

1. はじめに

特定の活性を有する薬物群に特徴的な構造フラグメントや骨格構造などを見出そうとする化学構造データマイニングのための様々な試みが行われている。

先に, 当研究室の大野らはNon-terminal Vertex Graph(NTG)[1]を定義し, 化学構造の骨格に注目した薬物構造特徴解析のためのアプローチを提案した。

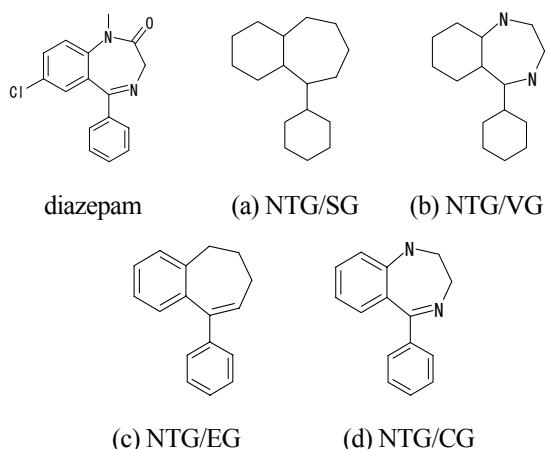
本研究では, NTGにもとづく構造データマイニングのシステム化を目的として, 複数の活性種間でのNTGの比較が可能なNTGデータベース(知識ベース)の自動生成のためのソフトウェアツールを作成するとともに, 市販の治験薬構造データベースを利用し, 種々の薬物群に対するNTG解析を試みた。

2. NTGとは

NTGとは環を中心とした骨格グラフで, 化学構造をグラフとみなしたとき, 次数が2以上の頂点のみを持つものと定義する。また, NTGには下に示す4つの異なるグラフ表現が定義されている。

- a) 単純骨格グラフ (NTG/SG)
- b) 頂点重み付グラフ (NTG/VG)
- c) 辺重み付グラフ (NTG/EG)
- d) 頂点および辺重み付グラフ (NTG/CG)

図1に, diazepamから導出される種々の表現レベルでのNTGの例を示す。



また, 前述のNTGの定義からは外れるが, NTG骨格に直接結合した2重結合の化学的重要性から, これを含むものもNTG/DGとして拡張定義されている。

3. NTGにもとづく構造データマイニング

大野らは, 注目する化合物群に対し, その結合表ファイルをもとに構造中のNTGを抽出するためのプログラムを実装した。同プログラムは, 抽出された各NTGの頻度情報の出力も可能である。しかし, それをもとに活性に特徴的なNTGを自動で抽出する機能は実装されていない。解析の対象となる化合物群の数が大きく, 構造的な多様性も高い場合には抽出・生成されるNTGの数は膨大になり, かつ, 対象を複数の活性種にまたがっておこなう必要があることからそのまま構造データマイニングに用いることは困難である。そこで本研究では, 先のプログラムを改訂し, 薬物構造のNTGと活性との関連付けを行ったNTGデータベースを自動生成し, 併せて, 注目する活性化合物群に特徴的なNTGのみを抽出・提示可能な解析ツールを開発した。これにより, 活性種の拡大にも容易に対応できるようになった。作成したNTG構造データマイニングツールの基本構成概念を図2に示す。

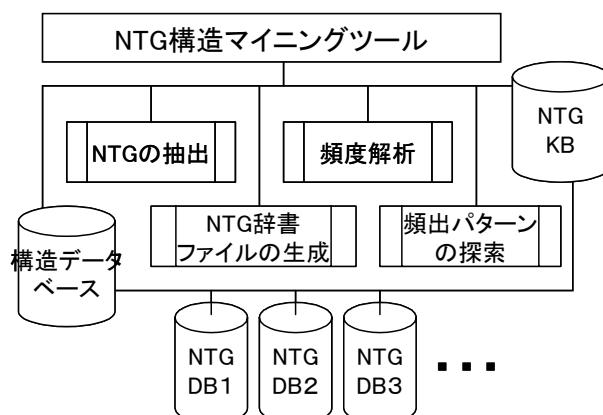


図2 NTG構造マイニングツールの基本構成概念

* taka@mis.tutkie.tut.ac.jp

4. 治験薬データベースのNTG解析

本研究では、治験薬データベースMDDR[2]の中で、収載件数の多い10種類の活性クラス(41534化合物)を対象に、NTG解析を試みた。はじめに、これら10種類の活性クラスに属する各化合物群に対し、前述の5つの異なる表現レベルでのNTGを抽出した。これらの結果をまとめて表1に示す。

