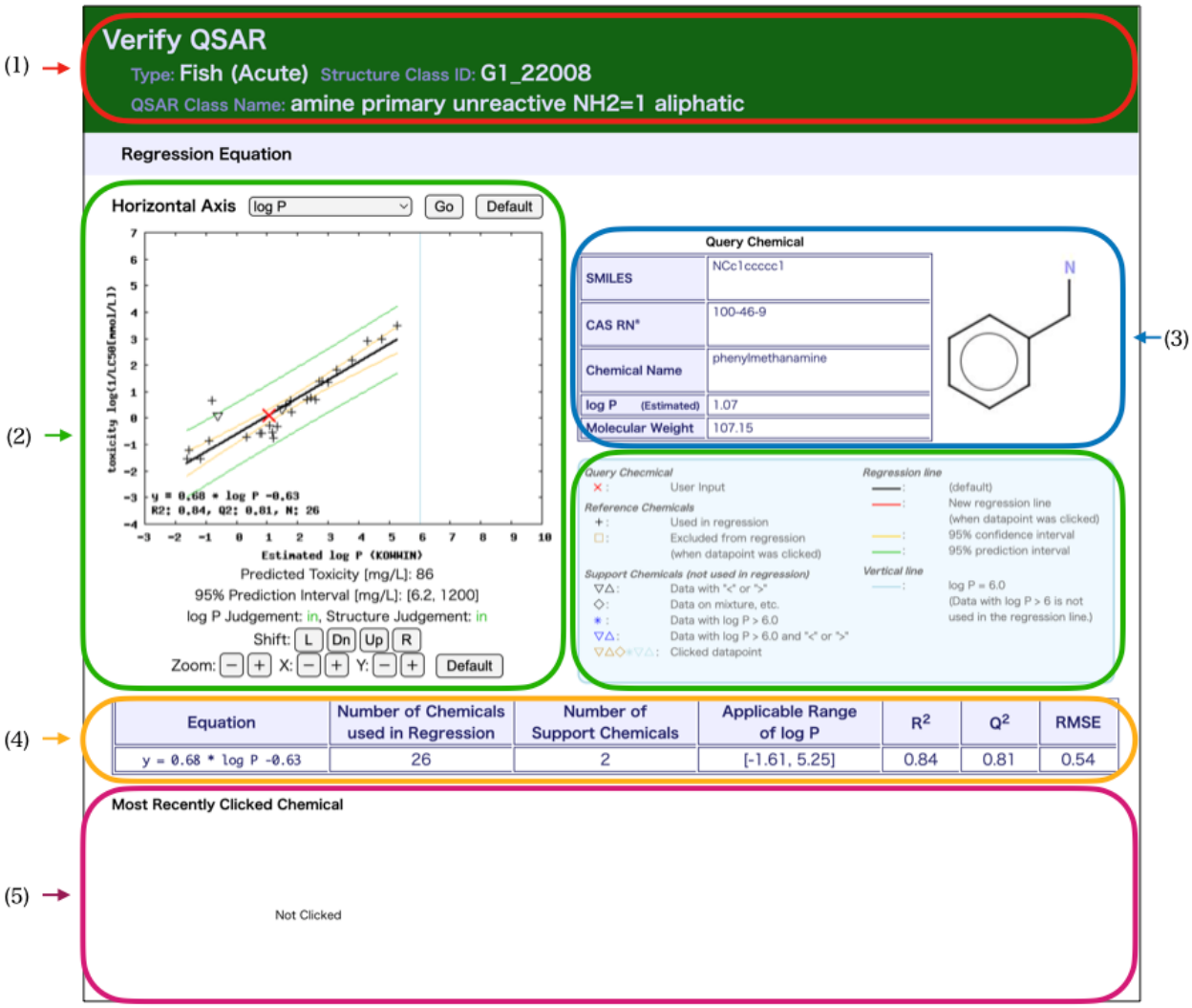


Figure 6-1. Verify QSAR screen (top)

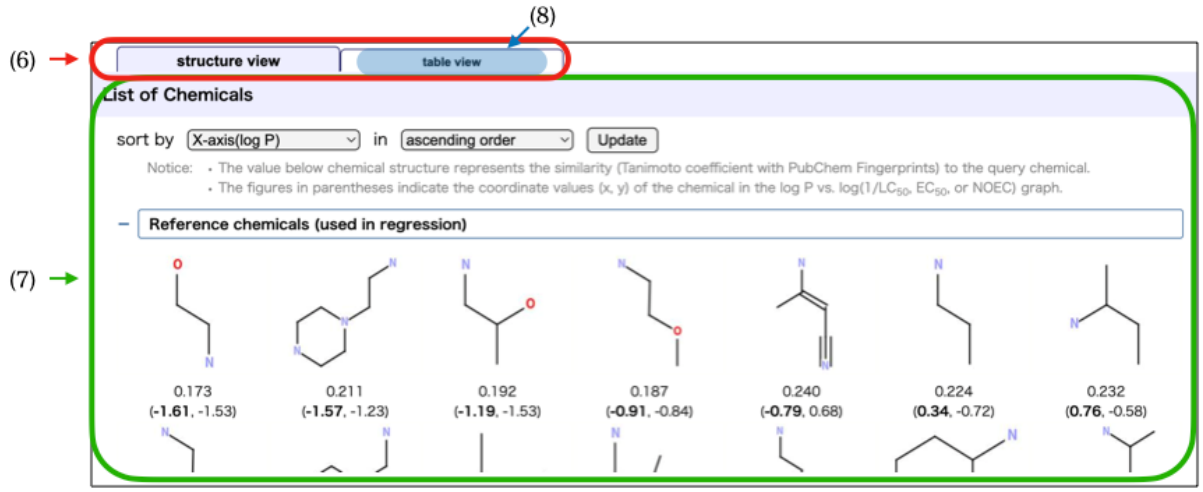


Figure 6-2. Verify QSAR screen (middle)

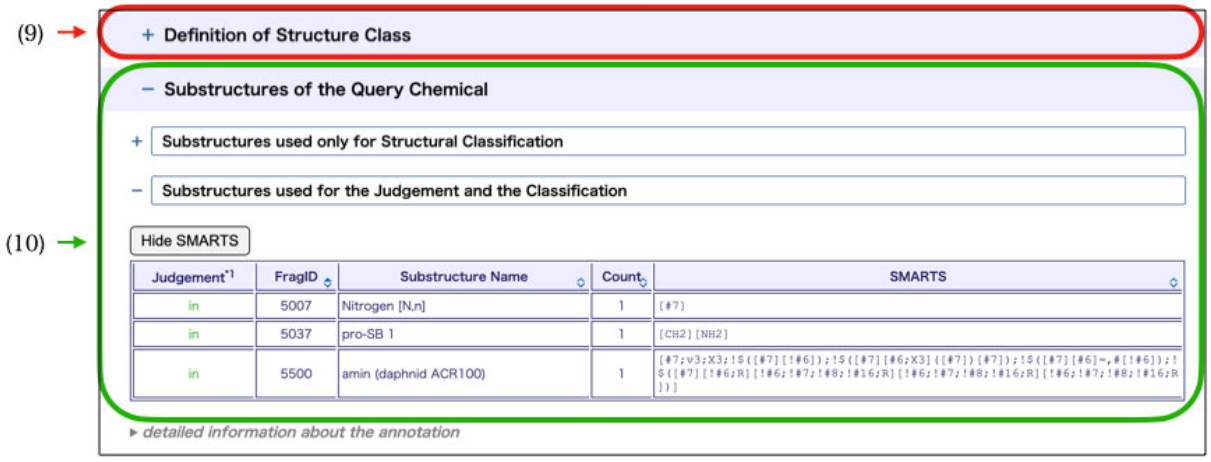


Figure 6-3. Verify QSAR screen (bottom)

(1) Summary of QSAR Class

The dark blue band at the top of the page provides summary information (Figure 6-4).

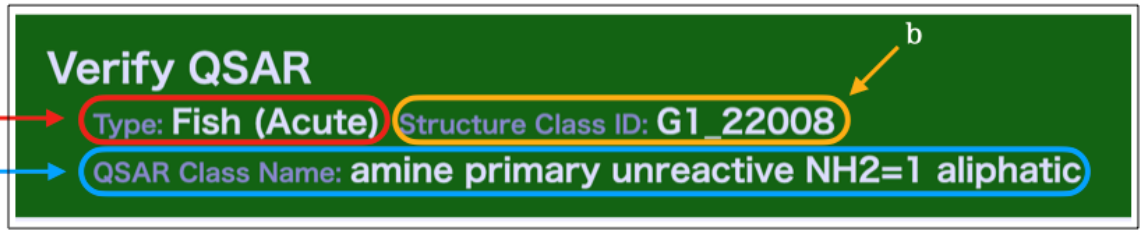


Figure 6-4. QSAR class basic information

a Type of predicted toxicity for the QSAR class
b Structure class ID corresponding to the QSAR class (see "(8) structure class definition")
c Name of the QSAR class

(2) Graph

The graph of the regression equation consists of four parts (Figure 6-5, 6-6).

