

戦略的創造研究推進事業 CREST
研究領域「科学的発見・社会的課題解決に向けた
各分野のビッグデータ利活用推進のための次世代
アプリケーション技術の創出・高度化」
研究課題「医薬品創薬から製造までのビッグデ
ータからの知識創出基盤の確立」

研究終了報告書

研究期間 2013年 6月～2019年 3月

研究代表者：船津 公人
(東京大学大学院工学系研究科 教授)

(1) 実施概要

化学分野では、何を作るかを考える際に、それをどう作るかを同時に考えることが強く求められる。医薬品開発でも同様である。本チームでは、創薬から製造までに関わる様々なデータをそれぞれの段階で求められる必要な知識形態に整え、この課題に取り組んできた。創薬にとって必要となる大規模な仮想ライブラリを合成経路情報や様々な水溶解度などの物性情報まで伴って創出格納し、そうして格納した仮想ライブラリから創薬の候補となる種構造群をその合成経路素案とともに取り出す。取り出された構造の合成経路の妥当性の評価を迅速に行い、実際に合成装置に実装する際の反応条件の最適化を行う。また、合成装置の安定した運転を確保するためにソフトセンサーによる監視と制御を行う。この際にことを円滑に操作できるように連携実証プラットフォームを構築した。つまり、仮想ライブラリ中のすべての構造に対して、化合物-タンパクの相互作用モデルを使ってアッセイ評価をあらかじめ行う、目的のアッセイを考えた際にそれに相当する構造群を仮想ライブラリ方から取り出す。取り出された構造群についてすでに合成経路の素案が伴っている。これは仮想ライブラリ中の構造生成の際に、反応予測、逆合成経路設計を模した形で構造を生成していることからこれが可能となる。この合成経路素案に対して遷移状態データベースを利用した第一原理計算を行うことで、提案された合成経路の妥当性の迅速な評価が行えるようになっている。この際に合成装置からの条件として、使用溶媒や副反応の許容範囲、合成原料の溶解性などの物性に関する要求事項が出てくるが、これは仮想ライブラリ探索の条件となったり、反応経路評価の際の条件となったりする。このような各機能の連携が取り易い連携実証プラットフォーム構築を進めてきた。

以下順に各チームの成果の概要を述べる

(京都大学 奥野グループ)

世界中で公表されているデータベース等からバイオ関連情報を保有する化合物データの網羅的収集を行い、これらをソースデータとして、ケミカル情報とバイオ情報を統合化するデータの構造化と高速処理アルゴリズムの開発を行なった。また、ケミカル情報とバイオ情報との相互作用解析・予測を実現する数理的モデル(統計・確率モデル)を開発した。さらに、構築したケミカル-バイオ相互作用ビッグデータに数理的モデルを適用することにより、新規化合物に対する標的タンパク予測と活性予測や、創薬指針となる創薬標的タンパク質の同定と医薬品候補化学構造の抽出を行なった。この考え方の妥当性を証明するために、実験研究者とともに実証実験も行った。

(国立研究開発法人理化学研究所 泰地グループ)

泰地グループでは、医薬品候補となりうる仮想的な大規模化学構造データを創出し、ビッグデータとしてチーム全体と共有してきた。また、化合物情報(化学構造、物性値、記述子、および合成経路情報など)から実用的な時間内で有用情報を検索する基本技術の開発を行い、チーム内外における有用情報の要求に即応するために必須となる技術を開発してきた。また、各グループが作成したデータやモデルの効果的な連携が図れるように、連携プラットフォームの構築も進めてきた。

(山口大学 堀グループ)

薬物候補化合物とその合成経路について、反応の遷移状態の有無によるその合成経路の妥当性、あるいは反応障壁の高さを評価して反応が進行するかどうかを、どのような副反応が起こるかについて定量的かつ実用的な時間内で判断するシステムを開発してきた。

(船津グループ)

化学製造装置に実装された化学反応が目的の製品純度できちんと合成されることをオンライン、リアルタイムの監視し、設定値から外れていくようであれば、それを迅速に制御して行かなければならない。ただ、ソフトセンサーはプロセスの状態変化によって、予測値が実際の値と会わなくなってくるために、ソフトセンサーの更新を行う必要があった。しかし、これは大きな負担が伴うために、この更新を自動で行う仕組みを開発した。またソフトセンサーの予測性能に影響するプラント運転データベースのメンテナンス法も開発してきた。

