

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 17 日現在

機関番号：82657

研究種目：挑戦的萌芽研究

研究期間：2013～2015

課題番号：25540017

研究課題名(和文) 化学物質のビッグデータ革新 埋蔵分子の理論的発掘とデータケミストリの新展開

研究課題名(英文) "Maizo"-Data Chemistry: Molecular Discovery from Chemical Big Data

研究代表者

佐藤 寛子 (Sato, Hiroko)

大学共同利用機関法人情報・システム研究機構(新領域融合研究センター及びライフサイ・新領域融合研究センター) 准教授

研究者番号：50291068

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：量子化学の理論にもとづいて化学反応経路を自動的に探索・蓄積し、そこから新規分子・反応「埋蔵分子」を発掘・発見するためのシステム開発と応用研究を実施した。化学反応経路データ自動創出・蓄積・可視化・検索の一連の基本動作が可能なシステムの核を構築し、可視化・検索のためのソフトウェアRMapViewを蓄積データの一部と合わせてフリーウェアとして公開した。糖分子の配座マップの自動探索と解析に応用し、本システムを利用することで、理論的に検索された化学反応経路データから、効率よく有効な情報を引き出すことができることを示した。さらに、新たな炭素分子ファミリーの存在を理論的に予測した。

研究成果の概要(英文)：The "Maizo"-Data Chemistry project is aimed towards molecular and reaction discovery based on big data of quantum mechanical (QM) global reaction route mappings. We have developed a primary system for automatic calculations and accumulation of global reaction route maps (r-maps) based on the GRRM (Global Reaction Route Mapping) method. We also have developed a system, RMapView, for interactively analyzing r-maps. The RMapView software is free to download together with part of the accumulated r-map data. We have demonstrated that using RMapView makes it possible to effectively derive valuable information from r-maps by applying this system to QM-based automatic conformational analysis of saccharides. Furthermore, we have theoretically predicted the existence of a new carbon family.

研究分野：計算化学、データ科学

キーワード：埋蔵分子 イニング 化学反応経路探索 可視化 量子化学計算 グラフィカルユーザインタフェース データケミストリ ケモインフォマティクス データマ

1. 研究開始当初の背景

化学物質の情報は、アメリカ化学会の一部局である CAS (Chemical Abstracts Service) によって網羅的に編集・蓄積されている。研究開始時点(2013年)の時点で存在が確認されていた化学物質は約 7000 万種類であったが、その後も加速度的に増加を続け、2015年6月には1億種類を突破した。化学物質の文献や分子構造・分光学・物性データや、供給方法は、化学物質を取扱う幅広い産業・学術分野にとって必要不可欠な情報である。化学は、化学情報の蓄積→データ解析と知識の蓄積→物質や供給方法の開拓・創成、を繰り返すことで発展してきた、化学情報と密接に結びついたデータサイエンスの一分野であるといえる。

化学物質の可能性と化学反応性は、量子力学の理論を化学に応用した量子化学によって原理的に解く(予測する)ことが可能である。化学物質とその供給方法は、ポテンシャル曲面上の極小点(反応物・生成物)と鞍点(遷移状態)を結ぶ化学反応経路を探索することによって得られる。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 GRRM (Global Reaction Route Map) と呼ばれる手法により初めて可能となった[引用文献①, ②]。GRRM は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手法であり、ポテンシャル曲面上のすべての極小点(反応物・生成物)が鞍点(遷移状態)を経た化学反応経路で連結された化学反応経路マップ(reaction route map: r-map)を自動的に生成する。種々の化学物質についての GRRM 探索によって、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の数は、存在が確認されている件数を遥かに凌駕することが明らかとなってきた。

近年、情報科学分野の機械学習やデータマイニング手法の発展と人工知能研究への応用研究の広がり、ならびに、コンピュータ性能の飛躍的な発展にともない、バイオインフォマティクスやマテリアルズインフォマティクスなど、自然科学分野においても、データ中心サイエンスの有効性が着目されてきている。研究代表者は、コンピュータケミストリを専門とし、特に、大量のケミカルデータを化学物質や化学反応の設計や予測に活用することを目指す、ケモインフォマティクス・データケミストリの分野で 25 年に及ぶ実績をもち、データサイエンスやバイオ・マテリアルズ・インフォマティクスという用語が生まれる前から、先駆的にデータケミストリに従事してきた。研究代表者は、GRRM という、まだ人類が手にしたことの無い化学反応経路ビッグデータを創出する最新の量子化学手法に着目し、これを情報科学・ケモインフォマティクスの最先端の技術と融合した、以下に記す「埋蔵分子」発掘研究のアイ