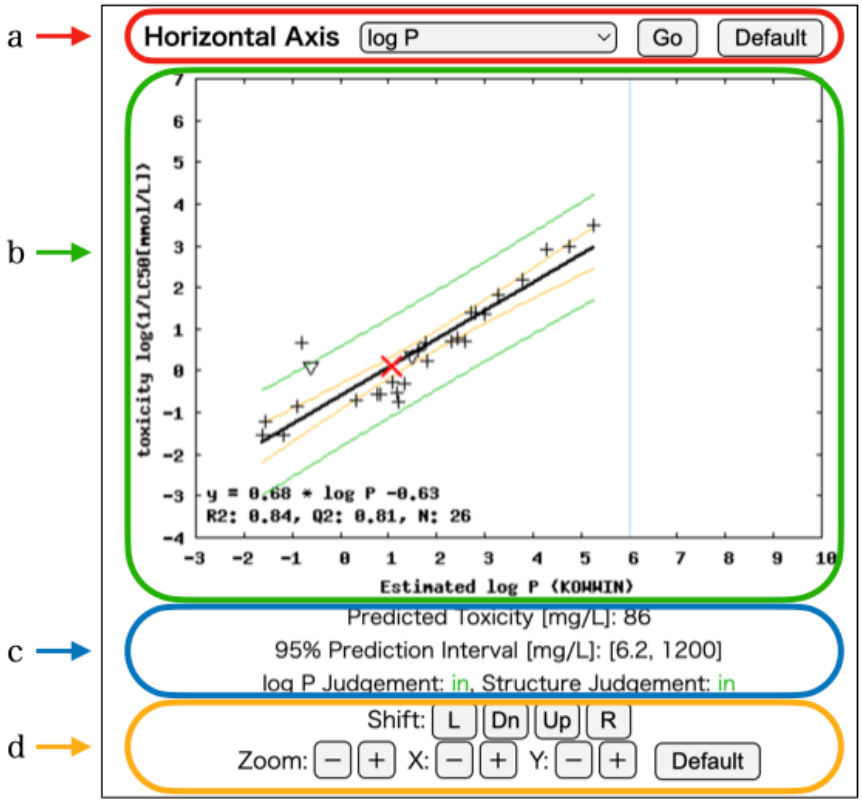


Figure 6-5. Graph of the regression equation (log P vs. Toxicity[$\log_{10}$(1/(mmol))])

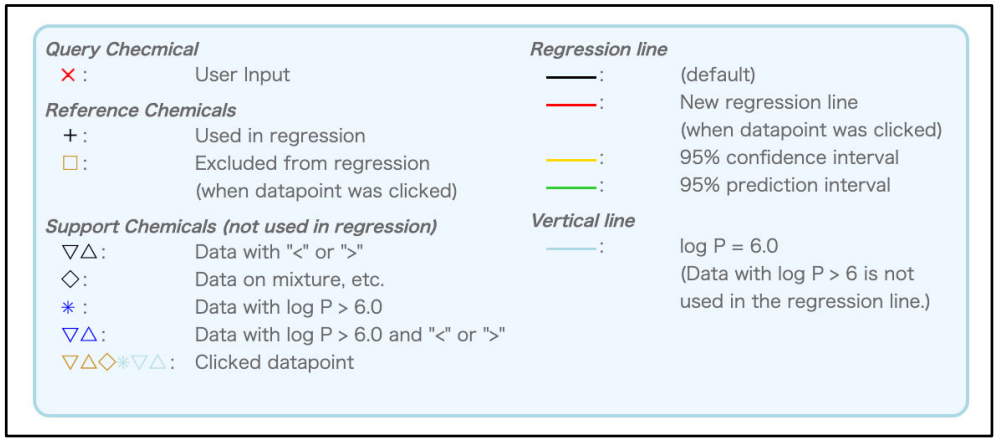


Figure 6-6. Graph legend

a Field to select the horizontal axis

Change from "log P" (default) to "Predicted Variable" and click the "Update" button to draw a graph of "measured toxicity value vs. predicted toxicity value", where toxicity value is presented as log(1/{LC_{50}, EC_{50}, or NOEC} in mmol/L). Click the "Default" button to return to the default display.

b Graph and legend

The graph of the regression equation is displayed on the left with the legend on the right. The following symbols are used in the legend:

×: Query chemical

+: Training set

Black line: Regression line

Orange curve: 95% confidence interval of the regression equation

Green curve: 95% prediction interval of log(1/{LC_{50}, EC_{50}, or NOEC})

------Support Data------

＊: Plot for a chemical with log P > 6.0

▽△: Plot for a chemical with an inequality sign

◇: Plot for an outlier

------When a training set (shown as "+" in the graph) is deleted by clicking on the symbol in the graph------

□: Deleted data

Red line: Regression line calculated without the deleted data

------Bottom-left of the graph------

First line Regression equation

Second line "R^2" Coefficient of determination of the regression equation

"N" Number of training set data (support chemical data are not included)

c Information for the query chemical

1st line: Predicted toxicity value (in mg/L]

2nd line: 95% prediction interval of predicted toxicity value (in mg/L]

3rd line: Log P judgment and structure judgment results

d Graph display buttons

First line

Shift L: Shift view left R: Shift view right

Dn: Shift view down Up: Shift view up

Second line

Zoom −: Zoom out whole graph +: Zoom in whole graph

X −: Zoom out the X axis +: Zoom in the X axis

Y −: Zoom out the Y axis +: Zoom in the Y axis

Selection of individual chemicals

When a point (+) on the graph (Figure 6-7) is clicked or one of the structures on the list is clicked (Figure 6-8), the chemical is excluded from the calculation, and an updated regression line is shown in red (Figure 6-6). The prediction interval (green curves) and confidence interval (orange curves) are also recalculated without the selected chemical. Multiple chemicals can be removed from the calculation.

When one or more training set data are removed from the calculation, the number of chemicals removed is displayed in the upper left of the graph (Figure 6-7) and, in addition to the original QSAR equation, the modified QSAR equation is displayed after an arrow (Figure 6-7). Updated information on R^2 and N is also displayed.

When a training set is removed, the + symbol for the chemical is changed to a □ (Figure 6-7), and the structure is highlighted by a violet frame in the training set data list (Figure 6-8). Click the point again to deselect.

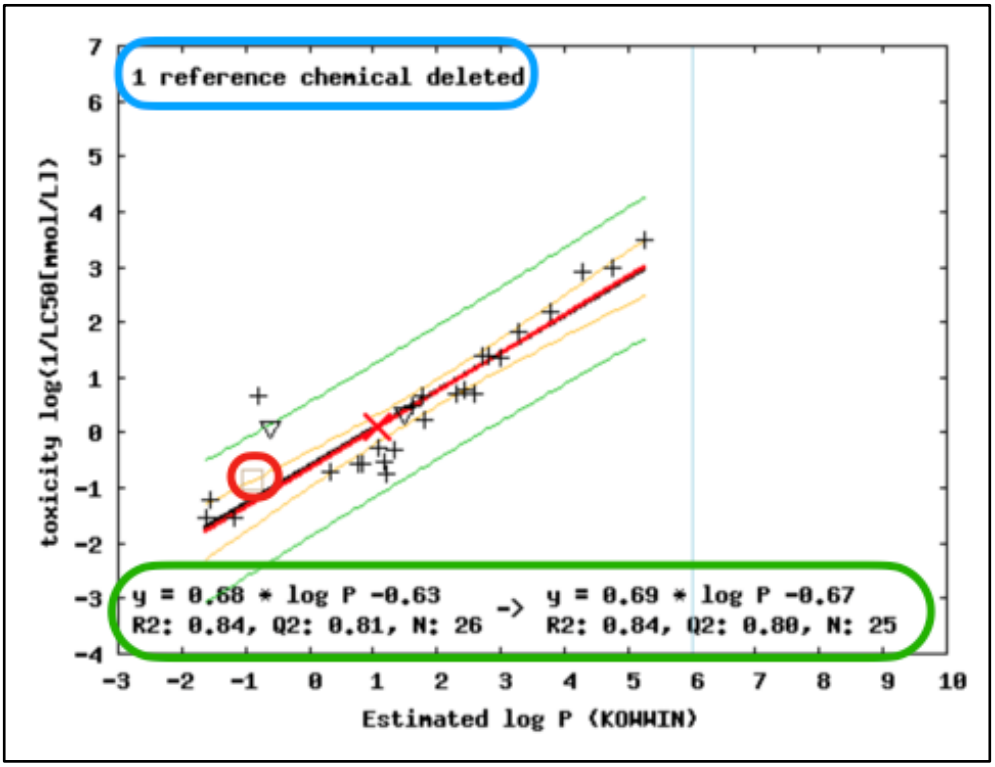


Figure 6-7. Selection of training set (graph)

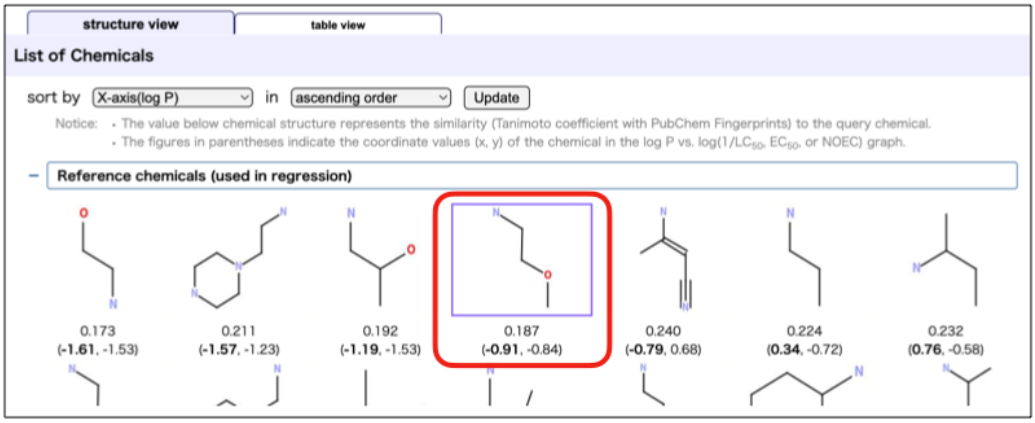


Figure 6-8. Selection of training set (Training set data dropdown)

(3) Information About the Query Chemical

The chemical structure and basic information about the query chemical are displayed to the

right of the graph legend (Figure 6-9).

Query Chemical

SMILES	NCc1ccccc1
CAS RN®	100-46-9
Chemical Name	phenylmethanamine
log P (Estimated)	1.07
Molecular Weight	107.15

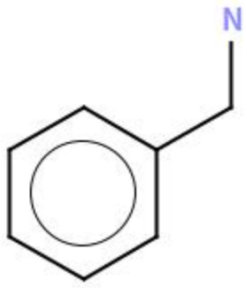


Figure 6-9. Query chemical information

SMILES: SMILES string of the query chemical

Chemical Name (User Input): Chemical name entered by the user

CAS RN (User Input): CAS number entered by the user

Molecular Weight: Molecular weight of the query chemical

log P: Log P value of the query chemical. "(User Input)" is appended for a user-defined value and "(Estimated)" is appended for a value estimated by KOWWIN™.

(4) Information About the Regression Equation

Information about the regression equation (QSAR equation) is displayed below the graph (Figure 6-10).

Equation	Number of Chemicals used in Regression	Number of Support Chemicals	Applicable Range of log P	R^2	Q^2	RMSE
y = 0.68 * log P -0.63	26	2	[-1.61, 5.25]	0.84	0.81	0.54

Figure 6-10. Information about the regression equation

Equation: Regression equation (QSAR equation)

Number of Chemicals used for Regression: Number of training set data used for the regression calculation

Number of Support Chemicals: Number of support chemicals identified

Applicable Range of log P: Minimum and maximum log P values for the training set data

R^2: Coefficient of determination of the QSAR equation

Q^2: Internal validation index by the leave-one-out method of the QSAR equation (for details, see the KATE2025 technical document)

RMSE: Root mean square error of the QSAR equation

(5) Chemical Clicked Last

The "Chemical Clicked Last" section provides information about the last clicked chemical within the Training set data or Support Chemicals list. In its default state (i.e., when no chemical has been clicked), the page reports "Not Clicked" (Figure 6-11).

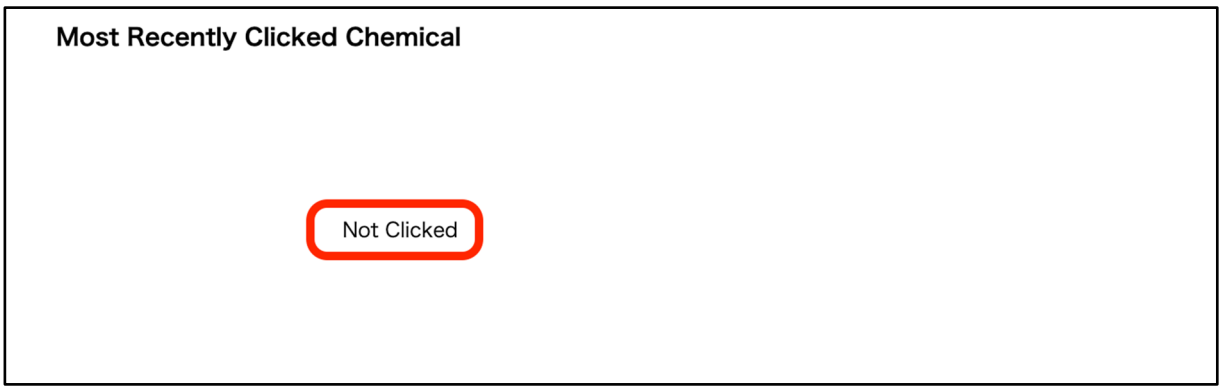


Figure 6-11. Chemical Clicked Last section: before a chemical has been clicked

When a chemical is clicked, the following information is displayed (Figure 6-12):

- Chemical structure
- SMILES string
- CAS number
- IUPAC name
- Coordinates on the regression graph
- Square of residual
- Molecular weight
- Measured toxicity value information ({LC_{50}, EC_{50}, or NOEC}; species; reference)
- Note (information about toxicity test etc., if applicable)

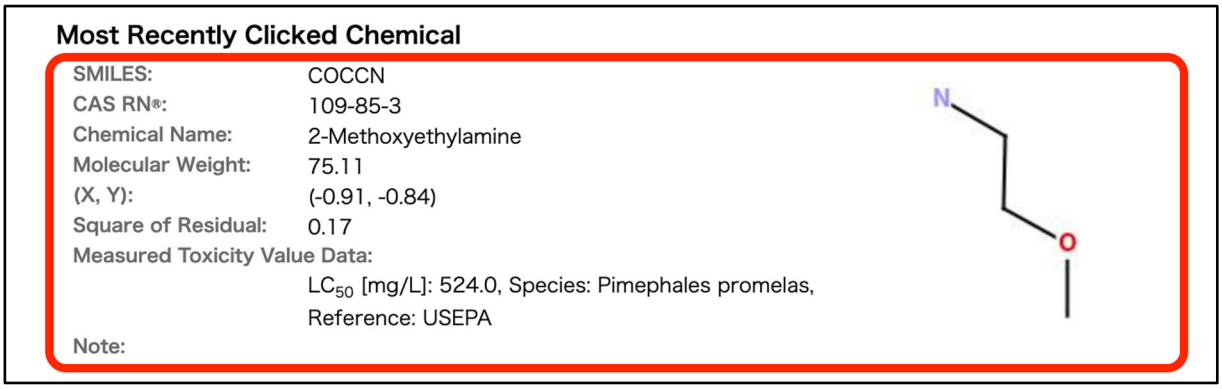


Figure 6-12. Chemical Clicked Last section: after a chemical has been clicked

(6) Functionality to switch display of chemicals information in QSAR class QSAR

In the middle section of Verify QSAR screen, you can use tabs to switch structure view (list of structural formulas) and table view (list of data on chemicals).

(7) List of Chemical Structures

The “Chemical list” dropdown provides lists containing the structural formulas of the training set and support chemicals within the QSAR class (Figure 6-13).

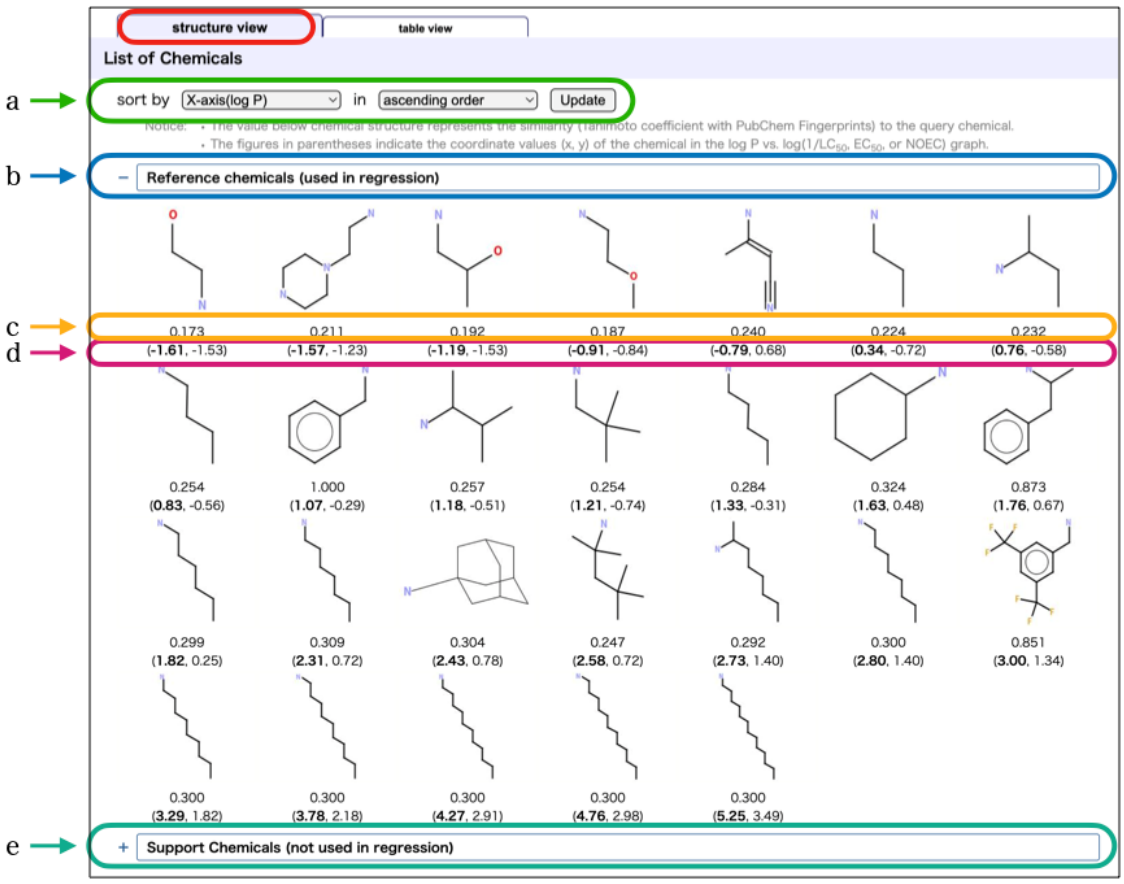


Figure 6-13. “Chemical List” dropdown

Each of these lists can be sorted using the dropdown menus and “Update” button at the top of the page (Figure 6-12a). The available options are sorted by CAS number, X coordinate (log P), Y coordinate (log (1/{LC_{50}, EC_{50}, or NOEC})), square of residual, and similarity, and each can be arranged in ascending or descending order. By default, the chemicals are arranged in ascending order of X coordinate (log P). Here, X coordinate, not X axis, is used in case “Horizontal Axis” (see “a” in (2)) is selected as “log P”; that way, even if the chemicals are sorted by X coordinate, when the graph of “Measured toxicity value vs. Predicted toxicity value” is displayed they are still sorted by log P value.