何を作るかを考える際に、化学製造装置からの制約条件として、使用溶媒、温度、圧力、拡

散律速などの制約が必然的にかかってくる。これらの制約条件を構造探索と合成経路検討の正の条件とすることが求められる。泰地グループで開発している、連携実証プラットフォーム上で、この制約条件をやり取りする仕組みを実現している。

(北海道大学 杉本グループ)

ビッグデータ応用研究領域の成果が、個別のプロジェクトの成果の単なる羅列に終わることなく、共通の応用技術の研究開発の促進に貢献できるように、本領域の成果を、サンプル・データセットも含めて、誰もが個々の分野における知識発見のためのデータ分析過程を体験できるような体験型ポータルとして公開することを目指し、船津プロジェクト、三好プロジェクト、吉田プロジェクト、越村プロジェクト、西浦プロジェクトの5プロジェクトの体験型ポータルを開発し、角田プロジェクトと平藤プロジェクトの体験型ポータル開発に着手した。

(2) 顕著な成果

<優れた基礎研究としての成果>

(東京大学 船津グループ)

1. Hiromasa Kaneko, Koji Muteki and Kimito Funatsu, "Improvement of Iterative Optimization Technology (for Process Analytical Technology Calibration-Free/Minimum Approach) with Dimensionality Reduction and Wavelength Selection of Spectra", *Chemometrics and Intelligent Laboratory Systems*, 147, 176-184, 2015.

概要: 従来の calibration model によるプロセス状態の監視とは異なり、calibration-free でプロセス状態を監視する手法を開発し、製剤連続生産プロセスにおけるコスト削減を実現した。アメリカ・ファイザーとの共同研究成果である。

2. Matheus de Souza Escobar, Hiromasa Kaneko and Kimito Funatsu, "On Generative Topographic Mapping and Graph Theory Combined Approach for Unsupervised Non-linear Data Visualization and Fault Identification", *Computers & Chemical Engineering*, Volume 98, 113-127, 2017

概要: 化学プラント運転中の異常発見のための可視化ツールの開発を行った。Generative Topographic Mapping(GTM)を利用し過去のプラント運転データの教師無し学習を通して得られたプロセス変化のクラスタを表わしたマップに現在の運転データを投影し、現在の運転データがどのような異常に向かっているかどうかの判断を行える。

3. Hiromasa Kaneko and Kimito Funatsu, "Adaptive Database Management Based on the Database Monitoring Index for Long-term Use of Adaptive Soft Sensors", *Chemometrics and Intelligent Laboratory Systems*, 146, 179-185, 2015

概要: ソフトセンサーモデルの精度を維持するには、ソフトセンサー構築に使われるデータの格納されているデータベース中のデータの多様性が会議となる。このデータベース中のデータの多様性を、データ数を一定としながら維持するためのインデックスの開発を行い成功した。このインデックスを用いてメンテナンスされたデータベースを利用したソフトセンサーの精度の検証実験を行ったところ、目的通りの成果を維持することが確認できた。

(京都大学 奥野グループ)

1. Masatoshi Hamanaka, Kei Taneishi, Hiroaki Iwata, Jun Ye, Jianguo Pei, Jinlong Hou, and Yasushi Okuno. "CGBVS - DNN: Prediction of Compound - protein Interactions Based on Deep Learning." *Molecular Informatics*, 2016

概要: 我々は、サポートベクターマシン(SVM)を用いたタンパク質と化合物の相互作用の予測手法(CGBVS)を提案した。しかし、CGBVSでは、計算時間とコンピュータメモリが指数関数的に増加するため、100万を超えるデータセットでは計算不可能となる。本論文では、Deep Neural Networksを用いた手法(CGBVS-DNN)を提案した。同じデータセットを用いた性能評価ではCGBVS-DNNはCGBVSを凌駕し、さらに400万のデータセットで学習した

CGBVS-DNN では 98.2%の精度を算出した。

2. Brown, J.B., Nakatsui, M., Okuno, Y.,* “Constructing a foundational platform driven by Japan’s K supercomputer for next-generation drug design” *Molecular Informatics* 33, 732-741, 2014

概要：スーパーコンピューター「京」を活用した世界最大規模の高速・超並列バーチャルスクリーニング、および大規模分子動力学シミュレーションを活用したタンパク質-化合物結合親和性の高精度予測について述べている。190 億ペアの化合物-タンパク質相互作用計算を行い、実験結果と比較すると実際の結合パターンと類似するパターンが得られた。また、分子動力学シミュレーション計算では、キナーゼやGPCRなどのタンパク質とそれに結合するとわかっている複数の化合物での検証により、計算値が実験値と相関のある結果を得た。

