



スーパーコンピュータ「富岳」を活用した 創薬DXプラットフォーム開発

研究・開発機関：理化学研究所、京都大学
利用施設：スーパーコンピュータ「富岳」
計算規模：数ノード～数千ノードを用いたフレキシブルな計算規模
利用ソフトウェア：kMoL、ChemTSv2、OpenFold、MP-CAFE、SBMolGen、INGOR等

Before

- 従来の新薬開発では、新規化合物の合成、標的分子の探索、スクリーニング（ふるい分け）等、実験を中心に進められることが多く、開発期間の長期化、開発費の膨大化が課題となっています。
- また、インシリコ創薬（コンピューター上で薬を創ること）では主として高い専門性を持った一部の専門家により、個別のソフトウェアを独自に組み合わせることで援用される形が取られていました。

After

- 多様なソフトウェア群を有機的に組み合わせることが可能な創薬DXプラットフォームを構築することにより、ワンストップのインシリコ創薬支援を期待でき、創薬分野のSociety5.0に寄与し創薬開発の超効率化が図れます。
- ユーザビリティに優れたインターフェースのもと、「富岳」の多ノード性を活用したハイスループット解析を提供することで、専門家でなくてもスパコンを利用したインシリコ創薬が可能になります。

背景と目的

医薬品開発は長い研究開発期間(10年程度)と多額の開発費(1,000億円程度)を要し、その低減が長年の課題となっています。その対策の一つとして、近年、進展著しいAIやシミュレーションによるインシリコ創薬に期待が寄せられています。このようなニーズを受け、私たちはスーパーコンピュータ「富岳」を基軸として、AIとシミュレーションを融合させた革新的な創薬DXプラットフォーム(図1参照)を構築し、創薬プロセスの超効率化・高速化を目指しています。

創薬DXプラットフォームは、創薬の第1ステップである「1. 創薬ターゲット探索」と第2ステップの「2. リード化合物(新薬候補化合物)創出」をシームレスに連結することで、疾患名・患者サンプル等の情報から、薬剤候補化合物の推定までを一気通貫で行う半自動のAI/シミュレーション計算プラットフォームになっています。具体的な活用事例としては、疾患名や患者のゲノム等オミクス情報からベイジアンネットワークやGCN等のAIによって標的遺伝子・タンパク質を推定します。さらに、分子生成AIや分子シミュレーションによって、その標的タンパク質に対するリード化合物を推定することが可能です。特に、「富岳」の高並列計算を活用することで時間と経費を節約するハイスループット解析(高速化合物評価)や、分子シミュレーションとAIの融合することで高精度化を実現する強みを有し、さらには、多様なソフトウェア群を有機的に組み合わせたワークフローを使いやすいインターフェースのもとで実施できる創薬計算基盤をHPCI環境*として提供しました。今後、さらなる評価と改良を重ねることで、創薬分野におけるSociety5.0(デジタル空間と現実社会の融合)の真の実現が期待できます。

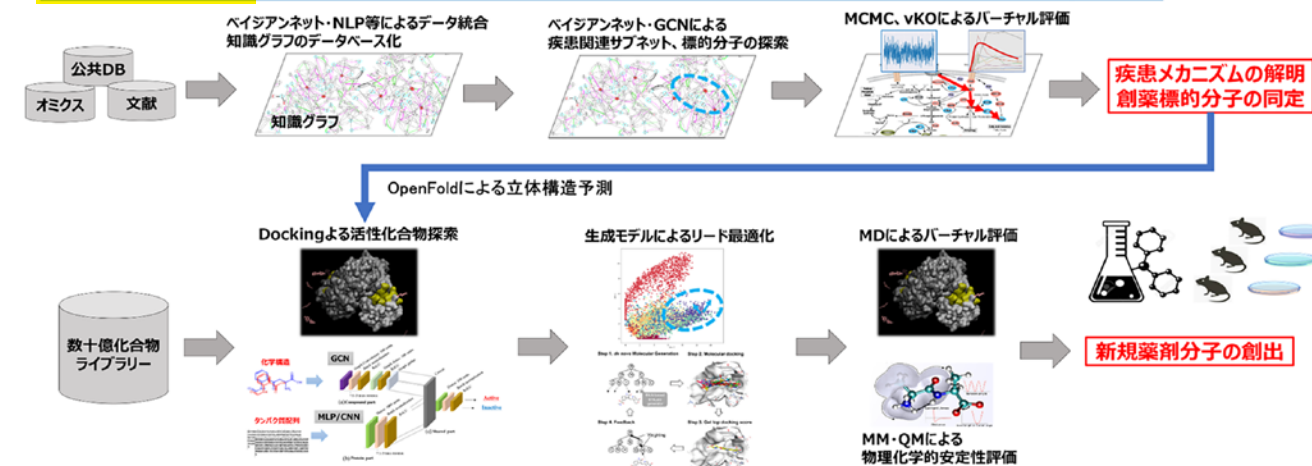
HPCI環境*：「富岳」を中核とし、全国の主要なスーパーコンピュータをネットワークでつなぎ、多様なニーズに応える革新的な計算機環境を実現するための基盤システム