デアを着想するに至った。

2. 研究の目的

我々は、理論的には存在しうるがまだ確認されていない化学物質を「埋蔵分子」と名付け、この革新をもたらすことが期待される「埋蔵分子」を GRRM 探索から得られるビッグデータ:r-map 群から発掘することを目指し研究を開始した。具体的には、以下に示す研究項目 1~5 を推進する。このうち、本研究期間(3年間)では、項目 1~3 に注力し、特に、自動データ創出・蓄積(項目 1)、解析・検索(項目 2)、利用(項目 3)の骨組を重点的に開発・公開し、研究期間終了後の機能拡充と応用研究の発展につなげることを目的とした。

(1) GRRM 探索と r-map 群の蓄積(データベース化)

(2) 可視化・解析ソフトウェアの開発

(3) r-map 群からの新規物質・反応の発見

(4) ウェブコミュニティシステムの開発

(5) 機械学習・ケモインフォマティクスの活用による分子・反応設計や新規発見への応用

本研究により、新規化学物質・化学反応の設計と予測を目指すデータケミストリの新たなフェーズを拓くことが期待される。以下に記述するとおり、本研究期間では、このための足がかりとなる基盤を構築することを目的として、具体的な目標と研究計画を設定した。

3. 研究の方法

本研究期間に実施した研究項目 1~3 の方法を以下に記述する。

(1) GRRM 探索と r-map 群の蓄積(データベース化)

種々の組成式・量子化学計算レベルについて GRRM 探索を実施し、得られる r-map 群を蓄積し、データベース化する。データベースには、安定構造、遷移状態構造、解離チャネルの構造、これらをつなぐ化学反応経路とその過程の幾何学構造に加えて、ポテンシャルエネルギー値や種々の物理化学パラメータを格納する。GRRM 探索を自動的に実施し、データを蓄積する仕組みも開発する。

(2) 可視化・解析ソフトウェアの開発

GRRM でポテンシャル曲面の全面探索を行うと一般に巨大な r-map が出力される。たとえば、 C_6H_6 の組成式について SCC-DFTB レベルで探索すると 7,000 件の極小点(安定構造)と 26,229 件の鞍点(遷移状態構造)が得られる[引用文献③]。r-map の規模は、構成原子数の増加により指数関数的に増加する。この巨大な r-map がもつデータ量は、人間の認知サイズを優に超えているため、「埋蔵分子」や「埋蔵反応」の発見のためには、解析のためのソフトウェアツールが不可欠である。

量子化学計算から得られた化学反応経路に含まれる分子構造やエネルギープロファイルなどを解析するためのソフトウェアは種々存在する。しかし、いずれも従来法から得られる小規模データを対象に設計されており、また、ネットワーク型の化学反応経路マップは想定外であるため利用できない。

そこで、ポテンシャル曲面の網羅的探索から得られる巨大な化学反応経路ネットワークをインタラクティブに可視化・解析するためのソフトウェアを開発する。

(3) r-map 群からの新規物質・反応の発見

項目 1 から得られる r-map を項目 2 で開発するソフトウェアで可視化・解析し、新規化学物質や反応経路の発見を目指した応用研究を行う。

本研究期間でシステムの核となる、上記の項目 1~3 の骨組みを開発・公開し、研究期間後の機能拡充と応用研究の発展につなげるとともに、システムの利用の幅を広げるためのウェブコミュニティシステムの開発を行う計画である(研究項目 4)。本研究期間においては、データベース化や解析システム的设计・開発(研究項目 1~2)を進めるにあたり、将来のウェブコミュニティシステム開発を志向し実施する。

本研究期間の研究項目 2 では、人間の認知能力の範囲を超えた巨大なサイズの化学反応経路データの解析をインタラクティブにアシストするものである。一方、この化学反応経路のビッグデータに、いわば人工知能的に挑もうとする研究が、研究項目 5 の、「データマイニング・ケモインフォマティクス」の活用による分子・反応設計や新規発見への応用」である。本研究期間におけるデータベース設計や解析ソフトウェアの設計・開発(研究項目 1~2)の際にも、このために資するデータや機能を考慮する。応用研究(研究項目 3)にあたっては、具体的な化合物や反応の解析を行うことで、このためのアイデアを獲得・蓄積する。

<引用文献>

- ① Ohno, K.; Maeda, S. *Chem. Phys. Lett.* **2004**, *384*, 277-282.
- ② Ohno, K. *Chem. Rec.*, **2016**.
DOI: 10.1002/tcr.201500284
- ③ Tokoyama, H.; Yamakado, H.; Maeda, S.; Ohno, K. *Chem. Lett.* **2014**, *43*, 702-704.

