

プロジェクト名(タイトル):

創薬プロセス効率化を目指した機械学習のための学習データの作成

利用者氏名:

○ 大田 雅照 (1)、高野 浩顕 (1)、千葉 峻太郎 (2)、池口 満徳 (2)、吉留 崇 (2)、津田 和実 (1)、下道 修 (1)、小山 秀 (1)、宮口 郁子 (2)、鹿島 亜季子 (2)、佐藤 美和 (2)、中川 寛之 (3)、小甲 裕一 (3)、馬 彪(3)、井阪 悠太(1)

理研における所属研究室名:

- (1) 計算科学研究センター HPC/AI 駆動型医薬プラットフォーム AI 創薬連携基盤ユニット
- (2) 計算科学研究センター HPC/AI 駆動型医薬プラットフォーム 分子デザイン計算知能ユニット
- (3) 計算科学研究センター HPC/AI 駆動型医薬プラットフォーム バイオメディカル計算知能ユニット

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

医薬品開発のためには多数のプロセスが存在し、承認薬はそのすべてを通過して初めて誕生する。承認までの10～20億US\$とも見積もられているコストと10年以上という開発期間を削減するための手段として、各プロセスにおいて深層学習を含む機械学習手法およびシミュレーションを利用することは検討の価値がある。本課題では、創薬開発のプロセスでの人工知能(Artificial Intelligence, AI)技術応用の可能性を調べるため、学習に必要となるデータの作成を実施している。

特に、本課題では「医薬品候補化合物の特性予測のための深層学習モデル」、「リガンド結合によるタンパク質周囲の水の置き換えの3D-RISMによる解析」、「タンパク質構造予測」などに注目し、量子化学計算、分子動力学計算、3D-RISM法、ホモロジーモデリング、電子密度計算などによってデータ作成および方法論開発を実施している。

本課題のメンバーは、ライフサイエンス分野でAI技術とビッグデータ利用を推進するコンソーシアム(ライフインテリジェンスコンソーシアム(LINC))、代表:奥野恭史、事務局:京都大学大学院医学研究科人間健康科学系専攻ビッグデータ医科学分野、理化学研究所計算科学研究センターHPC/AI駆動型医薬プラットフォーム)に所属し、各機関と連携の上で課題を実施している。

2. 具体的な利用内容、計算方法

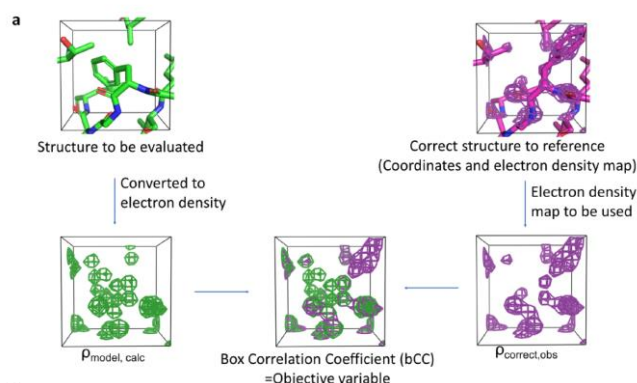
2-1 Machine learning to estimate the local quality of protein crystal structures

低解像度電子密度マップは、タンパク質構造の決定とその利用に大きな障害をもたらす可能性がある。我々は、

電子密度マップに基づくタンパク質局所構造の品質評価を、深層学習法の一つである3D-Convolutional Neural Networkを用いて実施する新しい方法QAEmapを開発した。

本方法の開発においては、HOKUSAIを用いて、Protein Data Bank (PDB)より入手した22個の1.5Å分解能以上のタンパク質構造を基に2.4Åおよび3.0Å分解能のモデル構造を作成し、そのモデル構造に基づき、AI用学習データとして低分解能電子密度を作成した。これらの座標と構造因子から電子密度化したものを記述子とし、高分解能の構造との相関を目的変数として、3D-CNNを用いて、アミノ酸単位で構造評価を行った。

詳細は添付論文(Sci Rep 11, 23599 (2021). <https://doi.org/10.1038/s41598-021-02948-y>)を参照のこと。



2-2 サイクリックペプチド AI 力場開発

サイクリックペプチドは創薬における新規モダリティ(分子形態)として注目されている。しかしながら、サイクリックペプチドは構造的柔軟度が高く、そのコンフォメーションは多様であり、N-Meアミノ酸や側鎖に天然型アミノ酸では存在しない元素を有するなどの多種多様な非天然型アミノ酸も用いられるため、従来の分子力場法では、そのエネル

