

「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^{†1} 小田朋宏^{†2} 中小路久美代^{†3}
宇野毅明^{†1} 田中宏明^{†4} 岩田 寛^{†4} 大野公一^{†5, †6}

量子化学に基づき理論的に自動探索される化学反応経路のネットワーク状の大規模データ(化学反応経路マップ)から、新規化学物質(埋蔵分子)やそれを生成する化学反応経路を掘り出し、有効な化学物質の設計や化学合成に応用することを目指すプロジェクトを開始した。本論では、本プロジェクトの目的を説明し、プロジェクトの一環として構築中の化学者が化学反応解析を行うための化学反応経路マップの可視化システムについて、化学者が化学反応経路マップを解析する際に必要とするビジュアルインタラクティブリティについての考察とシステムの現状を報告する。

Treasure Hunting for Novel Chemical Compounds -Towards Interactive Visualization for Chemical Reaction Route Maps

HIROKO SATOH^{†1} TOMOHIRO ODA^{†2} KUMIYO NAKAKOJI^{†3}
TAKEAKI UNO^{†1} HIROAKI TANAKA^{†4} SATORU IWATA^{†4} KOICHI OHNO^{†5, †6}

We launched a project for the discovery of new chemical compounds and reaction routes from big data of reaction route networks, called chemical reaction route maps, which are automatically retrieved based on quantum mechanics. Applications for designing useful chemical compounds and reaction routes are part of the project. We report here the overview of the project and the first results of a visualization system especially for chemists who perform reaction analysis based on quantum mechanical calculations.

1. はじめに

現在までに存在が確認されている化学物質は約 7000 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、理論的には存在しうるが、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。我々は、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、最初の成果を報告する。

2. 分子構造と化学反応経路の探索

2.1 トポロジカルな分子構造の数え上げ

指定された組成式を満たす分子構造を数え上げる方法としては、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(ここではトポロジカル法と呼ぶ)がある。トポロジカル法としては、ドイツ Bayreuth 大学の A. Kerber, R. Lane らにより開発されている MOLGEN¹ が代表的である。トポロジカル法の利点は極めて高速に数え上げを可能とする点にある。得られる分子は原子価を満たすもののみであり、分子構造は平面である。

2.2 量子化学に基づく分子・反応経路探索

一方、量子化学に基づき、ポテンシャル曲面上の極小点と鞍点を結ぶ化学反応経路を探索することで、分子構造とその合成方法を得る数え上げ方法(ここではポテンシャル法と呼ぶ)も原理的に可能である。これを自動的に探索することは不可能であると考えられていたが、2004 年に大野公一・前田理(当時東北大学)らにより発表された SHS(Scaled Hypersphere Search)と名付けられた画期的な手法により、従来の常識が破られた。^{2,3,4} 本手法は、GRRM(Global Reaction Route Map)プログラムに実装され、ポテンシャル曲面を超球面探索法により網羅的に探索する。現在までに種々の化学物質に適用され、従来知られていた化学物質や反応経路を遥かに凌駕する可能性が発掘されることが明らかとなってきた。たとえば、組成式 C_6H_6 の場合、トポロジカル法では 217 種の分子構造が得られるが、GRRM

†1 国立情報学研究所
National Institute of Informatics

†2 株式会社 SRA
SRA Inc.

†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
Kyoto University

†4 東京大学情報理工学系研究科
Tokyo University

†5 東北大学大学院理学研究科
Tohoku University

†6 量子化学探索研究所
Institute for Quantum Chemical Exploration

では、半経験的量子化学計算のレベルで制限探索を行った場合でも 2073 種の平衡構造が出力される(制限を設定しない場合にはこの数を遥かに超える(計算実行中))。ポテンシャル法の利点は、分子の性質や反応性、物性を議論するのに不可欠な、分子構造が 3 次元であること、電子状態や分子軌道の情報を含むこと、遷移状態を含む反応経路の情報が提供されることである。

2.3 「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。システム構想を図 1 に示す。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・化学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。