4. 研究成果

本研究期間 3 年間で、r-map データの自動創出→蓄積→可視化・検索の一連の基本動作が可能なシステムの核を構築し、可視化・検索のためのソフトウェア RMapView をフリーウェアとして公開した。蓄積したデータについても順次公開を行ってきている。応用研究として、糖分子の配座マップの自動探索と解析に応用し、本システムを利用することで、理論的に検索された化学反応経路データか

ら、効率よく有効な情報を引き出すことができることを示した。さらに、新たな炭素分子ファミリーの存在を理論的に予測した。当初の計画に沿って順調に研究を遂行し、システムの核となる部分の構築、データの蓄積について、目標を達成することができた。応用研究については、新規物質や反応経路の理論的な発見まで達成することができ、当初の予定よりも大きく前進することができ、本研究期間後の研究の進展につながる成果を得ることができた。各研究成果についての詳細を以下に記す。

(1) GRRM 探索と r-map 群の蓄積(データベース化)

研究開始から現在に至るまで、「小分子から順次 GRRM 探索計算を継続して実施し、r-map を蓄積してきている。これまでの計算では、主に HF 法、密度汎関数法を用いている。蓄積したデータの一部は、RMapView と合わせて公開している。

(2) 可視化・解析ソフトウェアの開発

r-map データをインタラクティブに可視化・解析するためのソフトウェア RMapView を開発した。r-map データから、有効な情報を効率的に引き出すための種々の機能を開発・実装した。また、ソフトウェアの基本設計においては、機械学習による発見やウェブコミュニティシステムにも対応することを考慮した。RMapView の主な機能を以下に示す：

- r-map データ全体図の表示
- エネルギープロファイル図表示
- 化学反応ムービーの表示
- 始点(反応物)から終点(生成物)までの反応経路の列挙と、遷移状態エネルギーと反応ステップ数にもとづくソート
- 分子構造の線形表記コード(SMILES, InChi, CAST)による検索
- 分子構造描画による検索

図 1~3 に、 CH_3NO_2 の r-map の RMapView による解析例を示す。

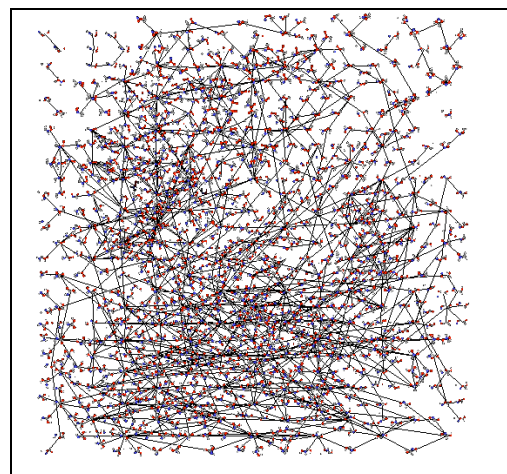


図 1. RMapView による解析例： CH_3NO_2 の r-map の全体図表示

図1はCH₃NO₂のr-mapの全体図表示である。RHF/6-31+G*レベルで全面探索を行った結果である。探索の結果、安定(EQ)構造:169個、遷移状態(TS)構造:616個、解離チャネル(DC)構造:191個からなるr-mapが得られた。