(理化学研究所 泰地グループ)

概要：論文による成果は特にないが、各グループの成果を連携して活用できる連携プラットフォームの設計し、容易に情報およびデータの活用ができる仕組みを開発した。このような仕組みは、創薬から製造までを対象としたものとしては国内外にはないと思われる。今後この内容を論文としてまとめたいと考えている。

(山口大学 堀グループ)

概要：理論計算を用いた反応解析は、実験結果を説明するために使われてきた。本研究では、これまで実験が行われていない反応に関して遷移状態の存在やその活性化自由エネルギー、反応自由エネルギーを計算することで、これまで試みられていない新しい化合物の合成反応の可能性を明らかにすることを目的とした、発想を逆転させたユニークな研究となっている。今後この内容を論文としてまとめたいと考えている。

(北海道大学 杉本グループ)

特になし

< 科学技術イノベーションに大きく寄与する成果 >

(東京大学 船津グループ)

日本学術振興会143委員会プロセスシステム工学のワークショップ No.32「ソフトセンサー実装」

概要：ワークショップ No.32では、ソフトセンサーを簡便に構築するツールと、そこで検討されたソフトセンサーを化学プラントなどに実装・運用するオンラインツールを企業参加者とともに開発をしている。これらは開発後、参加各企業で利用されるが、他の企業に対してもソフトウェアメンテナンスを目的に有償で頒布する予定である。

(京都大学 奥野グループ)

ライフインテリジェンス コンソーシアム(LINC)

概要：製薬企業のニーズと IT 企業の AI 技術を結び付け、ライフサイエンス分野における AI 技術を開発・活用するためのコンソーシアムの枠組みにより、本プロジェクトで開発した技術を産業や社会への展開・実装がスムーズに進むことが期待できる。

(国立研究開発法人理化学研究所 泰地グループ)

概要：大規模な仮想的化学構造データを創出し扱うためのソフトウェア基盤として、化学反応のトランスフォーム、構造変換エンジン、創出した構造のデータベースへの変換、データベースの検索システムを一体として整備した。これにより、製薬企業等が試用できる体制の基盤を整えた。

(山口大学 堀グループ)

概要: 量子化学計算結果に関するデータベース(DB)は多数存在するが、反応機構の情報を含む化学反応に特化するとともに、そのデータを新たな反応解析に利用できるプログラムとともに提供しているDB(TSDB)はほとんど見られない。さらに構築したTSDBシステムは、新たな合成ターゲットと反応物のどちらかまたは両方が類似する反応を見出す機能を有しており、その検索結果は反応の可否を評価する理論計算の時間短縮に貢献する。

(北海道大学 杉本グループ)

概要: 論文による成果は特にないが、本領域の成果を、サンプル・データセットも含めて、誰もが個々の分野における知識発見のためのデータ分析過程を体験できるような体験型ポータルとして公開することを目指し、船津プロジェクト、三好プロジェクト、吉田プロジェクト、越村プロジェクト、西浦プロジェクトの5プロジェクトの体験型ポータルを開発し、角田プロジェクトと平藤プロジェクトの体験型ポータル開発に着手した。

<代表的な論文>

(東京大学 船津グループ)

1. Hiromasa Kaneko, Koji Muteki and Kimito Funatsu, "Improvement of Iterative Optimization Technology (for Process Analytical Technology Calibration-Free/Minimum Approach) with Dimensionality Reduction and Wavelength Selection of Spectra", *Chemometrics and Intelligent Laboratory Systems*, 147, 176-184, 2015.
2. Matheus de Souza Escobar, Hiromasa Kaneko and Kimito Funatsu, "On Generative Topographic Mapping and Graph Theory Combined Approach for Unsupervised Non-linear Data Visualization and Fault Identification", *Computers & Chemical Engineering*, Volume 98, 113-127, 2017
3. Hiromasa Kaneko and Kimito Funatsu, "Adaptive Database Management Based on the Database Monitoring Index for Long-term Use of Adaptive Soft Sensors", *Chemometrics and Intelligent Laboratory Systems*, 146, 179-185, 2015

(京都大学 奥野グループ)

1. Masatoshi Hamanaka, Kei Taneishi, Hiroaki Iwata, Jun Ye, Jianguo Pei, Jinlong Hou, and Yasushi Okuno. "CGBVS - DNN: Prediction of Compound - protein Interactions Based on Deep Learning." *Molecular Informatics*, 2016
2. Brown, J.B., Nakatsui, M., Okuno, Y.,* "Constructing a foundational platform driven by Japan's K supercomputer for next-generation drug design" *Molecular Informatics* 33, 732-741, 2014

(理化学研究所 泰地グループ)

特になし

(山口大学 堀グループ)

1. K. Hori, M. Sumimoto, T. Murafuji, "Quantum chemistry-assisted synthesis route development ." *AIP Conf. Proc.* 2015, 1702, 090019.

