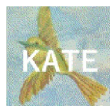


生態毒性予測システム「KATE2020」インターネット版 操作マニュアル(2023年3月30日版)



※ KATE2020は、化学物質の生態毒性に関する

- ・ 魚類急性毒性試験における半数致死濃度 (LC50)
- ・ ミジンコ急性遊泳阻害試験における半数影響濃度 (EC50)
- ・ 藻類生長阻害試験における半数影響濃度 (EC50)
- ・ 魚類初期生活段階毒性試験における無影響濃度 (NOEC)
- ・ ミジンコ繁殖試験における無影響濃度 (NOEC)
- ・ 藻類生長阻害試験における無影響濃度 (NOEC)

を予測するシステムです。

※ 本システムで得られた予測結果は、「化学物質の審査及び製造等の規制に関する法律」に基づく届出に必要な生態毒性試験結果として利用することはできません。化学物質の生態毒性影響の程度についての参考としてご利用ください。

ご質問等がございましたら、下記までお問い合わせ下さい

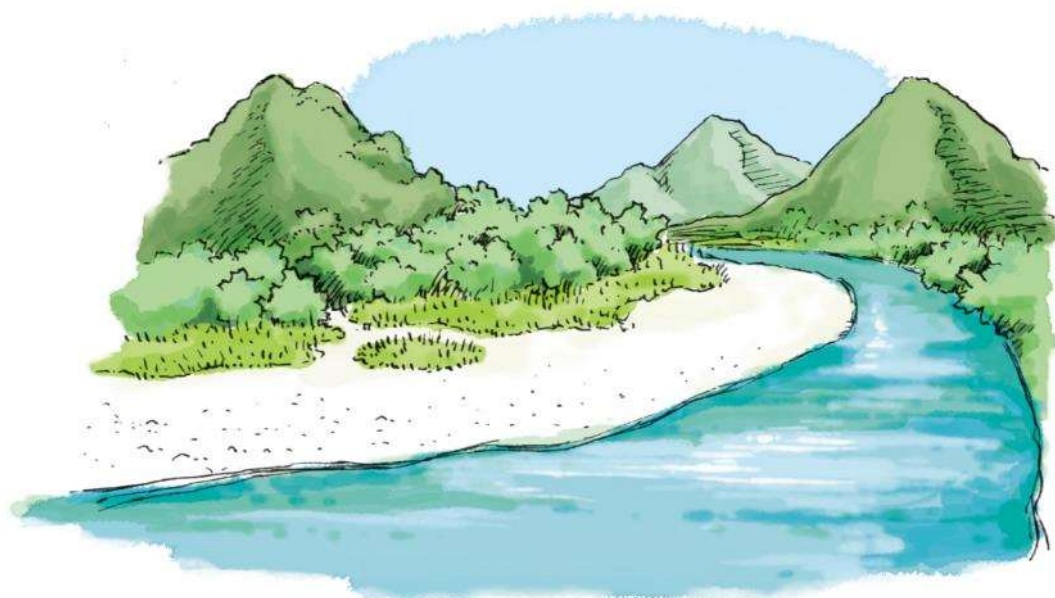
国立研究開発法人 国立環境研究所 環境リスク・健康領域 KATE担当 kate@nies.go.jp

Copyright(C) 2019-2023 Ministry of the Environment, Government of Japan.

All Rights Reserved

マニュアル改訂履歴

バージョン	発行日	改訂履歴
第0.1版	2018年3月29日	KATE2017 on NET 6版向けマニュアル
第0.8版	2019年1月30日	KATE2017 on NET 正式版向け暫定版
第0.9版	2019年3月29日	KATE2017 on NET 正式版向け暫定版
第1.0版	2019年5月24日	KATE2017 on NET 正式版対応マニュアル
第1.0.1版	2019年6月4日	説明文を一部修正
第1.0.2版	2019年7月30日	JSME Editorに関する記述を一部変更
	2020年2月3日	KATE2020 (version 1.0) 公開
第2.0版	2020年3月19日	KATE2020 (version 1.0) 向け対応マニュアル
	2021年1月28日	KATE2020 (version 2.0) 公開
第3.0版	2021年1月28日	KATE2020 (version 2.0) 向け対応マニュアル
第4.0版	2022年3月30日	KATE2020 (version 3.0) 向け対応マニュアル
	2022年3月30日	KATE2020 (version 3.0) 公開
第5.0版	2023年3月30日	KATE2020 (version 4.0) 向け対応マニュアル
	2023年3月30日	KATE2020 (version 4.0) 公開



目次

略語一覧.....	5
1. はじめに.....	7
(1) 生態毒性予測システム「KATE」: KAshinhou Tool for Ecotoxicity とは.....	7
(2) KATE2020 版とは.....	7
(3) KATE2020 version 3.0 から version 4.0 への主な変更点.....	7
(4) KATE2020 version 2.0 から version 3.0 への主な変更点.....	7
(5) KATE2020 version 1.1 から version 2.0 への主な変更点.....	8
(6) KATE2020 version 1.0 から KATE2020 version 1.1 への主な変更点.....	8
(7) KATE2017 on NET から KATE2020 version 1.0 への主な変更点.....	8
(8) KATE2017 on NET までの開発経緯.....	8
(9) サポートケミカルについて.....	9
(10) log P について.....	9
(11) 免責事項.....	9
(12) 謝辞.....	10
(13) 参考文献.....	10
2. KATE2020 の概要.....	11
(1) QSAR 予測手順の概要.....	11
(2) QSAR 式、毒性予測値の決定、及び適用領域判定に関する詳細.....	14
3. ログイン方法.....	18
4. 化学物質情報の入力 (Input 画面).....	19
(1) 予測できない化学物質について.....	20
(2) 入力方法 1 : SMILES の直接入力.....	20
(3) 入力方法 2 : 描画した構造式からの SMILES への変換.....	20
(4) 入力方法 3 : CAS 番号や物質名からの SMILES への変換.....	21
(5) 入力 log P の利用.....	23
(6) 入力 CAS 番号および入力物質名の利用.....	23
(7) KOWWIN™ 計算のスキップ.....	23
5. QSAR 予測結果の表示 (Results 画面).....	24
(1) 予測対象物質の基本情報.....	25
(2) QSAR 予測結果.....	26
6. QSAR クラス情報の詳細表示 (Verify QSAR 画面).....	29
(1) QSAR クラスの基本情報.....	30

(2) グラフ関連部分.....	30
(3) 予測対象物質情報.....	33
(4) 回帰式情報.....	33
(5) 最後にクリックされた物質.....	34
(6) 構造式一覧.....	35
(7) 物質データ.....	36
(8) 構造クラス定義.....	37
(9) 予測対象物質の部分構造一覧.....	38
7. 複数の化学物質の予測.....	40
(1) 入力ファイル「SMILES list」について.....	40
(2) 予測方法手順.....	41
(3) 動作について.....	42
8. 一括印刷フォーマット表示 (Print Format 画面)	43



略語一覧

EC50: 50% Effective Concentration (半数影響濃度)

試験水に溶解した化学物質などによって、半数（50%）の試験生物に対して影響を与えると考えられる濃度

KATE: KAshinho Tool for Ecotoxicity

国立環境研究所 環境リスク・健康研究センターにおいて研究・開発された生態毒性QSARシステム」の通称。「ケイト」と読む。

KOWWIN™:

US EPAなどが開発しているEPI Suite™ (Estimation Programs Interface : 化学物質を迅速にスクリーニングするためのアプリケーションなどで使用されることを目的としたツール) に含まれる化学物質のlog P推定プログラム。

LC50: 50% Lethal Concentration (半数致死濃度)

試験水に溶解した化学物質などによって、半数（50%）の試験生物を死亡させる濃度。

log P: The logarithm of the octanol/water partition coefficient

(オクタノール/水分配係数)

ある化学物質について、1-オクタノールと水の2つの溶媒中の平衡状態における濃度比を常用対数で表したもの。化学物質の疎水性を表す指標とされている。

<http://www.eic.or.jp/ecoterm/?act=view&serial=295>

(EICネットの用語解説。2023年03月01日アクセス)

NOEC: No Observed Effect Concentration (無影響濃度)

対照区と比較して統計的に有意な（有害）影響が認められなかった最高濃度であり、LOEC（最小影響濃度）のすぐ下の濃度区である。

<http://www.env.go.jp/chemi/report/ierac18/1-ref2.pdf>

<http://www.env.go.jp/chemi/report/ierac18/index.html>

(環境省 化学物質の環境リスク評価 第18巻 第1編 参考2 用語集等より。2023年03月01日アクセス)

(Q)SAR: (Quantitative) Structure-Activity Relationships ((定量的)構造活性相関)

化学物質の構造上の特徴又は物理化学定数と生物学的活性（毒性等）の相関関係を構造活性相関（SAR: Structure-Activity Relationship）といい、定量的なものを定量的構造活性相関（QSAR: Quantitative Structure-Activity Relationship）という。両者を併せて(Q)SARと記載することもある。構造活性相関は、例えば、特定の官能基の有無から物質の有害性の多寡を推測することを指し、構造を手掛かりに毒性等を定量的に算出する仕組みをいわゆるQSARモデルと呼ぶ。

<http://www.env.go.jp/chemi/report/ierac18/1-ref2.pdf>

(環境省 化学物質の環境リスク評価 第18巻 第1編 参考2 用語集等より。2023年03月01日アクセス)



SMARTS: SMiles ARbitrary Target Specification

SMILESを拡張した部分構造を表現するための識別子。

http://www.daylight.com/dayhtml_tutorials/languages/smarts/

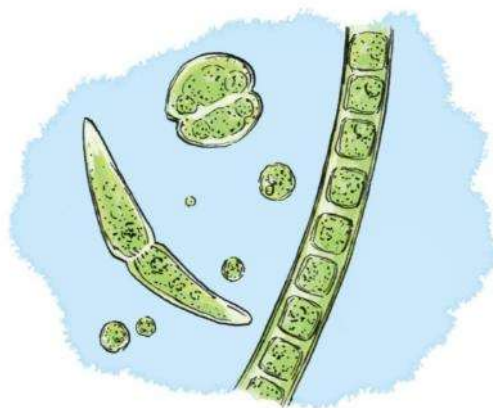
(Daylight社詳細解説。2023年03月01日アクセス)

SMILES: Simplified Molecular Input Line Entry System

化合物の分子構造等を印刷可能な文字で線形表記した識別子。

<http://www.daylight.com/smiles/index.html>

(Daylight社詳細解説。2023年03月01日アクセス)

US EPA: United States Environmental Protection Agency (米国環境保護庁)

1. はじめに

(1) 生態毒性予測システム「KATE」: KAshinhou Tool for Ecotoxicityとは

国立研究開発法人 国立環境研究所 環境リスク・健康領域において、環境省の請負業務により研究・開発された生態毒性 QSAR システムです¹⁾。

現在の最新バージョンである KATE2020 (version 4.0)では、化学物質の構造から以下の毒性値の種類²⁾に対する予測結果が得られます。

- ・ 魚類急性毒性試験 (OECD TG 203) における半数致死濃度 (LC50)
- ・ 魚類初期生活段階毒性試験 (OECD TG 210) における無影響濃度 (NOEC)
- ・ ミジンコ急性遊泳阻害試験 (OECD TG 202) における半数影響濃度 (EC50)
- ・ ミジンコ繁殖試験 (OECD TG 211) における無影響濃度 (NOEC)
- ・ 藻類生長阻害試験 (OECD TG 201) における半数影響濃度 (EC50) 及び無影響濃度 (NOEC)

化学物質の CAS 番号検索や構造式エディタを用いた作図等を用いて SMILES 記法による入力を行い、log P を記述子とした QSAR 予測を行います。

KATE2020 の構築に当たっては、環境省が実施した生態毒性試験結果 (魚類急性毒性試験、ミジンコ急性遊泳阻害試験、魚類初期生活段階毒性試験、ミジンコ繁殖試験、藻類生長阻害試験)³⁾及び US EPA のファットヘッドミノール・データベースの魚類急性毒性試験結果⁴⁾をトレーニングセットデータとして用いています。今後、試験結果が追加された場合には QSAR 式の見直しを行う予定です。

(2) KATE2020版とは

2019年1月から正式版として公開していた KATE2017 on NET 版に対して幾つかの修正を行い、2020年2月に KATE2020 (version 1.0) として公開し、2023年3月には version 4.0 へバージョンアップを行いました。インターネット上のブラウザ画面で操作を行います。

(<https://kate.nies.go.jp/onnet2020.html>)

(3) KATE2020 version 3.0からversion 4.0への主な変更点

ユーザーインタフェースの改良

- ① ログイン方法を変更し、ユーザ ID とパスワードを使用せずに KATE2020 にログインできるように変更
- ② 予測結果画面や QSAR クラス詳細画面等の修正

QSAR モデルの更新

- ① 魚類急性に関する生態影響試験結果の見直しを行い、毒性値の再計算が必要なものは最新のテストガイドラインに基づいて再計算を実施し、トレーニングセットとしての使用が不適切なものはサポートケミカルに変更
- ② 2021年度に生態影響試験が実施されたチオール2物質 (魚類急性、甲殻類急性) とイミド1物質 (甲殻類急性) をトレーニングセットデータに追加
- ③ ①②の結果を受けて QSAR クラスの回帰式の再計算を実施

QSAR クラス名の更新

- ① すべての QSAR クラスの名称を精査し、一部の QSAR クラスについて名称を変更

(4) KATE2020 version 2.0からversion 3.0への主な変更点

QSAR モデルの更新

- ① 甲殻類急性に関する生態影響試験結果の見直しを行い、毒性値の再計算が必要なものは最新のテストガイドラインに基づいて再計算を実施し、トレーニングセットとしての使用が不適切なものはサポートケミカルに変更



- ② ①の結果を受けて QSAR クラスの回帰式の再計算を実施

QSAR クラス名の更新

- ② 統計値基準 ($R^2 \geq 0.7$ and $Q^2 \geq 0.5$ and $n \geq 5$) を満たすすべての QSAR クラスの名称を精査し、一部の QSAR クラスについて名称を変更

(5) KATE2020 version 1.1からversion 2.0への主な変更点

QSAR モデルの更新

- ③ デフォルトで表示する QSAR クラスの基準を「 $R^2 \geq 0.7$ and $Q^2 \geq 0.6$ and $n \geq 5$ 」から「 $R^2 \geq 0.7$ and $Q^2 \geq 0.5$ and $n \geq 5$ 」に変更
- ④ 毒性値 1 つを修正 (転記ミスによる)。このことに付随して、藻類慢性の QSAR クラス「CNOS_X basic aromatic n unreactive」の QSAR 式を変更

表示・操作の改良

- ③ 部分構造に対する構造判定の追加
- ④ 印刷フォーマット表示の追加
- ⑤ 複数化学物質予測でエラーがあった場合に途中で止まる問題の修正

(6) KATE2020 version 1.0からKATE2020 version 1.1への主な変更点

- ① 予測毒性値が 10^6 以上もしくは 10^{-5} [mg/L]未満の場合は指数表記 (例: $2.3e-7$) で表示するように変更
- ② 予測毒性値を有効数字 2 桁で表示するように変更
- ③ 化学物質入力の際に、正しい情報を入れたときにもエラーになる場合がある問題の修正