このr-mapを、縦軸をポテンシャルエネルギー値として、最安定構造付近を拡大して表示した結果が図2である。

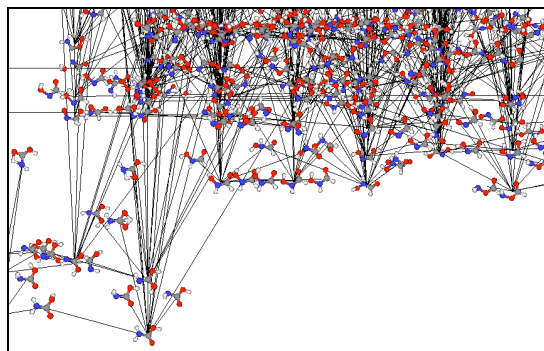


図2. RMapViewerによる解析例: CH₃NO₂のr-mapを、縦軸をポテンシャルエネルギーとして部分的に拡大表示した結果

この最安定構造はcarbamic acidである。そこで、これを出発物質として選択し、生成物質として(aminooxy)methanoneを選択すると、2物質間の化学反応経路が自動的に列挙される(図3)。この例では2つの最低エネルギー経路が計算された(図3の一番したのウィンドウに2本表示されている)。このうちの1つの経路を選択すると、r-map上でこの経路がハイライト表示される。ここで、ムービー表示を選択すると、分子モデル図(一番左上)とエネルギープロファイル図(一番右上)が表示され、分子モデル図上で、エネルギープロファイルと連動した化学反応ムービーが表示される。

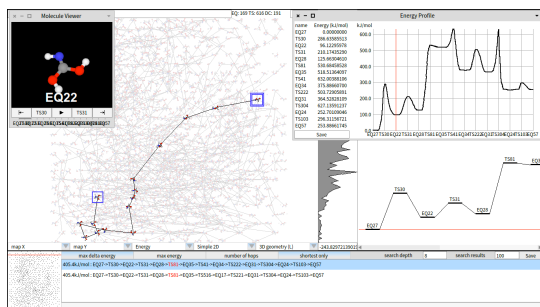


図3. RMapViewerによる解析例: CH₃NO₂のr-mapについて、carbamic acidを出発物質、(aminooxy)methanoneを生成物質として、最安定反応経路を計算した結果

RMapViewerはSmalltalk開発言語により先行版を開発し、2014年7月より、公開サイト:
<http://sourceforge.net/projects/rmapviewer/>
より無償公開を開始した。公開後も新規機能を開発・実装し、バージョンアップを行うとともに、より汎用的な開発言語であるSCALA

への移植も行った。Smalltalk版、SCALA版ともにオープンソースとして近く無償公開の予定である。

(3) r-map群からの新規物質・反応の発見

① 分子の立体配座解析への応用研究

応用研究の1つとして、GRRMによる探索計算とRMapViewerを用いた立体配座解析研究を行った。グルコース等の単糖を対象とし、量子化学計算による立体配座のr-mapを自動的に生成することに成功した。さらに、RMapViewerにより2種の代表的な立体配座間に複数の最安定反応経路が存在することを明らかとした。

得られた結果の一部として、β-D-グルコースの椅子型配座⁴C₁の最安定構造を出発物質とし、もう一種の椅子型配座¹C₄の最安定構造を生成物として計算された3種の最安定反応経路を図4に示した。経路中の最も高い活性化エネルギーを示したステップは、3経路ともに同じであった。図4に、相当する遷移状態構造(半椅子型²H₃配座)とその前後の安定構造(椅子型⁴C₁配座と舟型B_{3,0}配座)を、それぞれ青枠と緑枠で示した。

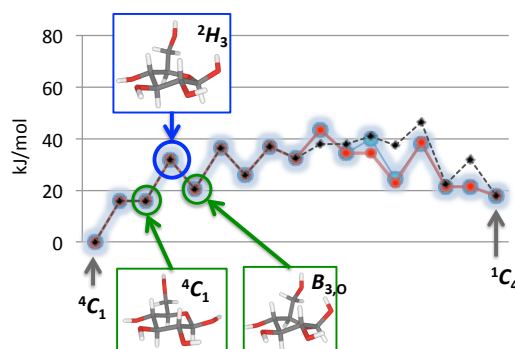


図4. 立体配座解析への応用例: β-D-グルコースの2種の椅子型配座間の最安定反応経路

② 新規炭素分子の発見への応用研究

もう1つの応用研究として、r-mapからの新規炭素分子の存在を理論的に予測する研究を行った。

炭素分子を構成する安定な基本骨格としては5員環や6員環が知られており、グラフェンやナノチューブ、フラーレンなどを構成する。本研究によって見出されたのは、4員環を基本骨格とする炭素分子ファミリーであり、プリズムカーボンと名付けた。これらは、高いポテンシャルエネルギーを持つが、周囲は高いエネルギー障壁で囲まれていることから、一旦生成すれば安定に存在しうる物質であると期待される。発見したプリズムカーボンとそのファミリーの一部を図5に示した。

