

ドーパミンアンタゴニスト活性のマイニング

岡田 孝[†] 山川 眞透[†]

[†] 関西学院大学情報メディア教育センター 〒662-8501 兵庫県西宮市上ヶ原1-1-155

E-mail: tokada@kwansei.ac.jp

あらまし 多数の生理活性に対して活性化化合物群の構造的特徴を明らかにする研究プロジェクトを開始した。効率的に知識を獲得するため、属性の生成、整理からカスケードモデルによるマイニング、さらに結果の解釈に至る各過程を支援する計算環境の整備を行った。また最初の課題として、4種のドーパミン受容体タンパク質 (D1 - D4) に対してアンタゴニスト活性を有する化合物群から、それぞれの活性に特徴的な構造群を抽出した。この過程で、データスケープ探索を可能とするルール群の表現が、知識の発見に有効であることを立証できた。得られた知見は専門家にも知られていない情報を多数含んでおり、しかもリーズナブルで薬品設計に応用できるとの高い評価を得ることができた。

キーワード 構造活性相関, マイニング, ドーパミン受容体, アンタゴニスト活性, カスケードモデル, データスケープ

Mining Characteristics of Dopamine Antagonists

Takashi OKADA[†] and Masumi YAMAKAWA[†]

[†] Center for Information & Media Studies, Kwansei Gakuin University

Uegahara 1-1-155, Nishinomiya, Hyogo, 662-8501 Japan

E-mail: tokada@kwansei.ac.jp

Abstract A research project started to clarify structural characteristics of pharmaceutical compounds for a variety of medicinal activities. A workbench has been developed to support the mining process efficiently. It includes the preparation of attributes, mining by the cascade model, and the interpretation of derived rules. As the first subject, we studied drugs with antagonist activities for four dopamine receptor proteins (D1 - D4) and discovered various lead structures. The results have shown the usefulness of mining by the cascade model and its capability of datascape survey. Pharmacologists evaluated the acquired knowledge, and remarked that the results contained many unknown, but reasonable pieces of knowledge that are applicable to the drug design.

Key words structure activity relationships, mining, dopamine receptor, antagonist, cascade model, datascape

1. はじめに

我々はアクティブマイニングプロジェクトの一環として、多様な構造を持つ化合物群から、活性発現に重要な構造的特徴のマイニングを試みている。このプロジェクトでは、多くの生理活性を対象とした解析を行って知識を獲得し、最終的に知識ベースとして公表することを目的としている。本報告ではまず、これらの作業を効率的に遂行するために構築したワークベンチの概要を最初に解説する。最初の解析対象としては、ドーパミンのアンタゴニスト活性を取り上げた。4種のドーパミン受容体タンパクに対してアンタゴニスト活性を有する化合物群を対象とし、それぞれの受容体に特徴的な構造をマイニングした。その解析プロセスと結果について述べる。

2. 方 法

2.1 対象データ

2次情報誌 Drug Data Report には、特許を中心とする開発薬品の構造式とその生理活性の情報が、毎年ほぼ1万件の割合で報告されている。この雑誌の情報をデータベース化したMDL社のMDDRデータベース Ver.2001.1 (23.01)には、ほぼ12万件の薬品データが蓄積されており、それらにはおよそ800種の活性が付されている。このデータベースでは、同一活性に対しても多様な化合物構造が存在しているため、活性毎の構造特徴を見出すために適当である。本プロジェクトではこのデータベースを用いてマイニングを行う。

4種のドーパミン受容体 (D1 - D4) に対して、アンタゴニスト活性を持つ化合物データ 1,364種類が記載されている。検証用に10%の化合物を除外し、残りの1,227化合物を解析対象

とした。これらの中で D1 - D4 に活性を示す化合物数は、それぞれ 154, 383, 234, 515 である。なお、同一化合物が複数の活性を示す場合がある。