(7) KATE2017 on NETからKATE2020 version 1.0への主な変更点

QSAR モデルの更新

- ① log P の推定方法を ClogP から KOWWIN™ に変更、および変更に伴う一部の QSAR モデルの修正
- ② 予測対象物質が使用する log P 値の優先順位を 1 ユーザ入力値、2 KOWWIN™ 推定値に変更 (以前は 1 ユーザ入力値、2 KOWWIN™ 実測値、3 KOWWIN™ 推定値であった)
- ③ log P > 6.0 のデータは全て QSAR モデルから除外
- ④ トレーニングセットと QSAR クラスの追加・削除
- ⑤ QSAR 式構築に使用されない物質 (サポートケミカル) を参考情報として表示するように変更

表示・操作の改良

- ① QSAR クラスに含まれるトレーニングセットおよびサポートケミカルのデータ一覧表の追加
- ② QSAR クラスに対応する構造クラスの定義一覧の追加
- ③ QSAR 式詳細画面の構造式一覧にソート機能の追加
- ④ QSAR 式詳細画面中グラフにおいて、トレーニングセットの一部を除外して計算される回帰直線と予測区間・信頼区間を連動させた
- ⑤ KOWWIN™ 計算をスキップする機能の追加

(8) KATE2017 on NETまでの開発経緯

現在もホームページ上で公開している KATE2011 版については、2008 年 1 月に試用版、2009 年 3 月に KATE2009 をインターネットで公開しました⁵⁾。2011 年 3 月には、更新版 KATE2011 を公開し、トレーニングセットデータの追加、部分構造の分類ルールの修正、



構造判定の変更と皮膚感作性に関する部分構造の追加等を行いました。

その後、現在の KATE2020 版と同じ系列の新しい生態毒性予測システムの開発を開始し、2018年3月に KATE2017 on NET β 版、2019年1月に KATE2017 on NET 正式版 (version 1.0) を公開しました。KATE2011 から KATE2017 への変更では主に、部分構造検索方式の FITS (KATE2011 で使用される部分構造表記法および検索プログラム) から SMARTS 記法および CDK を利用した検索プログラムへの変更、藻類や慢性の毒性値予測の追加、不等号付き毒性値データの追加、構造クラスの導入、部分構造・QSAR クラスの大幅な変更、log P 計算モジュールの ClogP から KOWWIN™ への変更、表示・操作の改良、英語化等を行いました。

(9) サポートケミカルについて

KATE2020 では、以下のデータは、QSAR モデル構築には使用せず、サポートケミカル (参考情報) として表示のみを行うようにしました。

① log P 推定値>6.0 の化学物質データ ②不等号付きデータ ③外れ値

なお、②の不等号付きデータは、log P の適用領域内にある場合、構造判定に使用しています。

(10) log Pについて

本システムは、化学物質の毒性を予測する際に使用する log P として、US EPA が著作権を有する log P 予測モデル KOWWIN™ を US EPA の許諾を得て使用しています⁶⁾。

利用者は下記に示す KOWWIN™ 使用許諾条件について遵守してください。

KOWWIN v1.69 (April 2015)

© 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.

(11) 免責事項

KATE の予測結果は十分な予測精度を保証できるものではありません。本システムは、化学物質の生態毒性影響の程度についての参考情報を得るためのツールの一つとしてご利用ください。環境省および国立環境研究所は KATE による毒性予測値を保証するものではなく、また、KATE による毒性予測値の使用により生じた損害については一切の責任を負いません。

また、現時点では KATE による毒性予測結果を「化学物質の審査及び製造等の規制に関する法律」(化審法) に基づく届出に必要な生態毒性試験結果に代替するものとして利用することはできません。

著作権、リンク等については、KATE ウェブサイト内のサイトポリシーをご覧ください。 (<https://kate.nies.go.jp/spolicy.html>)



(12) 謝辞

KATE2020 は下記のソフトウェアまたはライブラリからの結果を使用させていただいております。ここに記して謝意を表します。

- Open Babel
 - <http://openbabel.org/wiki/Category:Installation>
 - 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™ (included in EPI Suite™)
 - <https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface>
- (上記 URL は全て、2023 年 03 月 01 日アクセス)

(13) 参考文献

- 1) <https://kate.nies.go.jp> (2023年03月01日アクセス)
- 2) <http://www.env.go.jp/chemi/sesaku/01.html> (2023年03月01日アクセス)
- 3) <http://www.env.go.jp/chemi/sesaku/seitai.html> (2023年03月01日アクセス)
- 4) https://archive.epa.gov/med/med_archive_03/web/html/fathead_minnow.html (2023年03月01日アクセス)
- 5) A Furuhashi, 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> (2023年03月01日アクセス)



2. KATE2020の概要

(1) QSAR予測手順の概要

KATE2020は、化学物質の毒性予測を下記の流れで行います（図2-1）。

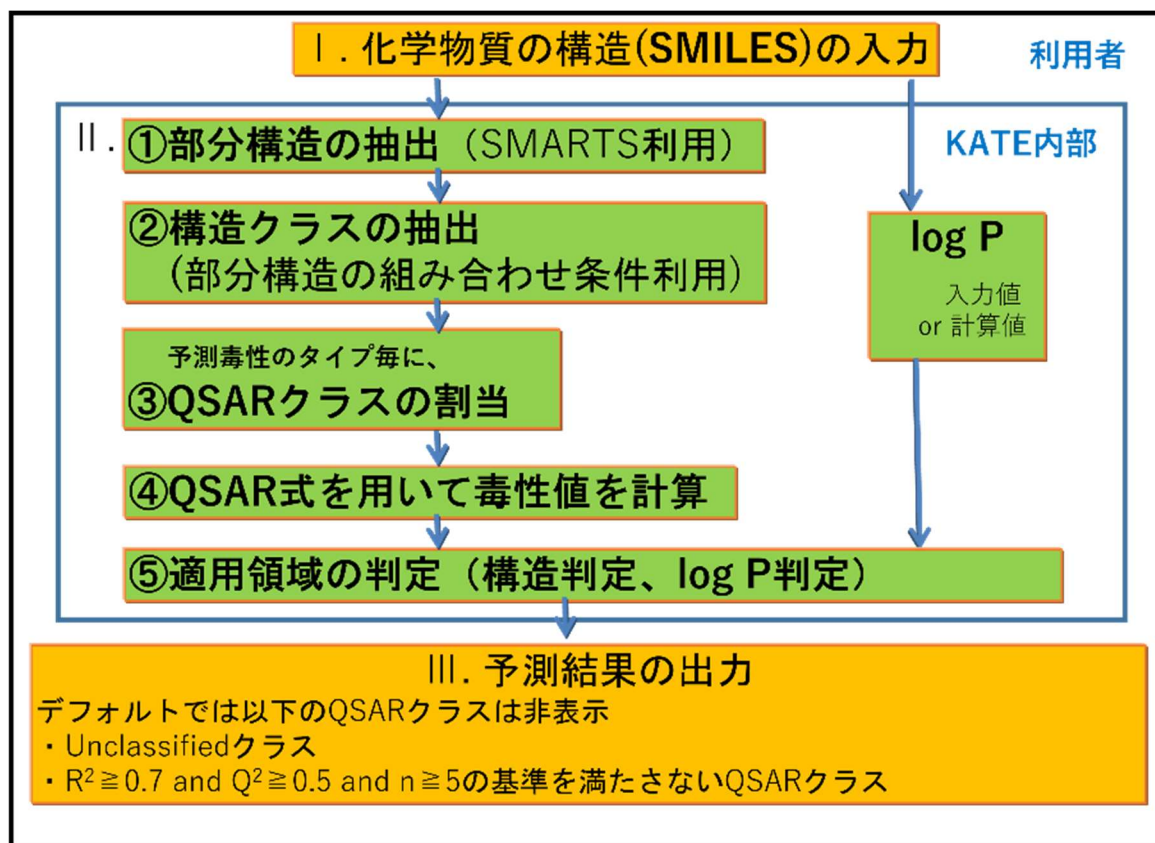


図 2-1 KATE2020 の QSAR 予測フロー

I. 利用者による化学物質の構造（SMILES記法による）の入力

II. QSAR式、毒性予測値の決定、及び適用領域判定に関する処理

- ① 予測対象物質（入力された化学物質）の部分構造を抽出
- ② 複数の部分構造の組み合わせによる構造クラス*1を抽出
- ③ 予測毒性のタイプごとに、構造クラスに対応するQSARクラス*2を割当（複数のクラスに割当てられることもあります。）
- ④ 割り当てられたQSARクラスごとに、QSAR式*3を用いて毒性値を計算
- ⑤ 適用領域の判定（構造判定とlog P判定）

*1 各部分構造の個数条件のAND/ORによる組み合わせにより定義した分類（37ページの「構造クラス定義」参照）

*2 各予測毒性のタイプでの物質の構造に基づいて定義した分類

*3 QSARクラスに含まれるトレーニングセットで形成されたモデル。ここではlog Pを記述子とする単回帰式



Ⅲ. 予測結果の出力

- ① 予測対象物質が何らかの予測毒性のタイプでどのQSARクラスにも分類されなかった場合、Unclassifiedクラス*4に割り当てられます。
- ② デフォルトではUnclassifiedクラス、及び統計値 $R^2 \geq 0.7$ and $Q^2 \geq 0.5$ and $n \geq 5$ を満たさないQSARクラス*5は非表示にしています。

*4 どのQSARクラスにも分類されなかった場合に割り当てられるQSARクラス

*5 R^2 , Q^2 , n はそれぞれ決定係数、内部バリデーションの指標（Leave-one-out法）、トレーニングセット数であり、各QSARクラスに対してあらかじめ計算されています。



具体的な例として、化学物質 1-pyridin-3-ylethanone を予測したときの予測フローを図 2-2 に示します。

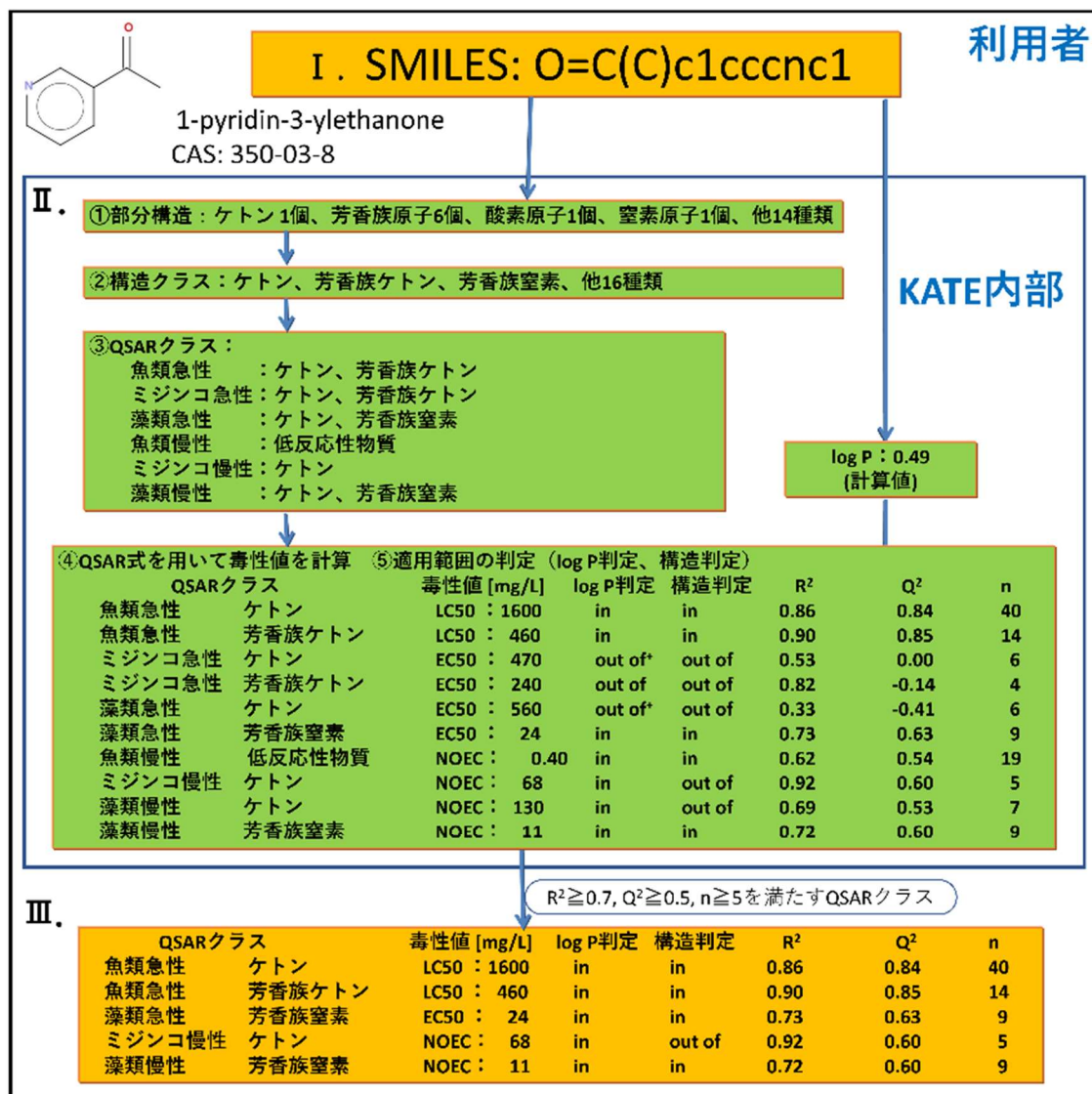


図 2-2 KATE2020 の QSAR 予測フロー (1-pyridin-3-ylethanone の予測例)

※ 図2-2中のケトン、芳香族ケトン、芳香族窒素、低反応性物質のKATE2020における実際の名称は以下になっています。

- ケトン : COS_X ketone unreactive
- 芳香族ケトン : COS_X ketone unreactive aromatic
- 芳香族窒素 : CNOS_X basic aromatic n unreactive
- 低反応性物質 : CNO_X unreactive (Fish chronic), excl. (CnosX w/o n+)



(2) QSAR式、毒性予測値の決定、及び適用領域判定に関する詳細

KATE2020 では、利用者が入力した化学物質の構造（SMILES 記法で作成）に基づき、以下の処理を行い QSAR 式及び毒性値を予測します。以下に、簡単な説明を行います。

① 部分構造の抽出

KATEが定義した部分構造一覧（SMARTS記法で作成）をもとに、予測対象物質に含まれる各部分構造の個数を計算します。SMARTSを利用した部分構造個数計算にはCDKライブラリを利用しています。