ギーを精度良く評価することができない。そこで AI 技術を用いて、量子化学レベルの高精度のエネルギー計算を瞬時に行うシステムを開発している。

方法: サイクリックペプチドのシーケンス(非天然型アミノ酸を含む)と各アミノ酸の2次元構造情報から、立体構造生成システムを用いて3次元構造を作成する。作成された3次元構造を基にソフトウェア OMEGA2 (OpenEye)を用いてサイクリックペプチドの多様なコンフォメーションを発生させる。HOKUSAI を用いて、これらの多様なコンフォメーションそれぞれについて、量子化学計算法であるFMO法により、そのエネルギー値を計算する。現在は1サイクリックペプチドあたり200コンフォメーションのFMO計算を実施している。この結果得られたFMOエネルギー値と、サイクリックペプチドの3次元構造座標から深層学習法を用いてAI力場を構築する。

2-3 Exon Skipping データベース開発および AI 予測モデル開発

Exon Skipping を用いた新規治療薬は新規モダリティーの成功事例として注目されている。標的となる核酸シーケンスと、その Exon Skipping 活性については、データベースが構築されていない状況であった。核酸シーケンスとその Exon Skipping 活性について、主に特許より、そのデータ収集、curation を実施した。HOKUSAI を利用し、収集したデータを用いて機械学習法により Exon Skipping 活性を予測するシステムを構築した。

詳細は添付論文(Nucleic Acids Res 49, 2021, <https://doi.org/10.1093/nar/gkab442>)を参照のこと。

2-4 3D-RISM法により計算されたタンパク質水和状態を高速に推算するAIシステムの開発

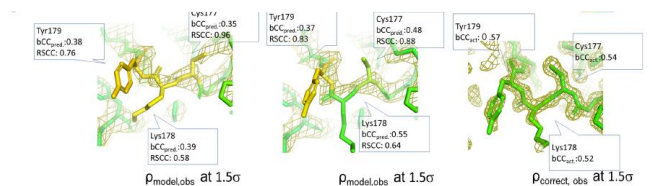
我々は3706個のタンパク質について3D-RISM法を用い、その水和状態を網羅的に計算し、その水和状態が、低分子リガンドのタンパク質との相互作用、結合ポーズなどに深く関わっていることを明らかにしてきた(J. Comput. Chem., 2020, 41, 2406-2419)。3D-RISM法を用いた水和状態情報は有用なものであるが、その計算にはHOKUSAIを利用したとしても1タンパク質あたり約2時間程度必要となる。そこでAI技術を用いて、タンパク質立体構造情報から、その水和状態を迅速に推算するシステムを開発することを試みた。既にHOKUSAIを用いて3D-RISM法により計算した水和状態を、3D-CNNを用いて学習させた。(現在、論文執筆中)

筆中)

3. 結果

3-1 Machine learning to estimate the local quality of protein crystal structures

3D-CNNを用いることにより、低分子電子密度マップからタンパク質局所立体構造を評価するAIシステムの開発に成功した。



詳細は添付論文(Sci Rep 11, 23599 (2021), <https://doi.org/10.1038/s41598-021-02948-y>)を参照。

3-2 サイクリックペプチド AI 力場開発

HOKUSAI上にHOKUSAI用FMO計算ソフトウェアを実装した。また、HOKUSAIを用いて、1サイクリックペプチドでFMO計算を実施し、その実行速度を確認した。

3-3 Exon Skipping データベース開発および AI 予測モデル開発

Exon Skippingの標的となる核酸シーケンスとその Exon Skipping 活性についての、データベースを構築し、公開した。

また、機械学習法を用いて核酸シーケンスより Exon Skipping 活性を予測するシステムを構築した。

詳細は添付論文(Nucleic Acids Res 49, 2021, <https://doi.org/10.1093/nar/gkab442>)を参照のこと。

3-4 3D-RISM法により計算されたタンパク質水和状態を高速に推算するAIシステムの開発

既に3D-RISM法により計算した水和状態とタンパク質立体構造を、3D-CNNを用いて学習させ、タンパク質立体構造情報のみから、水和状態を高速に推算するAIシステムの開発を行った。(現在、論文執筆中)

