

TFS 差スペクトルによる薬物構造データマイニング

Structure Data Mining of Drugs using Differential Spectra of The TFS

藤島 悟志
Satoshi Fujishima

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

We introduced the Differential spectra between the TFS for mining chemical structural features of Drugs that have a particular activity. The Differential spectra make us possible to characterize a particular chemical class with a set of TFS peaks that are appeared in the chemical class but not appeared in another class. A software tool was developed, which can generate the Differential spectra, and it also can find structural fragments assigned to the class-specific TFS peaks obtained by the Differential spectral analysis. The Detail of this approach will be discussed with examples featuring the tool.

1. はじめに

筆者らは、先に、構造類似性を基礎とした事例ベースの活用による薬理活性の推定やリスクレポートの可能性について検討した[Takahashi 02, 高橋 03]。ここでは、事前の部分構造知識を必要としない構造特徴のプロファイリング手法である Topological fragment spectra (TFS) 法[Takahashi 98]を用い、化学構造に対応する固有の TFS を生成し、TFS 空間における類似性をもとに、類似構造薬物の検索を実現した。また、活性既知の薬物分子をプローブとした活性候補化合物の推定や目的外作用を有する例外分子の推定における有用性も示した。同様な観点から、薬物活性クラスのカテゴリ・識別問題における TFS 法の応用を試み、TFS を入力シグナルとした人工ニューラルネットワーク (Artificial Neural Network, ANN) やサポートベクターマシン (Support Vector Machine, SVM) の有用性を明らかにした [Fujishima 04, 錦織 03]。の有用性を明らかにした。これらの研究から、TFS を基礎とした構造類似性検索や活性クラス分類が有効であることは明らかであり、TFS が化合物の構造特徴を良く表現していると考えられる。

ところで、TFS の各ピークは、列挙された部分構造をもとに生成されている。従って、TFS の生成時に生成フラグメントと各ピークの対応付けの情報を保持しておけば、ピークの裏側に隠されている具体的な構造フラグメントを容易に同定・確認 (以下、ピーク同定) することができる。また、ピーク同定が可能になれば、TFS を記述子とした種々のデータ解析についてもより詳細な考察が可能となる。TFS ピークの同定の問題は知識発見の観点からも極めて重要であることは明らかであり、筆者らは別途、そのシステム化についても検討し、報告している [Fujishima 04]。

以上の成果を踏まえ、本研究では、化学構造の TFS 表現に対して、新たに TFS の差スペクトルの概念を導入した。そして、異なる活性クラスに属する化合物間での差スペクトルをもとに、各活性クラスに特徴的な TFS ピークの抽出と、そのピーク同定を基礎とした薬物構造データマイニングのためのアプローチについて検討を行った。以下にその詳細を述べる。

2. 方法

2.1 TFS と差スペクトル

TFS とは、化学構造の定量的な類似性評価を目的に、高橋らによって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 [Takahashi 03]。

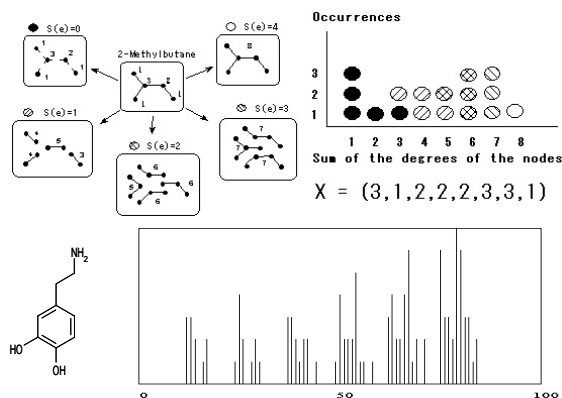


図1 TFS の生成手順と生成例

今、構造が異なる任意の 2 つの化合物 a, b について、それぞれの TFS を TFS(a), TFS(b) で表すとき、TFS(b) に対して TFS(a) にのみに出現するピークから構成される TFS をその TFS の“差スペクトル”と定義し、次のように表す。

$$TFS(a-b) = TFS(a) - TFS(b) \quad (1)$$

ここでの差スペクトルは、対応するピークの有無に注目し、自身にしか存在しないピークのみが残されたスペクトルとして定義される。このとき、2 つの TFS の次元数が異なる場合は次元数の