2.2 処理の流れ

MDDR データベースの検索から始まり、得られたルールの解釈に至るまでの過程を図 1 に示す。図の上部では、各活性毎に検索で得られた化合物セットに対し、塩の除去と同一化合物のチェックを行った後、Cache システムを用いて MM-AM1-geo 法により HOMO, LUMO, dipole moment 値を計算する。その後、各活性毎のファイルを併合して、活性クラスの付与、ClogP 値の計算、再び同一化合物の除去を行っている。なお、検証用に指定した比率の化合物群をテストセットとして除外し、学習用とは別に出力している。各処理の間に記された SDF とは化学構造表現用のフォーマットで化学構造式および化合物の性質を記したテキストファイルである。

後半の処理では、まず構造式中の線形フラグメント群を作成し、さらにこれを利用したカスケードモデルによるルールの生成が行われる。フラグメント群の作成については次節で説明する。生成されたフラグメント群中で、データセット中での存在比率が過大または過小なものを除外して属性群として利用する。なお、keep-del list には、存在比にかかわらず必ず利用/除外するフラグメントを指定する。

次いで、DISCAS (ver.3) で得られたルール群を用い、これらを支持する化合物群の構造を視察し、仮説としてまとめる。この段階では、ルールに出現する属性と化合物構造のデータベースから、Spotfire ソフトウェアを利用して化学構造式を表示する。なお、最終的に得られる仮説は出発点のルールとは大きく異なるため、再度化学構造データベースの検索を行って、その妥当性を検証する。

以上の処理の流れはかなり複雑なものであり、中にはアルゴリズムの存在しないステップもあるため、専門家による作業が欠かせない。これらの作業をできる限り合理化するため、Excel と ChemDraw システムが統合された環境において多数のマクロを作成し、創薬の専門家が効率的に作業を遂行できる環境を構築した。

2.3 フラグメント記述子の生成

構造式からのフラグメント記述子生成法は、先の報告 [1] を発展させたものであり、下記の方法によっている。

- (1) 指定した種類の元素および結合両端の原子を起点として、最短 path 長が max-length 以内のすべての原子との間で線形フラグメントを取り出す。
- (2) この線形フラグメントをその構成元素と結合の種類、各原子の配位数と付随水素原子の有無、および分岐構造で枝上の最初の原子により特徴づける。
- (3) 利用者の指定した詳細度に従い、これらの線形フラグメントを記述する。

詳細な記述形式を採用すると、多種類の記述子が生成され、個々の記述子を支持する事例数が少なくなる。その結果として、偏った分布を示す記述子が多くなり、活性の識別に有効なもの