② 構造クラスの抽出

構造クラス定義一覧をもとに、予測対象物質の構造に合致する全ての構造クラスを抽出します。

③ QSARクラスの割当

各予測毒性のタイプについて、QSARクラス定義一覧をもとに予測対象物質の構造クラスに対応するQSARクラスを割り当てます。KATE2020では、同じ予測毒性タイプに対して複数のQSARクラスが割り当てられる場合があります。また、どのQSARクラスにも分類されなかった場合は、Unclassified クラスに割り当てられます。

④ QSAR式による毒性予測値の計算

各QSARクラスにはQSAR式が割り当てられており、予測対象物質のlog Pの値をQSAR式に代入することにより、 $\log(1/\text{毒性値}[\text{mmol/L}])$ を計算します。次に、予測対象物質の分子量を用いて毒性予測値 $[\text{mg/L}]$ に単位変換します。

⑤ 適用領域の判定

KATE2020では、予測対象物質の毒性予測値が、予測結果として適用できる範囲内にあるかどうかを判定します。A) 構造による判定とB) log Pによる判定の2つを行い、両方とも適用領域内の場合に、KATE2020での毒性予測値が予測結果として適用可能と判定されます。

A) 構造判定

KATE2020では、「構造判定用部分構造」*1の比較により、予測対象物質の構造が当該QSARクラスの適用領域内であるかどうかについて判定します（図2-3）。判定には下記の3つの場合があり、「in」又は「in (conditionally)」の場合に当該QSARクラスを、構造に関して適用領域内と判定しています。

in : 適用領域内

予測対象物質に含まれる「構造判定用部分構造」の全てが、当該 QSAR クラスの「構造判定用部分構造リスト*2」に含まれる場合（図 2-3 におけるピンクとオレンジの範囲）、または予測対象物質が持つ部分構造に「構造判定用部分構造」が1つも含まれていなかった場合。

in (conditionally) : 条件付き適用領域内

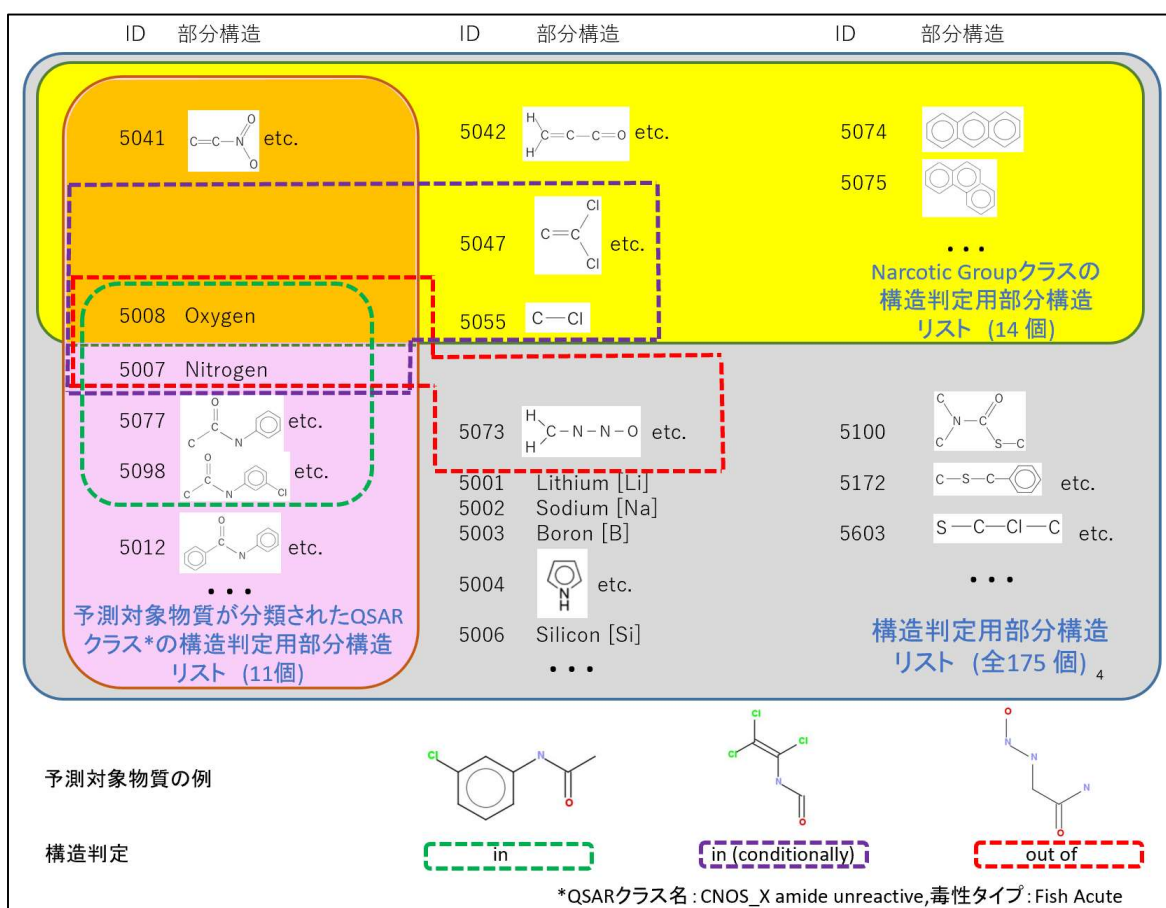
「in」の条件には合致しないが、予測対象物質の「構造判定用部分構造」全てが、当該 QSAR クラスの「構造判定用部分構造リスト」、あるいは Narcotic Group*3クラスの「構造判定用部分構造リスト」に含まれる場合（図 2-3 におけるピンクとオレンジとイエローの部分）。



out of : 適用領域外

「in」と「in (conditionally)」の何れの条件にも合致しない場合。すなわち、予測対象物質の「構造判定用部分構造」に、当該 QSAR クラスの「構造判定用部分構造リスト」と Narcotic Group クラスの「構造判定用部分構造リスト」の何れにも含まれない部分構造がある場合。(図 2-3 においてグレー部分の構造が含まれる場合)

- *1 KATE が定義した部分構造一覧の中で、構造判定にも使用されるもの。KATE2020 では 175 個の部分構造が該当する(詳細は KATE2020 技術文書参照)。
- *2 当該 QSAR クラスに含まれる、トレーニングセットおよび log P 判定が "in" (適用領域内) である不等号付きデータの物質が持つ、部分構造のリスト。
- *3 特異的な生理活性作用に基づかないベースライン毒性(麻酔作用)。KATE2020 では、脂肪族炭化水素、スルホキド、脂肪族・芳香族エーテル、脂肪族・芳香族ケトン、アルコールといった単純な麻酔作用のみで毒性が説明できると考えられる QSAR クラスが予測毒性のタイプ毎に用意されており、これらをまとめた QSAR クラスを Narcotic Group として各予測毒性のタイプで再定義しています。



※ 緑枠は「予測対象物質の例」の一番左の物質で予測して緑枠の部分構造が抽出されることにより構造判定が緑枠の「in」となることを示しています。紫枠と赤枠も同様です。

※ “etc.”が付くものは、同じIDでも複数の部分構造があり、1つの例のみ示しています。

図2-3 構造判定の例



B) log P判定

KATE2020では、予測対象物質のlog P値が当該QSARクラスのトレーニングセットのlog Pの最小値と最大値の間にあるかどうかで適用領域内にあるかどうかを判定します。なお、KATE2020では、高疎水性で予測精度が低い $\log P > 6$ の物質は全て適用領域外としています（KATE2020版における変更点）。

in : 適用領域内（図 2-4）

out of : 適用領域外。ただし、下記の out of+になる場合は除く（図 2-5）。

out of+ : 適用領域外。ただし、予測対象物質の log P 値は当該 QSAR クラスの参考情報であるトレーニングセットとサポートケミカルを併せた全物質の log P の最小値と最大値の内側に存在します（図 2-6）。

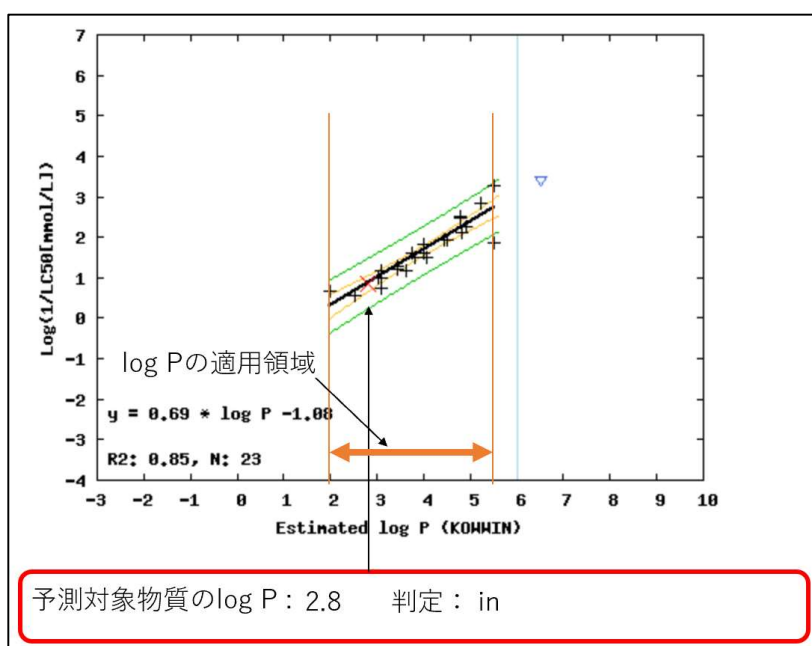


図 2-4 log P 判定の例 (in)

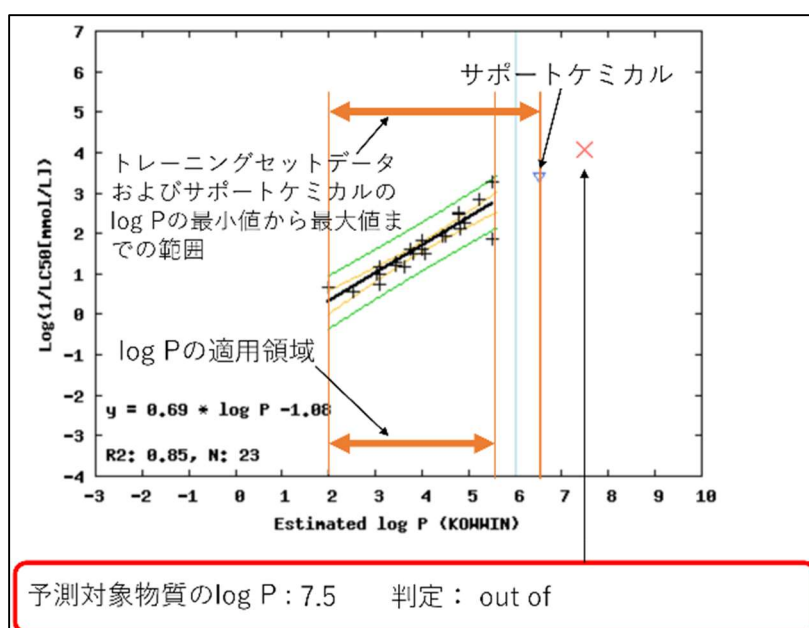


図 2-5 log P 判定の例 (out of)



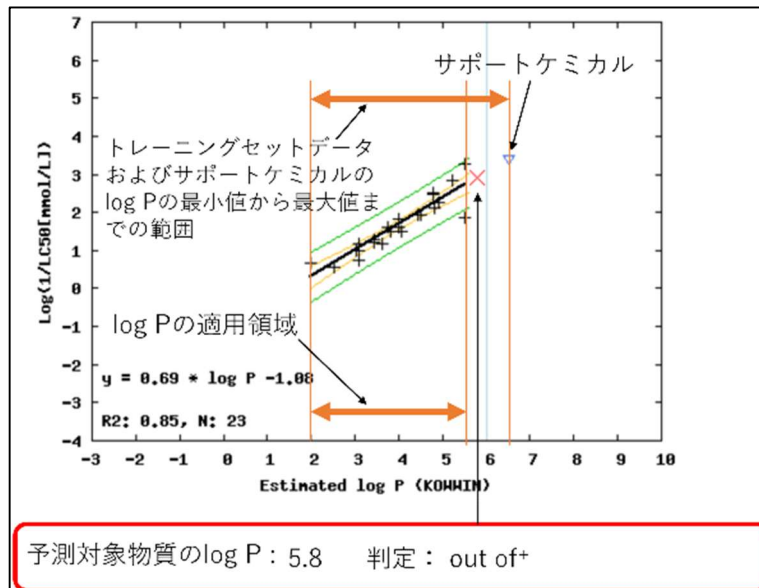


図 2-6 log P 判定の例 (out of+)



3. ログイン方法

2023年3月にリリースされたKATE2020 ver. 4.0より、ログインするためのユーザIDとパスワードは不要になりました。ユーザは図3に示すKATE2020のログイン画面 (<https://kate2.nies.go.jp/nies/index.php>) にアクセスし、①免責事項に同意するチェックボックスにチェックを入れ、②”Start the session” ボタンをクリックするとログインすることができます。SMILES等のユーザが入力したデータおよび出力結果（予測毒性値やQSARクラス等）はKATE2020サーバではなくセッションにのみ保持され、セッションが切れると自動的にそれらは削除されます。セッションはウェブブラウザを閉じたりKATE2020の操作を一時間停止したりすると自動的に切れます。また、予測を行う度に過去の情報は上書きされ、削除されます。

KAshinhou Tool for Ecotoxicity
KATE2020 version 4.0

the Terms of the Agreement

KOWWIN v1.69 (April 2015)

© 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.

agree to and accept the terms of the agreement above.

Start the session

図3 ログイン画面



4. 化学物質情報の入力（Input画面）

KATE2020にログインすると以下の画面（Input画面）が表示されます（図4-1）。

i. SMILES which does not contain carbon or nitrogen atoms.
ii. SMILES which includes elements other than H, C, N, O, F, Si, P, S, Cl, As, Br, Sn, and I.
iii. SMILES which includes ions other than ammonium [N+] or [n+].
iv. SMILES which includes ".", i.e. SMILES which expresses a mixture.

The strings such as [Na], [K], [Li], [Na+], [K+] and [Li+] in SMILES should be replaced by the protonated forms. For example, SMILES "c1ccccc1O".
Glossary is here.
Thanks to Chemical Identifier Resolver Service provided by NCI/CADD Group.

Output from <https://cactus.nci.nih.gov> may be shown here.

Input SMILES of your chemical

CAS RN: Name:

SMILES (* Required):

SMILES can be generated by using molecular editor [JSME Editor](#).

Optional:
log P:
 Skip KOWWIN Calculation
* When any error occurs in log P calculation by KOWWIN, you can skip KOWWIN Calculation.

Prediction of Multiple Chemicals

SMILES List:

filename: *Not Selected*

Caution: KATE2020 can accept up to 100 chemicals at present.

