

KP10 インテグレートドインシリコスクリーニング (V):

ヒト P 4 5 0 予測モデルの構築と考察 (クラスモデル)

[富士通, 富士通九州システムエンジニアリング]

湯田 浩太郎, Ciloy Martin Jose, 北島 正人

1. はじめに

我々は失敗の少ない創薬を行う¹⁾という観点で、単に薬理活性のみのスクリーニングを行う1次元スクリーニングでは不十分であると考え、薬理活性に加えてADME / 毒性 / 物性等を総合的に評価することの可能な2次元インシリコスクリーニング、すなわち「インテグレートドインシリコスクリーニング」⁽²⁾を提案した。

今回はADME分野で注目されているP450に関する予測モデル(クラスモデル)の作成を試み、この過程で様々な有意義な知見を得たので報告する。

2. 本論

2.1 インテグレートドインシリコスクリーニング概要

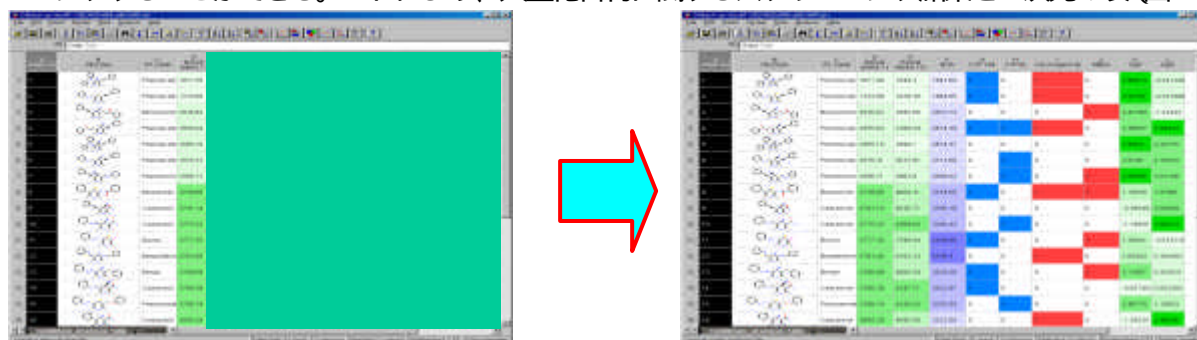
我々が行うインシリコスクリーニングは多変量解析 / パターン認識を基本とする。本アプローチによるインシリコスクリーニングの3大特徴は以下のようなものである。

実行速度が速いので、大量化合物のスクリーニングを行うことが可能である。

構造変化性の高い(基本骨格等の異なるレベル)化合物群を扱うことが可能である。

薬理活性 / ADME / 毒性 / 物性等の様々な分野に関する同時予測も可能。

これらの特徴により、理想ではあるが従来技術では実現困難であった“複数項目の同時スクリーニング”の実現が可能となる。すなわち薬理活性、ADME、毒性、物性等をそれぞれ複数項目について同時スクリーニングすることができる。これにより、大量化合物に関するスクリーニング結果を2次元の表(図1(b))と



* 1次元スクリーニング (a)

複数化合物 × 1薬理活性

* 2次元スクリーニング (b)

複数化合物 × 複数薬理活性 × 複数毒性 × 複数物性

図1. 薬理活性のみの(一次元)スクリーニングと薬理活性/ADME/毒性/物性等を同時評価するインテグレートド(二次元)インシリコスクリーニングの違い

して俯瞰でき、化合物に関する種々特性の総合評価が可能となる。この2次元インシリコスクリーニングの実施により 開発最終過程でADME特性の不適合や毒性発現等による開発中断の起こりにくい創薬を実現できる。これが我々の提案する「インテグレートッドインシリコスクリーニング」の本質である。

2.2 科学に基づいたデータ解析アプローチと単純分類アプローチ

解析の基本 :情報等価原理

インシリコスクリーニングを行うためには予測モデル (判別関数 重回帰式、他)が必要である。この予測モデルを作成する過程で科学的な情報を盛り込んで作成するのか、あるいは科学的な情報とは関係なく、与えられたデータの分布情報のみに基づいて単純な分類モデルとして作成するかで予測モデルの予測信頼性 (特に外挿性)や、予測モデルから得られる情報の種類や質が大きく異なってくる。

