

# 骨格類似性を優先した分子構造の類似性評価

## Evaluation of Structural Similarity Using NTG-based Molecular Skeleton

和田 雅宏  
Masahiro Wada

高橋 由雅  
Yoshimasa Takahashi

豊橋技術科学大学 工学部 知識情報工学系  
Department of Knowledge-based Information Engineering, Toyohashi University of Technology

In the former work, the authors proposed a quantitative evaluation method for structural similarity analysis based on Topological Fragment Spectra (TFS). In the method, whole molecular graphs were employed for the evaluation. But the method doesn't give us the results that have priority to the same molecular skeleton like a particular ring system. This paper describes an alternative approach to structure similarity analysis of molecules, which gives priority to basic skeletons of molecules. The basic skeleton was described with Non-Terminal vertex Graph (NTG) which does not have any terminal vertex or any isolated vertex. The NTG is obtained by pruning the terminal vertexes of a molecular graph. Evaluation of the similarity between NTGs is carried out with the TFS-based method. The details of the NTG-first approach with TFS method will be discussed with an illustrative example.

### 1. はじめに

先に、筆者らは化学構造の Topological Fragment Spectra (TFS) [Takahashi 98] 表現を用いた構造類似性の定量的評価手法について提案している。TFS 法は構造類似性検索においても有用であることが実データを用いて実証されている[高橋 03]。同法は、局所的な構造情報から分子全体にわたる多様な構造特徴の記述を目的としたものである。しかしながら、これまでに提案された方法は同一の骨格構造(特に環システム)を有する化合物群に対して、常にこれを優先的に反映した類似性評価を与えるものではなかった。

そこで、本研究では、別途、化学構造の骨格プロフィールに注目した構造データマイニングのために提案された Non-Terminal vertex Graph (NTG) [大野 03] の概念を応用し、必要に応じて骨格プロフィールを優先的に反映できる、新たな構造類似性評価手法の開発と計算機への実装を試みた。

### 2. Topological Fragment Spectra (TFS)

TFS は、当研究室で考案された構造情報の数値的な記述手法の一つである。その生成手順は、(1) 対象とする化学構造式から可能なフラグメントをすべて列挙し、(2) 列挙したそれぞれのフラグメントに対して数値的な特徴付けを行う。そして、(3) その特徴付けの値と出現頻度のヒストグラムを生成する。このヒストグラムが TFS であり、これを多次元パターンベクトルとして用いることで、化学物質の構造特徴を数値的・定量的に表すことができる。この手法は、フラグメントの定義ファイルを必要とせず、また、生成されたフラグメントの特徴付けの方法を工夫することによって、様々な特性スペクトルを生成することができる。

### 3. Non-Terminal vertex Graph (NTG)

NTG とは頂点次数が 1 となるような頂点および孤立頂点を持たないグラフと定義する。従って、化学構造のグラフ表現(分子グラフ)から抽出される NTG は環を中心とした基本骨格を表すものと言える。また、原子や結合タイプの区別の有無に応じて様々な表現レベルを考えることができる。NTG の 5 つの異なる

表現レベルを表 1 に示す。また、diazepam の化学構造とそこから抽出される各表現レベルでの NTG の例を図 1 に示す。

表 1 NTG のいろいろな表現

略称	グラフ表現
NTG/SG	単純グラフ
NTG/VG	原子の種類のみつきグラフ
NTG/EG	結合の種類のみつきグラフ
NTG/CG	原子・結合の種類のみつきグラフ
NTG/DG	NTG/CG 成分に直接 2 重結合を有する分子グラフ