Maintained by: Health and Environmental Risk Division, National Institute for Environmental Studies
Copyright(C) 2019-2023 Ministry of the Environment, Government of Japan, All Rights Reserved

図 4-1 化学物質情報の入力画面

KATE2020における予測は、SMILESに基づいて行います。化学物質のSMILESを入力するには以下の方法があります。

入力方法 1 : SMILES を直接入力する方法

入力方法 2 : 描画した構造式を SMILES に変換して入力する方法

入力方法 3 : CAS 番号や物質名から SMILES を取得して入力する方法



(1) 予測できない化学物質について



KATEにおいて予測対象となる化学物質は、基本的には有機化合物です（例外として一部の無機窒素化合物（ヒドラジン等）も含まれます）。
KATE2020では、以下 i)~v)に該当する化学物質は予測できません。

- i) C、Nの何れの元素も含まない SMILES
- ii) C, N, O, F, Si, P, S, Cl, As, Br, Sn, I 以外の元素を含む SMILES
- iii) イオンを含む SMILES（ただし、ammonium [N+] と [n+] は入力可能）
- iv) 混合物を表す（記号「.」を含む） SMILES
- v) [Na], [K], [Li], [Na+], [K+], [Li+]等を含む SMILES の場合、プロトン化した形式に置き換える必要があります。例えば、"c1ccccc1O[Na]"は"c1ccccc1O"に置き換える必要があります。

入力 SMILES が上記 i)~v)に該当する場合、エラーメッセージ画面が表示されます。

(2) 入力方法1：SMILESの直接入力

ログイン後、前ページ図4-1の画面が表示されるので、赤枠で囲んだ入力ボックスに、毒性予測を行いたい化学物質のSMILESを入力します。

SMILESを入力したら、予測（Predict）ボタンをクリックします（図4-2）。この操作を行うとQSAR予測結果画面に進みます。ここでは予測対象物質の例として、NCc1ccccc1を入力します。

The image shows a web interface for entering SMILES. It features a text input field containing the SMILES string "NCc1ccccc1". To the right of the input field is a button labeled "Predict". The "Predict" button is circled in red. Above the input field, the text "SMILES (* Required):" is displayed in red.

図 4-2 SMILES 入力ボックス横の予測（Predict）ボタン

(3) 入力方法2：描画した構造式からのSMILESへの変換

JSME Molecular Editor（構造式エディタ、以下、及びウェブ上ではJSME Editorと表記します）を利用して、毒性予測を行いたい化学物質の構造式を描画し、それをSMILESに変換することができます。

手順 1：図 4-3 の赤枠の「JSME Editor」のリンクをクリックします。

The image shows a web page with a text area at the top containing instructions about replacing ion strings in SMILES. Below this is a "Glossary is here." link and a "Thanks to Chemical Identifier Resolver Service provided by NCI/CADD Group." message. A box contains the text "Output from https://cactus.nci.nih.gov may be shown here." Below this is a section titled "Input SMILES of your chemical" with three buttons: "CAS RN to Name, SMILES", "Name to SMILES, CAS RN", and "SMILES to Name, CAS RN". There are input fields for "CAS RN:" and "Name:". At the bottom of this section is a "SMILES (* Required):" label, a text input field containing "NCc1ccccc1", and a "Predict" button. At the very bottom, a link "SMILES can be generated by using molecular editor JSME Editor." is circled in red.

図 4-3 「JSME Editor」のリンク



手順2: JSME Editor 画面 (図 4-4) が起動するので、構造式 (ここではフェノールの例を示します) を描画し、「Submit smiles to KATE」ボタンをクリックします。JSME Editor については、JSME Editor のウェブページを参照ください。(https://jsme-editor.github.io/ , 2023 年 03 月 01 日アクセス)

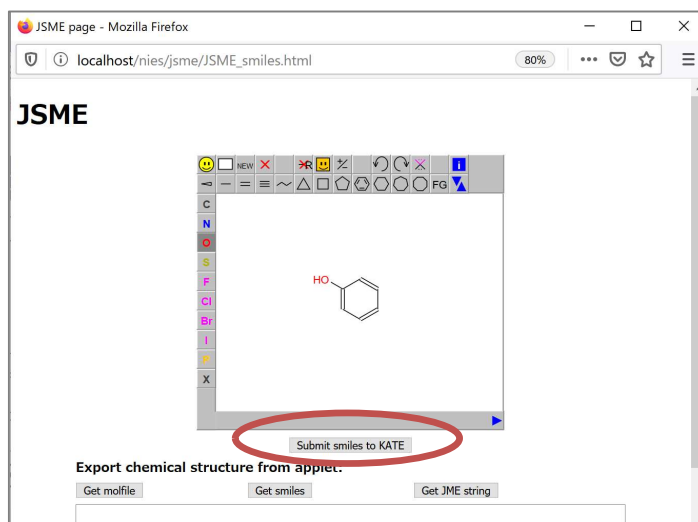


図 4-4 JSME Editor の画面

クリックすると、描画した構造式が SMILES に変換され、化学物質の入力画面の SMILES 入力ボックスに入力されます (図 4-5)。その後、予測 (Predict) ボタンをクリックすると予測結果画面に進みます。



図 4-5 構造式から SMILES に変換した結果

(4) 入力方法3: CAS番号や物質名からのSMILESへの変換

「CAS to SMILES, IUPAC Name」、「Name to SMILES, CAS」を使用して、CAS番号や物質名の入力から、SMILESへの変換を行うことができます。

なお、「CAS to SMILES, IUPAC Name」のボタンでは、SMILESとともにIUPAC名も出力され、「Name to SMILES, CAS」ボタンではSMILESに加えCAS番号も出力されます。



手順 1：毒性予測を行いたい化学物質の CAS 番号を CAS 入力ボックスに入力し、「CAS から SMILES と物質名への変換（CAS to SMILES, IUPAC Name）」ボタンをクリックします（図 4-6）。

Thanks to Chemical Identifier Resolver Service provided by NCI/CADD Group.

Output from <https://cactus.nci.nih.gov> may be shown here.

Input SMILES of your chemical

CAS RN to Name, SMILES **Name to SMILES, CAS RN** **SMILES to Name, CAS RN**

CAS RN: Name:

SMILES (* Required):

SMILES can be generated by using molecular editor JSME Editor.

図 4-6 CAS 番号入力ボックスと「CAS to SMILES, IUPAC Name」ボタン

クリックすると、CAS 番号が SMILES に変換され、SMILES 入力ボックスに入力されます（対応する構造式も出力され、変換された物質名（IUPAC 名）も Name ボックスに入力されます。図 4-7）。この後”Predict”ボタンをクリックすると予測結果画面に進みます。

Thanks to Chemical Identifier Resolver Service provided by NCI/CADD Group.

CAS: 100-46-9
IUPAC_NAME: phenylmethanamine
SMILES: NCC1ccccc1

Input SMILES of your chemical

CAS RN to Name, SMILES **Name to SMILES, CAS RN** **SMILES to Name, CAS RN**

CAS RN: Name:

SMILES (* Required):

SMILES can be generated by using molecular editor JSME Editor.

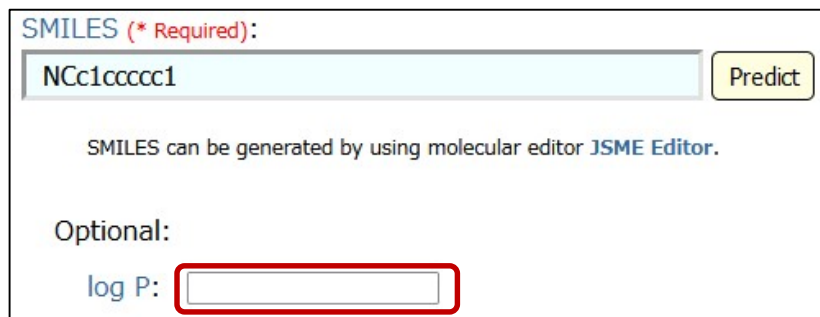
図 4-7 CAS 番号から SMILES と構造式に変換した結果

物質名の入力から SMILES への変換も上記と同様に「Name to SMILES, CAS」を利用して行うことができます。



(5) 入力log Pの利用

予測対象物質のlog P値として、入力値を指定することが出来ます（図4-8）。入力は任意です。入力したlog P値が毒性予測の際に優先的に利用されます。予測対象物質のlog P値が明らかな場合などにご利用ください。



SMILES (* Required):

NCc1ccccc1 Predict

SMILES can be generated by using molecular editor [JSME Editor](#).

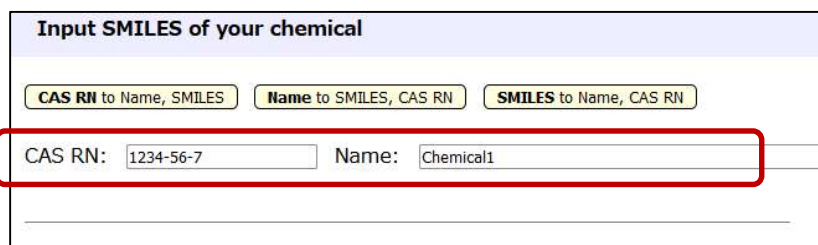
Optional:

log P:

図 4-8 log P の入力

(6) 入力CAS番号および入力物質名の利用

CAS番号や物質名も、ユーザが任意で指定することができます（図4-9）。入力は必須ではありません。入力した場合は、結果画面には入力した情報がそのまま表示されます。



Input SMILES of your chemical

CAS RN to Name, SMILES Name to SMILES, CAS RN SMILES to Name, CAS RN

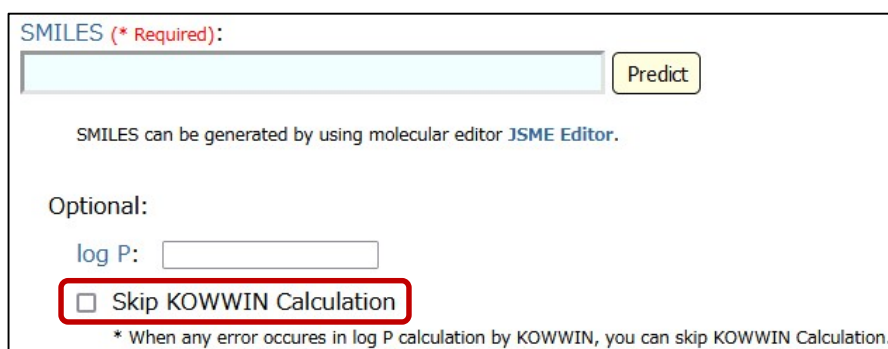
CAS RN: 1234-56-7 Name: Chemical1

図 4-9 CAS 番号と物質名の入力

(7) KOWWIN™計算のスキップ

KOWWIN™に対応していないSMILESが入力されることにより、KOWWIN™によるlog P計算中にエラーになることがあります。この場合には、“Skip KOWWIN Calculation”のチェックボックスにチェックをしてPredictボタンをクリックすることで、KOWWIN™によるlog P計算をスキップすることが出来ます（図4-10）。

log P計算をスキップした場合、予測毒性値を計算するにはlog Pのユーザ入力値が必要となります。log Pを入力しない場合でもQSARクラス分類は行われますが、予測毒性値は計算されません。スキップ後に、予測結果画面の中でlog P入力値を入れて予測毒性値を計算することも可能です。



SMILES (* Required):

Predict

SMILES can be generated by using molecular editor [JSME Editor](#).

Optional:

log P:

Skip KOWWIN Calculation

* When any error occurs in log P calculation by KOWWIN, you can skip KOWWIN Calculation.

図 4-10 KOWWIN™ 計算のスキップ





5. QSAR予測結果の表示 (Results画面)

予測 (Predict) ボタンをクリックした後、計算が終了すると、予測結果画面が表示されます (図5-1)。予測結果画面では、予測対象物質の基本的な情報 (CAS番号、物質名、SMILES、分子量、log P、構造式等)、予測対象物質に割り当てられたQSARクラス、予測対象物質の各QSARクラスで予測結果 (予測毒性値、95%予測区間、適用領域の判定)、各QSARクラスの統計値等を得ることができます。

Results

CAS RN®	100-46-9		
Chemical Name	phenylmethanamine		
SMILES	NCc1ccccc1		
Molecular Weight	107.15		
log P	User Input Value	<input type="text"/>	<input type="button" value="Re-calculate"/>
	Estimated Value by KOWWIN	1.07	
	Measured Value in KOWWIN Database	1.09	

Include(Acute): Fish Daphnid Alga
 Include(Chronic): Fish Daphnid Alga
 Exclude(): R² < 0.7 Q² < 0.5 n < 5

QSAR Results

Print Detail	QSAR Class Name <small>Click the name to see details of the QSAR model</small>	Type of Predicted Toxicity*		Predicted Toxicity [mg/L]	95% Prediction Interval	log P		Applicability Domain Judgement			Statistics of QSAR Class				
		Organism	Acute or Chronic			Value	Type	log p ³³ [Range]	Structure ³⁴	R ²	Q ²	RMSE	n ³⁵		
<input checked="" type="checkbox"/>	amine primary unreactive NH2 = 1 aliphatic	Fish	Acute	100	[13, 800]	1.07	Estimated	in	in	[-1.61, 5.25]	in	0.92	0.90	0.40	25(2)

*1 The query chemical may be classified into multiple QSAR classes.

*2 "Duration" and "Indicator" of each "Type of Predicted Toxicity"

Type of Predicted Toxicity		Duration	Indicator
Organism	Acute/Chronic		
Fish	Acute	96 h	LC50
Daphnid	Acute	48 h	EC50
Alga	Acute	72 h	EC50
Fish	Chronic	embryonic stage and 30 d after hatching	NOEC
Daphnid	Chronic	21 d	NOEC
Alga	Chronic	72 h	NOEC

*3 KATE2020 judges whether the log P value falls within the applicability domain or not, based on the log P value of the query chemical:
in: Within range between the minimum and the maximum log P values of the reference chemicals of the QSAR class.

図 5-1 QSAR 予測結果画面



(1) 予測対象物質の基本情報

結果画面の上部には入力した SMILES と補足情報 が表示されます (図5-2)。

CAS RN®	100-46-9		
Chemical Name	phenylmethanamine		
SMILES	NCc1ccccc1		
Molecular Weight	107.15		
log P	User Input Value	<input type="text"/>	Re-calculate
	Estimated Value by KOWWIN	1.07	
	Measured Value in KOWWIN Database	1.09	

図 5-2 予測対象物質の基本情報表示

- a 予測対象物質の構造式
- b CAS 番号 (ユーザが入力した場合のみ表示)
CAS 番号チェック (チェックディジット*が正しいかどうかの確認のみ) が行われ、間違っていた CAS 番号はその後に "incorrect" が追加されます。
* CAS 番号が正しいかどうかをチェックするための一番右側の一桁の数字のこと。入力ミスがないかを検算するのみで、SMILES と CAS 番号が一致しているかを確認するわけではありません。
- c 物質名 (ユーザが入力した場合のみ表示)
- d 予測対象物質の SMILES
- e 予測対象物質の分子量の計算値 (Open Babel プログラムを使用)
- f 予測対象物質の log P のユーザ入力値 (ユーザが入力した場合のみ表示されます)
- g 予測対象物質の log P の KOWWIN™ による推定値
- h 予測対象物質の log P の実測値 (KOWWIN™ データベースにある場合のみ表示)
注) KOWWIN™ データベースに予測対象物質の log P 実測値が複数ある場合は、データベース上の全ての値が表示されます。
- i 「Re-calculate」 ボタン: クリックすると、f 欄に入力されている log P 値を用いて QSAR 予測結果を更新します。



(2) QSAR予測結果

結果画面の中段には、QSARクラス名、予測毒性タイプ、毒性予測結果（緑字）、予測毒性値の95%予測区間、予測対象物質に対して使用したlog P、適用領域の判定、統計値が表示されます（図5-3）。以下で表示内容の説明をします。

The screenshot shows the QSAR Results page. At the top, there are filter options for 'Include (Acute/Chronic)' and 'Exclude' with checkboxes for Fish, Daphnid, and Alga. Below this is a table with columns for Print Detail, QSAR Class Name, Type of Predicted Toxicity, Predicted Toxicity, 95% Prediction Interval, log P, Applicability Domain Judgement, and Statistics of QSAR Class. A red box highlights the 'Update' button. Red arrows labeled 'a' through 's' point to various elements: 'a' points to the filter checkboxes, 'b' to the 'Exclude' section, 'c' to the 'Update' button, 'd' to the 'Print Detail' checkbox, 'e' to the 'Create Print Format' button, 'f' to the QSAR Class Name, 'g' to the Organism, 'h' to the Acute or Chronic type, 'i' to the Predicted Toxicity, 'j' to the 95% Prediction Interval, 'k' to the log P Value, 'l' to the log P Type, 'm' to the log P^33 Range, 'n' to the Structure^4, 'o' to the Structure^4, 'p' to the R^2, 'q' to the Q^2, 'r' to the RMSE, and 's' to the n^5.

図 5-3 QSAR 予測結果表示

チェックボックス

- a Include チェックボックス：チェックされている予測毒性タイプの QSAR クラスのみが表示されます。チェックを外すことにより、表示する予測毒性タイプを絞ることが出来ます。デフォルトでは全てにチェックが入っています。
- b Exclude チェックボックス：統計値 R^2 （決定係数）、 Q^2 （内部バリデーション指標）及び n （トレーニングセット数）の条件をいずれか一つでも満たす QSAR クラスは表示されません。チェックを外したり、値を変更したりすることにより、表示される QSAR クラスを調整することが出来ます。デフォルトでは全てにチェックが入っており、 $R^2 < 0.7$ 、 $Q^2 < 0.5$ 、 $n < 5$ の条件に 1 つでも当てはまる QSAR クラス（および Unclassified クラス）は非表示となっています。

※一番左のチェックボックスにチェックを入れると、右側三つの全てのチェックボックスにチェックが入ります。例えば、「Exclude」の右の括弧内のチェックボックスにチェックを入れると、 R^2 、 Q^2 および n のチェックボックスに同時にチェックが入ります。

全ての予測毒性タイプに対して QSAR クラスが表示されない場合、「No applicable results.」と表示されます（図5-4）。

Print Detail	QSAR Class Name Click the name to see details of the QSAR model	Type of Predicted Toxicity ²		Predicted Toxicity [mg/L]	95% Prediction Interval	log P		Applicability Domain Judgement		Statistics of QSAR Class			
		Organism	Acute or Chronic			Value	Type	log P ³³ [Range]	Structure ⁴	R ²	Q ²	RMSE	n ⁵
No applicable results. Change the criteria above(R ² , Q ² or n).													

図 5-4 QSAR 予測結果表示（QSAR クラスが 1 つも表示されない場合）

- c 「Update」ボタン：クリックすることにより、変更したチェック状態に対する QSAR 結果に更新します。

一括印刷フォーマット用

- d 一括印刷フォーマット用チェックボックス。デフォルトでは、QSAR クラスが $R^2 \geq 0.7$ and $Q^2 \geq 0.5$ and $n \geq 5$ を満たし、適用領域内（log P 判定が in 且つ、構造判定が in または in (conditionally)）であるものにチェックが入っています。



- e 一括印刷フォーマット画面表示用ボタン。クリックすると予測結果画面および d でチェックが入った QSAR クラスの詳細情報画面を一括印刷するためのフォーマット画面が表示されます。

QSAR クラスの名称とリンク

- f QSAR クラスの名称。クリックすると、Verify QSAR 画面に移動します。

予測毒性タイプ

- g 生物群（魚類、ミジンコ類、藻類）
h 急性／慢性

e, f で示される予測毒性タイプには、以下のものがあります。

予測毒性タイプ		試験法	試験期間	毒性指標
生物群	急性／慢性			
魚類	急性	魚類急性毒性試験 (OECD TG 203)	96 h	LC50
ミジンコ類	急性	ミジンコ急性遊泳阻害試験 (OECD TG 202)	48 h	EC50
藻類	急性	藻類生長阻害試験 (OECD TG 201)	72 h	EC50
魚類	慢性	魚類初期生活段階毒性試験 (OECD TG 210)	胚期および ふ化後 30 日間*	NOEC
ミジンコ類	慢性	ミジンコ繁殖試験 (OECD TG 211)	21 d	NOEC
藻類	慢性	藻類生長阻害試験 (OECD TG 201)	72 h	NOEC

* 魚類初期生活段階試験は魚種やふ化日数によって試験期間が異なりますが、環境省が実施した生態毒性試験で用いられているメダカでは「胚期およびふ化後 30 日間」となっています。

予測値

- i 予測毒性値
j 予測毒性値の 95% 予測区間

log P

- k 予測対象物質の log P 値
l 予測対象物質に使用される log P の種類。以下の優先順位で決定されます。
1. User Input : log P のユーザ入力値
2. Estimated : log P の推定値 (KOWWIN™ 利用)

適用領域の判定

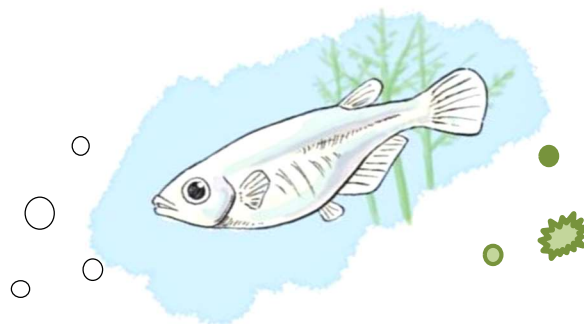
- m log P 判定結果
この欄の表示項目は、「2. (2)⑤ B) log P 判定」参照。
n [Range]
当該 QSAR クラスのトレーニングセットの記述子 log P の[最小値, 最大値] (log P の適用領域)。
o 構造判定結果
この欄の表示項目は、「2. (2)⑤ A) 構造判定」参照。



統計値

- p QSAR 式の R^2 (決定係数)
- q QSAR 式の Q^2 (Leave-one-out による内部バリデーションの指標。詳細は KATE2020 技術文書※参照)
- r QSAR 式の RMSE (平方平均二乗誤差)
- s QSAR 式のトレーニングセット数 (サポートケミカルを含まない)。かっこ内はサポートケミカルの数

※https://kate2.nies.go.jp/nies/doc/KATE_TechnicalDocument.pdf



6. QSARクラス情報の詳細表示（Verify QSAR画面）

QSAR予測結果でQSARクラス名をクリックすると、そのQSARクラスの詳細情報の画面（Verify QSAR画面）が表示されます（図6-1～6-3）。そのQSARクラスに関する回帰式のグラフ、トレーニングセットやサポートケミカルの詳細情報、構造クラスの定義、および予測対象物質の部分構造情報等が得られます。

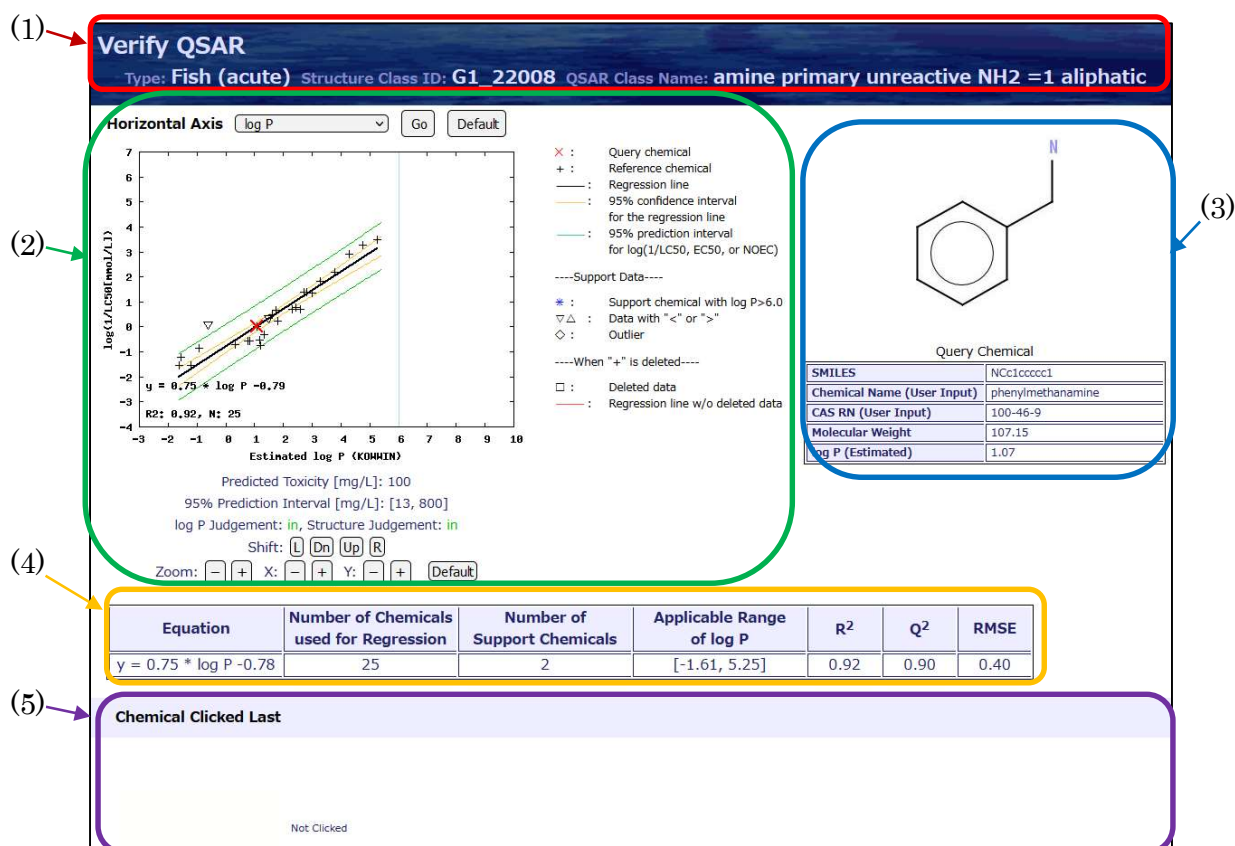


図 6-1 Verify QSAR 画面（上段）

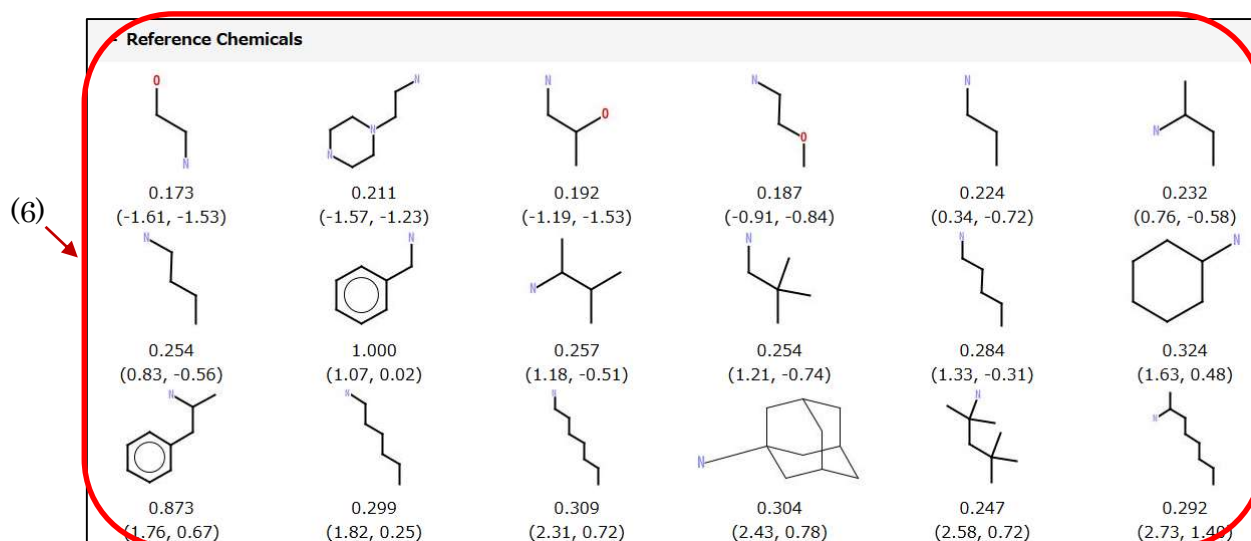


図 6-2 Verify QSAR 画面（中段）



(7) → + Chemical Data

(8) → + Definition of Structure Class (ID: G1_22008)

(9) →

Substructures of the Query Chemical

+ Substructures used only for Structural Classification

- Substructures used for the Judgement and the Classification

Hide SMARTS

Judgement*1	FragID	Substructure Name	Count	SMARTS
in	5007	Nitrogen [N,n]	1	[#7]
in	5037	pro-SB 1	1	[CH2][NH2]
in	5500	amin (daphnid ACR100)	1	[#7;v3;X3;\$([#7][!#6]);!\$([#7][#6;X3][!#7]);!\$([#7][#6]=,#[!#6]);!\$([#7][!#6;R][!#6;!#7;!#8;!#16;R][!#6;!#7;!#8;!#16;R][!#6;!#7;!#8;!#16;R])]

*1 The "Judgement" column is detailed information on the structure judgement result.
 in: the substructure is found in the "substructures for structure judgement" extracted from the reference chemicals in the QSAR class.
 in (conditionally): the substructure does not meet the condition of "in", but the substructure is found in "substructures for structure judgement" extracted from the reference chemicals in the Narcotic Group class.
 out of: the substructure does not meet the condition of "in" nor "in (conditionally)", that is, the substructure is found in neither the "substructures for structure judgement" extracted from the reference chemicals of the QSAR class nor those from Narcotic group class.

図 6-3 Verify QSAR 画面 (下段)

(1) QSARクラスの基本情報

上部に表示される紺色の帯の各項目は以下のようになっています (図6-4)。

図6-4 QSARクラスの基本情報

- a 当該 QSAR クラスに対する予測毒性タイプ
- b 当該 QSAR クラスに対する構造クラス ID
- c QSAR クラスの名称

(2) グラフ関連部分

グラフ関連部分は、以下のようになっています (図 6-5)。

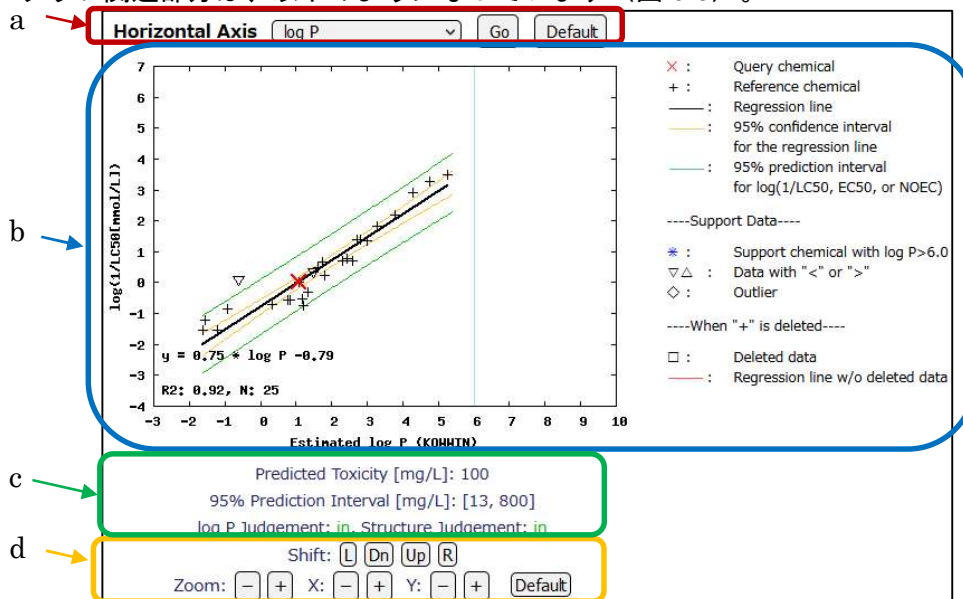


図 6-5 グラフ関連 (毒性値と log P の関係)



a 横軸の選択

デフォルトの「log P」から「Predicted Variable」に変更し、「Update」ボタンをクリックすることにより、 $\log(1/\{LC50 \text{ または } EC50 \text{ または } NOEC\}[\text{mmol/L}])$ ({ } 部分は、以下 EC50 等と記す) の「実測毒性値 vs. 予測毒性値」のグラフに切り替えることができます。「Default」ボタンをクリックするとデフォルトに戻ります。

b グラフと凡例の説明

左側はグラフ（デフォルトでは毒性値 $\log(1/EC50 \text{ 等 } [\text{mmol/L}])$ と log P の関係）、右側は（主にデフォルトの際の）グラフの凡例となっています。

×：予測対象物質

＋：トレーニングセットデータ

黒色の直線：回帰直線

橙色の双曲線：回帰式の 95%信頼区間

緑色の双曲線： $\log(1/EC50 \text{ 等 } [\text{mmol/L}])$ の 95%予測区間

-----サポートケミカルデータ-----

*：log P>6.0 の物質に対する毒性試験データ（不等号付きデータは含まない）

▽△：不等号付きデータ

◇：外れ値

-----「+」（トレーニングセットデータ）をクリックして削除した場合-----

□：削除されたデータ

赤色の直線：「□：削除されたデータ」を除外した回帰直線

-----グラフ内左下の表示-----

1 行目 回帰式

2 行目 R2：回帰式の決定係数

N：トレーニングセット数（サポートケミカルは含めない）

c 予測対象物質情報

1 行目 予測毒性値 [mg/L]

2 行目 予測毒性値 [mg/L] の 95%予測区間

3 行目 log P 判定と構造判定の結果

d グラフの移動・縮小・拡大ボタン

1 行目 Shift L：左に移動 R：右に移動

Dn：下に移動 Up：上に移動、

2 行目 Zoom -：全体の縮小 +：全体の拡大

X -：X 軸のみの縮小 +：X 軸のみの拡大

Y -：Y 軸のみの縮小 +：Y 軸のみの拡大

物質の選択

「グラフ上のポイント（+）」（図 6-6 の赤枠）または「Chemical List」の Reference Chemicals 内の構造式（図 6-7 の赤枠）をクリックして選択すると、その物質が回帰式の計算から除外され、その物質を省いた回帰直線（赤色）を追加で表示させることが出来ます。その際、予測区間（緑色の曲線）と信頼区間（橙色の曲線）については、その物質を省いて再計算されます。複数の物質を削除することも可能です。

グラフの左上に、選択（削除）した物質数についてのメッセージが表示されます



(図 6-6 の青枠)。また、グラフ内の左下(図 6-6 の緑枠)には物質を削除する前の QSAR 式に加えて、「->」印の後に削除後の QSAR 式が表示されます。R², N の更新情報も表示されます。

グラフ上では選択(削除)した物質は で表示されます(図 6-6 の赤枠)。構造式一覧上では選択(削除)された物質は紫の枠に囲まれて表示されます(図 6-7 の赤枠)。もう一度そのポイントをクリックすると選択が解除されます。

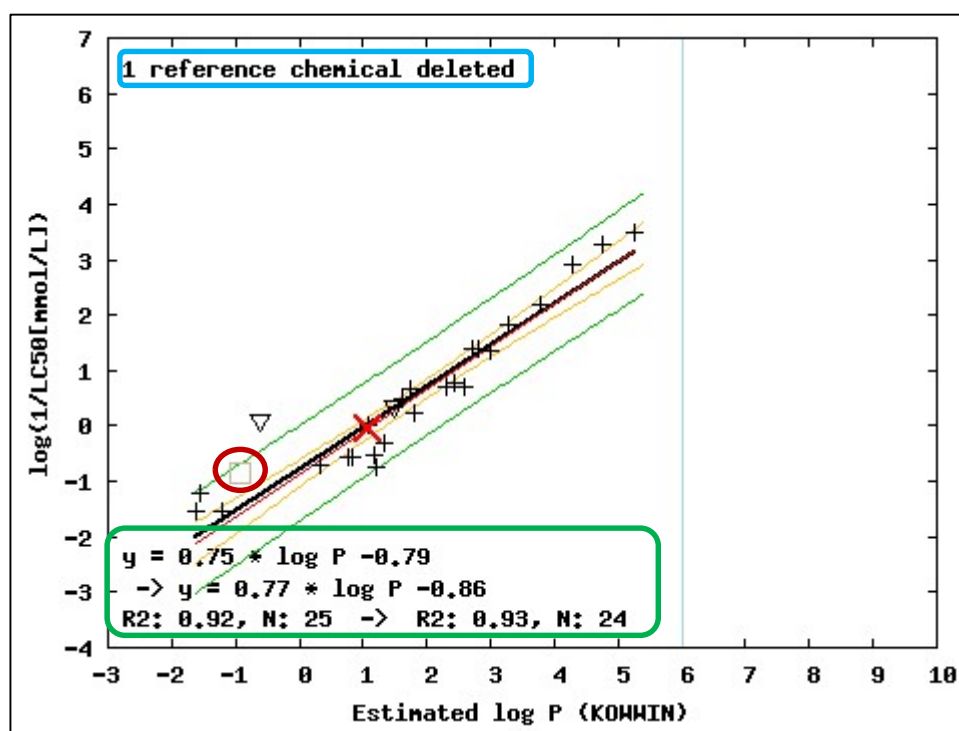


図 6-6 物質の選択(グラフ)

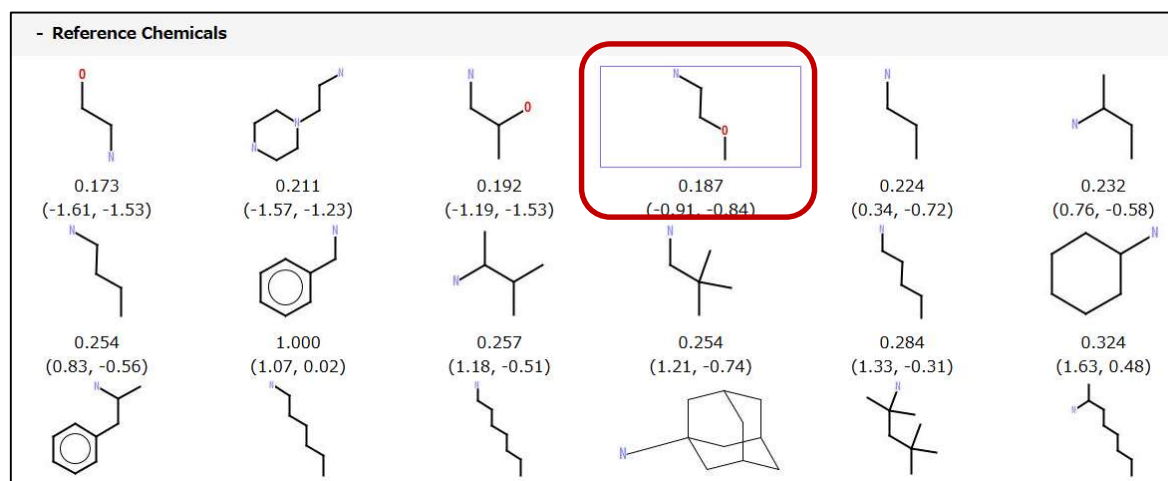


図 6-7 物質の選択(構造式一覧)



(3) 予測対象物質情報

グラフの凡例の右側に、予測対象物質の構造式および基本的な情報が表示されます（図 6-8）。

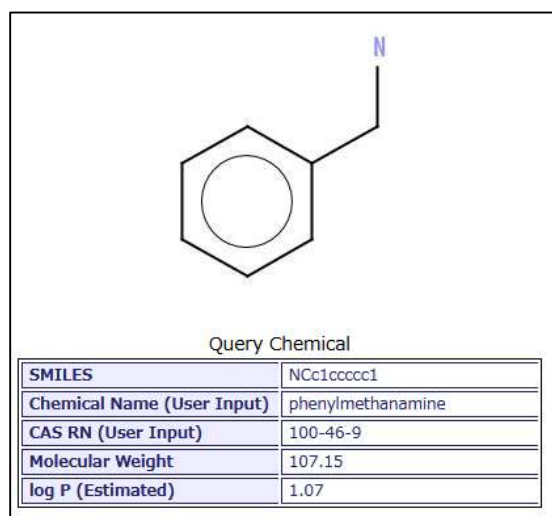


図 6-8 予測対象物質情報

表内の各行は以下のようになっています。

SMILES : 予測対象物質の SMILES

Chemical Name (User Input) : 予測対象物質に対するユーザが入力した物質名

CAS RN (User Input) : 予測対象物質に対するユーザが入力した CAS 番号

Molecular Weight : 予測対象物質の分子量

log P : 予測対象物質の log P 値。値の後ろには、ユーザ入力値の場合”(User Input)”、KOWWIN による推定値の場合”(Estimated)”が表示される。

(4) 回帰式情報

グラフ関連部分の下に回帰式 (QSAR式) 情報が表示されます (図6-9)。

Equation	Number of Chemicals used for Regression	Number of Support Chemicals	Applicable Range of log P	R ²	Q ²	RMSE
$y = 0.75 * \log P - 0.78$	25	2	[-1.61, 5.25]	0.92	0.90	0.40

図 6-9 回帰式情報

各列は以下のようになっています。

Equation : 回帰式 (QSAR 式)

Number of Chemicals used for Regression : トレーニングセット数

Number of Support Chemicals : サポートケミカル数

Applicable Range of log P : トレーニングセット中の log P 範囲 (最小値と最大値)

R² : 決定係数

Q² : Leave-one-out による内部バリデーションの指標

RMSE : 二乗平均平方根誤差



(5) 最後にクリックされた物質

「Chemical Clicked Last」の下に「最後にクリックされた物質」の情報が表示されま
す（図6-10、図6-11）。

本ページ（Verify QSAR画面）が表示されてから、まだ構造式一覧（35ページ（6）参照）
またはグラフ上の物質がクリックされていない間は、「Chemical Clicked Last」の下に
「Not Clicked」と表示されています（図6-10）。

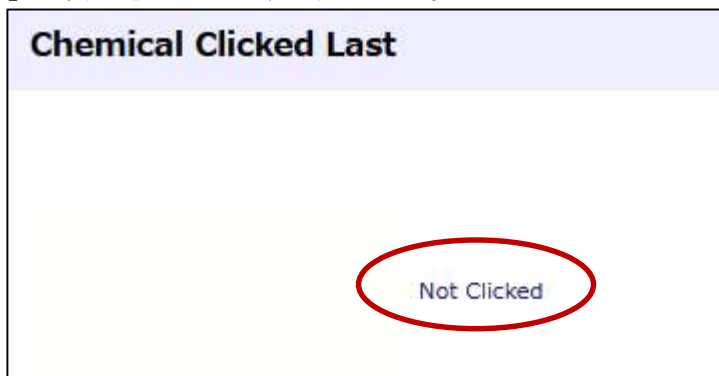


図 6-10 最後にクリックした物質の情報（物質のクリック前）

物質をクリックすると「Chemical Clicked Last」の下に「最後にクリックされた物質」
の情報：

- ・ 構造式
- ・ SMILES
- ・ CAS 番号
- ・ 物質名
- ・ グラフ上の座標
- ・ 残差の二乗
- ・ 分子量
- ・ 実測毒性値、生物種、出典
- ・ 補足情報（試験に関する情報等がある場合のみ）

が表示されます（図6-11）。

A screenshot of a software interface. At the top, there is a light blue header bar with the text "Chemical Clicked Last". Below this header, the main area is white and contains detailed information for a chemical substance. On the left, there is a chemical structure diagram of 2-methoxyethylamine. To the right of the diagram, the following information is displayed: SMILES: COCCN, CAS: 109-85-3, Name: 2-Methoxyethylamine, (X, Y): (-0.91, -0.84), Square of Residual: 0.39, Molecular Weight: 75.11, Measured Toxicity Value Data: LC50 [mg/L]: 524, Species: Pimephales promelas, Reference: USEPA. The entire content area is enclosed in a red rounded rectangle.

図 6-11 最後にクリックした物質の情報（物質のクリック後）



(6) 構造式一覧

「- Chemicals list」の下には、このQSARクラスのトレーニングセットおよびサポートケミカルの構造式一覧が表示されます（図6-12）。

The screenshot shows a web interface for a QSAR model. At the top, there is a 'Chemical List' section with a sorting menu (a) set to 'X-axis' and 'ascending order'. Below this is a text box (b) explaining that the first value under each structure is the similarity (Tanimoto coefficient) and the values in parentheses are the log P and log(1/EC50) coordinates. The main area displays a grid of chemical structures (c) with their similarity values (c) and coordinate pairs (d). At the bottom, there is a '+ Support Chemicals' section (e).