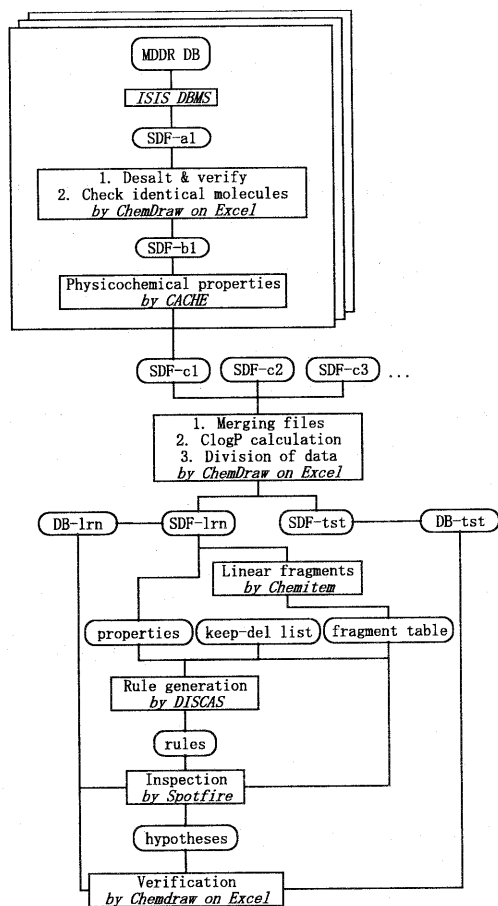


図 1 処理の流れ

Fig.1 Flow of mining

が減少する。反面、簡単な記述形式では、マイニング結果の解釈が困難となる。

ドーパミンアンタゴニストの場合 4 種の記述形式を検討した結果、線形フラグメントの両端から 2 個の原子のみに配位数と付随水素原子の有無を記載した形式を採用した。なお、線形フラグメントの長さは原子数が 10 以内に限った。例えば、C3H:C3-C-C4H-N3 は 3 配位の CH と C の芳香環 (は aromatic bond) に、C-C-N がつながった構造を示す。3 番目の C には配位数と水素の有無は記載されない。また、3 配位の N に H が記されていないことから、この N は 3 級アミンであることが分かる。

2.4 カスケードモデルとデータスケープの探索

マイニングの方法論としては、カスケードモデルを用いる [2]。この方法は相関ルールを発展させたものであるが、真の相関を表現できること、ルールの強さを BSS 値という一つの値で示し得ること、等の利点が存在する。カスケードモデルで発掘されるルールは、下記の形式により説明変数値の組み合わせと活性との間の局所的な相関を表現する。

Rule: 化合物数 U ==> 化合物数 D; BSS 値
IF 主条件 added on 前提条件群

THEN 目的変数: 分布 U --> 分布 D

THEN 説明変数: 分布 U --> 分布 D

ここで主条件, 前提条件には説明変数とその値が記述され, また主条件の適用前後を U, D で示している. 主条件の適用により, 対象化合物数, 目的変数値の分布がどのように変化するか記されるとともに, ルール全体の強さを表現する BSS 値が記載される. また, 説明変数の中に主条件適用に伴って大きな分布変化を示すものがあれば, それも付加的相関として示される.

意義ある解析結果を得るためには, 利用者がデータの全体像から詳細な分布に至る各レベルで, データ分布の特徴を把握できる必要がある. 我々はこれをデータスケープ (datascape) と呼んで, その把握を支援するルール表現を開発した. 詳細は文献 [3], [4] を参照されたい.

この新しいシステムを利用することにより, 下記の効果を得ることができる. (1) 全体のルール数が削減される. (2) ルール群が少数の主ルールとそれに付随する関連ルールに分けて示されるため, 全貌を理解しやすい. (3) ルール強度を示す BSS 値が急激に低下する説明変数の領域を ridge として表すことにより, 解析に有効な特徴ある化合物群を容易に探すことができる.

2.5 マイニング計算の遂行

D1-D4 とともに各活性の有無を目的変数として, 下記のパラメータ値を用いた DISCAS ver.3 による計算を行った.

$thres = 0.10, minsup = 0.01$ (lattice 展開の詳細度)

$thr-BSS = 0.015$ (ルール候補リンク強度の最低値)

$MDL-para = 0.2$ (ルール表現の簡易度)

$min-rlv = 0.7$ (ルール間関連性判断の閾値)

なお, D3, D4 活性の解析では出力されるルール数が多すぎるため, $thr-BSS$ の値を 0.02 に設定した. また, 数値変数のカテゴリー化規範値としては, HOMO(-9.0, -8.5), LUMO(-1.0, -0.5, 0.0), dipole moment (2, 4, 6) および LogP (2, 4, 6) を採用した.

例えば D1 アンタゴニスト活性の有無を目的変数とした解析では, およそ 11.5 分の計算で 88,883 個のラティスノードが生成され, 大きな BSS 値を有する 37 個のルール候補リンク群が検知された. これらのリンクは最適化により 8 種のルールに収斂し, さらに 1 種の主ルールとそれに付随する 5 種の関連ルールに組織化することができた.