高い方にゼロ詰めで作合わせるものとする。また、 $TFS(a-b) \neq TFS(b-a)$ であることは明らかである。

一方、クラス A に属する n 個の化合物 $a_i (i=1, \dots, n)$ についてその集合を A であらわすとき、次式で示される $TFS(A)$ をクラス A に属する化合物群に対する TFS の合成スペクトルと定義する。

$$TFS(A) = TFS(a_1 + a_2 + \dots + a_n) = \sum_i TFS(a_i) \quad (2)$$

前述した TFS の差スペクトルの概念は上記の TFS 合成スペクトルに対しても適用できる。すなわち、異なる活性クラスに属する化合物群の TFS 合成スペクトルの差をとることにより、相手クラスからみて特徴的なピークの抽出が可能となる。

ここでは、デジタル化された特性インデックス値でのピークの有無に注目し、各特性量でのピークの有無によって差スペクトルを生成する。これは自身にしか存在しないピークは、相手からみて特徴的であることを意味する。TFS の差スペクトルによる特徴抽出の基本概念を模式的に図 2 に示す。ここでは、化合物 a ($a \in A$) の化学構造に対する TFS を $TFS(a)$ 、化合物 b ($b \in B$) の化学構造に対する TFS を $TFS(b)$ とし、化合物 a の b に対する TFS の差スペクトルを $TFS(a-b)$ として考える。

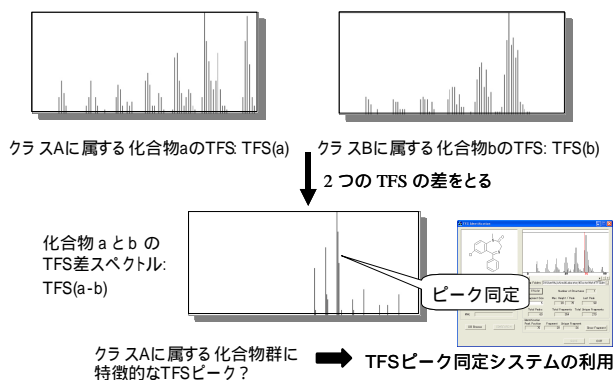


図 2 TFS の差スペクトルによる特徴抽出の基本概念

図 2 の差スペクトル $TFS(a-b)$ は、化合物 b から見て化合物 a の TFS に特徴的なピークということが出来る。そして、この特徴的なピークに帰属される生成フラグメントもまた、その化合物が属する活性クラスに特徴的なものであるとの考えから、データマイニングのためのピーク同定処理についても検討を行った。これには、著者が先に開発した TFS の各ピークに対応した生成フラグメントの同定が可能で、TFS ピーク同定システムを基礎とし、差スペクトルの自動生成とその対応するピークの自動同定機能を実装した改良システムを開発し、利用した。

尚、本研究では、TFS の生成において、各フラグメントの特徴付けには構成原子の質量数の和(フラグメント重量)を用いた。

2.2 TFS 差スペクトルのピーク同定

TFS の各ピークに対応するフラグメントの同定を行うことにより、構造活性相関の観点からも、部分構造情報にもとづくより具体的な考察が可能になると同時に、新たな知識の発見が期待できる。このことから、本研究では、先に報告した TFS ピーク同定システム[藤島 03]に対し、TFS 差スペクトルの生成および解析機能を追加するとともに、システムの改訂を行った。以下、本研究における差スペクトルの生成ならびに生成スペクトルのピーク同定はすべてこの改良システムを用いて行った。