(北海道大学 杉本グループ)

特になし

(1) 研究チームの体制について

① 「船津」グループ

研究代表者: 船津 公人 (東京大学大学院工学系研究科 教授)

研究項目

- ・製造プラントの安定運転・リスク事前管理・品質安定化のための知識抽出
- ・製造プラントにおける大規模運転データベースの管理手法の開発
- ・プラント運転モニタリングのための自動的モデル構築システムの開発
- ・モデルの自動的メンテナンス手法の開発
- ・運転監視・プロセス制御のための知識抽出

② 「奥野」グループ

主たる共同研究者: 奥野 恭史 (京都大学大学院医学研究科 教授)

研究項目

- ・ケミカル情報とバイオ情報の統合化と高速処理を可能にするデータ構造とアルゴリズムの開発
- ・ケミカル情報とバイオ情報の相互作用ビッグデータ解析を可能にする数理的モデルの開発

③ 「泰地」グループ

主たる共同研究者: 泰地 真弘人 (理化学研究所生命機能科学研究センター計算分子設計研究チーム チームリーダー)

研究項目

- ・大規模仮想化合物ライブラリの高度化
- ・仮想大規模ライブラリの拡充
- ・超大規模ライブラリからの有用情報検索技術の開発
- ・超大規模仮想ライブラリのコンテンツ可視化技術の開発
- ・グループ間連携実証プラットフォームの開発

④ 「堀」グループ

主たる共同研究者: 堀憲次 (山口大学大学院創成科学研究科 教授)

研究項目

- ・TSDB/QMRDB の構築と遷移状態データの充実
- ・DB データ自動登録システムの実装
- ・置換基法を用いた TS 探索のための基本構造の開発と計算
- ・TS 初期構造の作成のためのアルゴリズムの開発と実装
- ・合成経路可能性の判定法の開発と実装
- ・合成経路スクリーニングクラウドシステムの構築と拡充・整備
- ・奥野・泰地グループによる提案された薬物候補化合物合成反応の解析

⑤ 「杉本」グループ

主たる共同研究者: 杉本 雅則 (北海道大学大学院情報科学研究科 教授)

研究項目

- ・ビッグデータ応用領域研究成果体験型ポータルの開発

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

(東京大学 船津グループ)

- ・日本学術振興会143委員会(プロセスシステム工学)において、平成28年10月からソフトセンサーを実装するためのワークショップを化学工業系企業、プラント企業などの20社と立ち上げた。

- いくつかの化学工業会社とは実際に個別のプラントの特性の応じたソフトセンサーの新規開発を進めている。

(京都大学 奥野グループ)

- アカデミアとの連携について、
化合物とタンパク質との相互作用予測の開発について、文部科学省「ポスト「京」で重点的に取り組むべき社会的・科学的課題に関するアプリケーション開発・研究開発」重点課題における創薬課題チーム（重点課題1）と連携し、当該研究成果のスーパーコンピュータ実装の検討を行っている。
- 産業界との連携について、
Intel社との共同開発で、Deep LearningのフレームワークTheanoのCPU最適化を行っている。さらに、製薬会社22社が所属するKBDDコンソーシアム（“K” supercomputer-based drug discovery project by Biogrid pharma consortium）のビッグデータ創薬ワーキンググループへの当該研究成果の公開を行うとともに、製薬現場での利用に関する検討を行っている。

(国立研究開発法人理化学研究所 泰地グループ)

- JST さきがけ研究者の田部井博士(現理研 AIP)との共同により大規模仮想化合物ライブラリを対象とした高速な類似度検索サービスを開発実装した。JSTフェアヘポスターを出展し多くの方々と交流する機会をいただいた。また、大規模仮想化合物ライブラリの質の向上を目的とした共同研究を民間製薬企業の研究グループとの間で行った。

(山口大学 堀グループ)

- 本グループで提案された化合物については、実際の合成とそのバイオアッセイを行うことを山口大学山崎教授グループとの共同研究で行うこととなっている。

(北海道大学 杉本グループ)

特になし