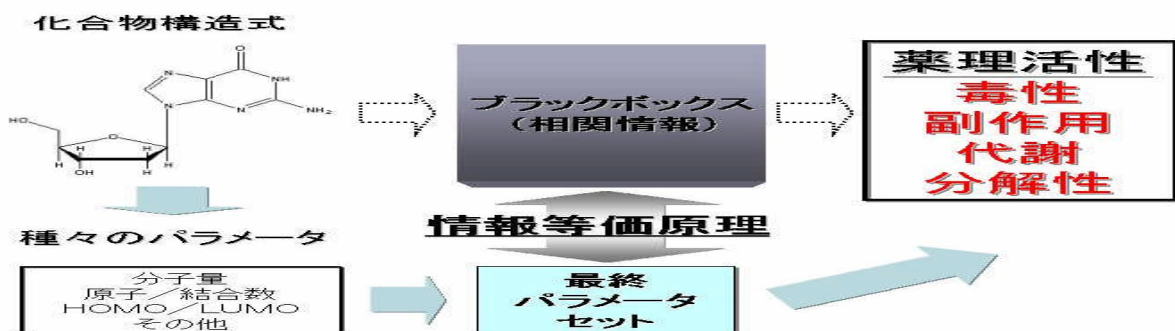


図2. 多変量解析/パターン認識による構造-活性/ADME/毒性/物性相関実施上の基本原理図

本分野の研究による基本原理は「情報等価原理」である(図2)。すなわち、図2中にブラックボックスとして表現される解析目標項目と相関する情報と、データ解析に用いる最終パラメータセットが有する情報が等価であるという原理である。従って、最終パラメータセットには解析目標項目である薬理活性/ADME/毒性/物性等と強い相関を有するパラメータ群が選択されていることが必要である。このパラメータセットに目標活性/ADME/毒性/物性等と何ら相関のないパラメータが含まれているならば、それらはデータ解析上での単なるノイズ情報であり、これらのノイズ情報を含んだままで要因解析や外挿を伴う予測を行うことは極めて危険である。以下に理由を述べる。

予測モデル構築時における、サイエンスに基づくデータ解析と、単純分類の問題

本分野で適用されるデータ解析は大きく、クラス分類およびフィッティングの二つに分類される。これらの問題のうちフィッティング(線形重回帰および非線形重回帰)に関する議論は本講演要旨のJ08で行っているので、そちらを参照されたい。本稿ではクラス分類問題に関して簡単にまとめる。

予測に用いるパラメータ群が予測目的(薬理活性/ADME/毒性/物性等)と何ら相関がない場合、これらのパラメータ(以下、ノイズパラメータ)群により形成されるD次元空間上にサンプルA(活性化化合物)およびN(非活性化化合物)が無秩序に分散して存在する状態となる(図3(a))。しかし、ノイズパラメータ群を取り除き(特徴抽出)解析目的と相関のあるパラメータ群のみにすると、サンプル化合物群はM次元空間上にA化合物群およびN化合物群との2グループに分かれて存在する状態となる(図3(b))。このパターン空間を形成するM個のパラメータは解析目的項目と化合物構造との間に存在する重要な相関情報を有することとなる。サイエンスに基づく信頼性の高い解析やスクリーニングを行うためには、図3(b)のようなパターン空間を形成するパラメータ群を、AやNが無秩序に分散する空間(図3(a))を形成するパラメータ群から取り出す手続き(特徴抽出)が最も重要なものとなる。

現在は様々なデータ解析手法が展開され、極めて強力な分類手法も多数展開されている。このような強力な分類手法を用いると(図3(a))のようなノイズを含んだ空間でも簡単に100%分類可能である(図3(a)中の点線(赤)部分)。たとえ100%分類出来たととしても、このようなサイエンスに基づかない空間の分類から得られる情報量は少なく、得られる情報の信頼性も外挿性も低い。最悪の場合は誤解釈する可能性もあるので、このような解析手法の適用には細心の注意が必要である。