Similarity	Coordinate (x, y)
0.173	(-1.61, -1.53)
0.211	(-1.57, -1.23)
0.192	(-1.19, -1.53)
0.187	(-0.91, -0.84)
0.224	(0.34, -0.72)
0.232	(0.76, -0.58)
0.254	(0.83, -0.56)
1.000	(1.07, 0.02)
0.257	(1.18, -0.51)
0.254	(1.21, -0.74)
0.284	(1.33, -0.31)
0.324	(1.63, 0.48)
0.873	(1.76, 0.67)
0.299	(1.82, 0.25)
0.309	(2.31, 0.72)
0.304	(2.43, 0.78)
0.247	(2.58, 0.72)
0.292	(2.73, 1.40)
0.300	(2.80, 1.40)
0.851	(3.00, 1.34)
0.300	(3.29, 1.82)
0.300	(3.78, 2.18)
0.300	(4.27, 2.91)
0.300	(4.76, 3.26)
0.300	(5.25, 3.49)

図 6-12 物質データの構造式一覧

「- Chemical list」の下（図 6-12 の a）はソート機能で、CAS 番号、X 座標（log P）、Y 座標（log(1/EC50 [mmol/L])）、残差の二乗（Square of Residual）、予測対象物質との類似度（Similarity）、に対して昇順と降順でソートすることができます。デフォルトでは X 座標（log P）の昇順でソートされています。

構造式の下に表示されている数値（図 6-12 の c）は、予測対象物質との類似度（PubChem fingerprint を用いた Tanimoto 係数）を示しています（詳細は KATE2020 技術文書参照）。数値は 0 と 1 の範囲を取り、予測対象物質との類似性が高いものほど 1 に近い値を取ります。

構造式の下に 2 段目に表示されている数値（図 6-12 の d）は、以下ようになります。

○内の左側：当該物質の X 座標（log P）

○内の右側：当該物質の Y 座標（log(1/EC50 等[mmol/L]))

「- Reference Chemicals」（図 6-12 の b）の下にトレーニングセットの構造式一覧が表示されており、その下の「+ Support Chemicals」（図 6-12 の e）をクリックすると、以下の行が表示され（図 6-13）、それらをクリックすると、対応するサポートケミカルの構造式一覧が表示されます。

The screenshot shows a dropdown menu titled '- Support Chemicals' with the following options:

- + Chemicals for Data with Inequality Sign
- + Outlier
- + Chemicals (log P > 6)

図 6-13 “Support Chemicals” ドロップダウン



(7) 物質データ

「+ Chemical Data」をクリックすると、「- Reference Chemicals」の下に当該 QSARクラスに属するトレーニングセットの詳細情報一覧が表示されます（図6-14）。さらに、「+ Support Chemicals」をクリックすると、当該クラスに属するサポートケミカルの詳細情報一覧が表示されます。






- Chemical Data										
- Reference Chemicals										
CAS No.	Chemical Name	SMILES	Structure Formula	Similarity	Molecular Weight	Estimated log P	Measured Toxicity Data			
							LC50 [mg/L]	log(1/LC50 [mmol/L])	Reference	Note
141-43-5	Monoethanolamine	NCCO		0.173	61.08	-1.61	2070.0	-1.53	USEPA	
140-31-8	1-(2-Aminoethyl)piperazine	NCCN1CCNCC1		0.211	129.21	-1.57	2190.0	-1.23	USEPA	
78-96-6	1-Amino-2-propanol	CC(O)CN		0.192	75.11	-1.19	2520.0	-1.53	USEPA	
109-85-3	2-Methoxyethylamine	COCCN		0.187	75.11	-0.91	524.0	-0.84	USEPA	
107-10-8	Propylamine	CCCN		0.224	59.11	0.34	308.0	-0.72	USEPA	

図 6-14 物質データ

- a CAS No.: CAS 番号
- b Chemical Name: 物質名
- c SMILES: KATE2020 で使用されている SMILES
- d Structure Formula: 構造式
- e Similarity: 当該物質と予測対象物質との類似度
- f Molecular Weight: KATE2020 で使用されている分子量
- g Estimated log P: KOWWIN™による log P 推定値
- h EC50*: 生物影響試験結果に基づく毒性値 [mg/L]
- i log(1/EC50*[mmol/L]):
※ h と i については、当該 QSAR クラスの予測毒性のタイプにより、LC50, EC50, NOEC の中から適切なものが表示されるようになっています。
- j Reference: 元データの出典（「MOE 試験実施年度」（例: MOE 2007）または「USEPA」）。ここで MOE のデータの出典は以下の URL となっています。

<http://www.env.go.jp/chemi/report/ierac18/2-1-2-1.pdf>

（化学物質の環境リスク評価 第18巻 第2編「(II-1) 環境省の生態影響試験（藻類、甲殻類、魚類）結果一覧（平成31年3月版）」。2023年03月01日アクセス）



「USEPA」のデータの出典は、米国環境保護庁（US EPA）のファットヘッドミノール・データベースの魚類急性毒性試験結果となっています。

k Note: 物質に関するその他の情報（生態影響試験に関する情報等）
 なお、物質データの並び順はソート機能と連動しています。

(8) 構造クラス定義

「+ Definition of Structure Class」をクリックすると、当該QSARクラスに対応する構造クラスの定義情報が表示されます（図6-15）。

表の先頭のID（図6-15の赤枠）が当該QSARクラスに対応する構造クラスのIDとなっており、「IDCode*1 or SMARTS」列にその定義（IDCode*1）が記載されています。IDCode*1内の要素のIDが次の行以降で定義されています。

「ID」列は、先頭文字がアルファベットの場合は構造クラスID、4桁の数文字の場合は部分構造IDとなっています。階層構造をわかりやすくするために、階層が深くなるごとに右にずらして表示しています。

- Definition of Structure Class (ID: G1_22008)

Hierarchy Depth	ID	Structure Class or Substructure Name	IDCode*1 or SMARTS
0	G1_22008	amine unreactive NH2 =1 aliphatic	G1_22001,>0,/4510,=0,/
1	G1_22001	amine unreactive NH2=1	G1_22000,>0,/3100,=1,/
2	G1_22000	NH2 amine unreactive	U_00033,>0,/G1_00010,=0,/
3	U_00033	amine primary unreactive	3100,>0,/R_00033,=0,/
4	3100	amine CNH2	[#7X3H2;!\$([#7]*v6)];!\$([#6]~[#7,#8,#16])]]
4	R_00033	amine primary reactive	3134,>0,[3114,>0,[3128,>0,[3135,>0,[3136,>0,[4714,>0 6099,>0,
5	3134	NH2, 2,5-halo	[#7X3H2;\$([#7]c1c([F,Cl,Br,1])ccc([F,Cl,Br,1])c1)
5	3114	hydrazine N-NH2	[Nv3H2;\$([Nv3H2]-[Nv3X3]);!\$([N(C=*)C=*)]]
5	3128	amine reactive 12,13,14 w H0	[N;H2;\$([Ncc([Nv3X3,OH1]),\$(Ncccc([Nv3X3,OH1]))
5	3135	NH2, hetero aromatic 3-, 4- n	[NH2;\$([NH2v3][c]1[c][c,n][c,n][c]1);!\$([NH2v3][c]1[c][c][c]1)
5	3136	NH2, mono-halo or unsubstituted	[NH2;\$([Nc1[cH,cR2][cH,cR2][cH,cR2][cH,cR2]1),\$(Nc1[cH,cR2][cH,cR2][c;\$(c[F,Cl])][cH,cR2][cH,cR2]1),\$(Nc1[c;\$(c[F,Cl])][cH,cR2][cH,cR2][cH,cR2]1),\$(Nc1[cH,cR2][c;\$(c[F,Cl])][cH,cR2][cH,cR2]1)
5	4714	o- or p-OH, NH2 substituted aromatic-NH2	[NH2,OH1;\$([OH1,NH2]aa[OH1,NH2]),\$([OH1,NH2]aaaa[OH1,NH2]),\$([OH1,NH2]a[ar2]a[OH1,NH2])]]
5	6099	hydrazine(NN, any)	[NX3v3R0;\$([NX3v3][NX3v3]);!\$([N;H2][#7;r5,r6][#6;r5,r6](=O))]]
3	G1_00010	oxoacid [C,c]CO2-, [C,c,O]SO3-	3034,>0,/4760,>0,
4	3034	carboxylic acid C(=O)O	[#6;\$([#6]([#8])([#6])([#8H1]))]
4	4760	-SO3H, Sulfonic Acid, sulfo-, -sulfonic acid	[C,c,O]s(=O)(=O)[O;\$(OH1)],\$(O[Na,Li,K])]]
2	3100	amine CNH2	[#7X3H2;!\$([#7]*v6)];!\$([#6]~[#7,#8,#16])]]
1	4510	aromatic NH2	c-[N;H2]

*1 See glossary.

a b c d

図 6-15 構造クラス定義情報

- a Hierarchy depth: 階層の深さ
- b ID: 構造クラス ID 又は部分構造 ID
- c Structure Class or Substructure Name: 構造クラスまたは部分構造の名称
- d IDCode*1 or SMARTS: 構造クラスまたは部分構造の定義。ID が構造クラス ID の場合は IDCode*1、部分構造 ID の場合は SMARTS となります。

*1 IDCode とは、KATE の構造クラスを定義するために、より低い階層で定義された構造クラスと部分構造の個数（構造クラスについては条件に合致する場合 1、合致しない場合を 0 とする二値をとる）の条件（例えば「4500,>0」）の AND/OR による組み合わせを線形表記した識別子です。一度定義した構造クラスは、より高い階層で定義する構造クラスの中にも使用できます。

IDCode の例： 4500,>0;6055,>2/6055,<5/4328,=0|F_00007,>0
 ここで「/」、「|」、「;」、「\$」はそれぞれ「AND」、「OR」、「AND」、「OR」を意味し、演算は「/」、「|」、「;」、「\$」の順で行います。上記の例で



は、以下のように()で囲まれたかのように解釈されます。

4500,>0 AND ((6055,>2 AND 6055,<5 AND 4328,=0) OR F_00007,>0)

上記の「4500,>0」は ID: 4500 の部分構造が 1 個以上存在することを意味します。

(9) 予測対象物質の部分構造一覧

「Substructures of the Query Chemical」の下の「Substructures used for the Classification and the Judgement」の下には「構造分類と構造判定の両方に使用される部分構造一覧」が表示されます(図6-16)。

f

Judgement*1	FragID	Substructure Name	Count	SMARTS
in	5007	Nitrogen [N,n]	1	[#7]
in	5037	pro-SB 1	1	[CH2][NH2]
in	5500	amin (daphnid ACR100)	1	[#7;v3;x3;\$([#7][!#6]);!\$([#7][#6;x3][!#7]);!\$([#7][#6]=,#[!#6]);!\$([#7][!#6;R][!#6;!#7;!#8;!#16;R][!#6;!#7;!#8;!#16;R][!#6;!#7;!#8;!#16;R])]

*1 The "Judgement" column is detailed information on the structure judgment result.
in: the substructure is found in the "substructures for structure judgement" extracted from the reference chemicals in the QSAR class.
in (conditionally): the substructure does not meet the condition of "in", but the substructure is found in "substructures for structure judgement" extracted from the reference chemicals in the Narcotic Group class.
out of: the substructure does not meet the condition of "in" nor "in (conditionally)", that is, the substructure is found in neither the "substructures for structure judgement" extracted from the reference chemicals of the QSAR class nor those from Narcotic group class.

a b c d e

図 6-16 予測対象物質の部分構造一覧 (構造分類と構造判定の両方に使用)

- a Judgement: 部分構造に対する構造判定結果 (KATE2020 version 2.0 からの新機能)。QSAR クラスの構造判定の詳細情報であり、QSAR クラスの構造判定が "out of" や "in (conditionally)" であった場合に、予測対象物質に含まれる部分構造のうちどれがその判定結果の原因であるかがわかります。判定には以下があります。

in : 適用領域内

当該部分構造が、当該 QSAR クラスの「構造判定用部分構造リスト」に含まれる場合。

in (conditionally) : 条件付き適用領域内

「in」の条件には合致しないが、当該部分構造が、当該 QSAR クラスの「構造判定用部分構造リスト」、あるいは Narcotic Group クラスの「構造判定用部分構造リスト」に含まれる場合。

out of : 適用領域外

「in」と「in (conditionally)」の何れの条件にも合致しない場合。すなわち当該部分構造に、当該 QSAR クラスの「構造判定用部分構造リスト」と Narcotic Group クラスの「構造判定用部分構造リスト」の何れにも含まれない部分構造がある場合。

- b FragID: 部分構造 ID。KATE 用に設定した 4 桁の番号であり、番号は開発の便宜上、任意に設定したものです。先頭が 5 で始まります (詳細は KATE2020 技術文書参照)。
- c Substructure Name: 部分構造の名称。
- d Count: 予測対象物質に対する SMARTS 合致数 (当該部分構造の個数)。
- e SMARTS: 部分構造の定義。



- f SMARTS の表示／非表示の切り替えボタン: デフォルトでは”Hide SMARTS で、クリックすると”SMARTS 列が非表示となり、ボタン名が”Display SMARTS”に変わります。もう一度クリックすると、再び SMARTS 列が表示され、ボタン名が”Hide SMARTS”に切り替わります。

また、「+ Substructures used only for Structural Classification」をクリックすると、「構造分類のみに使用される部分構造一覧」が表示されます（図 6-17）。

- Substructures used only for Structural Classification			
FragID	Substructure Name	Count	SMARTS
3001	elements other than CX	1	[!#6;!#9;!#17;!#35;!#53]
3003	elements other than COX	1	[!#6;!#8;!#9;!#17;!#35;!#53]
3004	elements other than CSX	1	[!#6;!#16;!#9;!#17;!#35;!#53]
3009	elements other than COSX	1	[!#6;!#8;!#16;!#9;!#17;!#35;!#53]
3011	elements other than COns	1	[!#6;!F;!Cl;!Br;!I;!n;!s;!o;!O]
3014	elements other than CnosX	1	[\$([!#6;!F;!Cl;!Br;!I;!n;!s;!o]),\$([n+])]
3022	Carbon	7	[#6]
3100	amine CNH2	1	[#7X3H2;!\$([#7][*v6]);!\$(N[#6](~[#7,#8,#16]))]
3121	amine Nv3 not hindered	1	[#7v3X3;!\$(NR0)[CR1][CR1]([CX4R0])[CX4R1];!\$(NR1)(C)(C)(C)(C);!\$(N C(=[CH2]));!\$(N[#6](~[#7,#8,#16]))]
4543	MF: not C,c,O,F	1	[!C;!c;!O;!F]
4711	aliphatic-NH2	1	[N;H2;v3;X3;!\$(NC=[S,N,O]);!\$(NCC(=O)O)][C]
4892	MF: not CHO (kPilotO)	1	[!C;!c;!O]
4893	MF: not CHOP	1	[!C;!c;!O;!P]
4910	aromatic	6	[a]

g
h
i
j

図 6-17 予測対象物質の部分構造一覧（構造分類のみに使用）

- g FragID: 部分構造 ID。KATE 用に設定した 4 桁の番号であり、番号は開発の便宜上、任意に設定したものです。先頭が 3,4,6,7 のものが存在します（詳細は KATE2020 技術文書参照）。
- h Substructure Name: 部分構造の名称。
- i Count: 予測対象物質に対する SMARTS 合致数（当該部分構造の個数）。
- j SMARTS: 部分構造の定義。