2.3 データセット

本研究での計算機実験には、米国 MDL 社の治験薬構造データベース MDDR (MDL Drug Data Report)[MDL 01]に収録されているドーパミン D1 受容体のアゴニスト(63 化合物)およびアンタゴニスト(169 化合物)を用いた。また、ノイズデータとして、何れの活性も持たない薬物構造データを別途用意し、合わせて用いた。ノイズ化合物は MDDR データベースよりランダムに抽出した。ノイズデータの数は D1 受容体アゴニスト、同アンタゴニストの合計と同じ 232 化合物とした。

3. 結果と考察

3.1 TFS 差スペクトルと改良ピーク同定システム

本研究では、先に報告した TFS ピーク同定システム[藤島 03]に対し、TFS 差スペクトルの生成および解析機能を追加し、システムの改良を行った。これにより、2 つの異なる化合物(群)の TFS (あるいは TFS 合成スペクトル)が同時に作成・表示できるようにするとともに、これらの差スペクトルの自動生成とそのピーク同定が可能となった。本研究で改良した TFS ピーク同定システムの実行画面例を図 3 に示す。

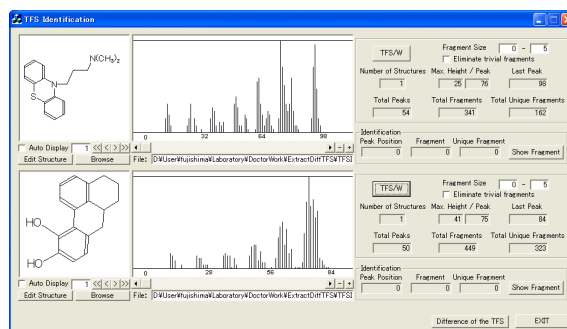


図 3 改良ピーク同定システムの実行画面例

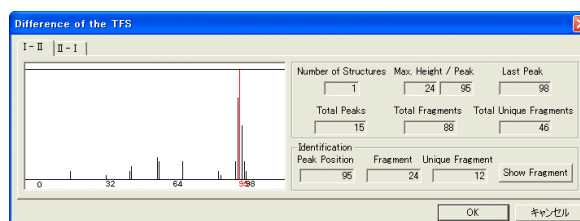


図 4 図 3 の TFS 差スペクトルの生成とピーク同定画面例

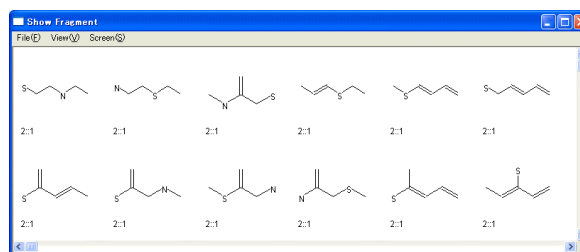


図 5 promazine の TFS 差スペクトルの質量数 95 に含まれるフラグメント

図 4 には図 3 で示した 2 つの TFS に対する差スペクトルを示した。図 3 の画面右下の[Difference of the TFS]ボタンを押すことにより、図 4 に示すように、入力した 2 構造の TFS 差スペクトルが生成・表示される。ここでは、promazine (図 3 上) と apomorphine (図 3 下) の TFS に対する差スペクトルを示した。図

から promazine には apomorphine が持たない独自の TFS ピークが存在することが分かる。この TFS 差スペクトルに対してピーク同定を行うことにより、apomorphine には存在しないピークに含まれるフラグメントを確認することができる。TFS 差スペクトル表示領域で任意のピーク(ここでは、質量数 95 のピーク)を選択し、フラグメント表示を行うと、質量数 95 のピークに対応する生成フラグメントが表示される(図 5)。図から、質量数 95 のピークには、apomorphine にはない窒素と硫黄によるフラグメントが含まれていることがわかる。このように、対象とする 2 つの構造に対して TFS 差スペクトルを適用することにより、構造の違いを容易に調べることができる。これらのアプローチは異なる活性クラスに属する化合物間の構造特徴抽出にも極めて有用であると考えられる。