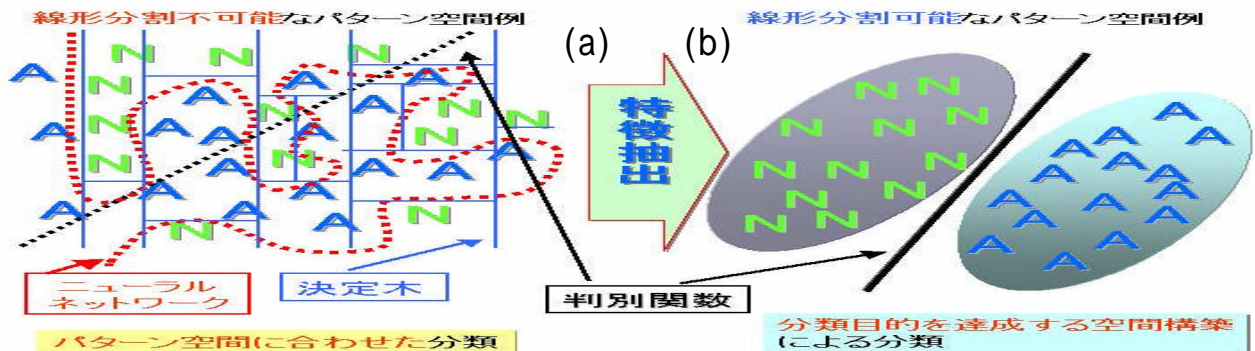


図3. 解析目標項目と相関のないパターン空間の単純分類と、相関のあるパターン空間の分類概念

データ解析を行う場合は解析手法の特徴(適用範囲、限界等)を十分に理解した上で予測モデルを構築することが、信頼性の高い予測モデルの構築や構造-活性/ADME/毒性/物性相関に基づく予測を行うという観点で極めて重要である。

我々は上記観点から、AやNがバラバラに存在するノイズの多い初期パラメータ(図3(a)の状態)空間を解析対象とせず、図3(b)のようにきれいな2群に分割されたパラメータ空間を解析対象とすることとした。この図3(b)のパラメータ空間を形成するパラメータ群を取り出すために、図3(a)を形成する初期パラメータ群に様々な特徴抽出手法を適用し、多数のノイズパラメータ群から解析目的に極めて相関の高い重要なパラメータのみを取り出した。最終的に図3(b)のような2群に分割されたパラメータ空間を形成したうえで予測モデルを構築した。この予測モデルより得られる情報を解析し、構造-ADME相関に関する要因解析を行った。

2.3 P450に関する予測上での問題点

P450は多数の酵素群から形成される酸化酵素の総称である。遺伝子配列の違いにより多数のファミリー群に再分割される。人P450ファミリー中、最多数の化合物群の酸化代謝に関与する酵素としてはCYP3A4がある。これはP450が関与する代謝化合物の約60%に関与すると報告されている。

2.4 解析目的

我々はCYP3A4のKm値を用いて線形重回帰による予測モデルを構築し、未知化合物のKm値予測に関する考察を行った(本講演要旨J08)。本発表では解析目的として新規化合物がCYP3A4の阻害剤となるか、あるいはCYP3A4以外のファミリー酵素の阻害剤となるかを判定するための予測モデルの構築に関する検討を行った。

2.5 使用データ

解析に用いたデータはザグレブ(Zagreb)大学のレンディッチ(Rendic)教授が構築したP450データベース⁽³⁾から取り出されたデータを用いた。(図4)

解析に用いたデータはCYP3A4阻害剤(112化合物)およびCYP3A4以外のファミリー酵素の阻害剤(122化合物)の、総計234化合物を用いて解析を行った。図5に解析に用いた化合物の一部を示す。