Clicking the “Training set data” dropdown reveals a list containing the structural formulas of the training set data (Figure 6-13 b). The values displayed directly below the chemical structures (Figure 6-13 c) are their similarity with the query chemical (Tanimoto coefficient using PubChem fingerprints; for details, see the KATE2025 technical document); the coefficient falls between 0 and 1, and a structure with higher similarity to the query chemical has a value closer to 1. The values displayed in parentheses (Figure 6-13 d) are the X coordinate (i.e., log P value) and Y coordinate

(log(1/{LC_{50}, EC_{50}, or NOEC})) of the chemical substance in the default graph.

Clicking the "Support Chemicals" dropdown (Figure 6-13 e) reveals three additional lists showing the structures of the support chemicals divided into those with data containing an inequality sign, outlier chemicals, and chemicals with a log P > 6 (Figure 6-14).

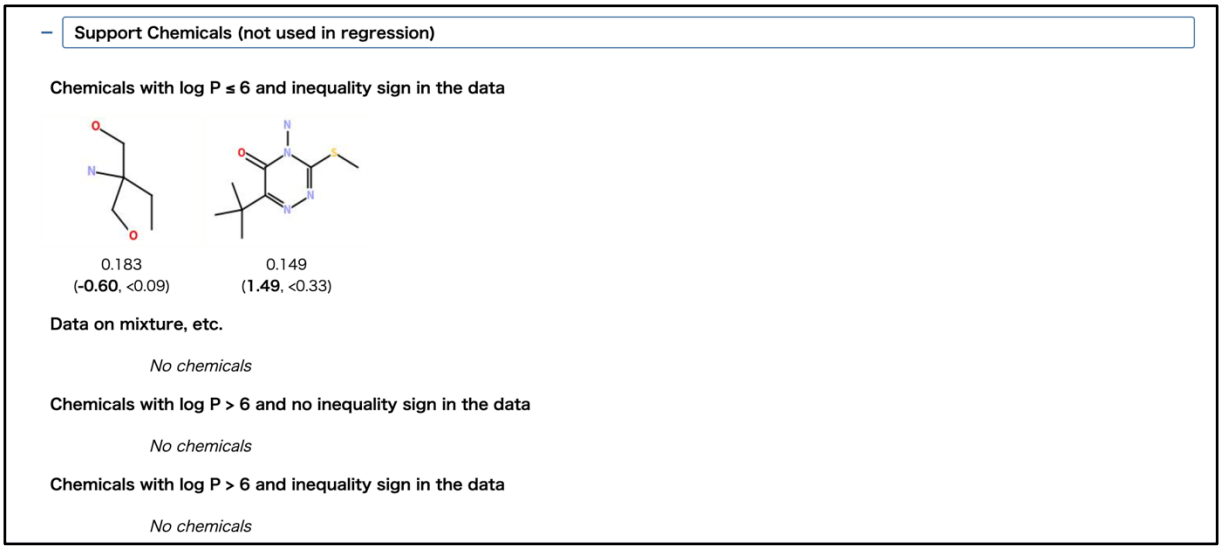


Figure 6-14. "Support Chemicals" dropdown

(8) Chemical Data

The "Chemical Data" dropdown provides the chemical details of the training set data (Figure 6-15) and support chemicals in the class. Both lists are formatted as shown below.

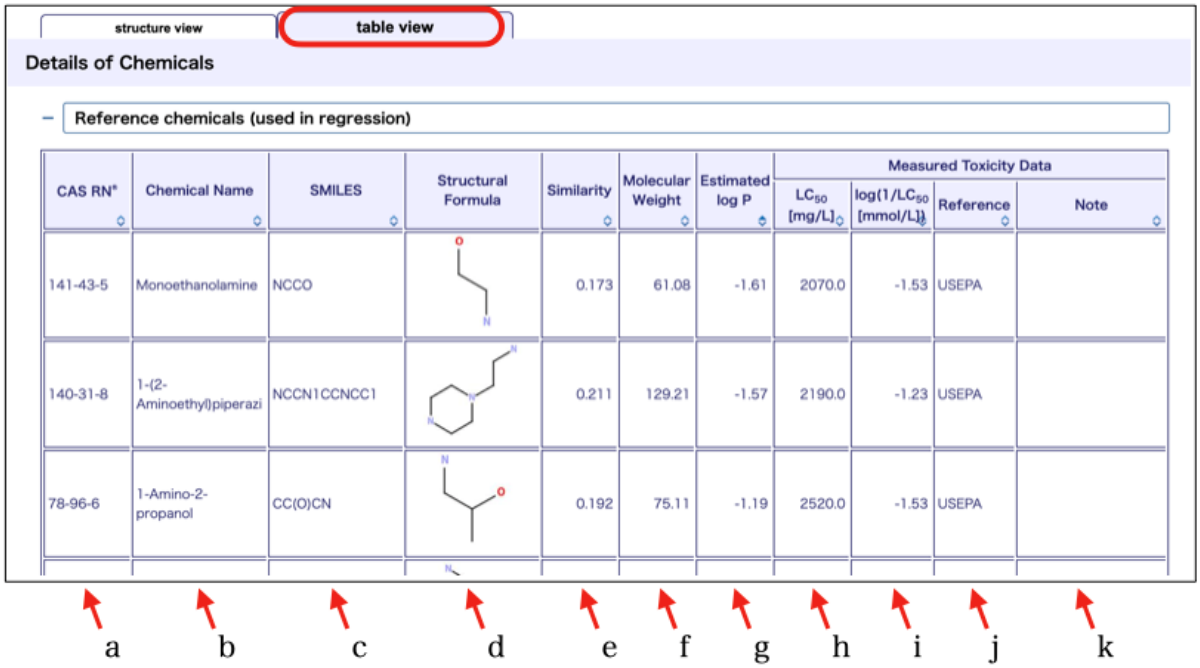


Figure 6-15. "Chemical Data" dropdown

a CAS No.: CAS number

b Chemical Name: Chemical name used in KATE2025

c SMILES: SMILES string used in KATE2025

d Structure Formula: Structural formula

e Similarity: Similarity between the chemical in the list and the query chemical

f Molecular Weight: Molecular weight

g Estimated log P: Log P value estimated using KOWWIN™

h EC_{50}*: Toxicity value (in mg/L) based on the results of ecotoxicity tests

i $\log(1/EC_{50}^{*}$ [mmol/L]):

*For h and i, the LC_{50}, EC_{50}, or NOEC corresponding to the type of predicted toxicity of the QSAR class is automatically displayed.

j Reference: Source of the original toxicity data. The year indicates the year the test was implemented. "MOE" means the data were obtained from the following webpage: https://www.env.go.jp/content/000212351.pdf (results of aquatic toxicity tests conducted by the Japanese MOE [in Japanese]; accessed March 1, 2025). "USEPA" means the data were fish acute toxicity test results obtained from the US EPA fathead minnow database as a training set data dataset.

k Note: Other information about the chemical substance

The order in which the data are displayed is determined by the sorting filter at the top of the page (see "a" in (6)).

Sorting Functionality

Some tables in KATE2025 implement a sorting function by clicking on the table header (columns) that can be sorted are marked with an upper and lower triangle in the bottom right-hand corner of the header (Figure 6-16).Each click on the header switches the ascending-descending order and fills in the corresponding triangle (Estimated log P column in the case of Fig. 6-16).

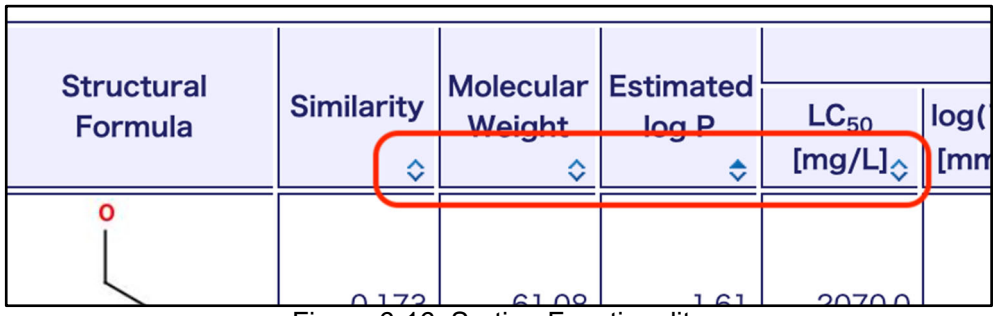


Figure 6-16. Sorting Functionality

※ Sorting by multiple columns is possible by holding down the Shift key and clicking.