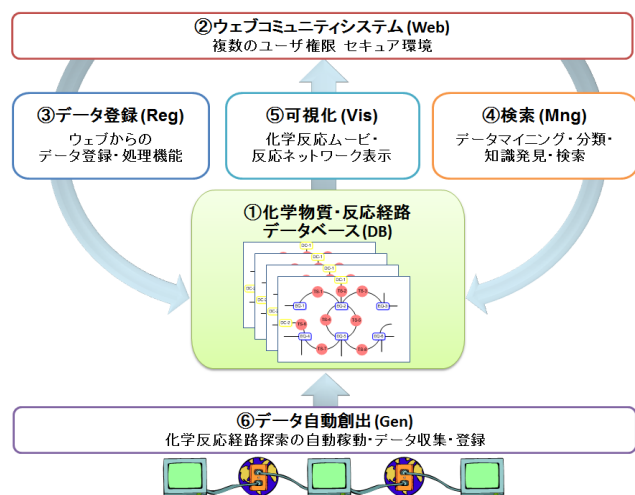


図 1 システムフロー

Figure 1 Block Diagram of the System

3. 化学者による化学反応経路マップの解析

3.1 GRRM からの出力：化学反応経路マップ

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的パラメータである。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。

3.2 化学反応解析の基本要件

化学反応経路を解析する手順や思考過程は、問題とする対象や内容、人によって異なるため、一概に 1 つのスキームを出すことはできない。しかし、どの場合にも必要な基本要件は存在すると考えられる。それらを以下にまとめる。

(1) 注目する属性

①分子構造

各 EQ, TS, DC の分子の構造情報は、化学反応解析にあたって、最も基本的な属性である。分子構造情報は、原子と結合のトポロジカルな関係を示す平面構造(二次元構造)から立体構造(三次元構造)まであり、目的に応じてそれぞれの分子構造情報が参照される。また、化学反応経路を経る過程で分子構造がどのように変化していくかを示す動画も重要である。

②平衡(EQ)構造のポテンシャルエネルギー

EQ 構造のポテンシャルエネルギーは、分子の内部エネルギーであり、その化合物がどの程度安定に存在しうるかを示す尺度である。つまり、エネルギー値が低いほど、現実によく存在しうる可能性が高いことを意味する。また、エネルギーが高い化合物は、高い反応性をもつことから、化学反応設計のキー反応物となるポテンシャルをもつ。そのため、EQ のポテンシャルエネルギーは、化学反応の解析を行う上で必須の基本的な物理化学的数値データである。

③遷移(TS)構造のポテンシャルエネルギー

ある EQ 構造が化学反応により別の EQ 構造へと変化するためには、エネルギーの壁を超えなければならない。

このエネルギー障壁は活性化エネルギーと呼ばれ、TS のポテンシャルエネルギー（遷移状態エネルギー）と反応物（EQ 構造）のポテンシャルエネルギーの差分で表される。1 つの EQ 構造が複数の EQ 構造へ変化する複数の化学反応経路が存在する場合には、最も低い活性化エネルギーを経る化学反応が最も起こりやすい。このため、遷移状態エネルギーは化学反応を解析する上で、重要な物理化学的数値データである。

④化学反応ステップ数

2 つの EQ 構造を結ぶ複数の反応経路のうち、どれが起こりやすいか、あるいはどの反応経路で化学合成の設計を行うかを考える場合には、いくつかの EQ 構造を経ているかの情報、つまり化学反応ステップ数も大事なキー情報となる。化学合成設計を行う場合には、よりステップ数の少ない反応経路の方が、より効率的な合成となるからである。

(2) 反応物を指定した検索

化学反応経路マップの基本的な解析方法の 1 つは、上記の属性に着目して、興味のある反応物から出発した検索と各ノードの詳細情報の解析を行うことである。出発点となる反応物を決める基準は、通常、分子構造、ポテンシャルエネルギー値である。そこから化学反応経路の繋がりを辿り、個々のノードの詳細情報（分子の構造情報、物理化学的パラメータ等）や反応経路に沿って分子構造が変化する様子へと解析を進める。

この検索は、出発点を反応物だとすれば、反応予測に

つながり、反応生成物だとすれば、化学合成設計につながるものである。

(3) 反応物と生成物を指定した検索

化学反応経路マップのもう 1 つの基本的な解析方法は、2 つのノードを指定して、その間を結ぶ全ての化学反応経路の検索と各ノードの詳細情報の解析を行うことである。始点と終点を決める基準は、通常、分子構造であり、ポテンシャルエネルギーも重要なファクターである。2 つのノードを結ぶ可能な全ての化学反応経路のうち、より低い遷移状態エネルギーを経るものが実際に起こりやすいと考えられる。また、反応ステップ数が少ないものは、化学合成の観点から有望な化学反応経路であると考えられる。これらの属性に着目しながら、化学反応経路の解析を行う。