表1 各活性クラスから抽出されたNTG数

活性クラス	件数	NTG				
		SG	EG	VG	CG	DG
1.抗高血圧	9047	2178	3075	3678	4181	4352
2.抗アレルギー	7628	1776	2456	3128	3497	3665
3.抗高脂血症	4556	999	1425	1690	1973	2078
4.抗炎症	4255	848	1209	1598	1776	1869
5.抗鬱	3011	812	1081	1312	1453	1501
6.抗精神病	3018	941	1267	1540	1701	1745
7.鎮痛	2501	698	940	1099	1225	1280
8.抗糖尿病	2367	625	897	1054	1179	1236
9.抗真菌	2151	464	651	694	816	862
10.抗けいれん	2151	445	596	786	859	899

表中の各NTGの数は重複するものは除外し、ユニークなもののみを示した数である。NTGの重複チェックにはMorgan法による規範表現を用いた。筆者らは当初、治験薬データベースに収載されている化合物群は、特定の骨格構造を基本とした一連の置換基誘導体が多数含まれていることを予想していた。しかしながら、もっとも包括的な表現である単純骨格グラフ(NTG/SG)のレベルでも(化合物数/ユニークなNTGの数)比は約4~5であり、NTG/CGあるいはNTG/DGレベルではすべて(化合物数/ユニークなNTGの数)<3であることがわかる。このことから、各活性クラスの化合物群は何れも構造多様がかなり高いものであることを示している。

次に、NTGと活性クラスとの関連に注目し、構造データマイニングの視点から、特定の活性クラスに属する化合物群に特徴的なNTGを発掘することを試みた。ここでは、特定の活性クラスに属する化合物群から抽出したNTGと、異なる活性を持つ化合物群から抽出したNTGを比較し、あるNTGを有する化合物群の全て、あるいは大部分が、特定の活性クラスに属するものであれば、そのNTGが注目する活性クラスに特徴的なNTGであるとした。解析の結果得られた各活性クラスに特徴的なNTG/SGの例を1件ずつ図3に示す。

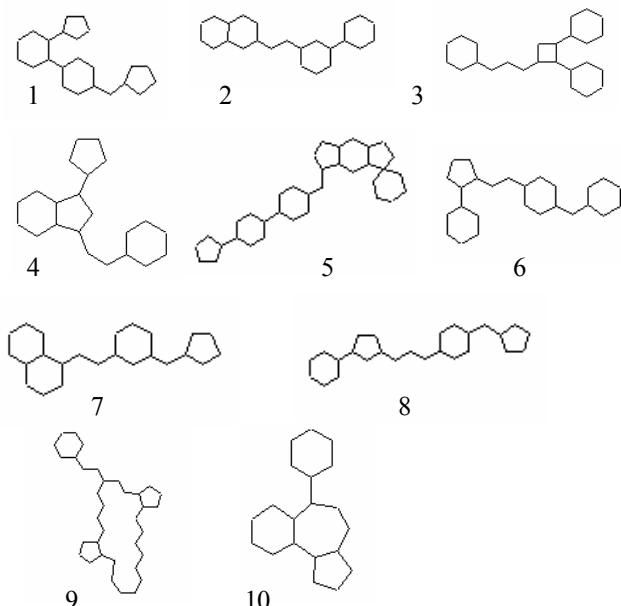


図3 各活性クラスに特徴的なNTG/SGの例

図3に示したNTGはいずれもここで対象とした全活性クラスの化合物中でそれぞれの活性クラスにのみ見出されたものである。この他にもさまざまなNTGが抽出された。また、支持度(注目クラスの構造中での度数/全出現度数)を基準に支持度の高いNTGを頻出NTGパターンとして考えることもできる。図3に示すNTGはこの場合の支持度は1.0(100%)となる。

5. おわりに

本研究では、薬物のNTGと活性との関連付けにもとづく構造データマイニングツールの開発を行うとともに、市販治験薬構造データベースのサブセットを対象にマイニングを試みた。先にも述べたようにNTGには様々な表現レベルが定義可能であり、それぞれ異なる情報を提供し得る。また、ここでは单一の活性クラスのみに特徴的なNTGに注目したが、複数の活性クラスにまたがっての特徴的なNTGのマイニングも考えられる。今後の課題としては、活性種の拡大を図りながら引き続き解析を進め、薬物化学構造からのNTGと活性クラスを関連づけたNTG知識ベースの構築、さらにはNTGの階層関係を利用した帰納データベースの開発などがあげられる。

参考文献

- [1] 大野貴生, 高橋由雅, 第30回構造活性相関シンポジウム公演要旨集, 41-42, 2002.
- [2] MDL Drug Data Report, MDL, ver 2001. 1(2001)