



生態毒性QSARモデル 「KATE」



Web版の解説とスタンドアロン版の公開

(独)国立環境研究所 環境リスク研究センター
白石寛明

大分大学教育福祉科学部
吉岡 義正(東京会場のみ)

はじめに

QSARとは

- ▶ Quantitative Structure–Activity Relationshipの略
- ▶ 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)といい、定量的なものを定量的構造活性相関(QSAR)といいます。
- ▶ 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モデルと呼びます。

生態毒性予測システムKATE



- ▶ KATE (KAshinhou Tool for Ecotoxicity) は、
 - 環境省の請負業務(平成16-19年度)として(独)国立環境研究所環境リスク研究センターにおいて、研究・開発された生態毒性QSARモデル
 - KATE は、平成19年7月より3省合同審議会に魚類及び甲殻類の予測結果を参考資料として提出し実績を重ねています
- ▶ 2008年1月に試用版がWeb公開:<http://kate.nies.go.jp>
- ▶ 化学物質の部分構造から毒性値を予測:
 - 魚類急性毒性試験における半数致死濃度(LC50)
 - ミジンコ遊泳阻害試験における半数影響濃度(EC50)
- ▶ **本システムで得られた予測結果は、化審法の届出に必要な生態毒性試験結果として利用することは出来ません**

今年度のスケジュール

現在:

Web版とスタンドアロン版の結果の一致



12月下旬:

スタンドアロン版のベータテストの実施



年度末:

