

# 「埋蔵分子」発掘プロジェクト

## —化学反応経路マップのインタラクティブ可視化に向けて

佐藤寛子<sup>1</sup>, 小田朋宏<sup>2</sup>, 中小路久美代<sup>3</sup>, 宇野毅明<sup>1</sup>, 田中宏明<sup>4</sup>, 岩田寛<sup>4</sup>, 大野公一<sup>5,6</sup>

1. 国立情報学研究所, 2. 株式会社SRA, 3. 京都大学 学際融合教育研究推進センターデザイン学ユニット,  
4. 東京大学情報理工学系研究科, 5. 東北大学大学院理学研究科, 6. 量子化学探索研究所

### Abstract

現在までに存在が確認されている化学物質は約700万種類であり、年間数十万~百万種オーダーで増え続けている。しかし、理論的には存在しているが、いまだ人類が手にしていない化学物質種数はこれを遥かに凌駕することが明らかとされつつある。私達は、この革新をもたらす「埋蔵分子」を理論的に探索・発掘し、これらを供給する化学反応経路を、分子のポテンシャルエネルギーや電子状態に関するパラメータとともにデータベース化し、化学系の研究者をはじめ、研究・教育機関や一般からも利用できるシステムを目指し、プロジェクトを開始した。まず、化学反応解析を専門とする化学者のための可視化システムの開発に着手したので、最初の結果を報告する。

### 分子構造と化学反応経路の探索

#### ◎トポロジカルな分子構造の数上げ（トポロジカル法）

原子（点）と結合（線）のトポロジカルな関係にもとづき数学的に列挙する方法 (ex. MOLGEN[1])

**利点:** ・高速に分子構造を生成することができる。

**欠点:** ・組合せの爆発が起こる。  
・分子構造は平面であり、3次元構造の取扱いが難。  
・結合長、結合角、イオン構造などの多彩な構造の可能性を取り込むのが難。  
・ポテンシャルエネルギーなどの物性値は構造を生成しただけでは得られない。

#### ◎両手法の比較

Input:  $C_6H_6$  (例: ベンゼン) とした場合に、出力される分子構造の個数

- トポロジカル法 (MOLGEN) : 217
- ポテンシャル法 (GRRM) : 2073

※GRRMでは、PM3法、LADD=4とした限定探索を実施。全面探索ではこれを遥かに上回る異性体が出力される(計算継続中)。

#### ◎量子化学に基づく探索（ポテンシャル法）

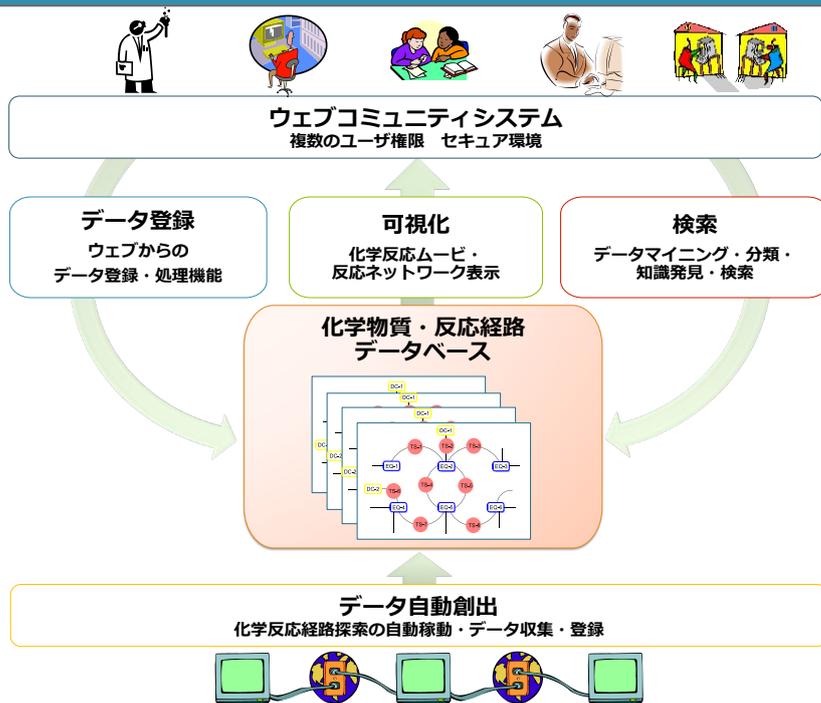
分子のエネルギー曲面を探索する方法 (ex. GRRM[2])

- 利点:**
- ・ポテンシャルエネルギーにより分子の存在の可能性を含めて理論的に評価できる。
  - ・反応が起こる際に越えなければならないエネルギー障壁(活性化エネルギー、遷移状態エネルギー)も得られる。
  - ・分子軌道や電子状態の情報が得られる。
  - ・結合長、結合角、イオン構造などの多様性を取り込むことができる。
- 欠点:**
- ・計算コストが膨大。
  - ・網羅的な探索は不可能だと考えられていた(→GRRM法により可能となった)。
  - ・計算条件(用いる理論や使用する基底関数、探索の方法(限定探索/全面探索)など)により、得られる分子構造の個数や種類が異なる場合がある。

MOLGENからは、原子価を満たす、トポロジカルな異性体が生成される。

これに対して、GRRMからは、原子価を満たさないものも含めた、立体異性体が生成される。GRRMで制限を与えない全面探索を行うと、より多くの構造(4000種以上)が生成される[3]。→ 未知化合物が含まれる可能性が高い

### 「埋蔵分子」発掘プロジェクト



そこで、私達は、GRRMを利用し、未知の化学物質を含む化学反応経路の網羅的探索とリポジトリ化を行い、新規物質の発見と創出のための化学情報資源を整備する『埋蔵分子発掘プロジェクト』を開始した。

GRRMによる分子構造・反応経路探索は、ユーザによるものと、分散型GRRMにより広範囲に自動的に行われるものの2種類を想定している。いずれもリポジトリ化され、計算結果の可視化と解析、データリポジトリの検索を可能とする。大量に蓄積された分子構造・反応経路データのマイニングにより、新規物質の発見と反応経路設計へと応用する機能も構想に含まれている。

現在までに、データベースのプロトタイプを開発するとともに、GRRMの計算結果を化学者が解析するための可視化ツールの開発を進めてきている。GRRMから得られる分子構造と化学反応経路の数は膨大なため、化学者が化学反応経路探索結果を検索・解析する作業を効果的にアシストできる可視化システムと、これと連動する効率的な検索アルゴリズムが必要である。可視化・検索システムを通じて、化学者が化学反応に関する新しい知見を得たり、新規化合物・化学反応を発見したりできるシナリオを想定し、このためのインタラクティブな可視化システムを目指して開発を進めている。

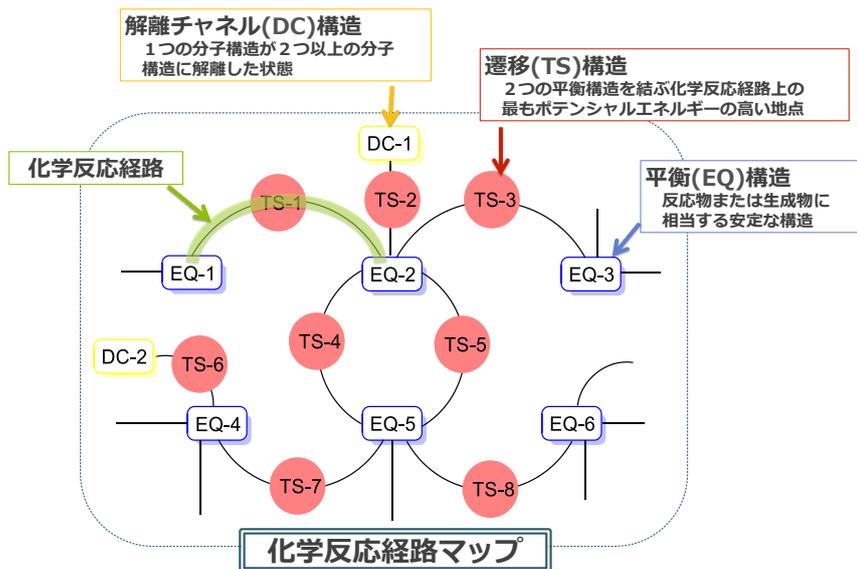
### References

[1] <http://molgen.de/?src=documents/molgen4.html> [2] Ohno, K.; Maeda, S. *A Chem. Phys. Lett.* **2004**, *384*, 277-282.; Maeda, S.; Ohno, K. *J. Phys. Chem. A* **2005**, *109*, 5742-4753.; Ohno, K.; Maeda, S. *J. Phys. Chem. A* **2006**, *110*, 8933-8941. [3] 大野公一, 佐藤寛子, 岩本武明, “反応経路自動探索法GRRMの特性評価と機能拡充”, 第6回分子科学討論会, 2012. [4] [http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)