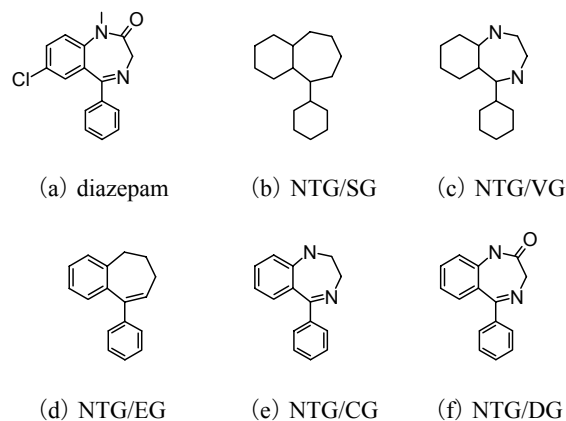


図 1 diazepam の異なるグラフ表現にもとづく NTG

### 4. 骨格類似性を優先した類似性評価

本研究では、化学構造の類似性評価を、全構造に対応する分子グラフ(全グラフ)間の類似性と、そこから抽出した NTG の類似性の両面から評価を行う。このとき、全グラフでの類似性よりも NTG にもとづく類似性の評価結果を優先することによって、骨格の類似性をより反映させた評価が可能となる。

本研究で検討した、骨格類似性を優先的に考慮した構造類似性評価の処理の流れを図 2 に示す。はじめに、対象とするすべての化学構造に対し、各分子グラフ(全グラフ)に対する TFS の生成を行う。次に、各分子グラフから NTG を抽出する。そして、個々の化学構造に対して抽出された NTG から、その TFS (NTG-TFS) を生成する。

続いて、全グラフから生成された TFS をもとに類似度の計算を行う。さらに、上で生成された NTG-TFS をもとに、NTG 間の類似度を計算する。最後に、全グラフと NTG の 2 種類の類似性評価の結果を用いて、最終的な評価を行う。この際の評価手順は、以下の通りである。

- 1) 注目する分子グラフについて、その NTG の類似度が高いものから順に優先順位をつける。
- 2) 同じ NTG を持つ分子グラフが複数存在する場合は、全グラフでの類似度が高いものを優先する。

このように全グラフと NTG から生成される2つの異なる TFS を用い、NTG での類似性解析の結果を優先することにより、骨格構造の類似性を優先的に反映した評価が可能となる。

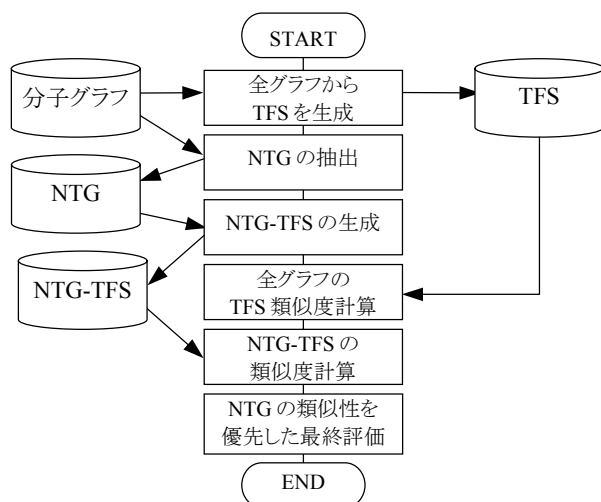


図2 骨格類似性を優先考慮した構造類似性評価

## 5. 計算機実験の結果と考察

ここでは、上記の考えを計算機に実装し、構造類似性検索の実験を通して従来法との比較を行った。実験には、図4に示す治験薬データベース MDDR (MDL Drug Data Report) [MDL 02] に記載されている6件の化学構造データを用いた。このうちの1件をクエリとして構造類似性検索を試みた。実験に用いたクエリ構造とそのNTGを図3に示す。

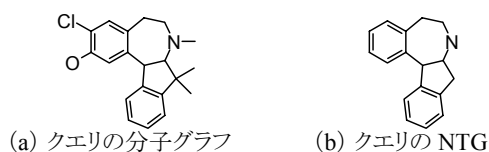


図3 検索実験に用いたクエリ分子グラフとNTG

本実験では、類似性評価のための TFS の生成におけるフラグメントの特徴づけにはすべて、質量数を用いた。また、TFS 間の類似度の評価関数には(1)式に示す Cosine 係数を用いた。

$$\text{Cosine 係数 } S_{A,B} = \frac{\sum_{j=1}^{j=n} x_{jA} x_{jB}}{\sqrt{\sum_{j=1}^{j=n} (x_{jA})^2 \sum_{j=1}^{j=n} (x_{jB})^2}} \quad (1)$$