2.6 ルールからの仮説生成

専門家がリード構造に対する合理的な仮説として認めるためには, ルールから出発したとしても, 専門家自身が活性値の分布や化学構造式の視察により, 十分に納得する過程が必要である. 本研究では, まずルールの条件記述に従って, 図 2 の右側に示すように, 説明変数の値に従って活性/不活性の分布を表示する. 専門家はこのような表示から, (1) 主条件の適用により大きく分布が変化すること, (2) 前提条件の有無によりその変化の度合いが異なること, を実感できる.

ついで, 棒グラフや円グラフをクリックすることにより, その部分を支持する化学構造式の一覧が図の左側のように現れるので, ルール中のフラグメントが実際に重要な意味を持つのか,

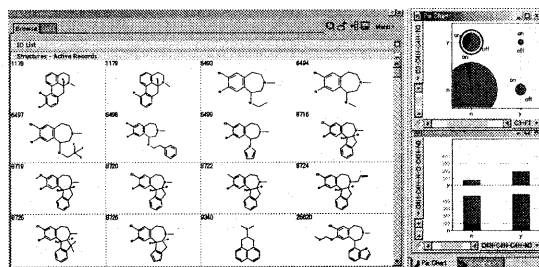


図 2 分布と化学構造の視察

Fig.2 Inspection of distributions and chemical structures

それとも単に一群の化合物に共通の大きな骨格の一部を表しているに過ぎないかを判断できる. 実際, 化学者にとっては構造式記述こそが想像力をはぐくむ源であり, 支持事例の構造式が提示されなければ, 決して仮説が納得されることはない. さらに軸となる変数を変えて視察を行うことにより, 当初のルールとは全く異なった仮説が導かれることもある.

専門家が導く多くの仮説は, 対応する活性値の分布を表示して, 検証を行う必要がある. しかし, 多くの仮説はシステムが採用したフラグメント群とは異なった部分構造により表現される. そこで仮説を部分構造式として表し, これを新規の属性として加えて活性値の分布を見る必要がある. 現段階では, このステップが専門家にとり煩雑な作業となっているため, さらなる効率化が必要である.

3. マイニング結果と考察

得られたルールについて, その主条件, 前提条件と目的活性値の分布から, そのルールの重要性を概観することができる. しかし, 開発薬品データベースを対象としているため, その化合物構造には偏りがあり, ルールの記述をそのまま受け取ると間違った, あるいは解釈不能の仮説を導くことがある. そのため, 意味ある構造的特徴を把握するには, (1) 他の説明変数との付加的な相関関係のチェック, (2) ridge 領域と関連ルールの調査, (3) 活性/不活性化合物群構造式の視察, 等の作業が必要となる. 以下では, 各受容体のアンタゴニスト活性毎に, DISCAS システムから出力されたルールで重要と考えられる部分を示し, さらに考察を進めて見出された構造的特徴を説明する.

3.1 D1 アンタゴニストの特徴

この場合, 下記の表現を持つただ 1 つの主ルールとその関連ルール 5 種が提示された.

```
Rule 1: Cases: 413 ==> 91; BSS = 29.00
IF [C3-C4H-C4H-N3: y]
added on [C4H-C4H-C4H-N3: n] [C3-F1: n]
[C4H-C4H-N-C-C4H-N3: n]
THEN D1: 0.81 0.19 --> 0.24 0.76 (off on)
THEN D4: 0.44 0.56 --> 0.84 0.16 (off on)
```

主条件に現れる C3-C4H-C4H-N3 の部分構造が存在するならば, D1 活性化合物の比率が 19%から 76%へと大きく上昇することが分かる. また, 同時に D4 活性の比率が大きく低下する点も興味深い.