# 化学者による化学反応経路マップの解析と可視化

## ◎ 化学反応経路マップの構成要素

- ① 分子構造（平衡構造、遷移構造、解離チャンネル）の幾何学情報
- ② 各分子のポテンシャルエネルギー等の物理化学パラメータ



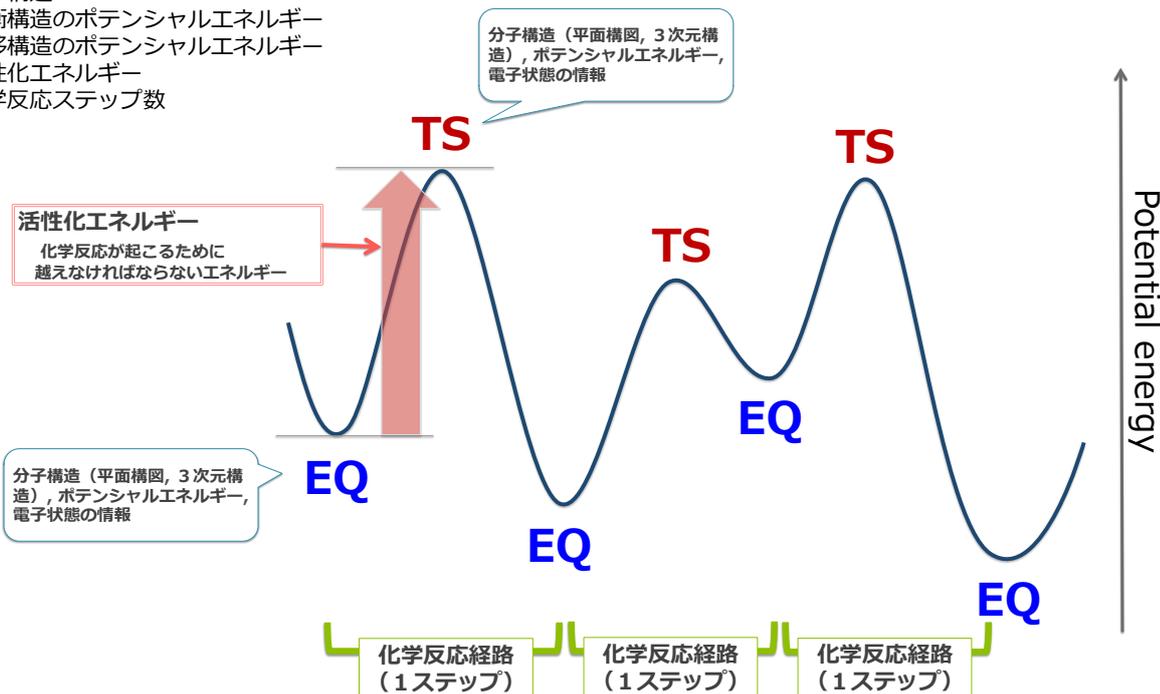
GRRMから得られる出力データは、同じ組成式（例えば $C_6H_6$ など）を持つ複数の分子構造（異性体）の幾何学情報（原子の種類と三次元座標値）と、そのポテンシャルエネルギー等の物理化学パラメータである。

ポテンシャル曲面の極小点に存在する安定な分子構造は**平衡構造(EQ)**と呼ばれ、化学反応における反応物または反応生成物に相当する。平衡構造が化学反応により別の化合物に変化する過程は化学反応経路とよばれるが、2つの平衡構造を結ぶ化学反応経路上の最もポテンシャルエネルギーの高い地点は**遷移状態 (TS: Transition State)**と呼ばれる。これは、ポテンシャルエネルギー曲面の鞍点に相当する。化学反応により、1つの分子構造が分解し、2つ以上の分子構造へと変化する場合もある。この際に得られる分離した状態は**解離チャンネル(DC: Dissociation Channel)**とよばれる。

したがって、GRRMの出力は、複数のEQ, TS, DCをノードとしてもつ、化学反応経路のネットワークデータであり、1つの組成式から得られるネットワーク全体を**化学反応経路マップ**とよぶ。

## ◎ 注目する属性一何を解析するか？

- ① 分子構造
- ② 平衡構造のポテンシャルエネルギー
- ③ 遷移構造のポテンシャルエネルギー
- ④ 活性化エネルギー
- ⑤ 化学反応ステップ数



### 解析のポイント

- ◎ 化学反応が起こりやすい経路を探す
- ◎ 化学反応がどのような分子構造の変化を経て起こるかを知る
- ◎ 化学合成しやすい経路を探す

### 可視化に求められる要素

- ◎ より低い活性化エネルギーを経る化学反応経路の表示
- ◎ 化学反応経路に沿った分子構造変化の履歴（ムービー）
- ◎ 化学反応経路マップ全体と局所的な表示（全体図から各部詳細へ）
- ◎ 反応ステップ数によるソート