※ In some tables, it is possible to return to the initial state by clicking on the header of a column that cannot be sorted.

(9) Structure Class Definition

Clicking the "Definition of Structure Class" dropdown reveals information about the structure class corresponding to the QSAR class (Figure 6-17). The assigned structure class is highlighted in yellow, with the structure class ID in the "Structure ID" column, the name of the structure class or substructure in the "Description" column The "Decision Tree" column contains the definition of the structure class or substructure.

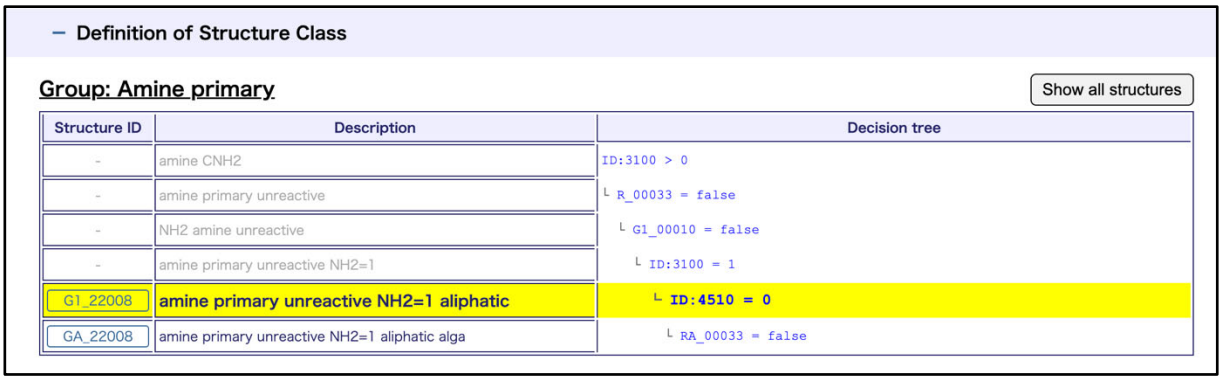


Figure 6-17. "Structure Class Definition" information

(10) Substructures of the Query Chemical

The "Substructures of the Query Chemical" section contains two dropdowns. Clicking the lower "Substructures used for the Judgement and the Classification" dropdown reveals a list of the substructures used in both the structural classification and the structural judgement (Figure 6-18).

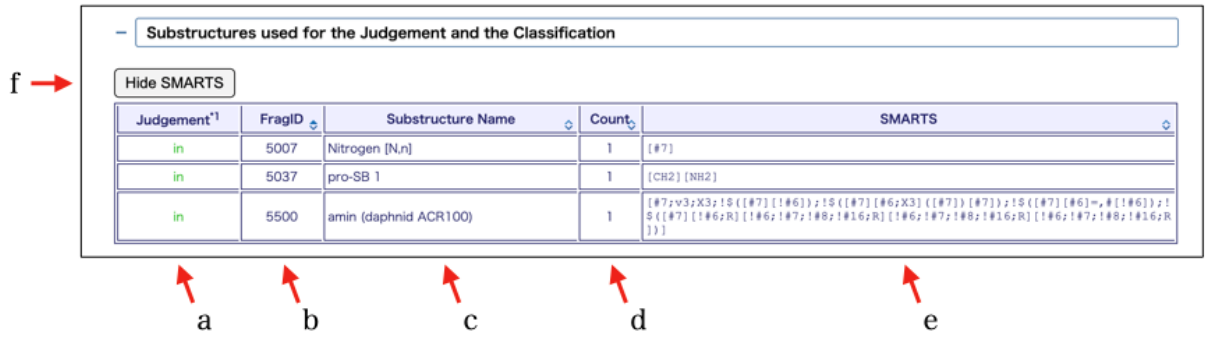


Figure 6-18. "Substructures used for the Judgement and the Classification" dropdown

a Judgement: This column provides detailed information about the structural judgment results, and you can see which of the substructures contained in the query chemical are the cause of the judgment results.

Three judgments are possible:

in: The substructure is found in the "substructures for structure judgement" extracted from the training set data in the QSAR class.

in(p): The substructure does not meet the condition of "in", but the substructure is found in "substructures for structure judgement" extracted from the training set data in the Narcotic Group class.

out: The substructure does not meet the conditions of "in" or "in (conditionally)"; that is, the substructure is found in neither the "substructures for structure judgement" extracted from the training set data of the QSAR class nor those from the Narcotic group class.

b FragID: ID of the substructure. This is a four-digit number and was arbitrarily set for convenience during the development of KATE. At present, the FragIDs used in this table all start with the number 5 (for details, see the KATE2025 technical document).

c Substructure Name: Name of each substructure (please note that the names may change in the future).

d Count: The number of substructures in the query chemical corresponding to SMARTS

e SMARTS: Definition of the substructure in SMARTS notation

f Show/Hide SMARTS Button: Click to hide or display the "SMARTS" column.

Clicking the "+Substructures used only for Structural Classification" dropdown reveals a list of the substructures used only for the structural classification (Figure 6-18).

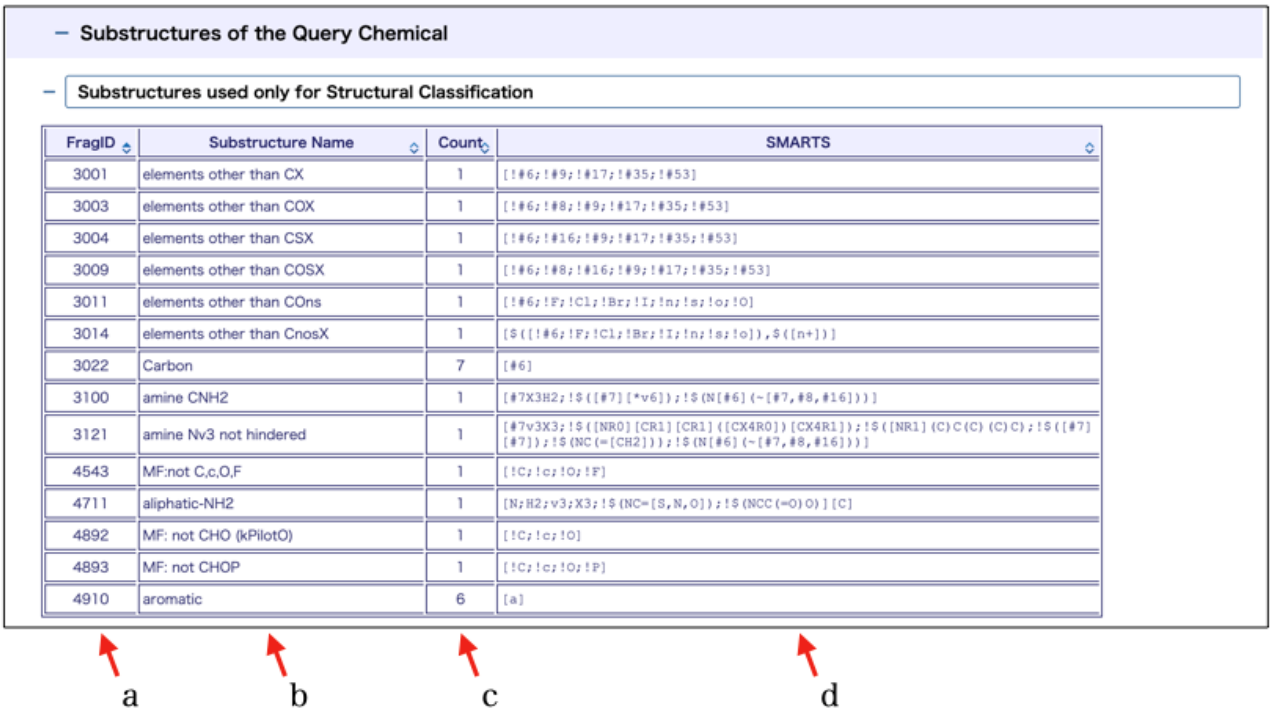


Figure 6-18. "Substructures of a Query Chemical" dropdown (used only for structural classification)

a FragID: ID of the substructure. This is a four-digit number and was arbitrarily set for convenience during the development of KATE. At present, the FragIDs used in the table all start with a number 3, 4, 6, or 7 (for details, see the KATE2025 technical document.

b Substructure Name: Name of each substructure (please note that the names may change in the future).

c Count: The number of substructures in the query chemical.

d SMARTS: Definition of the substructure in SMARTS notation.

7. Prediction of Multiple Chemical Substances (Input screen)

Files containing multiple SMILES strings can be entered to perform sequential prediction for multiple chemical substances. Up to 1,000 chemicals can be predicted at a time.

(1) Input File format: “SMILES List”

In KATE2025, to process multiple SMILES strings consecutively, we use a custom file format called "SMILES List".