ドーパミン自体の構造式を図3のIに示す。主条件中のフラグメントはIのNH₂が3配位のNに変わった部分であるが、ベンゼン環を指示しているわけではない。実際にルールの条件を満たす活性化化合物群の構造を視察したところ、活性分子69の中でフラグメント中C3の部分は、ただ1つの含窒素5員環を除き他はすべてベンゼン環であった。従って図3のIIを、D1アンタゴニスト活性のリード構造と解釈するのが合理的である。なお、主条件を満たしながら、不活性な化合物にはC3の部分が複素環やオレフィンであるものが多い。

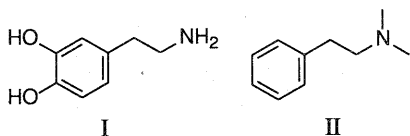


図3 ドーパミンとD1活性のリード構造
Fig.3 Dopamine and lead structure of D1 activity

3種的前提条件の中で第1と第3の条件はN原子を含むフレキシブルなフラグメントが存在しないことを示しているが、このようなフラグメントが存在すれば活性点への配位競合により、アンタゴニスト活性の発現を妨害すると考えられる。なお、2番目の前提条件のC3-F1フラグメントの解釈は不明であるが、他の前提条件と主条件を満たし、これを満たさない化合物9種には、すべて2配位のNを含む芳香環が存在していた。

このようにD1活性では、ルールの表現から非常に素直にリード構造の仮説を導くことができた。しかし、このルールはD1活性分子154種中の69分子を説明するに過ぎず、ridge情報や関連ルール群からさらに他のリード構造が発見される可能性がある。

3.2 D2アンタゴニストの特徴

この活性の解析では、活性と不活性を導く主ルール各1種が出力された。活性を導くRule2の重要な部分を下に示す。

```
Rule 2: Cases: 1227 ==> 277; BSS= 24.86
IF [C4H-C4H-O2: y] added on [ ]
THEN D2: 0.31 0.69 --> 0.61 0.39 (on off)
THEN C3-O2: 0.42 0.58 --> 0.88 0.12 (y n)
THEN C3H:C3-O2: 0.34 0.66 --> 0.70 0.30 (y n)
New:inside [C3H:C3H:C:C3-O2: y]
D2: 0.53 0.47 / 238 --> 0.81 0.19 / 128
New:inside [C3:C3-N3: y]
D2: 0.49 0.51 / 208 --> 0.93 0.07 / 60
New:inside [C3-C4H-N-C-C4H-N3: y]
D2: 0.14 0.86 / 183 --> 0.83 0.17 / 29
```

このルールには前提条件がなく、片側にアルキル鎖をもつエーテル(C4H-C4H-O2)が存在すれば0.61の割合で活性が現れることを述べている。しかし、その精度は低く、また専門家にとって、このようなエーテルだけでD2活性が発現しないことは常識である。そこで、この主条件の適用に伴って出現頻度が大きく変化するフラグメントを見ると、C3-O2とC3H:C3-O2の出現頻度がそれぞれ88%、70%へと上がっており、エーテルの他方は芳香族炭素へ連結していることが示唆される。

次にルール下部のridge情報を見てみよう。最初のridge情報は、芳香環を片側に持つエーテルが、主条件適用前後でそれぞれ238、128化合物存在し、その中ではそれぞれ53%、81%が

活性であることを示している。これは上記の付加的相関から推測された仮説を裏付けるデータである。また、2番目のridgeから芳香環にアミンが置換した構造が93%と高い活性を示し、さらに3番目のridgeのフラグメントを持つ化合物でも、主条件適用後の29化合物中で83%が活性を示す。まず、3番目のridge中の化合物群構造を視察したところ、活性なものはすべて下記IIIの構造を有していた。2番目のridge情報を考え合わせると、ortho位にアミンとエーテルが置換したIVがリード構造の候補として推測できる。実際に調べるとこの構造を持って主条件を満たす化合物49中45が活性であり、これがD2リード構造の1つであると結論できる。しかし、これだけでは説明できない活性化化合物群が2番目のridgeには含まれている。残った12の活性化化合物中で11種は、OとN間が6原子以上隔てられた下記Vの構造をもっていた。この構造はIVと同じ配位を可能とするものであり、これらが合理的なリード構造と考えられた。

