

学 位 論 文 題 名

Software development of ligand-binding analysis for
structure-based drug design

(SBDD のためのリガンド探索ソフトウェアの開発)

学位論文内容の要旨

Structure-based drug design (SBDD) is one of several methods in the rational drug design toolbox for modern drug discovery industry. SBDD uses the known 3D geometrical structure of proteins, which are most often derived from X-ray crystallography or nuclear magnetic resonance (NMR) techniques at an atomic resolution, to assist in the development of new drug compounds. At this level of resolution, researchers can precisely examine the interactions between atoms in protein targets and atoms in potential drug compounds that bind to the proteins. This ability to work at high resolution with both proteins and drug compounds makes SBDD one of the most powerful methods in drug design.

As experimental methods such as X-ray crystallography and NMR develop, the amount of information concerning 3D structures of protein targets has increased dramatically. For processing the dramatically increased data fast and efficiently, in the last decade, computer-assisted SBDD method has increased considerably and played an important role in drug design. The most fundamental goal of it is to predict whether a given "molecule" usually referred as "ligand" will bind to a target protein and if so how strongly. In this study, a new software was developed for the ligand-binding analysis for a target protein, including several parts as follows: binding-site prediction, feature analysis of predicted binding sites, ligand-database establishment, ligand selection. In SBDD, also in our software, binding site identification is the first and prerequisite step. In this study, we presented a new algorithm, Roll, implemented in a program named POCASA, which can predict binding sites by detecting pockets/cavities of proteins with a rolling sphere. To evaluate the performance of POCASA, a

test with the same data set as used in several existing methods was carried out. POCASA achieved a higher success rate of 77% than the other methods. In addition, the test results indicated that POCASA can predict meaningful shapes of ligand binding sites. Based on the predicted pockets/cavities, geometric and chemical analysis of putative binding sites was performed. For geometric analysis, two features were considered in this study: volume and shape. The number of grid points in predicted pockets/cavities was used to represent the volume of putative binding sites. A well-established technique, Principal Component Analysis (PCA), which is a useful tool for finding patterns in data of high dimensions, was applied to the shape description. For chemical analysis, the chemical features of surrounding atoms in protein molecule were extracted as the binding-site chemical features. A ligand database including 8641 ligands in Protein Data Bank (PDB) was established for ligand selection. For the binding analysis of one ligand candidate in the database, a similar way to analyzing putative binding sites was carried out. A grid-based method was used to calculate van der Waals volume of the ligand candidate. All stable conformations of the ligand candidate were generated and PCA was adopted to analyze the ligand shapes and only several representative conformations were remained. The comparison of the geometric and chemical features between ligand candidates and putative binding sites in protein target will give an estimation of ligand binding affinity. Several tests were accomplished and the results substantiated the usefulness of our software in the ligand-binding analysis.

学位論文審査の要旨

主査	教授	田中	勲
副査	教授	出村	誠
副査	准教授	姚	関

学位論文題名

Software development of ligand-binding analysis for structure-based drug design

(SBDD のためのリガンド探索ソフトウェアの開発)

構造情報に基づく薬剤設計(SBDD: structure-based drug design)は、現代の薬剤開発にとって主要な方法の一つである。SBDD 法による新薬開発では、X線結晶構造解析あるいは核磁気共鳴法 (NMR) で決定された蛋白質の原子座標を使うことになる。原子分解能の解像度を持った結晶構造を利用することで、研究者はターゲットとする蛋白質と、それに結合する薬剤候補化合物との間の相互作用の詳細を調べることができる。蛋白質を構成する原子と薬剤候補化合物を構成する原子との間の相互作用が詳細にわかることが、薬剤設計において SBDD 法の最も重要な特徴であり、また、この方法を最も強力な方法の一つにしている。

X線結晶学や NMR 法が発達して、蛋白質の三次構造に関する情報の量は飛躍的に増加した。それらのデータを、迅速に、効率的に処理するために、この 10 年間、コンピューターを利用した SBDD 法が発展し、薬剤設計の分野で重要な役割を演じてきた。その方法の最も基本的なゴールは、ある分子が「リガンド」となりうるかどうか、もしそうならばどれだけ強く蛋白質に結合するかを予測することである。

本研究では、リガンド結合解析のために、新しいソフトウェアを、以下の 4 つのパーツとして開発した。それらは、結合部位予測、予測した結合部位の解析、リガンドデータベースの構築、そしてリガンド選択である。SBDD 法においては、結合部位を同定することが、最初の不可欠なステップであり、それは本ソフトウェア開発でも同様である。この研究では、そのために新しいアルゴリズム (Roll) を開発し、そして POCASA と名づけたプログラムに組み込んだ。Roll は、蛋白質の分子表面にそって球を回転させ、蛋白質のポケット/キャビティーを発見し、結合部位を推定する。

POCASA のパフォーマンスを評価するために、既存のプログラムを評価するのに使われたのと同じデータセットを使ってテスト計算を実施した。その結果、POCASA は、他のどの方法より高い 77% という成功率を達成した。さらに、テスト計算の結果は、

POCASA がリガンド結合部位の形を正しく予測できることを示した。このためには、求めたポケット/キャビティーから推定した結合部位の幾何学的、化学的解析を実施した。幾何学的な解析のためには、2つの性質、体積と形を考慮した。体積は、ポケット/キャビティーを構成するグリッド点の数として求めた。また、形を表現するのには、Principal Component Analysis (PCA) (高次元のデータからある種のパターンを発見するのに有用な手法である) を用いた。一方、化学的解析のためには、蛋白質分子表面の化学的性質を抽出した。リガンド選択のためには、Protein Data バンク (PDB) から 8641 のリガンドを含むリガンドデータベースを用意した。データベースから選んだリガンド候補化合物の結合を評価するには、結合部位を解析するのと同様の方法で行った。すなわち、体積を求めるためには、グリッド点からファンデルワールス体積を求めた。リガンド候補化合物のすべての安定なコンフォメーションを発生させ、そして、リガンドの形を解析するために PCA を適用し、いくつかの代表的なコンフォメーションのみを残した。リガンド候補と蛋白質ターゲットの幾何学的、化学的特徴の比較により、推定した結合部位とリガンドの結合の評価をした。

以上、本研究では、X 線結晶構造で求められた蛋白質の座標からリガンド結合部位を予測するための新しい方法を開発した。さまざまなテストの結果、開発したソフトウェアは、リガンド結合解析において非常に有効であることを示した。本研究が生命科学に及ぼす貢献には多大なものがあり、よって審査員一同は、申請者が北海道大学博士（生命科学）の学位を授与される資格あるものと認めた。