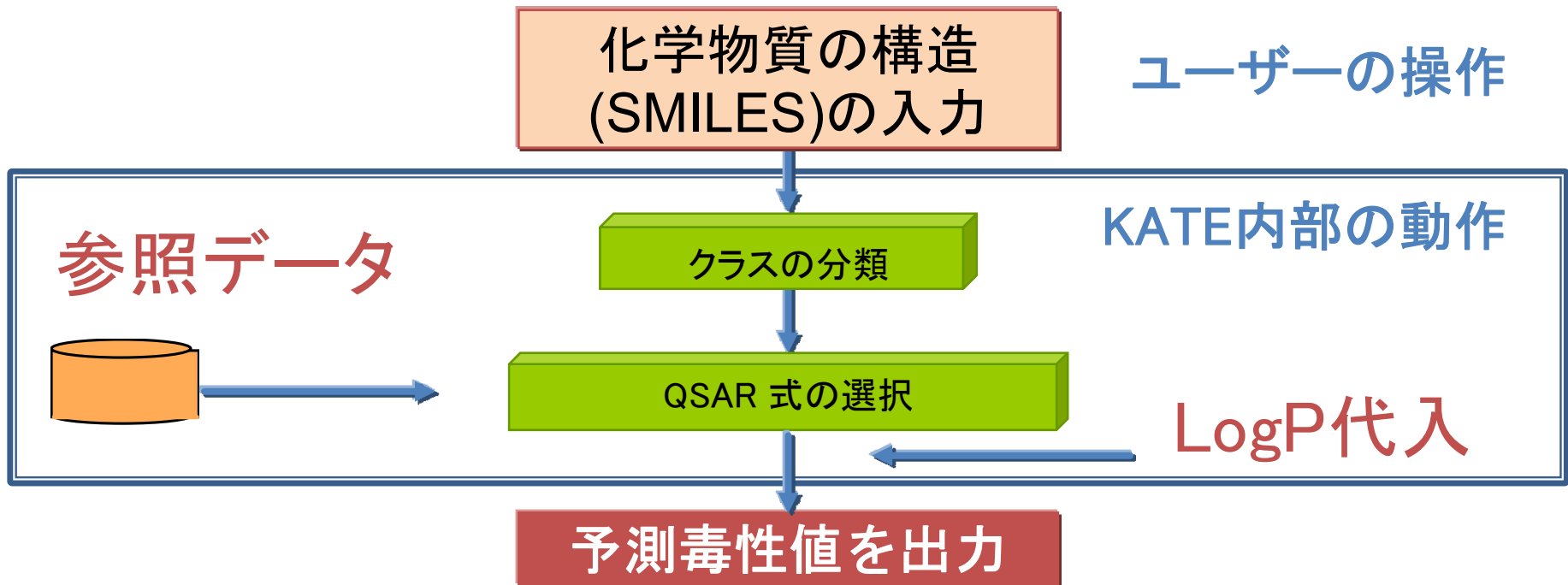
更新Web版KATEの公開

+

スタンドアロン版の配布

KATEの解説

KATEでの予測方法 1/2

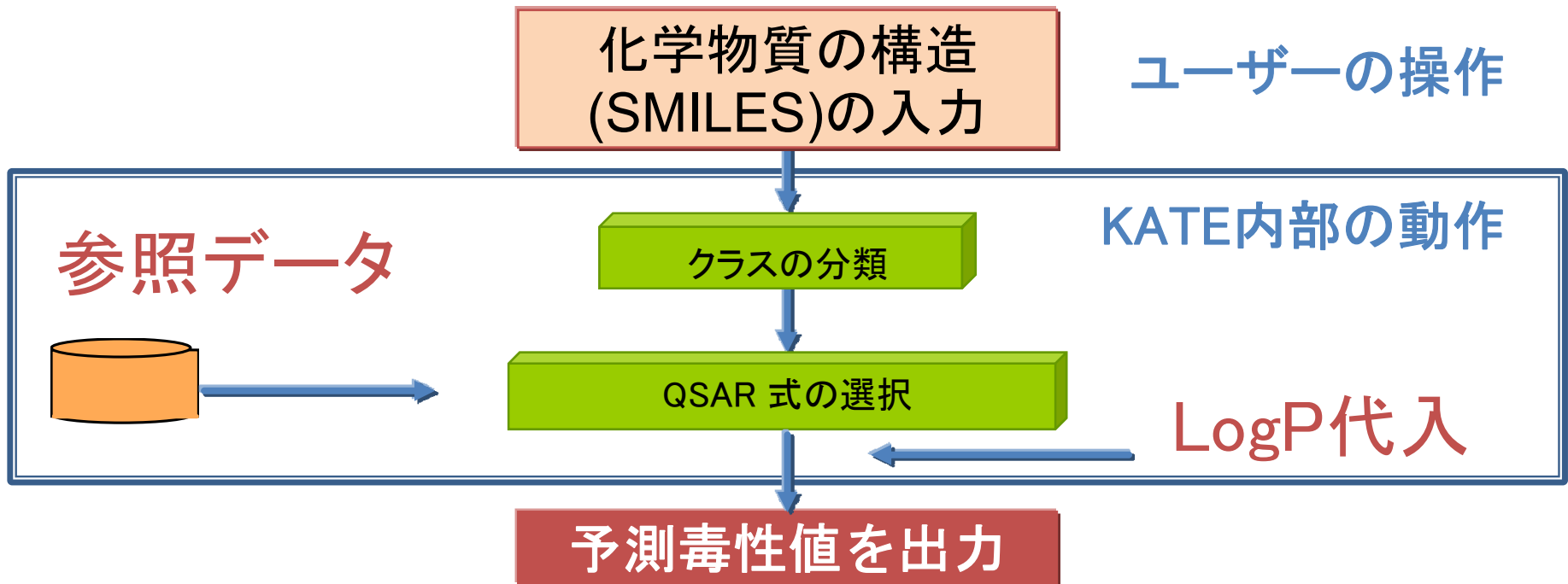


- ▶ 化学物質の構造に基づくクラス分類した線形式を使用
- ▶ 水-オクタノール分配関数(logP)の値から単相関で毒性値を予測
- ▶ 現バージョンでは入力されたSMILESから計算値ClogPを算出し、オプションでユーザー入力による実測LogP予測も可能

(システムの一部でDaylight社製のソフトウェアを使用、SMILESは後ほど解説)

<http://www.daylight.com/>

KATEでの予測方法 2/2



▶ 参照データについて

- 環境省が実施した生態毒性試験結果(魚類急性毒性試験 96hr LC50、ミジンコ遊泳阻害試験 48hr EC50)
- 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(96hr LC50)
- 今後試験結果が更新された場合に、QSAR式見直しを行う予定