3.2 ドーパミン受容体アゴニスト/アンタゴニストの TFS スペクトル解析とデータマイニング

ここでは、活性に特徴的なピークやフラグメントの抽出の検討を行った。ドーパミン D1 受容体アゴニスト活性を有する 63 種の化合物と、同受容体に対してアンタゴニスト活性を有する 169 化合物について、それぞれの活性クラスごとに TFS 合成スペクトルを作成した。さらに、これらのデータをもとに活性クラス間の TFS 差スペクトルの生成を行った。作成したピーク同定システムを利用し、得られた差スペクトルの各ピークに対するピーク同定を試みた。

図 6 にドーパミン D1 受容体アゴニスト(以下、単に D1-アゴニストと呼ぶ)63 化合物とドーパミン D1 受容体アンタゴニスト(以下、単に D1-アンタゴニストと呼ぶ)169 化合物に対する TFS 合成スペクトルを示す。また、図 7 には D1-アゴニストの D1-アンタゴニストに対する TFS 差スペクトル(図 7 左)と D1-アンタゴニストの D1-アゴニストに対する差スペクトル(同右)をそれぞれ示す。差スペクトルの図から分かるように、D1-アゴニストから D1-アンタゴニストの TFS を引いた差スペクトルは 1 本のみとなった。一方、D1-アンタゴニストから D1-アゴニストの TFS を引いた差スペクトルは複数あることが確認できた。

このように、2 つのクラスに注目する場合、どちらか一方のクラスに特徴的なピークが存在すれば、それが活性クラス間の相違を表すものとなり、それをもとに各活性クラスに固有の構造特徴を調べることは十分に可能であると考えられる。

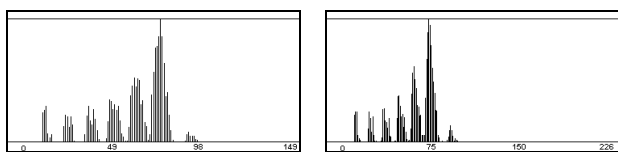


図 6 ドーパミン D1-アゴニスト(左)と D1-アンタゴニスト(右)の TFS 合成スペクトル

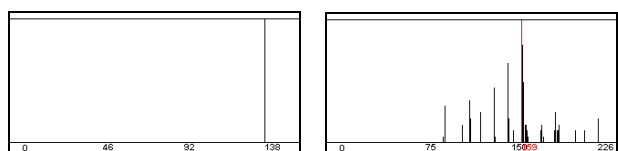


図 7 TFS 差スペクトル(左: D1-アゴニスト, 右: D1-アンタゴニスト)

すなわち、差スペクトルを構成しているピークが各活性に特徴的である場合、そのピークに含まれているフラグメントもまた、その活性に特徴的であると考えられることができる。そこで、各 TFS 差スペクトルのピーク同定を行った。D1-アゴニストの差スペクトルは唯一残ったピーク(質量数=138)を同定したところ、1 つのフラ

グメントのみ含まれていることが分かった。また、D1-アンタゴニストの差スペクトルでは最大ピーク(質量数=159)を同定した結果、5 つのユニークなフラグメントが対応していることが分かった。ここでは、5 つのフラグメントを図 8 に示す。