The SMILES List has the following features:

1. It is a tab-delimited text file.
2. Column names are limited to SMILES, CAS, NAME, LOGP, and ID, representing SMILES, CAS Number, chemical name, log P, and ID, respectively. Column names are case-insensitive.
3. Each line must include a SMILES string. Columns other than SMILES are optional.
4. The ID column can only contain alphanumeric characters, '-' and '_'.
5. The number of tabs must be the same in all lines.
6. The column order can be customized according to your preferences.

Two examples are provided below.

Example 1: Single column

```
SMILES
CCCCOC(=O)CS
CC(=C)CS
CC1(CC2(C)CC3(Br)C1)CC(Br)(C2)C3
CCCCCCCCCBr
CCCCCCCCC(Br)CBr
CCCCCCCCCCBr
```

Example 2: Multiple columns

```
ID          NAME        LOGP        SMILES      CAS
A01         name1       -0.8        CCCCOC(=O)
A20         name2                   CC(=C)CS
B06         name3       1.3         NCc1cccnc1  3731-52-0
```

* "   " and "   " denote tab characters.
For ID: A01, CAS Number is empty. For ID: A20, both log P and CAS Number are empty.

In KATE2025, to see how to format the file, click “about the SMILES List format” below the “Prediction of Multiple Chemicals” section(Figure 7-1).

Figure 7-1. Link to details about formatting a SMILES List

(2) Prediction Procedures

Step 1. Click the “Select” button on the Input screen (Figure 7-2).

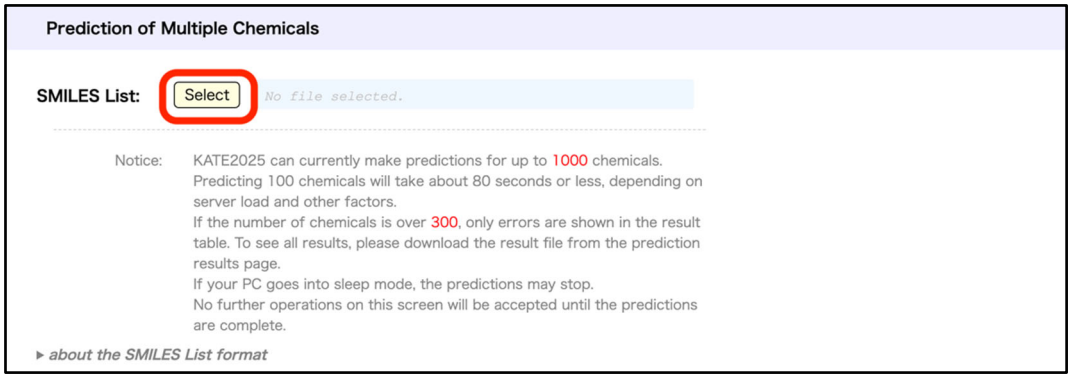


Figure 7-2. Select Button for “Prediction of Multiple Chemicals”

Step 2. Select the input SMILES list and click the “Open” button (Figure 7-3).

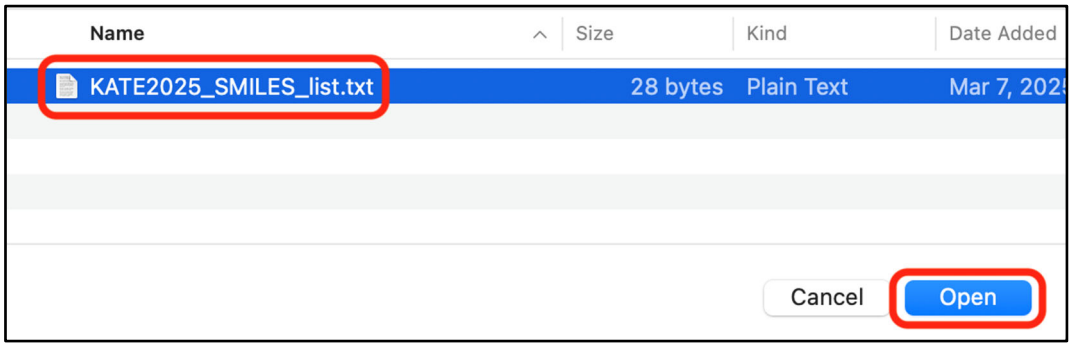


Figure 7-3. Selection of SMILES list

If the SMILES list file contains the following issues, the alert will be out.

1. File other than text was given.
2. The file format does not match the SMILES list (e.g., column name is wrong, the number of tabs at each line does not match, etc.)
3. The file contains more than 1,000 chemicals.

.Step 3. Click the “Predict” button to start the prediction (Figure 7-4).

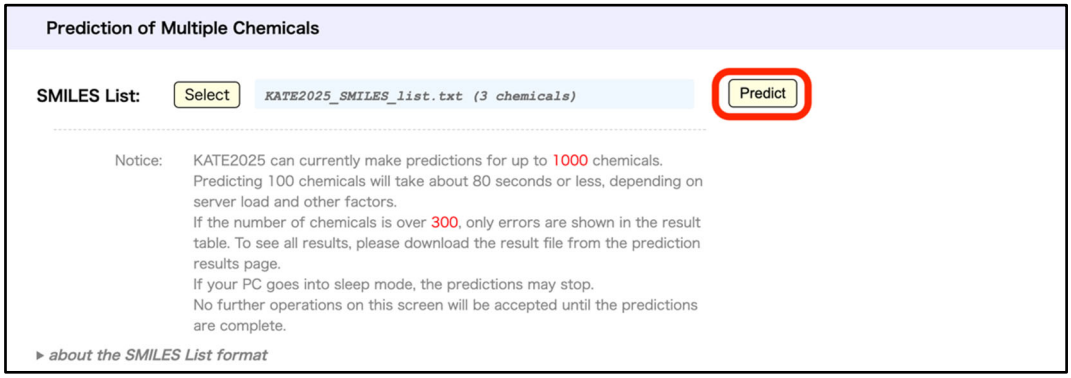


Figure 7-4. “Predict” Button for using a SMILES list

Once prediction starts, progress bar is shown on the right side (Figure 7-5). “Total time” is displayed, showing approximate time needed for completing prediction (the time does not count down along with progress bar). This time is calculated based on the number of given chemicals and does not show precise prediction time. The time will change according to the structures of chemicals, but normally the prediction will finish in about 80 minutes for 100 chemicals.

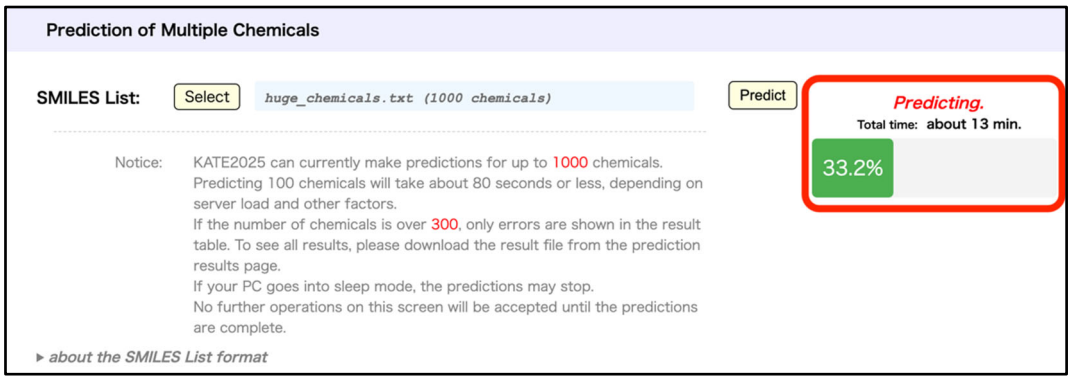


Figure 7-5. Progress of prediction

(3) QSAR Prediction Results (Output screen)

After the calculation, the prediction results are displayed (Figure 7-6).