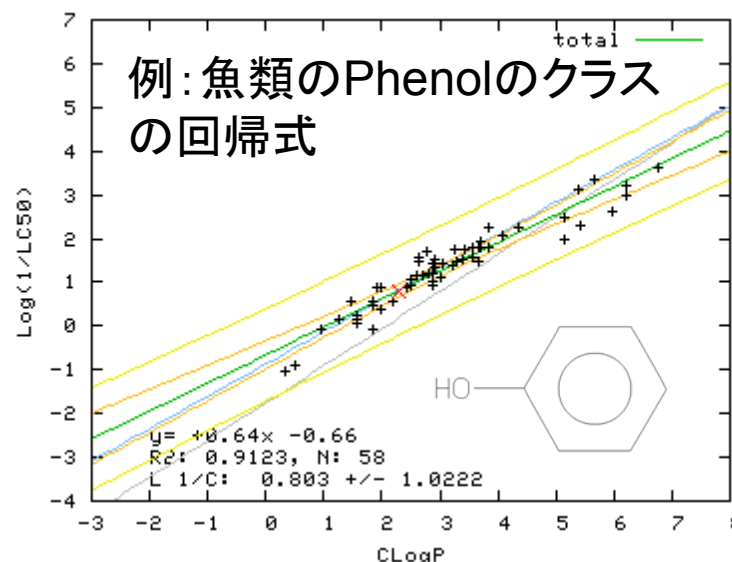
KATEのクラスについて

▶ 現在約68種類のQSAR式・クラスが存在

○ 魚類で参照データが20以上の例:

phenols; aliphatic amines; aliphatic alcohol ether;
 Aromatic Amines; aromatic HC; aliphatic halogen;
 Aromatic Amines 2; aromatic aldehyde; aliphatic ketone

	魚類	甲殻類
式の数 (参照データ 2以上)	66	48
データ数が十分な式の数 (参照データ 5以上)	49	19
信頼できる式の数 (参照データ 5以上 かつR ² が0.7以上)	33	10



* R²とは決定係数のことで0から1の間の値をとり、当てはまりのよさを示す
 一般に単回帰式の場合、R²が0.7以上で参照物質5点以上だと信頼が高い

KATEのNeutral Organics



- ▶ 脂肪族炭化水素、スルホキンド、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分子種の一覧が生物種ごとに用意されており、これらはNeutral Organicsというクラスとして再定義されています。
- ▶ KATEでNeutral Organicsに分類されるクラス群：
 - sulfoxides; aliphatic HC (hydrocarbon);
 - aliphatic ether; aromatic/aliphatic ether;
 - aliphatic ketone; aromatic ketone;
 - aliphatic alcohol ether

使用方法について

- ▶ URL: <http://kate.nies.go.jp>
- ▶ 予めユーザー登録が必要です。



KATE

KAshinhou Tool for Ecotoxicity
生態毒性予測システム (試用版)

[Login](#)

更新履歴

サイトポリシー

よくある質問
FAQ

 独立行政法人
国立環境研究所

 環境リスク研究
センター

 環境省

 環境省
化学物質審査室

English

生態毒性予測システム (KATEver0.1) 試用版の利用を開始しました。2008.01.31

ケイト
生態毒性予測システム「KATE」について

生態毒性予測システム (通称: **KATE**^{※1}) は、環境省の請負業務 (平成16年度から平成19年度) として (独) 国立環境研究所環境リスク研究センターにおいて、研究・開発された生態毒性 **QSAR** モデルです。

化学物質の部分構造から **魚類急性毒性試験** における半数致死濃度 (LC₅₀) 及び **ミジンコ遊泳阻害試験** における半数影響濃度 (EC₅₀) を予測するシステムです。化学物質の入力は、CAS番号^{※2} 検索や構造式エディタを用いた作図等による SMILES^{※3}、またはユーザ自身が作成した MOL ファイル^{※4} によって行います。

※1 KAshinhou Tool for Ecotoxicity
 ※2 化学物質を特定するための最大10桁の数値からなる識別子
 ※3 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
 ※4 MDL社が開発した分子座標のフォーマット

KATE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、ミジンコ遊泳阻害試験) 及び米国環境保護庁 (US EPA) の **ファットヘッドミニョーデータベース** の魚類急性毒性試験結果を参照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。

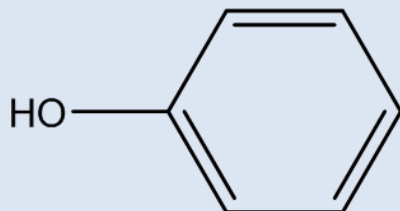
初めて利用する方は、まず [新規ユーザ登録](#) を行ってください。KATEの使用法、予測結果等については [操作マニュアル](#) をご覧ください。

[KATE Login](#)