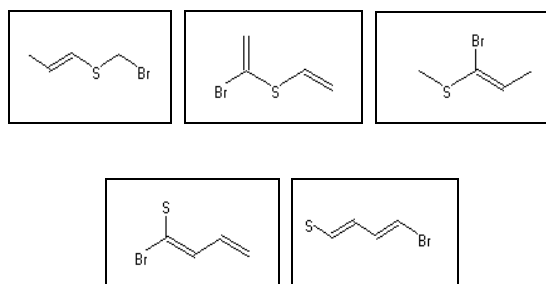


図 8 D1-アンタゴニストの差スペクトルに対して質量数=159 のピークを同定して得られたフラグメント

3.3 ドーパミン受容体相互作用活性の有無と TFS ピーク同定による構造特徴解析

ここでは、活性クラスの差ではなく、注目活性の有無に注目した正例と負例による TFS 差スペクトルの解析を試みた。正例データとしてドーパミン D1 受容体アゴニスト/アンタゴニスト(以下、まとめて D1 受容体相互作用薬と呼ぶ)合計 232 件を、また負例データとしてどちらの活性も持たないノイズ化合物 232 件を使用した。

前節の場合と同様に、それぞれのデータセットに対して TFS 合成スペクトルを生成した。D1 受容体相互作用活性を有する化合物群およびノイズ化合物群に対する TFS 合成スペクトルをそれぞれ図 9 に示す。この 2 つの TFS 合成スペクトルをもとに、その差スペクトルの生成を試みた。D1 受容体相互作用活性を有する化合物群のノイズ化合物群に対する TFS 差スペクトルを図 10 に示す。得られた差スペクトルから D1 受容体相互作用活性を有する化合物群には、ノイズ化合物群に見られない特徴的なピークが 9 つ存在することが確認できる。そこで、これら 9 本のスペクトルに対してピーク同定処理を行い、対応する生成フラグメントについて調べたところ、全部で 16 種のフラグメントが得られた。これらのフラグメント中には相互に内包関係にあるものもあり、最終的にはこれらを除いた 12 種のユニークなフラグメントを同定した。これらの構造フラグメントの一覧を図 11 に示す。

これらのフラグメントは、ランダムに抽出されたノイズ化合物群の構造中には見られない部分構造であり、ここでの解析に用いた D1 受容体の相互作用を有する化合物群に特徴的な部分構造であることは明らかである。

また、図 11 の 3-7 については 3.2 節で示した D1-アンタゴニストの差スペクトルの質量数=138 のピークに見られたフラグメントとすべて一致した。このことはこれらのフラグメントについてはノイズ化合物も含め、改めて D1-アンタゴニストに固有のユニークな構造フラグメントであることが推察される。

また、その他のピークについても同じピークが残っている。そしてこれらのピークに含まれるフラグメントは、図 11 に示したフラグメントの一部もしくは派生したものであることも確認した。これらことから、ノイズデータ存在下においても、硫黄を含む 5 員環とそれに結合する臭素からなるフラグメントとそれを含む化合物は D1-アンタゴニストに特徴的であると推察することができる。

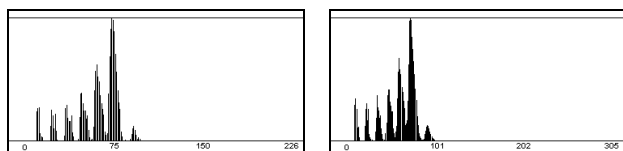


図9 D1 受容体相互作用薬(左)とノイズ化合物群(右)のTFS合成スペクトル

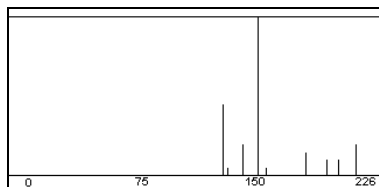


図10 D1 受容体相互作用薬からノイズデータの差スペクトル

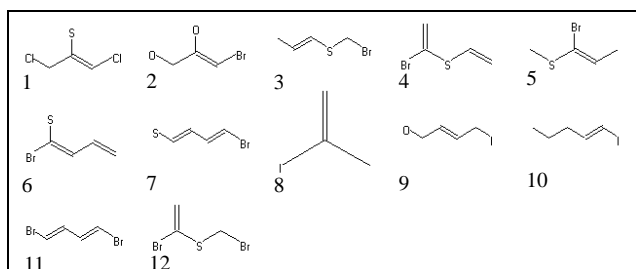


図11 D1 受容体相互作用活性を有する化合物群に特徴的な構造フラグメント

4. まとめ

これまでの研究から、化学構造の TFS 表現が、薬物分子の構造類似性評価や活性クラス識別における構造特徴記述子として有用であることは明らかである。また、結果として得られた TFS のピークの化学的な意味を解析することは新たな知識の発見という観点からも重要となる。