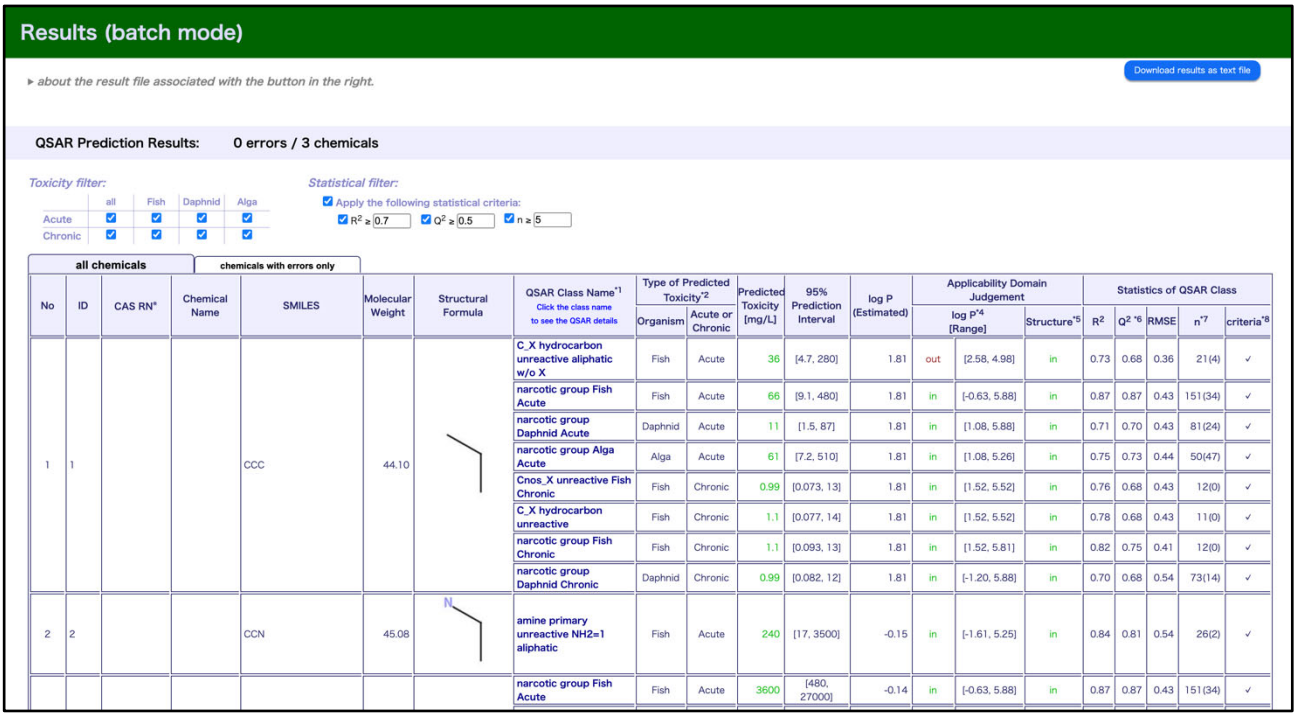


Figure 7-6. Results obtained from using a SMILES list

Click on the QSAR class name in the QSAR Class Name column to go to the Verify QSAR screen. The results screen and Verity QSAR screen. SMILES with errors whose toxicity could not be predicted are shown in the table with error message. You can switch the results with tabs (Figure 7-7).

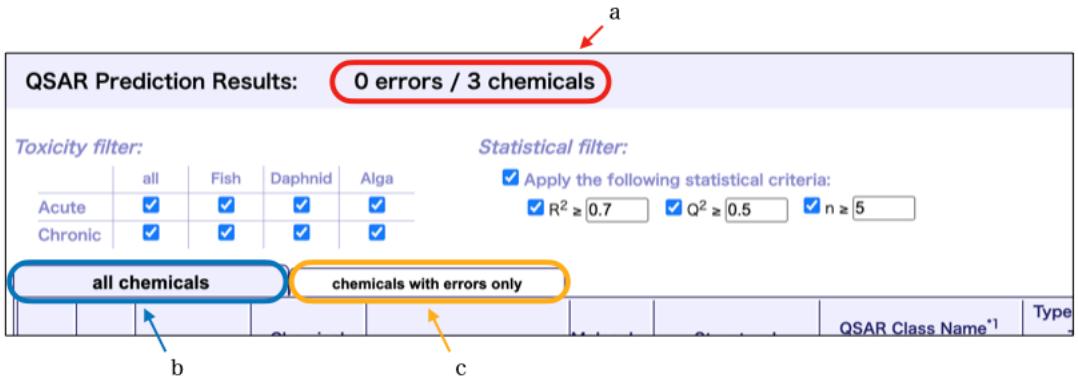


Figure 7-7. Tabs to switch results of multiple chemicals prediction

a Number of chemicals and number of errors
If error(s) occurred during prediction, the number of chemicals with error is shown in red.

b Tab for all chemicals
Prediction results for all chemicals are shown.

c Tab for chemicals with error
Only chemicals with error are shown (Figure 7-8).

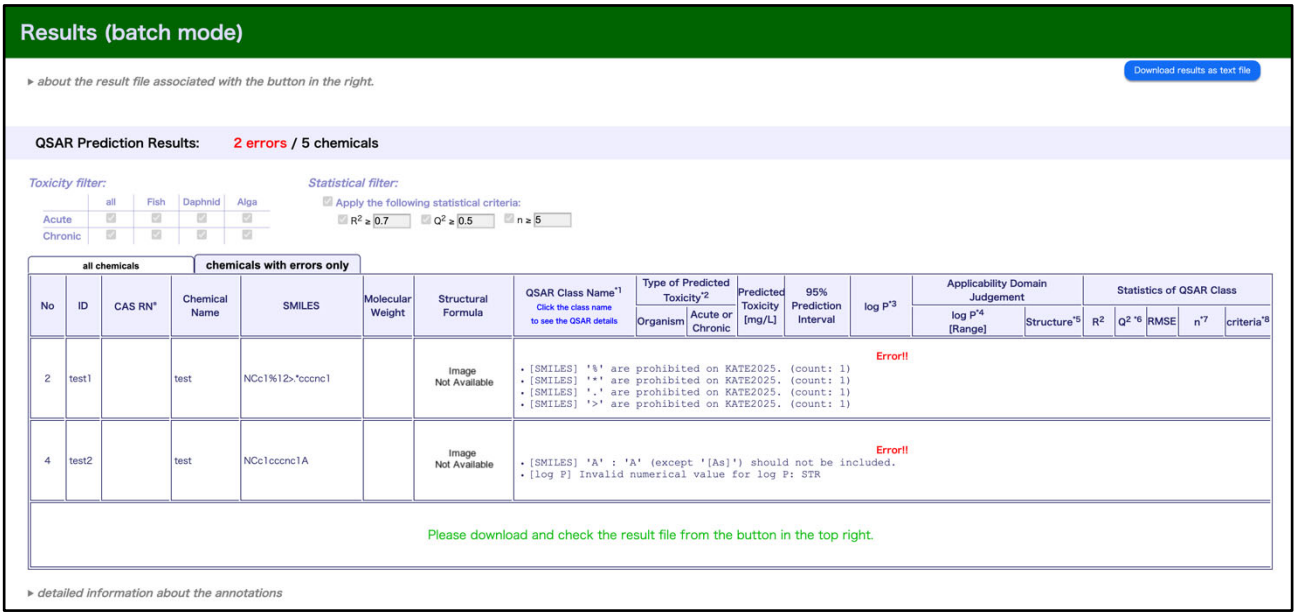


Figure 7-8. Example of showing only chemicals with error

If more than 300 chemicals were predicted, prediction results are not shown on web browser (Figure 7-9). You can download prediction results as tab-separated value (tsv) text file (Figure 7-10) by clicking blue button on the right top side (in case of less than 300 chemicals as well).

Figure 7-9. In case of more than 300 chemicals predicted

If more than 300 chemicals were predicted, prediction results are not shown on web browser. You can download prediction results as tab-separated value (tsv) text file (Figure 7-10) by clicking blue button on the right top side (in case of less than 300 chemicals as well).