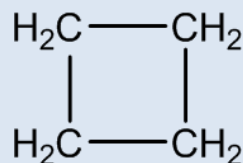
化学物質の入力形式

- ▶ SMILES: 化合物の分子構造を線形表記した識別子*

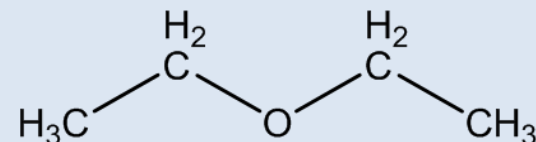
フェノール: c1ccccc1O



シクロブタン: C1CCC1



ジエチルエーテル: CCOCC



- ・ 小文字“c”は芳香族炭素
- ・ 環構造は数字を用いて始点を表記

- ・ 大文字“C”は脂肪族炭素
- ・ 水素は自動的に付加

- 詳細はDaylight社ページ <http://www.daylight.com/smiles/index.html>

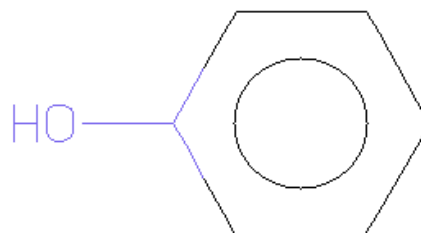
* Weininger D. 1988. SMILES, A Chemical Language and Information Systems. 1. Introduction of Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.*, 28: 31–36.

予測結果画面 1 / 2

Logout > Input molecule > View result > Verify QSAR

生態毒性予測システム (KATE ver0.1) ようこそ さん

CAS	108-95-2 WebKis-Plus	
物質名	PHENOL	
SMILES	Oc1ccccc1	
分子量	94.12	
水溶解度 (mg/L)	入力(実測値)	<input type="text"/> 更新
	DB(実測値)	1.47
LogP	入力(実測値)	<input type="text"/> 更新
	構造より予測	化学物質特性予測システム
色の説明	黒	部分構造
	青	毒性構造
	灰	未検索



- ▶ ページの上部に入力したCAS番号、物質名、SMILESを表示
- ▶ 分子量は入力した情報から値を表示
- ▶ ClogP計算値はライセンス上の理由で提供できません
- ▶ SMILESは内部変換されるため入力と異なる場合あり
- ▶ LogP値を入力すると実測値による予測が可能
- ▶ 溶解度を入力すると毒性値の溶解度判定が可能

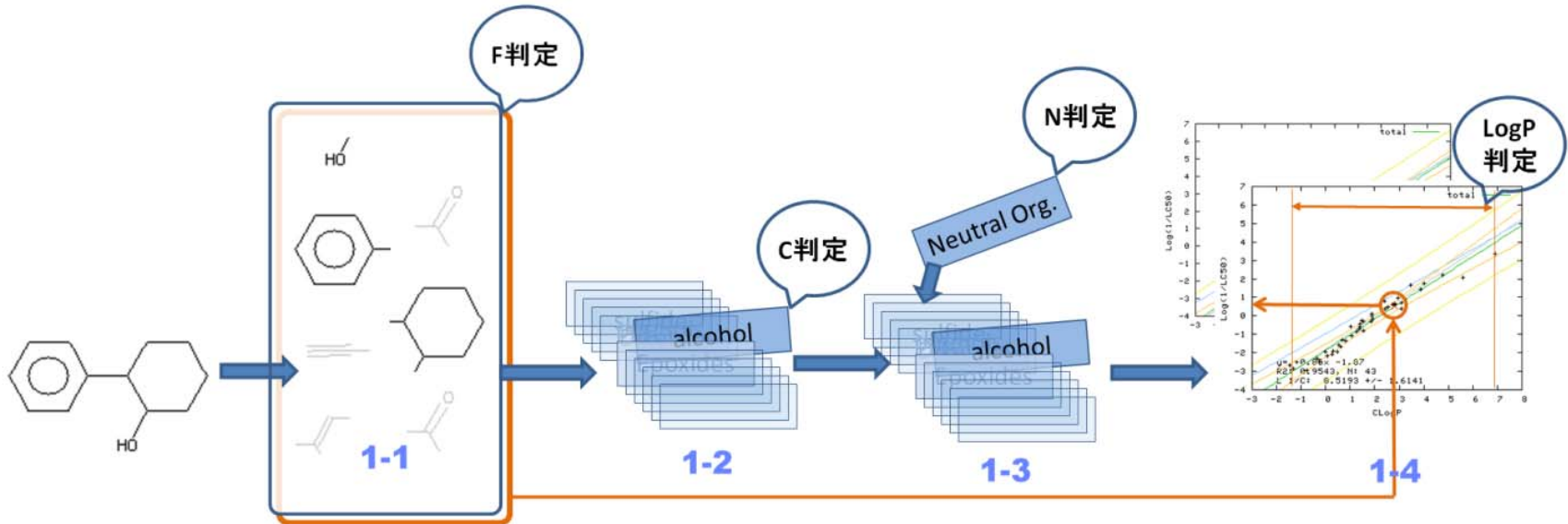
予測結果画面 2/2

QSAR結果

クラス※1	エンドポイント			LogP	予測結果(mg/L)	判定				
						構造N※2	構造C※3	LogP※4	溶解度※5	既知F※6
 phenols	Fish	96hr	LC50	CLogP予測値	4.870e+1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	未入力	<input type="radio"/>
 phenols	Daphnid	48hr	EC50	CLogP予測値	1.582e+1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	未入力	<input type="radio"/>

- ▶ ページ中段に、化学物質のクラス分類、Fish(魚類)、Daphnid(ミジンコ類)の予測結果等が表示されます。
- ▶ 複数のクラスに分類される場合には、全てのクラスにおける結果が表示されます。判定の項目に表示される○の数が多いクラスを優先してください。
 - 但し、Neutral OrganicsとNeutral Organics以外のクラスで○の数が同じ場合には、Neutral Organics以外のクラスを採用するか、Verify QSARの画面で参照物質やQSAR式の決定係数などを参考に判断してください。

判定の項目について



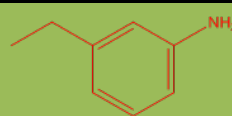
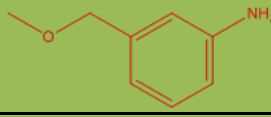
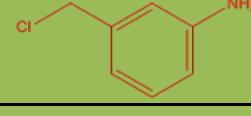

入力された構造式は、1-1部分構造一覧となり、1-2クラス分類が行われ、1-3必要ならNeutral Organicsが追加され、1-4毒性値が産出され、1-5適用ドメイン判定が下される。

入力物質の判定について

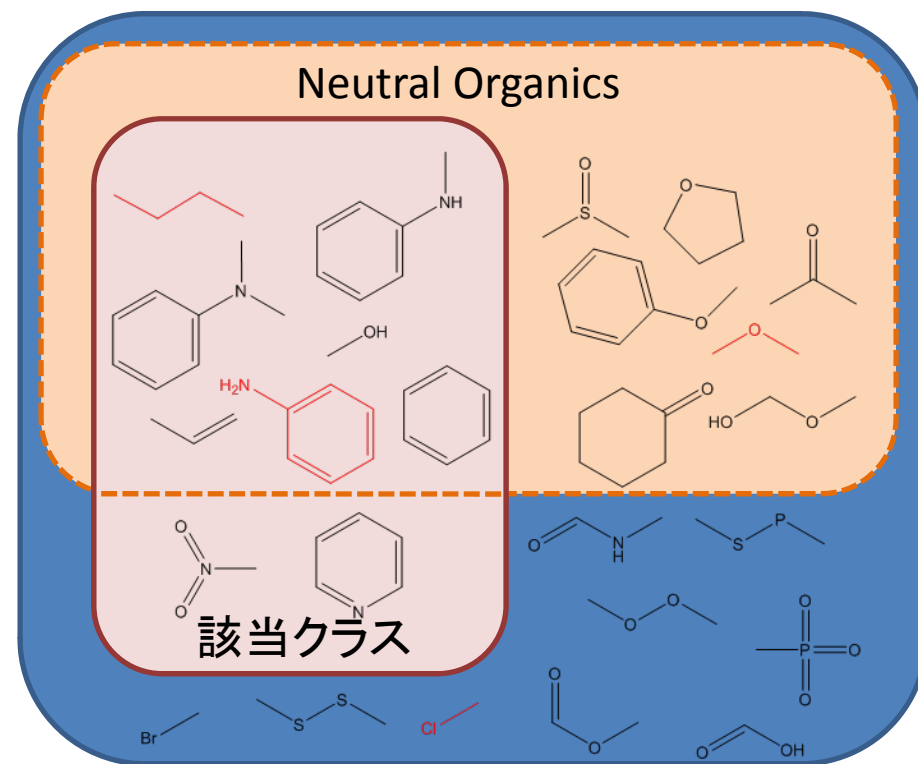
構造C判定: 該当クラスを構成する参照物質に含まれない部分構造がある場合に×

構造N判定: Neutral Organicsと該当クラスを構成する参照物質に含まれない部分構造がある場合に×

既知F判定: 未検索構造がある場合に×

フラグメントの組み合わせ例 該当クラス: Aromatic Amines	C 判定	N 判定	F 判定
該当クラス + 脂肪族C-C 	○	○	○
該当クラス + 脂肪族C-O-C 	×	○	○
該当クラス + 脂肪族C-Cl 	×	×	○
該当クラス + Ag 	○	○	×

F判定×の時、結果の信頼性極めて低い



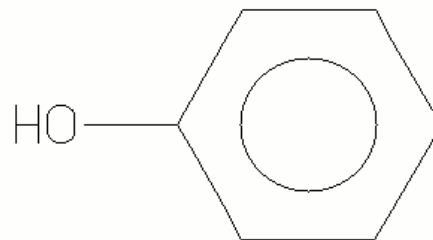
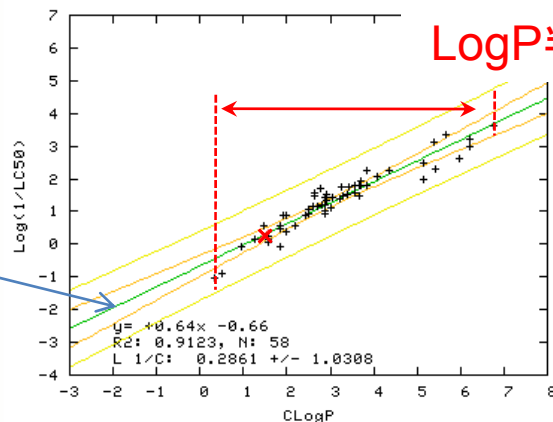
KATEの回帰式

Logout > Input molecule > View result > Verify QSAR

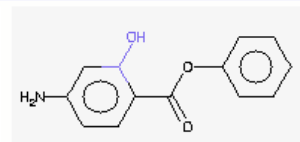
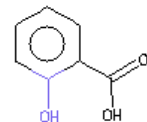
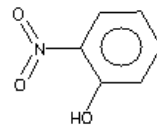
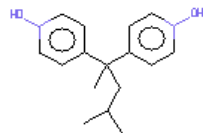
フラグ

式の詳細 生物種: 魚類 ID: 10020041 Name: phenols

グラフには
回帰式 R^2
参照データ数 n
が表示される。



参照物質一覧



- ▶ 予測結果画面で、分類されたクラスの 🔍 をクリックすると、そのクラスの回帰直線 (total、緑色) 及び参照物質一覧が表示されます。入力した化学物質はグラフ上 × です。

KATEのデモ



- ▶ ここからは実際の画面をご覧ください

FAQ Web版のよくある質問

▶ 質問

- ログイン時に「セッションがタイムアウトになりました。再度ログインしてください。」のメッセージが出ます。
- 予測結果(mg/L)の「e⁺²」や「e⁺⁴」等の意味を教えてください。
- ログイン名、パスワードを忘れてしまいました。

以下のページを参照ください:

<http://kate.nies.go.jp/faq.html>

その他のWeb版KATEに関する問い合わせ先:

kate@nies.go.jp

よくある質問より 1/2



情報の機密性は保持されますか？

- ▶ 構造式情報(SMILES)はログアウトをすると作業ファイルは削除され、サーバーには残りません。ログアウトを忘れた場合には、再度、同じユーザ名でログインし、ログアウトすると作業ファイルは削除されます。作業中にはSMILESが記載されたファイルが作成されていますが、他のユーザがファイルの覗き見をできないように管理されています。ただし、ユーザ名とパスワードを盗まれると作業中の様子がview resultの画面で閲覧可能になりますので、ご注意ください。
- ▶ 複数の化合物を予測する「SMILESリストファイルからの入力」は、情報の遺漏に関する検証が不足しております。機密性を考慮する場合は、使用しないようお願い致します。

よくある質問 2/2



情報の機密性は保持されますか？

- ▶ 通常のhttp通信による構造式情報(SMILES)の傍受を懸念される場合は、

<https://kate.nies.go.jp>

として、httpsプロトコルを利用ください。SSL(Secure Sockets Layer)暗号化通信と、サーバー認証を使用して、より安全性の高いログインが可能です。

- ▶ SSL通信によって機密性は向上しますが、インターネット通信での完全な機密保持は不可能ですので、ご利用に当たっては御社の機密情報の管理規定に基づき判断してください。当方では、通信上の情報の遺漏に関して責任を負いません。
- ▶ **スタンドアロン版の利用をご検討ください。**

スタンドアロン版について

スタンドアロン版開発の背景

- ▶ 秘密保持の問題

Web上では秘密の漏えいの恐れがあり、スタンドアロン版での予測システムが必要

- ▶ 透明性の確保

独自システムの構築実績がある大分大学と共同開発

- ▶ ライセンスの問題上

商用プログラムを含むWeb版KATEを無料配布することは不可能

Web試用版の仕様変更

- ▶ **入力には化学構造式 (SMILES) とLogPが必要**
 - 実測LogPによる毒性予測を基本
 - 実測LogP値がない場合には予測値を利用
- ▶ **構造分類の基本プログラムの変更**
 - 大分大学との共同研究によるシステムの導入



スタンダードアロン版の技術

大分大学 吉岡義正

(大阪会場では白石寛明が講演)

KATE on PAS(PC版)について



- ▶ PAS (Platform for Assessment from Structure)は、構造分類に基づく物性や毒性を予測するための独自のシステムです。
- ▶ PASは、部分構造の取得プログラム (FITS; Fragment Identification by Tree Structure)、構造図の表示・入力プログラムなどからなる統合システムです。
- ▶ KATE on PASは、KATE用に作成されたPASです。
- ▶ 開発言語(VB.net 2005)
- ▶ スタンドアロン版(PC版)と Web版はFITS記述で統一されます。

KATE on PAS の概要



PAS:技術的共用部

**SMILES — 2次元構造
相互変換プログラム**

SMILES

入力

部分構造の取得

**部分構造の定義
FITS**

**クラス分類
QSAR式などの選択**

クラス分類の定義

出力

**毒性値
物性値など**

FITSの概要

- ・ FITSは部分構造の規定に独自のルールを用いています。
- ・ 主体部分は、1次元構造を基本としたFITS記述です。

F/01211/**C=CNC=C**/1JnC=O,3V3,3B3,2Cy,3Cy,4Cy,2Rs4,/|

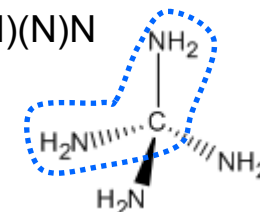
- ・ 部分構造の数え方

Order,Index,Account,Exclude,Reset,**Fits**,Pilotで規定します。

ID	Order	Index	Account	Exclude	Reset	Fits	Value	Pilot
319	300	1	-S [sulfur +4 or +6 valence;miscellaneous]	0	0	S/1/S/1V4,1=00,1#0,	0	16
319	310	1	-S [sulfur +4 or +6 valence;miscellaneous]	0	2	S/0/[Sb]// 1>0/02/[Sb]S// Specific/	0	16
316	320	1	-S- [aliphatic sulfur, 2 nitrogen attach]	0	0	F/010/NSN/2=00,	1.2	16
293	330	1	-ON02 [aromatic attach]	0	0	F/11/ON/2=02,1Ar1,2B3,	0.5	24
292	340	1	-ON02 [aliphatic attach]	0	0	F/11/ON/2=02,1Ar0,2B3,	-0.02	24
283	350	1	-OH [hydroxy, nitrogen attach]	0	0	F/01/NO/2H1,	-0.043	0
359	360	1	-S-P [sulfur, phosphorus attach]	0	0	F/10/SP/1V2,	0.627	48
314	370	1	-S- [di-,poly- sulfur attach]	0	0	F/010/SSS/2B2,	0	16
360	380	1	-SS- [disulfide]	0	0	F/20/SS/1V2,2=00,2=S0,	0.5497	16

例: NC(N)(N)Nの構造でNCNの構造の数を、目的に応じて1-6個まで定義できます。

SMILES: NC(N)(N)N



スタンドアロン版操作のデモ



- ▶ 百聞は一見にしかず

利用する立場から見れば, 便利に使えます!