本研究では、異なる活性を有する化合物群にはそれぞれ特徴的な TFS ピークが存在するという仮定のもと、新たに TFS の差スペクトルを定義し、その自動生成とピーク同定を基礎とした薬物構造データマイニングのためのアプローチを提案するとともに、先に開発、報告している構造特徴解析のための TFS ピーク同定システム上に実装した。

ドーパミン D1 受容体アゴニスト/アンタゴニストに対し、それぞれの TFS をもとに TFS 差スペクトルを求めたところ、ドーパミン D1 アンタゴニストにおいて、特徴的なピークが複数得られた。それらの TFS ピークは、ノイズ化合物群から生成した TFS 合成スペクトルとの差スペクトルにおいても見出されることを確認した。このように TFS 差スペクトルを適用することにより、注目する活性クラスの化合物群に特徴的な TFS ピークの抽出が可能であり、また、そのピークを同定することにより、個々のクラスに特徴的な構造フラグメントを調べることができる。以上のことから、本研究で提案した一連の操作は、薬物構造データマイニングにおける知識発見に対して一つの有効な手段を提供するもの確信する。

参考文献

[Fujishima 04] S. Fujishima, Y. Takahashi : Classification of Pharmacological Activity of Drugs Using TFS-Based Neural Network, *J. Chem. Inf. Comput. Sci.*, accepted.

[藤島 04] 藤島悟志, 高橋由雅: “化学構造データマイニングのための TFS ピーク同定システムの開発”, *J. Comput. Chem. Jpn.*, **3**, 49-58 (2004).

[藤島 03] 藤島悟志, 高橋由雅: “TFS を用いた化学構造データマイニング”, 第 17 回人工知能学会全国大会論文集, 2F2-03 (2003).

[藤島 04] 藤島悟志, 錦織克美, 加藤博明, 岡田孝, 高橋由雅: “ノイズデータを含むドーパミン受容体アゴニスト/アンタゴニストの活性クラス分類”, 人工知能学会 第 64 回知識ベースシステム研究会, pp.125-128 (2004).

[MDL 01] MDL: Drug Data Report, MDL, ver. 2001.1, (2001).

[錦織 03] 錦織克己, 高橋由雅: “薬物活性クラス分類へのサポートベクターマシン (SVM) の応用”, 第 17 回人工知能学会全国大会論文集 (2003).

[Takahashi 98] Y. Takahashi, H. Ohoka, and Y. Ishiyama: Structural Similarity Analysis Based on Topological Fragment Spectra, In: R. Carbo and P. Mezey (Eds), *Advances in Molecular Similarity 2*, pp.93-104, JAI Press, Stamford CT, (1998).

[Takahashi 02] Y. Takahashi, S. Fujishima and K. Yokoe: Chemical Data Mining Based on Structural Similarity, Proceedings of International Workshop on Active Mining, IEEE ICDM 2002, 132-135 (2002).

[Takahashi 03] Y. Takahashi, S. Fujishima, H. Kato, Chemical Data Mining Based on Structural Similarity, *J. Comput. Chem. Jpn.*, **2**, 119-126 (2003).

[高橋 03] 高橋由雅, 藤島悟志, 横江恭子: “TFS を利用した薬物活性クラス分類とリスクレポート”, 電子情報通信学会「人工知能と知識処理」研究会・情報処理学会「知能と複雑系」研究会・人工知能学会「人工知能基礎論」研究会・人工知能学会「知識ベースシステム」研究会, 「アクティブマイニング合同研究会」, (2003).