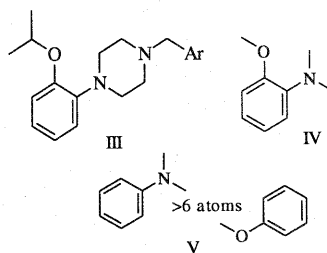


図4 D2活性のリード構造
Fig.4 Lead structures of D2 activity

次に不活性を導くRule1を下に示す。

```
Rule 1: Cases: 649 ==> 295; BSS= 26.14
IF [C3H:C3-C4H-N3: y] added on [C3-C4H-C4H-N3: n]
[C3H:C3H:C-N-C-C4H-N3: n]
THEN D2: 0.36 0.64 ==> 0.06 0.94 (on off)
New:inside [C4H-C4H-C-C-N-C3=O1: y]
D2 0.70 0.30 / 89 --> 0.21 0.79 / 24
New:inside [C3:C3H:C:C3-F1: y]
D2 0.65 0.35 / 118 --> 0.12 0.88 / 25
New:inside [C3H:C3H:C3-F1: y]
D2 0.62 0.38 / 124 --> 0.10 0.90 / 30
New:inside [C4H-C4H-O2: y]
D2 0.65 0.35 / 180 --> 0.12 0.87 / 40
```

このルールはD1活性のリード構造が存在しない場合で、かつ芳香環と3級アミン間のアルキル鎖長が1の場合には、D2活性がほとんど消失することを示している。これ自体も興味深い結果であるが、D2活性解析の立場からは、最初のridge領域に現れるアミド基が面白い。主条件適用前の89化合物中で62化合物が活性を示し、それらはVI、VIIの部分構造を持つものが大部分であった。これらはリード構造を示唆するものと言える。また、前記のRule2で活性を導くC4H-C4H-O2が最後のridge条件として現れており、ここでは本ルールの主条件がRule2のD2活性をうち消すように働いている。なお、F原子があらわれるridgeの意味は、D1活性の場合と同様さらなる検討が必要である。

3.3 D3アンタゴニストの特徴

この活性の解析からは、3種の主ルールとそれぞれに付随し

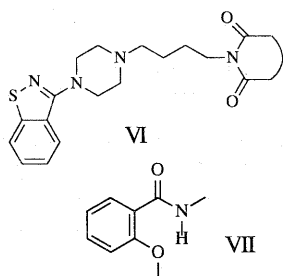


図5 D2 活性を示唆する構造

Fig. 5 Structures related to D2 activity

た 11, 5, 4 種の関連ルールが現れた。これらの中でもっとも明確なパターンを示したルールを次に示す。

Rule 1-UL4: Cases: 285 ==> 62; BSS= 25.90
 IF [C4H-C4H-C-C-N-C3=01: y]
 added on [C3H:C3-C-C4H-N3: y]
 THEN D3: 0.69 0.31 --> 0.05 0.95 (off on)
 THEN D1: 0.64 0.36 --> 1.00 0.00 (off on)

このルールの前提条件はまさに D1 活性のリード構造であり、VIII の様にフレキシブルなアルキル鎖につながったアミド基が共存すれば D3 活性が発現すると共に、D1 活性が消失することを示している。また、支持事例群構造の視察からは、アミド基と D1 リード構造が近すぎたり、コンフォメーションの柔軟性がないならば、D3 活性が発現していないことが判明した。

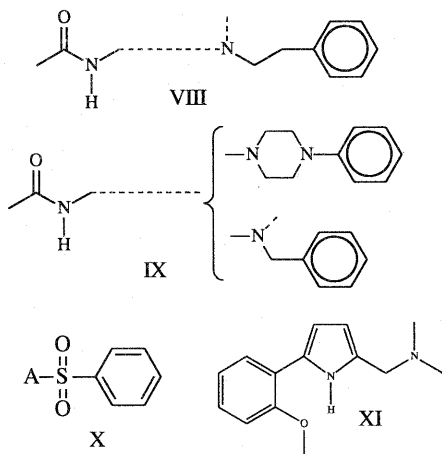


図6 D3 活性のリード構造

Fig. 6 Lead structures of D3 activity

主ルールの Rule1 は、Rule1-UL4 と同一の主条件を持ち、適用前後で 232, 54 化合物がそれぞれ重なっている。これら以外の活性化合物群の構造式を視察したところ、IX の構造が活性の要因であることが判明した。また、この場合には主条件と共に D4 活性が完全に消失している。なお、VIII, IX の構造をもっている、ベンゼン環の代わりに複素環がある場合、および NH₂ や CF₃ が置換しているものは不活性であった。