■使用データ(Zagreb Univ.)
 Prof. Slobodan Rendicのヒト薬物代謝データベース
 -7, 000件以上のCYP代謝データ
 -3, 800件以上の化合物
 -2, 200件以上の文献
 -*in vivo* and/or *in vitro*データ

■使用システム
 ADAPT (Automated Data Analysis and Pattern recognition Toolkit)

■実施手法
 多変量解析/パターン認識によるアプローチ

図4.P450データベース

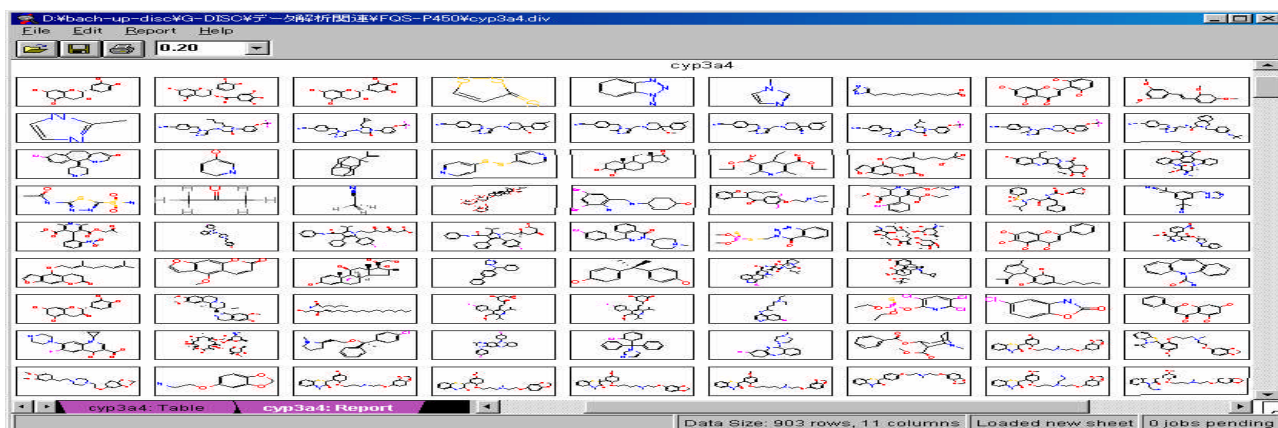


図5.モデル構築に用いたサンプル化合物群

3. データ解析

実際のデータ解析はADAPT⁽⁴⁾ (Automated data Analysis using Pattern recognition Toolkits)を用いて行った。最初にサンプル化合物構造式に基づいて約35種類のパラメータを発生させた。この初期パラメータの殆どは解析目的である化合物をCYP3A4とCYP3A4以外のファミリー酵素の阻害剤か否かを判別することとは何の相関もないパラメータ群(図3(a)参照)である。引き続いて、様々な特徴抽出手法をこの初期パラメータ群に適用し、サンプル化合物群を前記2グループにきれいに分割する最終パラメータ群(図3(b)参照)を取り出した。この重要なパラメータ群を用いてサンプル化合物群を2分割する線形判別関数を求めた。

[結論]

化合物群をCYP3A4の阻害剤かCYP3A4以外のファミリー酵素の阻害剤となるのかを識別する予測モデルを構築することが出来た。本アプローチによる解析では、最終的に得られた線形判別関数を形成するパラメータ群の要因解析をすることで、化合物構造との相関関係を議論することが可能である。これらの詳細な結果はポスターにて詳しく発表する。

[参考文献]1) T. Kennedy, Drug Discovery Today, 2, p.436 (1997).

2) 北島 正人、スイロイ ホセ、湯田 浩太郎 : ‘薬理活性およびADME-Toxを同時評価するインテグレートド高速/仮想インシリコスクリーニング(I)’、第3回CB学会大会講演要旨集、B5-4、東京、2002年

3) S. Rendic and F. Di Carlo : Human Cytochrome P450 Enzymes: A status report summarizing their reactions, substrates, inducers, and inhibitors. Drug Metabolism Reviews, 29, p.413 (1997).

4) Jurs研究室ホームページ :<http://research.chem.psu.edu/pcjgroup/index.html>