

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。

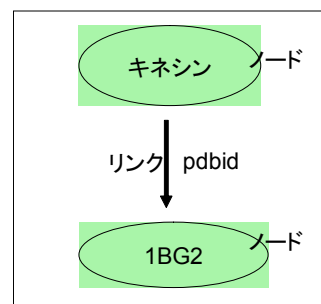


図 1: ノード間のリンク構造

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

Node クラス, Link クラス, SemanticNet クラスの3つのクラスを作成した。

- Node クラス・・・ノードの名前を name という変数に持ち, 自分から出て行くリンクと, 自分に入ってくるリンクをおのこの departFromLinks と arriveAtMeLinks の Vector 型の変数として保持する。
- Link クラス・・・3 変数 label, tail, head を持ち, label は Link が表現する関係 (例えば”pdbid”)を示し, tail は下位の Node, head は上位の Node を現す。
- SemanticNet クラス・・・ノードとリンクの管理や, 問い合わせを行うためのクラスである。リンクを links という Vector にすべて登録し, ノードは nodes という Vector に登録する。また Node の実体を得るためのハッシュテーブル nodesNameTable 変数を保持する。

4.2 管理のための GUI

図2は, 分子情報の管理を行うためのGUI (Graphic User Interface) である。PDBデータベースから取得してきた分子データと, 分類を行った分子情報を管理することが可能である。ウィンドウの左側の部分は, 取得したPDBデータ用のパネルと分類した分子データ用の2つのパネルを切り替えることができる。

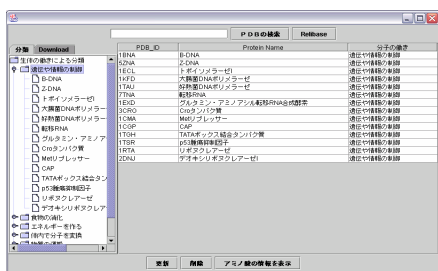


図 2 : 分子情報の管理

Tree の部分にそれらのデータの一覧が表示され Table の部分には ID, タンパク質の名前, ファイルのサイズ, 分子の働きについての情報が表示される。

5. 分類した分子情報の視覚化

5.1 分類を行った分子データ

まず初めに生体内の分子 84 個に対して分類を行った。この 84 個の分子は, 参考文献[2]より引用した分子である。生体内での分子の働きについて 7 つの項目を用意し, 84 個の分子をそれぞれの項目に分類した。7 つの項目については, ”

遺伝や情報の制御”, “食物の消化”, “エネルギーを作る”などである。

5.2 分類情報の視覚化

図 3 は, 分類された分子情報の視覚化を行ったものである。中心の円の周辺に置かれた 7 つの円は, 分子の生体内での働きを表している。それらの円を中心にしてさらに小さく表示された円がそれぞれの分子を表現している。そしてそれぞれの色の違いは分子の働きを表している。

またセマンティックネットを視覚化したものとしてもみることができる。各ノードは円で, ノード間のリンクは直線で表現され, どのようなセマンティックネットが構築されているかが分かる。

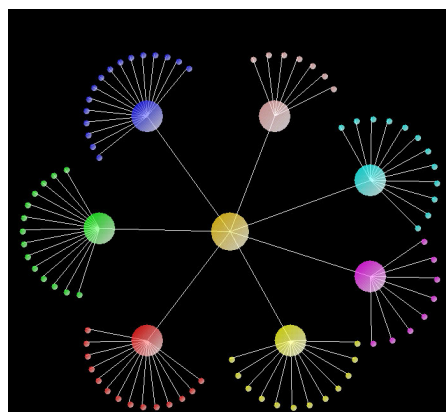


図 3 : 分類した分子情報の視覚化

5. おわりに

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。またプログラム上で分子データを管理するためにセマンティックネットを利用した。

この分類された分子情報を基にして, 例えばコンピュータ上で分子のドッキングを行う際に, 興味のある分子情報を効率よく選択できるようになる。

6. 参考文献

- [1] 平山令明, “分子レベルで見た体のはたらき”, 講談社
- [2] 新谷虎松, “Javaによる知能プログラミング”, コロナ社
- [3] Arthur M. Lesk, “Introduction to Bioinformatics”, Oxford University Press