Rule 1: Cases: 727 --> 94; BSS= 42.91
 IF [C4H-C4H-C-C-N-C3=01: y]
 added on [C3:C3-02: n] [C3-N3-C4H: n]

THEN D3: 0.80 0.20 --> 0.13 0.87 (off on)
 THEN D4: 0.49 0.51 --> 1.00 0.00 (off on)

2 番目の主ルールとその関連ルールでは、下記の 2 種が興味深い。Rule2 で選択される化合物はすべて X を持ち、複素環で 2 配位の N を持たない。Rule2-UL2 が対象とする化合物群は上記と重なるものが多いが、X の化合物群を除外すれば、活性化合物のほとんどが XI の構造を持つ。これらは D3 活性のリード構造と解釈できる。3 番目の主ルールは、アミド基に関連するパターンを発掘しているが、Rule1 と重なる事例がかなり多く、未だ明確な結果を提示できる段階には至っていない。

Rule 2: Cases: 536 ==> 67; BSS= 41.33
 IF [01: y] added on [C3=01: n] [C3H:C3-N3: n]
 THEN D3: 0.82 0.18 --> 0.03 0.97 (off on)
 THEN C3:N2: 0.37 0.63 --> 0.01 0.99 (y n)

Rule 2-UL2: Cases: 1024 ==> 65; BSS= 29.00
 IF [C3:N3H: y] added on [C3:C3:N3H: n]
 THEN D3: 0.79 0.21 --> 0.12 0.88 (off on)
 THEN D2: 0.35 0.65 --> 0.05 0.95 (on off)

3.4 D4 アンタゴニストの特徴

この活性の解析からは、活性と不活性を導くそれぞれ 3 種と 2 種のルール (関連ルールは 1 種のみ) が得られた。以下活性を導くルールについて説明する。なお、不活性ルールでは、アミド基や D1 活性のリード構造 II の存在が主条件として現れている。活性を導くルール中でもっとも明確なパターンを示したものは、下記の Rule2 である。

Rule 2: Cases: 631 ==> 162; BSS= 40.41
 IF [C3-C4H-N3: y] added on [C3H:C3H:C-C4H-N3: n]
 [C3-C4H-C4H-N3: n]
 THEN D4: 0.57 0.43 --> 0.07 0.93 (off on)
 THEN D2: 0.44 0.56 --> 0.04 0.96 (on off)
 New:inside [C3H:C3:C3:N2: y]
 D4 0.44 0.56 / 135 --> 0.00 1.00 / 41

主条件を満たす化合物群の構造を視察したところ、XII のフラグメントを中心として全体として左右に長くかつ両側に bulky なグループを持つものがほとんどであった。ここで左側の共役系には、methoxy 基による置換や複素環等を含んで良い。ただし、アミド基の存在は活性を阻害する。2 番目の前提条件は D1 活性のリード構造が存在しないことである。これを持ちかつ主条件も満たす化合物群は丸い形をした分子であるか、あるいはアミド基を持つために不活性である。1 番目の前提条件に現れるフラグメントは主条件を部分構造として含んでいる。この非存在が前提条件として現れていることは、ridge で現れるフラグメントと考え合わせるなら、含窒素複素芳香環が D4 活性に関与している可能性を示している。なお、この主条件適用で D2 が不活性となることは、すでに D2 活性の項で述べた。

