

## 汎用生態リスク評価管理ツール MeRAM のための化学構造辞書の構築

○泉原拓<sup>1</sup>, 桂樹哲雄<sup>1</sup>, 林彬勸<sup>2</sup>, 高橋由雅<sup>1</sup> ( <sup>1</sup>豊橋技科大院・工, <sup>2</sup>産総研)

## 【はじめに】

現在, 化審法や REACH 法, 生物多様性保全への対応として, 化学物質のリスク評価管理のニーズが高まっている. 一方, 化学物質の環境毒性にもとづく生態への影響等, リスク評価には多大な時間と高度な専門知識が必要とされる. 産業技術総合研究所では, 収集した毒性データを背景に, 誰でも容易に, 注目する化合物のリスク評価が行えることをねらいとした汎用生態リスク評価管理ツール MeRAM[1]の開発を進めている. 一方, 筆者らは, 同様に毒性評価済みデータから, 構造類似化合物のデータを訓練集合とし, QSAR モデルを生成・利用して生態毒性の推算/予測を行うためのシステム (PEACH) の開発を進めている[2]. 後者は化学構造式から注目するエンドポイントに対する毒性値の予測が可能であり, 欠損データの補完などに利用できる. しかし, MeRAM データベースには化学構造式の情報が入蔵されていない.

そこで本研究では, MeRAM データベースに対する構造情報そのものの補完や, 欠損データの予測になど, PEACH との連携を図るうえで重要となる化学構造辞書を作成するとともに, 同辞書を活用するためのソフトウェアツールの開発を行った.

## 【化学構造辞書の構築】

(i) 構造データの収集: 産総研で開発を進めている化学物質のリスク評価・管理システム MeRAM には約 3900 物質, 27 万件を越える毒性データが収録されている. PEACH との連携を図る上では構造データが不可欠である. そこで, はじめに MeRAM に収録されているデータの詳細を調査した. 1 物質に対して様々なエンドポイントに対する複数の試験結果が収録されていることから, CAS 番号を手掛かりに, 表 1 に示す 3 つのエンドポイントごとに化合物リストを作成した. 化合物リストの CAS 番号をもとに, 外部データベースから構造式の情報収集した. 構造式の収集には日化辞 Web を利用し, CAS 番号で検索し, 各々の MOL ファイルをダウンロードした.

(ii) 化学構造辞書の構築: MOL ファイル形式で収集された各化合物の構造情報はすべて独自の原子結合表に変換し, 構造辞書に収録した. また, 別途作成した類似構造検索のため構造特徴記述子としてすべての化合物について TFS (Topological Fragment Spectra) [3]を生成し, 収録した. 作成した構

表 1 エンドポイントと収録試験データと化合物数

エンドポイント	試験データ数	化合物数
魚類96h-LC50	21,225	2,461
ミジンコ48h-EC50	7,236	1,734
藻類72h-EC50	2,042	926

```
<compounds>
  <compound cas='CAS番号'>
    <name>化合物名</name>
    <ct_file>Molin結合表データ</ct_file>
    <tfs feature='特徴付け' max_size='最大サイズ' min_size='最小サイズ'>
      TFSデータ</tfs>
    <molecular_formula>分子式</molecular_formula>
    <molecular_weight>分子量</molecular_weight>
  </compound>
  ...
</compounds>
```

図 1 構造辞書ファイルの例 (CIML形式)

## Construction of Chemical Structure Dictionary for MeRAM

Taku Izumihara<sup>1</sup>, Tetsuo Katsuragi<sup>1</sup>, Bin-Le Lin<sup>2</sup>, Yoshimasa Takahashi<sup>1</sup><sup>1</sup>Department of Computer Science and Engineering, Toyohashi University of Technology,<sup>2</sup>National Institute of Advanced Industrial Science and Technology.

造辞書への収録項目は, CAS 番号, 名称 (複数を収録の場合はセミコロンで区切る), 分子式, 分子量, 原子結合表および TFS からなる. 構造辞書ファイルの仕様は, 上述の生態環境毒性予測システム (PEACH) との連携を勘案し, XML を基礎とした独自のテンプレートを定義して用いた (図 1). このテンプレートには CAS 番号, 化合物名, 分子式, 分子量, 物性値, 毒性値などの化合物情報に加え, 化学構造式を格納することができる. また, 階層構造を持つため, 1 化合物に対して項目 (物性値, 毒性値等) ごとに複数のデータを格納する際にも柔軟に対応できる.

### 【化学構造辞書ツールの開発】

上で作成した化学構造辞書の参照および積極的な活用をねらいとした化学構造辞書ツールの開発を行った. 本ツールでは, これまでに CAS 番号, 化合物名, 分子量による一般的な検索機能に加え, 化学構造式をクエリとして検索する全構造検索および構造類似性検索の機能を実装している. クエリの入力と検索結果の表示部を同じ部分で表示するユーザーインターフェースとなっており, クエリを入力して, 検索ボタン



図 2 化学構造辞書ツールの実行画面 (例)

を押すだけで結果が表示されるシンプルな設計となっている. また, 検索結果が複数存在する場合は, 別途検索結果のリストを表示し, そのリストを選択することで, 対応する結果が表示される. 開発した辞書ツールの実行画面例を図 2 に示す. 本ツールで実装されている構造類似性検索には, 当研究室で考案された TFS 法を利用している.

### 【まとめ】

MeRAM データベースにおける構造情報の補完および欠損データの予測計算のための PEACH との連携を念頭に, MeRAM が収録する魚類 96h-LC<sub>50</sub>, ミジンコ 48h-EC<sub>50</sub>, 藻類 72h-EC<sub>50</sub> のエンドポイントの毒性データをもつ化合物の化学構造式を収集し, 化学構造辞書を構築した. また, 作成した化学構造辞書を利用するため, シンプルで分かりやすいユーザーインターフェースを備えた辞書ツールを開発した. 検索機能としては, CAS 番号, 化合物名, 分子式による一般的な検索に加えて, 構造検索機能として全構造検索, 構造類似性検索を可能とした. 今後は予定として, 部分構造検索の実装並びに, PEACH との連携のためのデータインポート/エクスポート機能についても検討して行きたい.

【謝辞】 本研究は (社) 日本化学工業協会新 L R I 研究助成を受けて実施したものであることを明記し, 謝意を表する.

### 【参考文献】

- 1) 産業技術総合研究所, 「AIST-MeRAM, 産総研一汎用生態リスク評価管理ツール」, URL=<https://www.aist-riss.jp/software/AIST-MeRAM/> (2017/3/22 閲覧)
- 2) 高橋由雅, LRI 研究中間報告書「化学物質の有害性予測および環境リスク評価・管理システムの高度化」, 2016 年 8 月, 日本化学工業協会.
- 3) Kentaro Kawai, Satoshi Fujishima, and Yoshimasa Takahashi, Predictive Activity Profiling of Drugs by Topological-Fragment-Spectra-Based Support Vector Machines, *J.Chem. Inf. Model.*, **48**, 1152-1160, 2008.