

分子情報の分類・管理システムの設計と実装

†清 一人 †嶺 行伸 †平石 広典 †溝口 文雄

†東京理科大学 理工学部

1. はじめに

近年、計算機能力の向上と共に膨大な生物に関する情報を扱うバイオインフォマティクスの分野の研究が盛んになってきている。バイオインフォマティクスは、医療や創薬の分野へと適用することで、オーダーメイド医療や新薬開発などへの期待も大きい。特に創薬の分野においては、Structure-based Drug Design (SBDD)、分子の構造に基づいたドラッグデザインの研究が盛んに行われている。コンピュータを利用してタンパク質の立体構造を解析して、それと結合する化合物を探索することを対象とした研究である。これによって膨大に存在する化合物から薬の候補となる化合物を探索するための時間とコストの大幅な削減が期待される。一般的にコンピュータ上で化合物の探索を行うためには、受容体とリガンドと呼ばれる2つの分子がどのように結合するかの予測を行うためのドッキングソフトウェアが利用される。

このように創薬の分野においては分子の情報が頻繁に利用され、必要となる分子データを効率よく取得することが重要になってくる。しかし分子に関するデータは膨大に存在しているため、必要なデータを取得するには生物や化学に関する深い知識が必要になる。

2. 目的

本研究では、ネットワーク上のデータベースから分子に関するデータを取得し、生体の重要な反応との関連に基づいて分類を行った。そして分類された分子データの情報を表示・管理するためのGUIを作成した。これによって、直感的に分子データを選択することができ、必要な情報を効率よく取得することを目的とする。興味のある分子を容易に選択できるようになれば、分子ドッキングを行うソフトウェアなどに対する入力データを効率よく準備することが可能と

なる。

3. 設計方針

本研究では、プログラム上で分子データを管理する際にセマンティックネット[2]を使用する。例えば、”キネシンの PDB ID は 1BG2 である”という事実は、図 1 のようなノード間のリンク構造で表現する。リンクは下位概念(キネシンノード)から上位概念(1BG2 ノード)に向けて張られる。

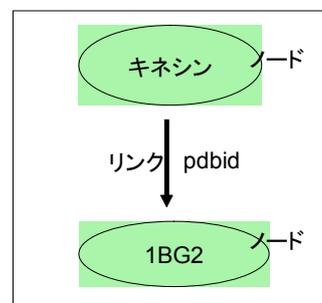


図 1: ノード間のリンク構造

セマンティックネットを構築するデータへの問い合わせは、”pdbid キネシン ?x”というクエリーによって行い、登録されているデータとクエリーとのマッチングを行うことによって、”?x = 1BG2”という結果を得ることができる。

4. 実装

4.1 データ取得

生体高分子の立体構造を収録したデータベースである PDB(Protein Data Bank)から分子の立体構造データを取得する。PDB は委託された構造のそれぞれに 4 文字の識別名を割り当てている(PDB ID)。ID があらかじめ分かっている場合は、その ID を含んだリクエストを PDB のサーバーに送信することでデータを取得することができる。ID が分かっていない場合には、タンパク質名などのキーワードを用いて検索を行うことにより、それと関連する ID を取得することが可能である。

4. 実装

4.1 セマンティックネットの実装

セマンティックネットを実装するために、

Design and mounting of molecule information of classification and management system

†Kazuto Sei, Yukinobu Mine, Hironori Hiraishi, Humio Mizoguchi

†Tokyo University of Science, science and engineering