次に下記の Rule4 と Rule1 を見てみよう。Rule4 の主条件は Rule2 と似ているが、対象をベンゼン環に限っている。前提条件は XIII, XIV の部分構造を持つ化合物群を除外するためであると解釈できた。なお付言すれば、最初にアミド基が D4 活性を阻害すると述べたが、アミド基中の NH 水素が前提条件中のエーテル酸素と分子内水素結合を作る場合は、活性阻害の効果が現れない。

Rule 4: Cases: 856 ==> 156; BSS= 33.82
 IF [C3H:C3H:C-C-C4H-N3: y]
 added on [C3H:C3:C3-02: n]
 THEN D4: 0.51 0.49 --> 0.04 0.96 (off on)

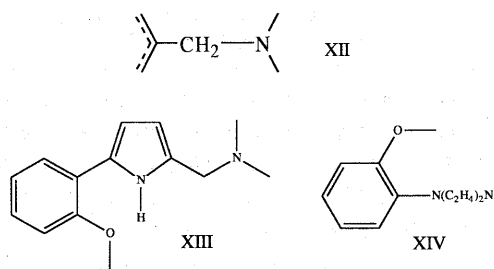


図 7 D4 活性にかかわる部分構造

Fig.7 Substructures related to D4 activity

```
Rule 1: Cases: 998 ==> 175; BSS= 44.27
IF [C3-C4H-N-C-C4H-N3: y] added on [C3-C4H-C4H: n]
THEN D4: 0.62 0.38 --> 0.11 0.89 (off on)
```

最後に Rule1 では主条件のフラグメントが、piperadino 基を含むように拡大されている。結果の構造の視察により、この場合もアミド基やエーテル酸素が活性を阻害する点は認められたが、前提条件の表現とは直接つながっていない。この条件の意味についてはさらなる検討が必要である。

4. おわりに

4種のドーパミン受容体に対してアンタゴニスト活性を有する化合物構造群を対象として、各受容体に対する活性発現に特徴的なリード構造を発見することができた。その結果、線形フラグメントの有無を属性として利用しカスケードモデルで解析することが、定性的な構造活性相関の理解に有効であることが示された。特に付加的な相関やridge情報は、ルール数の少なさとあまって、専門家が知識を発見するためのヒントを与えてくれることは確かである。

他方、ルール表現を使用した知識発見を効率化する観点から見ると、本報告での解析過程から明らかのように、ルールは知識に到達するための入り口に過ぎない。出力された情報を専門家が利用して解析作業を進めるわけであるが、専門家の知識に頼らざるを得ない点が多いのはやむを得ないとしても、専門家の作業手順さえ未だ明瞭でなく、試行錯誤を続けている状態である。今後、この過程を考察することにより、データマイニング技術自体を発展させ、効率的な知識発見を可能としていく予定である。

応用面では、今後 GPCR タンパク質群に関連したオピオイド、セロトニン等の生理活性を始めとして、MDDR データベース所載の多数の生理活性群に対して本方法を適用する。また解析結果を知識ベースとして蓄積し、WWW を通じて公開する予定である。さらに、本研究で構築したワークベンチを広く公開し、薬学者の参加をおおぐことによって、知識の集積を加速化することも計画している。

文 献

- [1] T.Okada, Discovery of Structure Activity Relationships using the Cascade Model: The Mutagenicity of Aromatic Nitro Compounds, *J.Comp.Aided.Chem.*, 2, 79-86 (2001).
- [2] 岡田孝, カスケードモデルとルール導出システム DIS-CAS, 関西学院大学情報科学研究, 15, 31-44 (2000).

<http://www.clab.kwansei.ac.jp/okada/www/contents2k/discas.pdf>

- [3] T.Okada, Datascape Survey using the Cascade Model, in K.Satoh (ed.) *Discovery Science 2002*, LNCS, Springer-Verlag (2002).
- [4] T.Okada, Topographical Expression of a Rule for Active Mining, in H.Motoda (ed.) *Active Mining*, 247-257, IOS Press (2002).