$S_{A,B}$ : 2つの TFS AB間の Cosine 係数  
 $x_{jA}$ : TFS A の j 番目 (j 次元目) の特性値  
 $n$ : TFS の次元数

図4は全構造に対応する分子グラフから生成した TFS による類似性評価の結果のみをもとに順位付けを行った(従来法)の結果を示す。ここでは、自分自身も含め、クエリ分子と同一の NTG 骨格を有する化学構造が3件あるにもかかわらず、1位のクエリ構造そのものを除く他の二つの評価順位は3位および6位となっていることが分かる。このように、従来法では基本骨格構造が同一の化学構造であっても構造類似性評価に直接反映されない場合がある。

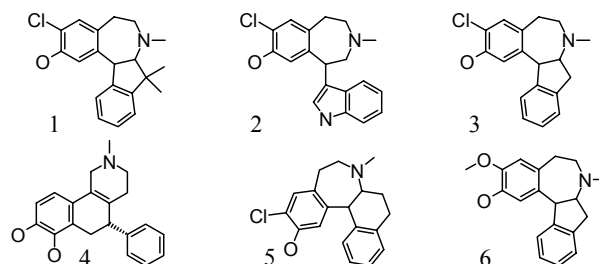


図4 従来法による構造類似性検索の結果(順位)

これに対して、全グラフとNTGを合わせて考慮し、NTGの類似度を優先して評価する本提案手法による実験では、クエリと同じ骨格構造を持つ分子グラフは1位から3位に集まり、異なる骨格構造を持つグラフは4位~6位に位置していることがわかる(図5)。以上のことから、本手法を用いることによって骨格構造の類似性を優先的に反映した構造類似性検索が可能であることが示された。

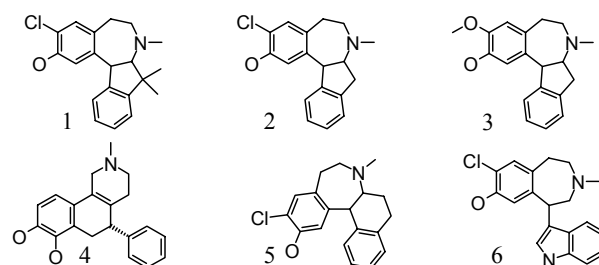


図5 NTGの類似度を優先した検索・評価

## 6. おわりに

本研究で提案した手法を用いることによって、骨格の類似性を類似性評価に優先的に反映できることが示唆された。しかしながら、実用的な見地からは引き続き種々の条件下での性能評価を行う必要がある。また、今後の課題としては複数の NTG の表現レベルを階層的に組み合わせた方法や、側鎖への対応に対する工夫などが考えられる。

## 参考文献

- [大野 03] 大野貴生, Non-terminal Vertex Graph (NTG)を利用した薬物の構造特徴解析, 豊橋技術科学大学 (2003).  
 [高橋 03] 高橋由雅, 藤島悟志, 加藤博明, 化学物質の構造類似性にもとづくデータマイニング, *J. Comput. Chem. Jpn.*, Vol. 2, No. 4, pp. 119-126 (2003).  
 [MDL 02] MDL Information Systems, Inc., MDL Drug Data Report.  
 [Takahashi 98] Y. Takahashi, H. Ohoka, and Y. Ishiyama, "Structural Similarity Analysis Based on Topological Fragment Spectra", In "Advances in Molecular Similarity", Vol 2, (Eds. R. Carbo & P. Mezey), JAI Press, Greenwich, CT, pp. 93-104, 1998..