この検索は、化学反応機構の解析や、化学合成設計につながるものである。

4. 化学反応経路の可視化

4.1 化学反応経路ネットワークの全体表示

我々は、まずは上述の基本的な検索と可視化を行うことができるシステムを目指して開発に着手した。現在までに開発した機能について述べる。

可視化システムに GRRM の出力データが読み込まれると、化学反応経路マップの全体図が表示される（図 2）。これにより、化学反応経路マップの大きさや個々の繋がりの様子を見ることができる。ズームイン・アウトの操作が可能であり、全体のどの部分が表示されているかは、ナビゲ

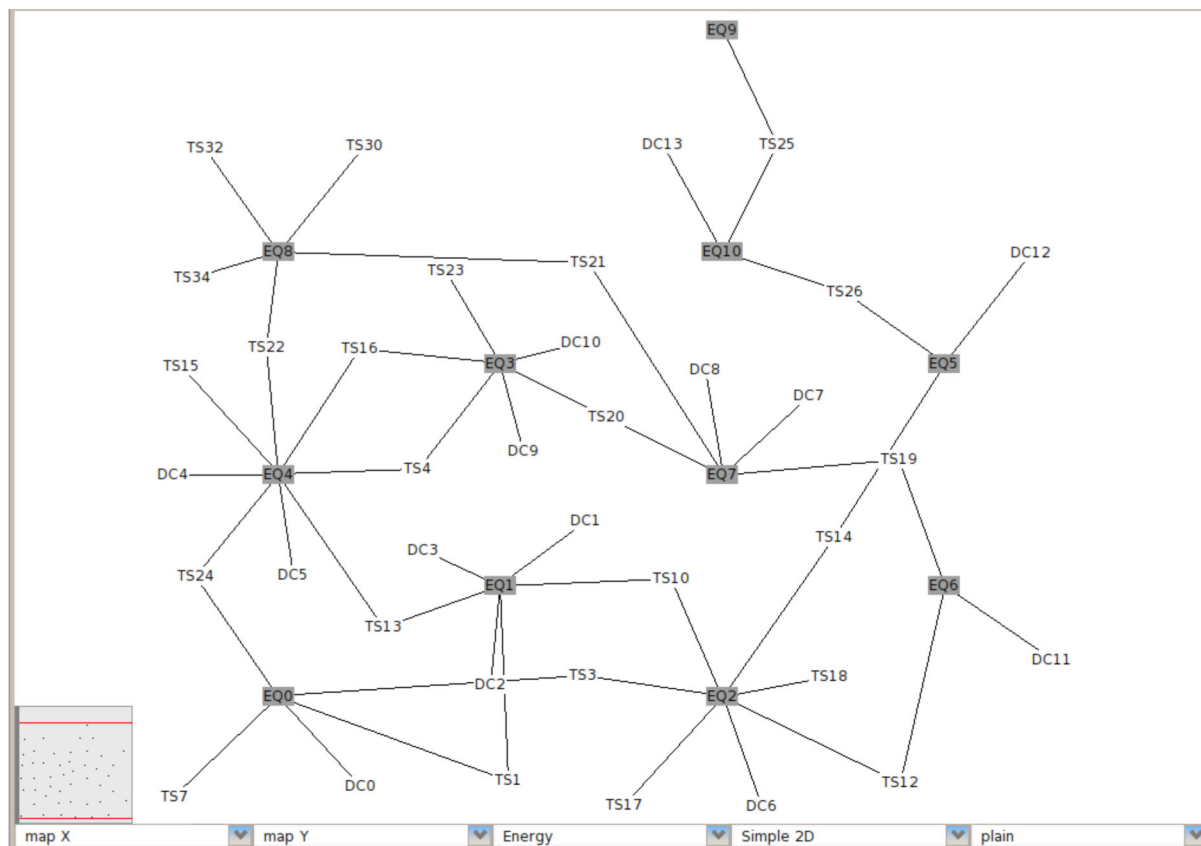


図 2 化学反応経路マップの全体表示 ($\text{C}_2\text{H}_2\text{O}$ についての探索結果)