4. まとめ

4-1 Machine learning to estimate the local quality of protein crystal structures

人工知能技術3D-CNNを利用することにより、低分解

2021年度 利用報告書

能構造の電子密度マップのタンパク質局所構造を評価するシステムの開発に成功した。

4-2 サイクリックペプチド AI 力場開発

HOKUSAIを利用した多種サイクリックペプチドの多様なコンフォメーションの FMO 計算を実施する環境が整った。

4-3 Exon Skipping データベース開発および AI 予測モデル開発

Exon Skipping の標的となる核酸シーケンスとその Exon Skipping 能についての、データベースを構築し、公開した。

4-4 3D-RISM 法により計算されたタンパク質水和状態を高速に推算する AI システムの開発

既に 3D-RISM 法により計算した水和状態を、3D-CNN を用いて学習させ、タンパク質水和状態を高速に推算する AI システムの開発を行った。

5. 今後の計画・展望

5-1 Machine learning to estimate the local quality of protein crystal structures

本方法は、リガンド、水の電子密度も含めた解析への拡張、低温電子顕微鏡電子密度マップへの応用が可能であり、これらを検討していく。

5-2 サイクリックペプチド AI 力場開発

多様なサイクリックペプチドそれぞれについて、大量のコンフォメーションを発生させ、その FMO エネルギー値を算出することにより深層学習用データを作成し、サイクリックペプチド AI 力場を構築していく。

5-4 3D-RISM 法により計算されたタンパク質水和状態を高速に推算する AI システムの開発

論文をまとめ発表する。

2021 年度 利用研究成果リスト

【雑誌に受理された論文】

Ikuko Miyaguchi, Miwa Sato, Akiko Kashima, Hiroyuki Nakagawa, Yuichi Kokabu, Biao Ma, Shigeyuki Matsumoto, Atsushi Tokuhisa, Masateru Ohta & Mitsunori Ikeguchi

“Machine learning to estimate the local quality of protein crystal structures”

Sci Rep 11, 23599 (2021). <https://doi.org/10.1038/s41598-021-02948-y>

S. Chiba, K. R. Q Lim, N. Sheri, S. Anwar, E. Erkut, M. N. A. Shah, T. Aslesh, S. Woo, O. Sheikh, R. Maruyama, H. Takano, K. Kunitake, W. Duddy, Y. Okuno, Y. Aoki, T. Yokota

“eSkip-Finder: a machine learning-based web application and database to identify the optimal sequences of antisense oligonucleotides for exon skipping”

Nucleic Acids Res 49, 2021, <https://doi.org/10.1093/nar/gkab442>

【会議の予稿集】

【口頭発表】

遠隔講義「計算生命科学の基礎8」

～未来の医療と創薬へつながる生命科学のデータサイエンスとシミュレーション技術～

「AIを用いた創薬の新展開」

大田 雅照 (理化学研究所)

2021 年 11 月 10 日 (オンライン開催)

千葉峻太郎, 奥野恭史, 邦武克彦, 青木吉嗣, 横田俊文 eSkip-Finder: エクソンスキッピングのためのアンチセンス核酸オリゴマーウェブデータベース、疾患モデル動物を基盤とした 筋ジストロフィーの新しい治療法開発 2021 年度 研究班会議 (ウェブ開催)、2021 年 11 月 27-28 日

千葉峻太郎, eSkip-Finder: エクソンスキッピングのためのアンチセンス核酸オリゴマーウェブデータベース、2021 年度 筋ジストロフィー研究班 合同班会議 (オンライン開催)、2022 年 1 月 14 日

【ポスター発表】

Kosuke Kawama, Yusaku Fukushima, Takashi Yoshidome, Mitsunori Ikeguchi, Masateru Ohta
Development of a Deep-Learning Model for Predicting the Hydration Structures around Proteins
65th Biophysical Society Annual Meeting, Feb 20, 2022

【その他 (著書、プレスリリースなど)】

“遺伝性神経・筋疾患の治療法開発を目的としたエクソン・スキップのデータベースおよび機械学習モデルに基づいた予測システムを世界で初めて開発・公開しました”、<https://www.r-ccs.riken.jp/outreach/topics/20210609-1/>, (2021 年 6 月 9 日)

著者: 千葉峻太郎, 大田雅照

「協調的 2 アミノ酸残基同時変異体の相互作用解析による新規抗体の Structure-Based Design」

医学のあゆみ 278 巻 6 号, 663-669 (2021)