7. 複数の化学物質の予測

複数のSMILESが書かれたファイルを入力することにより、複数の化学物質に対するQSAR予測を行うことができます。

(1) 入力ファイル「SMILES list」について

KATEでは、複数のSMILESを連続して処理するために、各行に1物質のデータの連続からなる入力ファイル「SMILES list」（ファイル名は任意）を使用します。KATE2020の「SMILES list」はKATE2017から仕様を変更しております。

入力ファイルはテキスト形式で、1行目にヘッダ（大文字・小文字を区別しません）を記入、2行目からそれぞれの予測対象物質の情報を記入してください。

ヘッダとしては 必須の“SMILES”の他に、任意で”ID”（ユーザ定義のID）、”LogP”（入力log P値）、”NAME”（物質名）、”CAS”（CAS番号）も入力可能です。各項目の区切りは「タブ」により区切ってください。

□ SMILES list の例 1

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

※ 上記では、1行目に列名「SMILES」が指定されており、2行目以降の各行で、予測対象物質の「SMILES」が指定されています。

□ SMILES list の例 2

```
NAME _____ LogP _____ SMILES _____ CAS _____ id
name1 _____ -0.8 _____ CCCCOC(=O) _____ A10
name2 _____ _____ CC(=C)CS _____ A50
name3 _____ 1.3 _____ Nc1ccnc1 _____ 3731-52-0 _____ F20
```



- ※ 列は交換可能です。
- ※ ヘッダおよび各値はタブにより区切ってください（例2では、`_____`はタブを表しています）。全ての行でタブの数をそろえる必要があります。
- ※ SMILES 列以外は省略可能です。

SMILES listについては、「Select」ボタンの上の「SMILES list」のリンクからも説明を見ることができます（図7-1）。

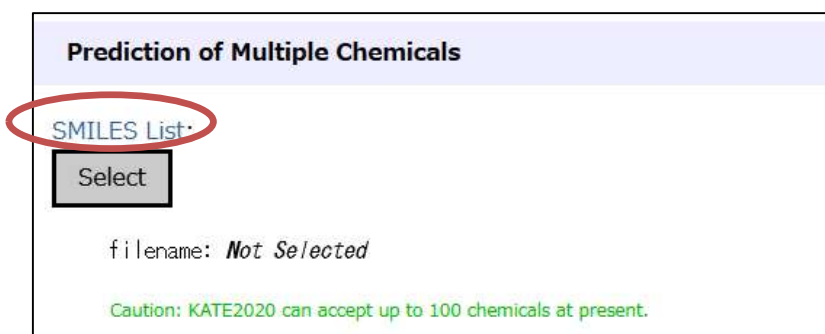


図 7-1 「SMILES list」説明画面へのリンク



(2) 予測方法手順

手順 1. 化学物質の入力画面で「Select」ボタンをクリックする（図7-2）。

図 7-2 「複数の化学物質予測」用の Select ボタン

手順 2. 毒性予測を行いたい化学物質の SMILES が書かれている入力ファイル「SMILES list」（(1)参照）を選択して「開く」をクリックする（図 7-3）。

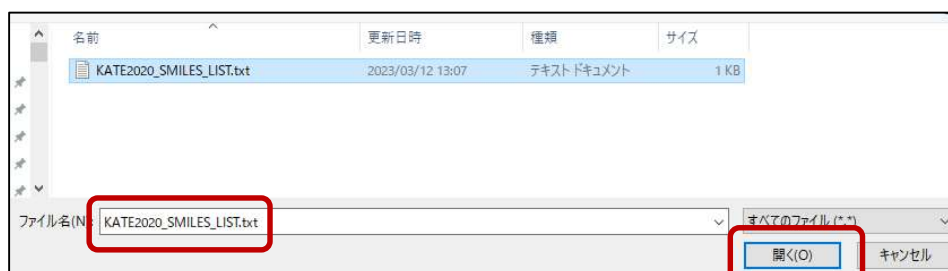


図 7-3 SMILES list の選択



手順3. 予測 (Predict) ボタンが現れるので、クリックすると計算が開始されます (図7-4)。

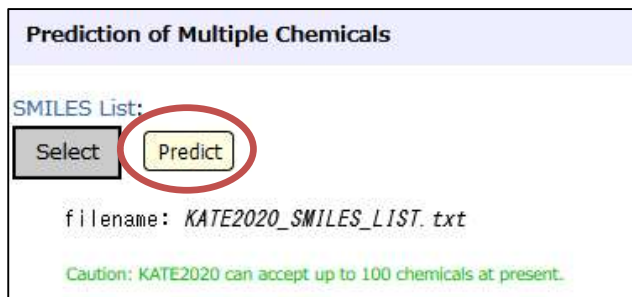


図 7-4 「複数の化学物質予測」用の予測 (Predict) ボタン

計算終了後、予測結果が表示されます (図 7-5)。

Results (batch mode)

Include: Acute Fish Daphnid Alga
 Include: Chronic Fish Daphnid Alga
 Exclude: R² < 0.7 Q² < 0.5 n < 5

Batch Results

No	ID	CAS RN [®]	Chemical Name	SMILES	Molecular Weight	Estimated log P	Structural Formula	QSAR Class Name <small>Click the name to see details of the QSAR model</small>	Type of Predicted Toxicity ²		Predicted Toxicity (mg/L)	95% Prediction Interval	log P	Applicability domain Judgement			Statistics of QSAR Class				
									Ingestion	Acute or Chronic				Value	Type	Structure ¹	R ²	Q ²	n	RMSE	n ¹
1				CCC	44.10	1.81		C_X unreactive aliphatic w/o X	Fish	Acute	36	[4.2, 310]	1.81	estimated	out of	[2.58, 4.98]	in	0.73	0.68	0.36	21(5)
								narcotic-group Fish Acute	Fish	Acute	64	[9.7, 430]	1.81	estimated	in	[-0.63, 5.88]	in	0.88	0.88	0.42	149(31)
								narcotic-group Daphnid Acute	Daphnid	Acute	12	[1.7, 79]	1.81	estimated	in	[1.08, 5.88]	in	0.75	0.74	0.41	79(22)
								narcotic-group Alga Acute	Alga	Acute	64	[7.2, 570]	1.81	estimated	in	[1.08, 5.26]	in	0.77	0.75	0.44	48(48)
								C_X HC unreactive	Fish	Chronic	1.0	[0.057, 19]	1.81	estimated	in	[1.52, 5.52]	in	0.78	0.68	0.43	11(0)
								Class_X unreactive Fish Chronic	Fish	Chronic	0.99	[0.056, 17]	1.81	estimated	in	[1.52, 5.52]	in	0.76	0.68	0.43	12(0)
								narcotic-group Fish Chronic	Fish	Chronic	1.1	[0.074, 17]	1.81	estimated	in	[1.52, 5.81]	in	0.82	0.75	0.41	12(0)
2				CCN	45.08	-0.15		amine primary unreactive NH2 =1 aliphatic	Fish	Acute	358	[4, 2800]	-0.15	estimated	in	[-1.61, 5.25]	in	0.92	0.90	0.40	25(2)
								No applicable results.													
3					62.13	1.27		No applicable results.													
4				CCO	46.07	-0.14		primary alcohol	Fish	Acute	3700	[350, 39000]	-0.14	estimated	in	[-1.75, 5.26]	in	0.92	0.90	0.44	22(15)
								CO_X alcohol unreactive w/o EO Fish	Fish	Acute	6700	[790, 57000]	-0.14	estimated	in	[-0.63, 5.81]	in	0.90	0.89	0.43	45(13)
								narcotic-group Fish Acute	Fish	Acute	3800	[550, 26000]	-0.14	estimated	in	[-0.63, 5.88]	in	0.88	0.88	0.42	149(31)
								CO_X alcohol unreactive w/o EO Daphnid	Daphnid	Acute	810	[38, 17000]	-0.14	estimated	out of*	[0.78, 5.81]	in	0.86	0.80	0.46	13(13)
								narcotic-group Daphnid Acute	Daphnid	Acute	270	[35, 2100]	-0.14	estimated	out of*	[1.08, 5.88]	in	0.75	0.74	0.41	79(22)
								primary alcohol	Alga	Acute	27000	[92, 8.1e+6]	-0.14	estimated	out of*	[2.31, 5.26]	in	0.91	0.79	0.36	6(16)
								CO_X alcohol unreactive w/o halogen, acid, EO	Alga	Acute	12000	[180, 770000]	-0.14	estimated	out of*	[1.08, 5.26]	in	0.95	0.90	0.35	9(14)
								narcotic-group Alga Acute	Alga	Acute	5500	[40, 66000]	-0.14	estimated	out of*	[1.08, 5.26]	in	0.77	0.75	0.44	48(48)
								narcotic-group Fish Chronic	Fish	Chronic	30	[1.0, 840]	-0.14	estimated	out of	[1.52, 5.81]	in	0.82	0.75	0.41	12(0)
								CO_X alcohol unreactive w/o EO Daphnid	Daphnid	Chronic	26	[1.0, 640]	-0.14	estimated	in	[-1.20, 5.81]	in	0.82	0.75	0.53	13(8)
								CO_X alcohol unreactive w/o halogen, acid, EO	Alga	Chronic	1100	[15, 85000]	-0.14	estimated	out of	[0.69, 5.26]	in	0.87	0.81	0.59	10(9)

図 7-5 「複数の化学物質予測」の予測結果

チェックボックスと表の内容は1つの化学物質予測の場合とほぼ同じ項目となっています。QSAR Class Name列の各項目をクリックすると、Verify QSAR画面に移動します。

(3) 動作について

- SMILES list 中、エラーがあった行については、計算をスキップして予測結果の当該行にエラーメッセージが表示されます。
- 一度に予測できる物質数は、現時点では100に制限しています。
- 予測時間は、SMILES list 中の化学物質の構造によって変化します。通常では最大30分程度で計算が終了すると考えられます。30分以上かかる場合は、入力したSMILES等が原因で計算に不具合を生じている可能性があります。その場合は、入力する化学物質 (SMILES) 数を減らす、あるいは単独物質での予測を行うなどの工夫をして、正常に動作するかご確認ください。なお、計算に不具合があるSMILES等を見つけた場合は、KATE担当までご一報頂けますと幸いです。



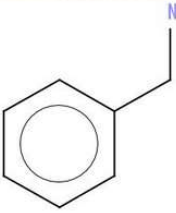
8. 一括印刷フォーマット表示 (Print Format画面)

QSAR予測結果で、ボタン **Create Print Format** をクリックすると、Results画面および「Print Detail」列でチェックが入ったQSARクラスのVerify QSAR画面を一括して印刷するためのフォーマット画面が表示されます (図8)。

(1) Ecotoxicity Prediction by KATE2020 version 4.0
March 12, 2023 at 13:28 (JST) [http://kate.nies.go.jp]

(2) **Results**

CAS RN®		
Chemical Name		
SMILES	NCc1ccccc1	
Molecular Weight	107.15	
log P	User Input Value	
	Estimated Value by KOWWIN	1.07
	Measured Value in KOWWIN Database	1.09

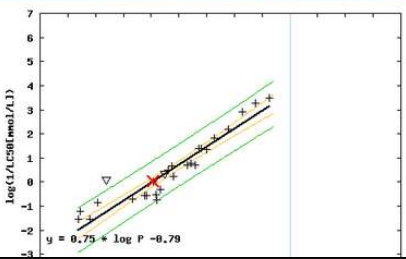


Include(Acute): Fish Daphnid Alga
 Include(Chronic): Fish Daphnid Alga
 Exclude(): R² < 0.7 Q² < 0.5 n < 5

QSAR Results

Print Detail	QSAR Class Name ^{*1}	Type of Predicted Toxicity ^{*2}		Predicted Toxicity [mg/L]	95% Prediction Interval	log P		Applicability Domain Judgement		Statistics of QSAR Class				
		Organism	Acute or Chronic			Value	Type	log P ³ (Range)	Structure ^{**}	R ²	q ²	RMSE	n ¹⁵	
<input checked="" type="checkbox"/>	amine primary unreactive NH2 =1 aliphatic	Fish	Acute	100	[13, 800]	1.07	Estimated	in	[-1.61, 5.25]	in	0.92	0.90	0.40	25(2)
<input checked="" type="checkbox"/>	amine unreactive NH2 =1 aliphatic	Daphnid	Acute	30	[0.21, 4300]	1.07	Estimated	in	[-1.61, 3.00]	in	0.78	-2.11	0.32	4(1)
<input checked="" type="checkbox"/>	amine unreactive NH2 =1 aliphatic (alga)	Alga	Acute	9.3	[0.00087, 100000]	1.07	Estimated	in	[-1.61, 3.00]	in	0.09	-3.92	0.59	4(0)
<input type="checkbox"/>	CNO_X unreactive (Fish chronic), excl. (CnosX w/o n+)	Fish	Chronic	0.22	[0.0097, 5.0]	1.07	Estimated	in	[-1.61, 5.99]	in	0.62	0.54	0.57	19(2)
<input type="checkbox"/>	N_X amine aliphatic NH2=1	Daphnid	Chronic	0.14	[0.0026, 7.8]	1.07	Estimated	in	[-1.61, 3.19]	in	0.45	0.29	0.76	19(0)
<input type="checkbox"/>	amine unreactive NH2 =1 aliphatic	Daphnid	Chronic	1.2	[0.020, 77]	1.07	Estimated	in	[-1.61, 1.63]	in	0.23	-2.19	0.25	4(0)
<input type="checkbox"/>	amine unreactive NH2 =1 aliphatic (alga)	Alga	Chronic	1.8	[0.00010, 31000]	1.07	Estimated	in	[-1.61, 3.00]	in	0.32	-2.65	0.62	4(0)

(3) Type: Fish (acute) Structure Class ID: G1_22008 QSAR Class Name: amine primary unreactive NH2 =1 aliphatic



Legend:
 × : Query chemical
 + : Reference chemical
 — : Regression line
 — : 95% confidence interval for the regression line
 — : 95% prediction interval for log(1/LC50, EC50, or NOEC)
 ----Support Data----
 * : Support chemical with log P>6.0
 ∇△ : Data with "<" or ">"
 ◇ : Outlier
 ----When "+" is deleted----
 □ : Deleted data
 — : Regression line w/o deleted data

図8 Print Format 画面

- (1) : タイトルおよび予測結果が出力された日時 (日本標準時)
- (2) : Results画面 (チェックは変更できない)
- (3) : チェックが入ったQSARクラスに対するVerify QSAR画面の主な部分 (log(1/EC50等[mmol/L]) vs log Pグラフ、予測対象物質の当該QSARクラスでの予測結果 (予測毒性値、95%予測区間、適用領域の判定等)、QSARクラスの回帰式と統計値、物質データ、構造判定用部分構造一覽等)