創薬DXプラットフォーム

- ・「創薬ターゲット探索」から「リード化合物創出」に至るHPC/AIフローを構築
- ・「富岳」を中心に、HPC/AIフローの自動化を図ることで、創薬の超効率化を実現



1. 創薬ターゲット探索：疾患名・患者サンプルデータ等を入力して、疾患メカニズムや標的タンパクを推定するHPC/AIフロー



2. リード化合物創出：標的タンパク質名を入力して、リード化合物を推定するHPC/AIフロー

図1 創薬DXプラットフォーム概要

利用成果

1. 創薬ターゲット探索

創薬ターゲット探索(図1上段)は、疾患名や患者サンプルデータなどを入力し、疾患メカニズムや標的タンパク質を推定するフローであり、20の中核ソフトウェアで構成されています。創薬ターゲット探索は3つの主要なプロセスにより実施されます。まず、公共データベース(DB)、文献、オミクスデータ(ゲノムなど生体分子についての網羅的な情報)を統合することで知識グラフを構築します。次に、知識グラフから疾患関連サブネットワークを抽出し、最後にサブネットワークから疾患メカニズムや標的タンパク質を推定します。少し専門的になりますが、オミクスデータに対しベイジアンネットワークを、文献には自然言語処理技術を、DBにはRDF表現法などを用いネットワーク構造(知識グラフ)として表現することで多様で膨大な公共データから統合的な知識グラフを構築します。そして、知識グラフやユーザーが入力したオミクスデータからベイジアンネットワークやGraph Convolutional Network(GCN)を活用することで疾患に関連するサブ分子ネットワーク抽出を行います。最後に抽出したサブネットワークから疾患に関連する重要なパスや分子を特定することで、疾患メカニズムや標的タンパク質を同定することができます。

2. リード化合物創出

リード化合物創出(図1下段)は、創薬ターゲット探索において、推定された疾患関連・標的タンパク質に対して、結合活性を示し、かつ薬物動態や安全性において良好なリード化合物を推定するフローであり、35の中核ソフトウェアで構成されています。創薬DXプラットフォームでは、リード化合物創出のために標的タンパク質の立体構造情報や既知化合物の活性情報の有無などの事前情報に合わせて、多様な個別アプリの組み合わせから最適な創薬計算フローを構築できます。例えば、実験的には標的タンパク質の構造が得られない場合でも、いくつかの活性・非活性化合物がわかっているような場合には、アミノ酸配列から標的タンパク質の立体構造をOpenFoldにより推定し、Autodock vinaを用いて既知の活性・非活性化合物のドッキングシミュレーションを通して標的タンパク質の最適な構造を得ます。この最適構造に対して、kMoL & ChemTSv2を用い、薬効、薬物動態、安全性を満たしながら多数の化合物を生成することで新規薬剤分子の創出を行うことができます。

文責 理化学研究所 計算科学研究センター 奥野 恭史、徳久 淳師、千葉 峻太郎、井阪 悠太
京都大学 医学研究科 岩田 浩明、大塚 教雄、鎌田 真由美、中澤 麻衣