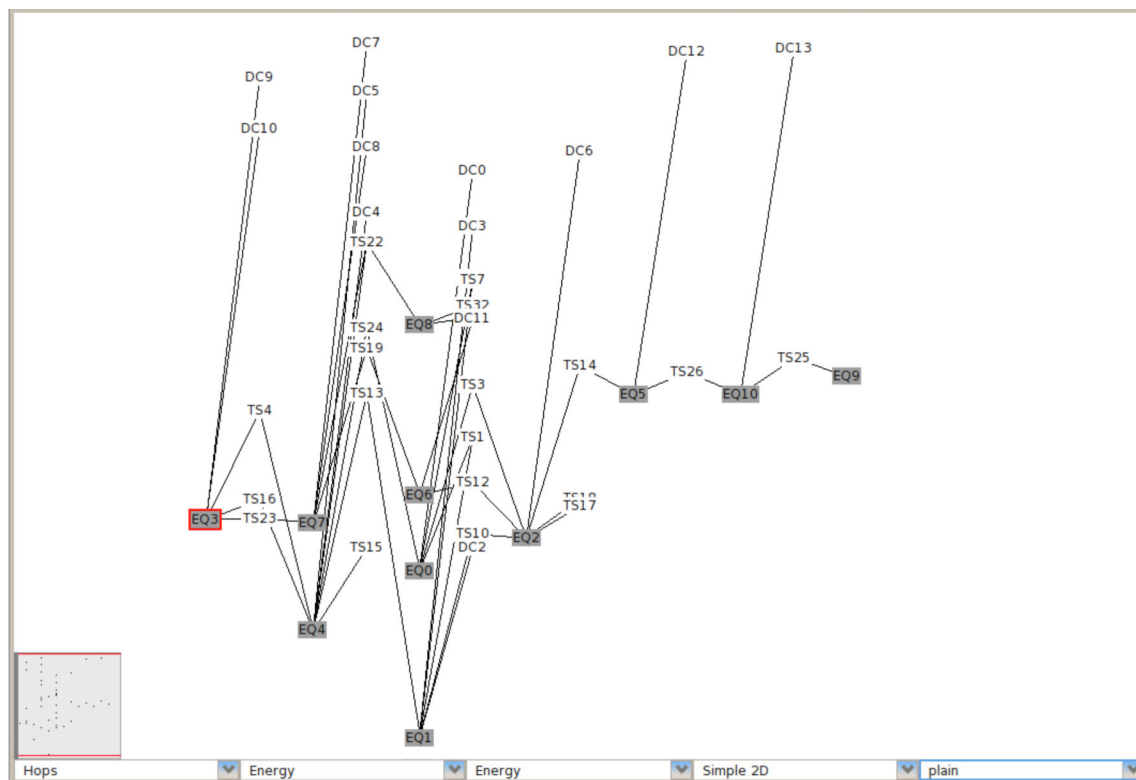
ーションウィンドウにより知ることができる。

4.2 ノードの属性

初期表示では、各ノードには EQ1, EQ2,..., TS1, TS2,..., DC1, DC2,...といった、平衡(EQ)構造、遷移(TS)構造、解離(DC)構造に ID が振られたラベルが表示されている。各ノードは、これ以外に、分子構造情報(原子と結合の個数と種類、原子の座標情報)、ポテンシャルエネルギー値等の、量子化学計算より出力される数値データを属性として有している。

析において一般的に利用される馴染みのあるグラフであり、化学者にとって直感的に理解しやすいものである。

ポテンシャルエネルギーと合わせて重要な属性は分子構造情報である。システムでは、表示するノード属性として幾何学構造(geometry)を選択すると、各ノードの分子の3次元構造を化学分野で頻繁に利用されるボール&スティック表示と呼ばれる分子モデル図で表示する(図4)。これらの3次元構造は化学者にとって極めて重要な可視化であり、これにより、化学者にとって興味のある分子構造を素早く



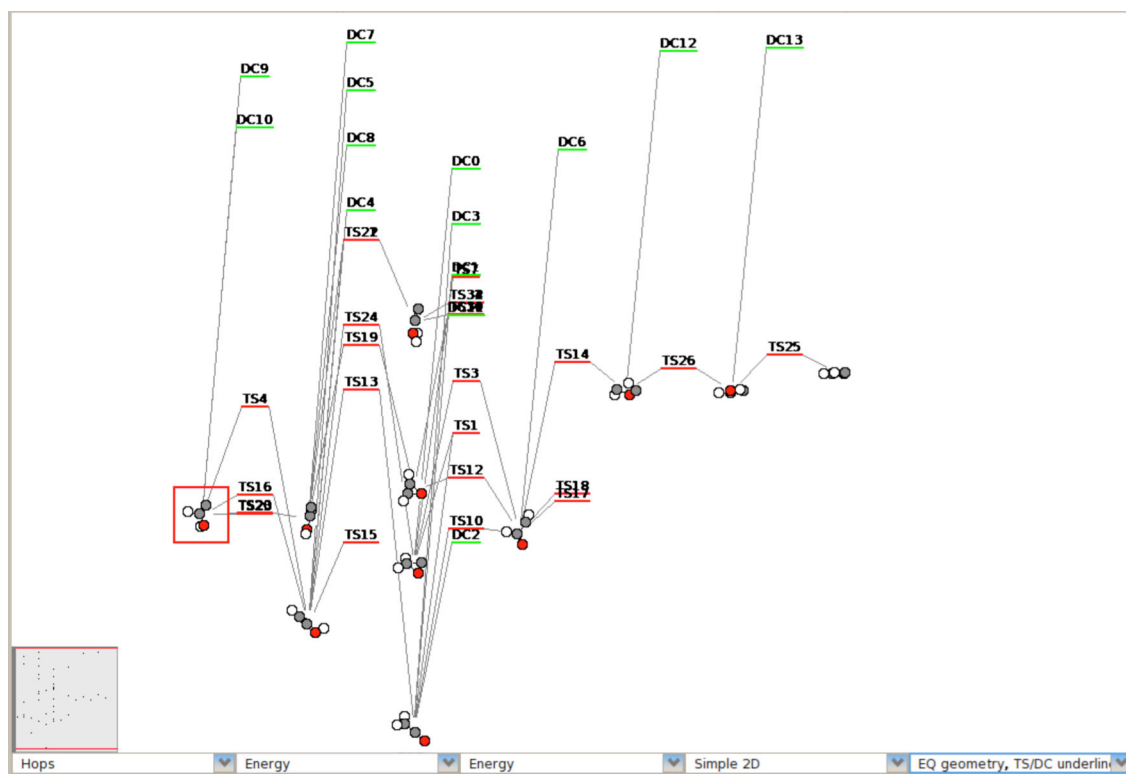


図 4 図3のノード上 (EQ のみ) に3次元分子モデルをボール&スティックで表示した図

Figure 4 The Reaction Route Map with Ball and Sticks Molecular Models in 3D (only for EQs).

5. おわりに

「埋蔵分子」発掘プロジェクトは始まったばかりである。ここに報告した可視化システムをはじめ、データベースシステムや分散型 GRRM との連携など、個々の機能の開発を進めている。これらの独立した機能を順次連結し、図1に示したシステム全体を形作っていく予定である。一方、個々の独立した機能は、単体でも動くものとし、今回報告した可視化システムは、化学者が個人のプラットフォームで利用できるものとして、2014 年春の公開・配布を目指して準備を進めている。

謝辞 本研究は、JSPS 科研費挑戦的萌芽研究 25540017、住友財団 2011 基礎科学研究助成、国立情報学研究所共同研究費、情報・システム研究機構データ中心科学リサーチコミュニティ基盤整備事業の助成を受けて実施した。

参考文献

- 1) URL: <http://molgen.de>
- 2) Ohno, K.; Maeda, S. A Scaled Hypersphere Search Method for the Topography of Reaction Pathways on the Potential Energy Surface, *Chemical Physics Letters*, Vol. 384, pp277-282 (2004).
- 3) Maeda, S.; Ohno, K. *J. Phys. Chem. A* **2005**, *109*, 5742-4753.
<http://office.microsoft.com/ja-jp/>
- 4) Ohno, K.; Maeda, S. *J. Phys. Chem. A* **2006**, *110*, 8933-8941.