ID	CAS RN	Chemical Name	SMILES	Molecular Weight	QSAR ID	QSAR Class Name	Organism	Acute or Chronic	Predicted Toxicity	95% Prediction Interval	log P Value	log P Type	log P Judgement	log P Range	Structure Judgement	R2	Q2	RMSE	n	criteria
1			CCC	44.1	12120341	C_X hydrocarbon unreactive aliphatic w/o X	Fish	Acute	36	[4.7, 280]	1.81	Estimated	out	[2.58, 4.98]	in	0.73	0.68	0.36	21(4)	yes
1			CCC	44.1	12899941	narcotic group Fish Acute	Fish	Acute	66	[9.1, 480]	1.81	Estimated	in	[-0.63, 5.88]	in	0.87	0.87	0.43	151(34)	yes
1			CCC	44.1	22120341	C_X hydrocarbon unreactive aliphatic w/o X	Daphnid	Acute	9.3	[0.67, 130]	1.81	Estimated	out	[2.58, 5.74]	in	0.66	0.51	0.42	15(3)	no
1			CCC	44.1	22899941	narcotic group Daphnid Acute	Daphnid	Acute	11	[1.5, 87]	1.81	Estimated	in	[1.08, 5.88]	in	0.71	0.7	0.43	81(24)	yes
1			CCC	44.1	32120341	C_X hydrocarbon unreactive aliphatic w/o X	Alga	Acute	370	[1.6, 87000]	1.81	Estimated	out	[2.58, 4.08]	in	0.66	0.38	0.47	6(10)	no
1			CCC	44.1	32899941	narcotic group Alga Acute	Alga	Acute	61	[7.2, 510]	1.81	Estimated	in	[1.08, 5.26]	in	0.75	0.73	0.44	50(47)	yes
1			CCC	44.1	12100151	C_X hydrocarbon unreactive	Fish	Chronic	1.1	[0.077, 14]	1.81	Estimated	in	[1.52, 5.52]	in	0.78	0.68	0.43	11(0)	yes
1			CCC	44.1	12500151	Cnos_X unreactive Fish Chronic	Fish	Chronic	0.99	[0.073, 13]	1.81	Estimated	in	[1.52, 5.52]	in	0.76	0.68	0.43	12(0)	yes
1			CCC	44.1	12899851	narcotic group Fish Chronic	Fish	Chronic	1.1	[0.093, 13]	1.81	Estimated	in	[1.52, 5.81]	in	0.82	0.75	0.41	12(0)	yes
1			CCC	44.1	22120351	C_X hydrocarbon unreactive aliphatic w/o X	Daphnid	Chronic	1.3	[0.067, 24]	1.81	Estimated	out	[2.58, 5.74]	in	0.58	0.32	0.44	11(2)	no
1			CCC	44.1	22899851	narcotic group Daphnid Chronic	Daphnid	Chronic	0.99	[0.082, 12]	1.81	Estimated	in	[-1.20, 5.88]	in	0.7	0.68	0.54	73(14)	yes
1			CCC	44.1	32100151	C_X hydrocarbon Alga Chronic	Alga	Chronic	2.8	[0.14, 54]	1.81	Estimated	in	[1.61, 5.52]	in	0.38	0.32	0.61	55(25)	no
1			CCC	44.1	32120351	C_X hydrocarbon unreactive aliphatic w/o X	Alga	Chronic	82	[0.26, 26000]	1.81	Estimated	out	[2.58, 4.20]	in	0.51	0.24	0.62	9(9)	no
1			CCC	44.1	32899851	narcotic group Alga Chronic	Alga	Chronic	6.2	[0.23, 160]	1.81	Estimated	in	[0.69, 5.81]	in	0.57	0.53	0.69	57(27)	no
2			CCN	45.08	12200841	amine primary unreactive NH2=1 aliphatic	Fish	Acute	240	[17, 3500]	-0.15	Estimated	in	[-1.61, 5.25]	in	0.84	0.81	0.54	26(2)	yes
2			CCN	45.08	22200841	amine primary unreactive NH2=1 aliphatic	Daphnid	Acute	32	[0.87, 1200]	-0.15	Estimated	in	[-1.61, 4.76]	in	0.78	0.38	0.55	7(1)	no
2			CCN	45.08	32200841	amine primary unreactive NH2=1 aliphatic alga	Alga	Acute	5.6	[0.0076, 4100]	-0.15	Estimated	in	[-1.61, 4.76]	in	0.43	-0.61	1.02	7(0)	no
2			CCN	45.08	12500251	CNO_X unreactive Fish Chronic, w/ N,O	Fish	Chronic	0.25	[0.012, 5.4]	-0.15	Estimated	in	[-1.61, 5.99]	in	0.62	0.54	0.57	19(2)	no
2			CCN	45.08	22200851	amine primary unreactive NH2=1 aliphatic	Daphnid	Chronic	0.69	[0.039, 12]	-0.15	Estimated	in	[-1.61, 1.63]	in	0.23	-2.22	0.26	4(0)	no
2			CCN	45.08	32200851	amine primary unreactive NH2=1 aliphatic alga	Alga	Chronic	0.86	[0.0039, 190]	-0.15	Estimated	in	[-1.61, 4.76]	in	0.6	-0.08	0.84	7(0)	no
3			CCO	46.07	10600041	CO_X primary alcohol	Fish	Acute	3700	[390, 34000]	-0.14	Estimated	in	[-1.75, 5.26]	in	0.92	0.9	0.44	22(15)	yes
3			CCO	46.07	12102041	CO_X alcohol unreactive w/o EO Fish	Fish	Acute	6200	[690, 54000]	-0.14	Estimated	in	[-0.63, 5.81]	in	0.89	0.88	0.45	46(13)	yes
3			CCO	46.07	12899941	narcotic group Fish Acute	Fish	Acute	3600	[480, 27000]	-0.14	Estimated	in	[-0.63, 5.88]	in	0.87	0.87	0.43	151(34)	yes
3			CCO	46.07	20600041	CO_X primary alcohol	Daphnid	Acute	1300	[150, 12000]	-0.14	Estimated	out(p)	[2.31, 5.26]	in	0.95	0.76	0.17	6(17)	yes
3			CCO	46.07	22102241	CO_X alcohol unreactive w/o EO Daphnid	Daphnid	Acute	750	[25, 22000]	-0.14	Estimated	out(p)	[0.78, 5.81]	in	0.78	0.72	0.55	14(13)	yes
3			CCO	46.07	22899941	narcotic group Daphnid Acute	Daphnid	Acute	240	[28, 2100]	-0.14	Estimated	out(p)	[1.08, 5.88]	in	0.71	0.7	0.43	81(24)	yes
3			CCO	46.07	30600041	CO_X primary alcohol	Alga	Acute	28000	[260, 2.9e+6]	-0.14	Estimated	out(p)	[2.31, 5.26]	in	0.91	0.79	0.36	6(16)	yes
3			CCO	46.07	32102041	CO_X alcohol unreactive w/o halogen, acid, EO	Alga	Acute	12000	[380, 360000]	-0.14	Estimated	out(p)	[1.08, 5.26]	in	0.95	0.9	0.35	6(14)	yes
3			CCO	46.07	32899941	narcotic group Alga Acute	Alga	Acute	4800	[420, 54000]	-0.14	Estimated	out(p)	[1.08, 5.26]	in	0.75	0.73	0.44	50(47)	yes

Figure 7-10. tsv file of prediction results of multiple chemicals

※ Structural formula images are not included.

※ In “criteria” column, “yes” is shown for QSAR classes satisfying statistical criteria ($R^2 \geqq 0.7$ and $Q^2 \geqq 0.5$ and $n \geqq 5$) and “no” for others.

※ For chemicals with error, “<<Error>>” is written in column “QSAR ID” and content of error is not shown. Please check the error contents on web browser.

If you find an issue with a particular SMILE, please contact the KATE Contact Desk (KATE@nies.go.jp).

8. Printing the Prediction Results

In the prediction results screen, click the button Create Print Format to format the prediction results for printing. The resulting screen will contain the prediction results for all the QSAR classes that were selected on the main results screen (Figure 8).

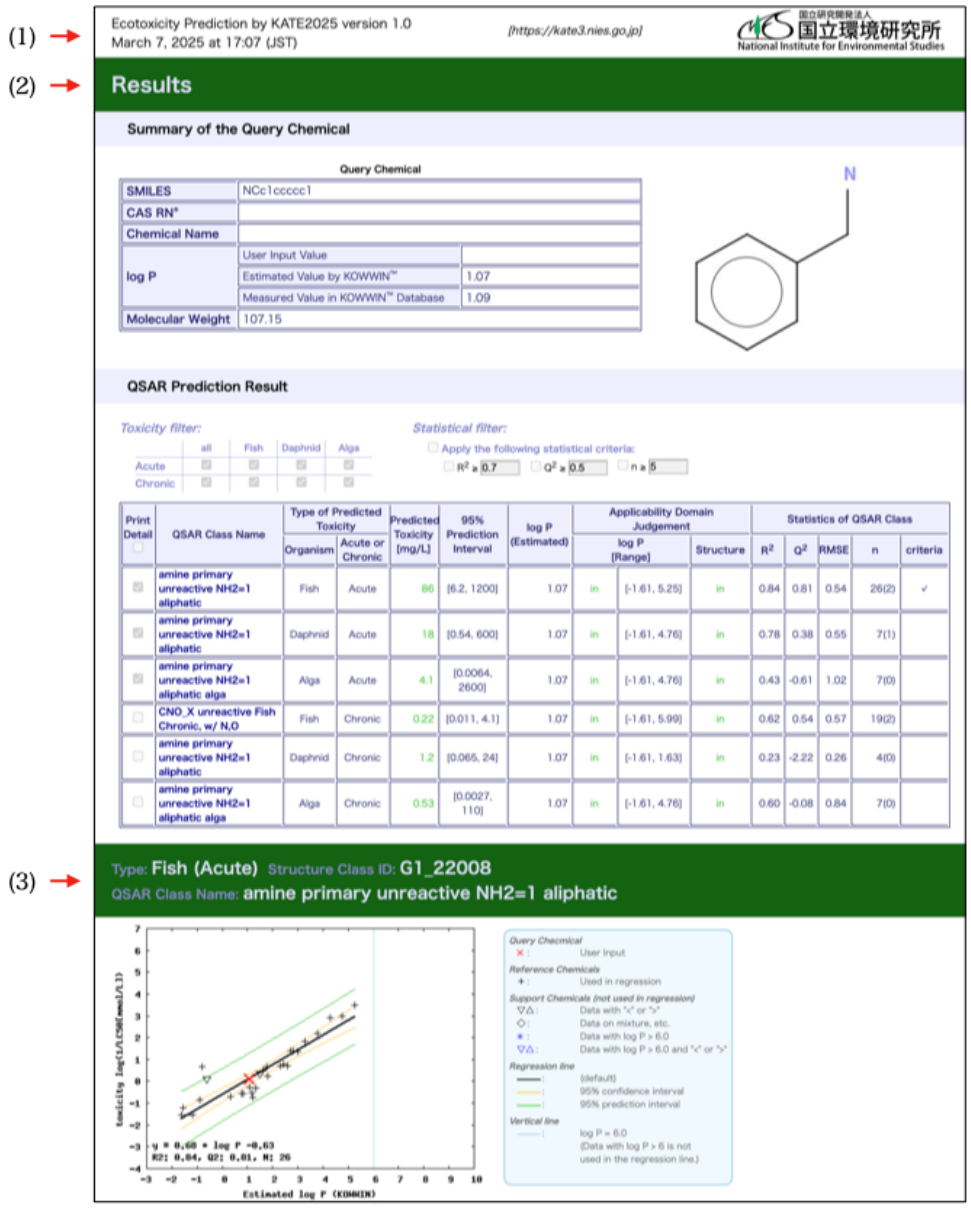


Figure 8. Print Format screen

(1) Title, Date and time when the prediction results were output (Japan Standard Time)

(2) QSAR results (Checkboxes are unchangeable)

(3) Information on individual QSAR classes such as plot of log P vs. toxicity, regression equation, chemicals etc. (Checked classes on the results page are shown)