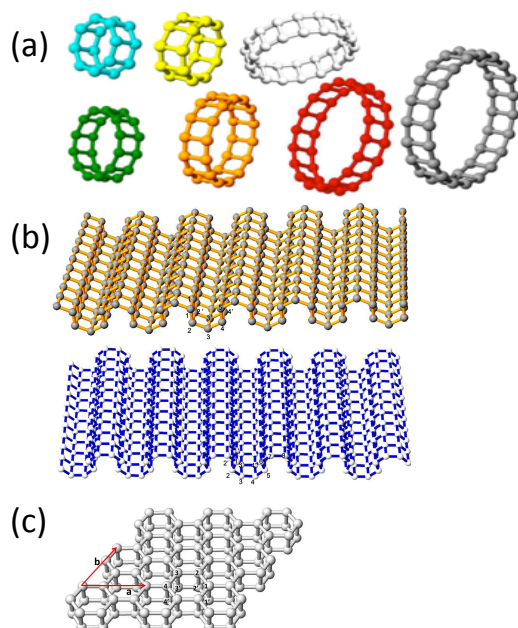


図 5. 理論的に存在が予測された新規炭素分子：プリズムカーボンファミリーの一部。(a) プリズムカーボン、(b) ウェーブ型プリズムカーボン、(c) シート型プリズムカーボン

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

- ① Koichi Ohno, Hiroko Satoh, Takeaki Iwamoto, Hiroaki Tokoyama, Hideo Yamakado, Wavy Carbon: A New Series of Carbon Structures Explored by Quantum Chemical Calculations, *Chemical Physics Letters*, 639 巻, 178–182, 2015, 査読有
doi:10.1016/j.cplett.2015.09.026
- ② Hiroko Satoh, Tomohiro Oda, Kumiyo Nakakoji, Takeaki Uno, Satoru Iwata, Koichi Ohno, “Maizo”-chemistry Project: Towards Molecular- and Reaction Discovery from Quantum Mechanical Global Reaction Route Mappings. *Journal of Computer Chemistry Japan*, 14 巻, 77–79, 2015, 査読有
doi:10.2477/jccj.2015-0048
- ③ Koichi Ohno, Hiroko Satoh, Iwamoto, T. Prism- C_{2n} Carbon Dimer, Trimer, and Nano-Sheets: A Quantum Chemical Study, *Chemical Physics Letters*, 633 巻, 120–125, 2015, 査読有
doi:10.1016/j.cplett.2015.05.024
- ④ Koichi Ohno, Hiroko Satoh, Takeaki Iwamoto, A Prism Carbon Molecule C_{20} , *Chemistry Letters*, 44 巻, 712–714, 2015, 査読有

[学会発表] (計 18 件)

- ① 大野公一, 佐藤寛子, 岩本武明, 時子山宏明, 山門英雄, 新型炭素構造の探索, 第 96 回日本化学会春季年会, 2016 年 3 月 24–27 日, 同志社大学京田辺キャンパス (京都)
- ② 大野公一, 佐藤寛子, 岩本武明, 時子山宏明, 山門英雄, “新炭素単体 Prism Carbon の構造とエネルギー”, 第 9 回分子科学討論会, 2015 年 9 月 16–19 日, 東京工業大学大岡山キャンパス (東京)
- ③ 大野公一, 佐藤寛子, 岩本武明, 時子山宏明, 山門英雄, 多角柱型炭素構造の反応経路と安定性, 化学反応経路探索のニューフロンティア 2015, 2015 年 9 月 15 日, 品川区立総合区民会館 (東京)
- ④ Hiroko Satoh, Tomohiro Oda, Kumiyo Nakakoji, Takeaki Uno, Hiroaki Tanaka, Satoru Iwata, Koichi Ohno, QM-based Data Chemistry: A Strategy for Computational Molecular/Reaction Analysis Based on The Global Reaction Route Mapping. PSI-K 2015 Conference, 6–10 September, 2015, San Sebastian (Spain)
- ⑤ 大野公一, 佐藤寛子, 岩本武明, 大域的ポテンシャル表面の量子化学自動探索に基づく埋蔵分子の発掘, 日本コンピュータ化学会 2015 春季年会, 2015 年 5 月 28–29 日, 東京工業大学大岡山キャンパス (東京)
- ⑥ 佐藤寛子, 小田朋宏, 中小路久美代, 宇野毅明, 岩田覚, 大野公一, 「埋蔵分子」発掘プロジェクト: 量子化学に基づくグローバル反応マップデータからの発見, 日本コンピュータ化学会 2015 春季年会, 2015 年 5 月 28–29 日, 東京工業大学大岡山キャンパス (東京)
- ⑦ 大野公一, 佐藤寛子, 岩本武明, 炭素の新単体 (Prism- C_{2n}) の 2 次元周期構造の探索, 第 18 回理論化学討論会, 2015 年 5 月 20–22 日, 大阪大学豊中キャンパス (大阪)
- ⑧ 佐藤寛子, 小田朋宏, 中小路久美代, 宇野毅明, 岩田覚, 大野公一, データケミストリ: 理論化学データを基盤とするインフォマティクスによる新規物質・反応経路の発見, 第 18 回理論化学討論会, 2015 年 5 月 20–22 日, 大阪大学豊中キャンパス (大阪)
- ⑨ 大野公一, 佐藤寛子, 岩本武明, GRRM 法による超原子価化合物候補の自動探索, 第 95 回日本化学会春季年会, 2015 年 3 月 26–29 日, 日本大学理工学部船橋キャンパス (千葉)
- ⑩ 大野公一, 佐藤寛子, 岩本武明, GRRM 法による未知化学の探索: H_4C_4 , 第 12

回京都大学福井謙一記念研究センター
シンポジウム-諸熊奎治先生傘寿記念
-,2015年1月23日,京都大学福井謙一
記念研究センター(京都)

- ⑪ 佐藤寛子,未知の埋蔵分子発掘プロジェクト,講演会「GRRMで拓く化学のニューフロンティア」,2014年11月30日,学士会館(東京),招待講演
- ⑫ 佐藤寛子,「埋蔵分子」発掘プロジェクトの展開について,JST さきがけ研究 21「情報と知」懇話会,2014年10月18日,東京大学本郷キャンパス(東京)
- ⑬ Hiroko Satoh, Tomohiro Oda, Kumiyo, Nakakoji, Takeaki Uno, Hiroaki Tanaka, Satoru Iwata, Hans Peter Luethi, Koichi Ohno, Chemoinformatics Meets Quantum Chemistry: A Strategy for Computational Molecular/Reaction Analysis Based on The Global Reaction Route Maps. 2014 Fall Meeting of the Swiss Chemical Society, 11th September, 2014, Zurich (Switzerland)
- ⑭ 佐藤寛子,小田朋宏,GRRMによる埋蔵分子発掘プロジェクトと化学反応経路マップの可視化,GRRMチュートリアル2014,2014年6月24日,八重洲あすか会議室(東京),招待講演
- ⑮ Hiroko Satoh, QM-Based Data Chemistry: Chemoinformatics Meets Quantum Chemistry, Asian International Symposium, 日本化学会第94春季年会,2014年3月27-30日,名古屋大学東山キャンパス(名古屋),招待講演
- ⑯ 佐藤寛子,小田朋宏,中小路久美代,宇野毅明,田中宏明,岩田覚,大野公一,「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,インタラクシオン 2014,2014年2月27日-3月1日,日本科学未来館(東京)
- ⑰ 佐藤寛子,小田朋宏,中小路久美代,大野公一,化学反応経路の全面探索の可視化とデータマイニングによる発見への取組み,化学反応経路探索のニューフロンティア 2013,2013年9月27日,京都大学福井謙一記念研究センター(京都)
- ⑱ 佐藤寛子,Stefano Borini,小田朋宏,中小路久美代,大野公一,理論化学-データケミストリ:超球面探索法より得られる化学反応経路データの蓄積と活用,第7回分子科学討論会,2013年9月24-27日,京都テルサ(京都)

[図書](計0件)

[産業財産権]

○出願状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

[その他]

ホームページ等

- ① 佐藤寛子,小田朋宏,公開ソフトウェア RMapView: A Tool for Visualization and Analysis of Global Reaction Route Maps,2014年7月公開
<http://sourceforge.net/projects/rmapviewer/>

6. 研究組織

(1)研究代表者

佐藤 寛子(Hiroko Satoh)
大学共同利用機関法人情報・システム研究
機構・新領域融合研究センター・特任准教授
研究者番号:50291068

(2)研究分担者

宇野 毅明(Takeaki Uno)
大学共同利用機関法人情報・システム研究
機構国立情報学研究所・情報学プリンシ
プル研究系・教授
研究者番号:00302977

新井 紀子(Noriko Arai)
大学共同利用機関法人情報・システム研究
機構国立情報学研究所・情報社会相関研究
系・教授
研究者番号:40264931

(3)連携研究者

()

研究者番号: