



構造活性相関部会・ニュースレター &lt;31 March 2004&gt;

# SAR News No.6

## 「目次」

### [記事]

#### /// QSAR 今昔///

私と QSAR ー傍観から実践への道程ー 西村 勁一郎 … 2

#### ///研究紹介///

能動学習法を利用したペプチドワクチン開発 宮川 知也 … 5  
 機械学習による薬物化学構造からの活性クラス識別 高橋 由雅  
 錦織 克己  
 藤島 悟志 … 10

### [報告]

- ・ 第 31 回構造活性相関シンポジウム 高橋 典子 … 14
- ・ 第 31 回構造活性相関シンポジウム・ポスター賞の研究紹介
  - KP14 分子重ね合わせに基づく活性化合物解析システムの開発  
 野中 はるみ, 富岡 伸夫, 板井 昭子 … 15
  - KP19 標的蛋白質の Induced Fit を考慮したリガンドドッキング(1):  
 ブラウン動力学法を用いた複合体構造最適化  
 山乙 教之, 広野 修一 … 18
  - KP28 SOM フィルターを用いた QSAR モデリング  
 秋元 紗恵, 佐々木 英史, 高橋 由雅 … 21

### [お知らせ]

- ・ 構造活性相関フォーラム 2004 … 22
- ・ 第 32 回構造活性相関シンポジウム … 23

//// QSAR 今昔 ////

## 私と QSAR

## —傍観から実践への道程—

大阪府立大学先端科学研究所 基礎科学部門 西村勁一郎

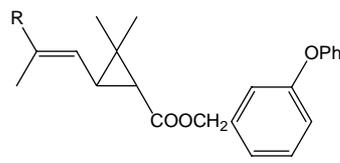
農薬の作用機構のことも知らず、「農薬の作用機構の実験をしたい」と言っただけで京大農学部農芸化学科農薬化学講座の門をたたいたのは、今から40年近くも前のことである。戴いた卒論テーマは、尿素系除草剤の合成と Hill 反応阻害活性の測定であった。学部を卒業して修士課程に入ることになってテーマを決めるとき、まだ懲りずに「農薬の作用機構の実験をしたい」と言ったものである。その年から藤田稔夫先生が同講座の助教授に就任され、筆者のテーマについての相談が先生方によって行われた。私自身はあまりよく解らないままに藤田先生のご指導で s-トリアジン類の合成から始めることになった。それなりに合成も進んで化合物も幾つかできていたが、どうも満足感がない。一体これって何のためにしているんだろうと。藤田先生は私が実験していたところとは少し離れた部屋におられ、訪れて聞くのも敷居の高い思いをしているままに夏頃まで日ちが過ぎ去った。その頃になって、筆者が4年生時代に行っていた実験の続きを行うことになった。あの頃に藤田先生にもっと教えをお願いしていたら今の私はなかったのは間違いない。どんな道が待っていたかは解らないが。歴史には「もし」がないとはよく言い得たものと思う。

話は飛んで、博士2年を終わった時点で農薬化学講座の助手に採用していただいた。博士課程の途中であったのでまずは博士号の取得を優先して実験を進め、2年後には学位を戴いた。博士論文の題目は農薬とは全く違ったため、それから2年半ほどは学位のための実験でやり残したところとか、修士課程の時にやり残した実験などを行っていた。その時の担当教授から「博士課程にいたU君も就職したし、彼が使っていた神経の電気生理の装置を使いこなせるようにアメリカへ留学して勉強して来なさい」と言われた。神経生理のことはちんぷんかんぷんだったが、これ幸いと二つ返事で先方の先生へのご連絡をお願いした。これが筆者のテーマの除草剤から殺虫剤への転換点となった。

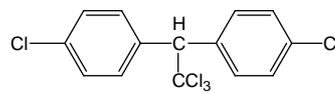
留学先から帰ったのが今から28年前。行く前に修士課程に入ったばかりだったN君が、天然物である殺虫性ピレスロイドの構造変換ですでに実用化されていたフェノスリン(1, R=Me)をリードとして、PhO基を他の置換基に変換した化合物や位置異性体を多数合成していた。合成は順調に進んでいるようだった。彼の殺虫試験や外部電極法による神経試験にはワモンゴキブリを使用し、他にもワモンゴキブリを使う学生が2人ほどいた。それまで昆虫は農学部内のある先生方のご厚意で提供を受けていた。しかし時には数が不足することもあったので自ら飼育することにし、その累代飼育が約20年間も続いた。

前書きが随分長くなったが、上記したピレスロイドを合成していた学生たちへの関わりが、筆者のピレスロイドとワモンゴキブリとの実際上の出会いであった。その時すでにN君は藤田先生のご指導で、QSARとして「構造-活性相関」と「活性-活性相関」の解析を行っていたことを想起する。しかし筆者には当時、その具体的な内容についてはほとんどわからなかった。ただ「構造-活性相関」の解析において立体パラメータとしては van der Waals 体積 (Vw) から評価した  $\Delta Vw$  (H原子を基準としたときの各置換基の Vw の増加分) がうまく使えることは聞き及び、それなりに理解していた。N君が博士課程に進学して半年ほどしてから、筆者とともに実験を開始した別の学生が、N君の合成したフェノスリン(1, R=Me)系化合物と、筆者が留学中に経験したザリガニ神経を用いる外部電極法の一手法の組み合わせで神経実験を開始した。データが揃い

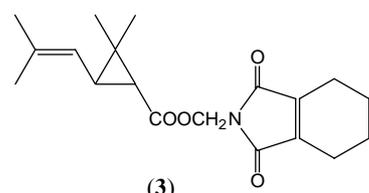
始めたとき、その学生とともに藤田先生のご指導を受けて QSAR を開始した。それが私の QSAR との実質的な出会いであった。当時はまだ京大では端末を使える時代ではなかったため、SAR News 第3号で山上知佐子先生が書かれたように大型計算機センターへ行ってデータをパンチカードに入れる作業が必要であった。というより、当時のデータの入力作業も結果の解釈もN君が行ってくれた。筆者は



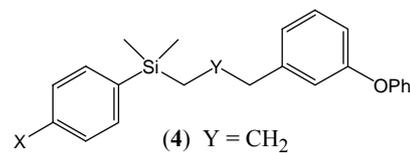
(1)



(2)



(3)

(4) Y = CH<sub>2</sub>

(5) Y = O

QSAR については付属品以外の何者でもなかった。実際に大型計算機センターへ行って自分で計算作業を行うようになったのは、N 君も卒業してさらにピレスロイドの計算作業を受け継いだ別の学生も卒業する直前であった。その時には筆者も自分でやらざるを得なくなったわけである。手ほどきしてくれたのはその学生と彼の仲間であった。その後間もなく研究室でも端末を使えるようになった。特に年度末には学生たちの卒業研究の完成期である。そのような時期にわざわざ大型計算機センターへ行ってデータの入力機を確保するために、朝早くから出かけて席を取る必要がなくなったのは感激であった。それでもそのうちに回線が込みだして、なかなか電話が通じないことも屢々であった。その後パソコンで計算をできるようになって、利用しか能のない筆者はさらに感激したものである。

相前後して、神経の電気生理学においてはピレスロイドと似た作用機構を持つ DDT (2) 誘導体の合成と生理活性の測定も並行して行った。化合物は藤田先生が助教授に就任された当初の学生が合成したものと、筆者と学生が新たに合成したものを使った。ある種のピレスロイドも含めて、それらのオクタノール-水系での log P の測定は UV 分光光度計を用いて自ら行う必要が生じた。拙著を繰ってみると、DDT 類のうち実測したものの中で最も大きな log P 値を持った化合物は無置換体で、その値は 4.87 であった。4 台後半の化合物は他にも 4 個あった。これほど大きな log P 値を持つとなかなか検量線を直線として得られなくて、測定を行うごとに異なったものになったのを悪夢のように回想する。その原因は操作中に混入する夾雑物が水層に残り、それが吸光度の正確な測定を妨げたのである。どれほどの期間をかけて四苦八苦したのか忘れてしまったが、とにかくスポイドやメスフラスコ、分配用の振とう瓶など使用するガラス器具類すべてを、あらかじめ蒸留したエタノールで何回も丁寧に洗うのが一番良かった。特にすりあわせ部分がくせ者であった。多分、洗剤などがごく少量残って測定に悪影響をもたらしたものと思われた。無置換の DDT 類縁体の log P の実測値に  $\pi(\text{Cl}) \times 2 = 1.42$  を足して DDT の log P 値を 6.29 とした。それ以前にある卒業生は標識炭素を持つ DDT を使って、振とうフラスコ法で log P 値を測定していた。その値は筆者が測定した値より低かった。その当時その卒業生と話してみると、放射性同位元素で標識した化合物を使った場合には log P の小さい夾雑物があるとそれが水層に移り、log P の実測値を小さくするというコメントを得た。目的は異なったが、後日になって筆者は標識された DDT の薄層クロマトグラフィーを行う機会があった。実際にその試料には標識された夾雑物がいろいろ含まれていることを認め、その数年前に経験した log P 値の食い違いを思い出したものである。

その頃に行った DDT 系列の神経活性についての思い出が一つある。DDT 系列と言ってもそのほかに DDD、プロラン、DCC などの系列も相前後して扱った。QSAR と生理活性の測定を並行して行っているとき、DDD とプロラン系列の解析は順調に進んだのに、DDT 系列のある化合物の神経試験において何回繰り返し測定しても効果が出ない。と言うことは活性値が出ない。やむを得ずこの化合物を除いて解析すると他の 3 系列の化合物とは統一して解釈できる結果が得られた。しかしこのままにしておくのも非常に気になることであったので、QSAR の結果から予想できる辺

りの濃度でしつこく実験を繰り返した。するとあるとき期待する効果が出た。それも一回だけではなく何回も繰り返して。このような効果はこの化合物で初めて経験したことではなく、ピレスロイドでも何回か経験していた。こんな経緯によって、この化合物もめでたく QSAR 解析に含めることができた。とは言うものの、今でもまだ釈然としないものが心のわだかまりとして残ったままである。当時も今も筆者は一応納得しているが、あまり後味の良いものではない。

他のピレスロイドの QSAR も多く経験した。フェノスリン (**1**, R=Me) の酸部分を第一菊酸と呼ぶが、それと同じ酸部分構造を持つテトラメスリン (**3**) の各種イミド類縁体。構造式 **1** において R=COOMe を持つ酸部分 (第二菊酸) の置換ベンジルエステル類。年代は前後するものもあるが、その他にカデスリンやデルタメスリンの酸部分や、シクロプロパンやシクロブタンを持つカルボン酸と置換ベンジルアルコールとのエステル類。さまざまなカルボン酸と *m*-PhO-ベンジルアルコールとのエステル類などなど。当初より立体パラメータとして有効に用いてきた  $\Delta V_w$  より、CPK モデルから算出される STERIMOL パラメータのうちの、置換基の最大幅や最大長の方が有効であることがわかった。そのような傾向は上記の DDT 系列についての解析の時から見え始めていた。上述した合成ピレスロイドの原型である化合物 **1** や **3** のようなエステル類だけでなく、その後開発された含ケイ素アルカン (**4**) やエーテル (**5**) についても解析した。時には合成原料を、また時には完成品を藤田先生を通じて多くの企業から提供を受けた。

ピレスロイドの QSAR を始めた当初から一つ不可解なことがあった。その殺虫活性-外部電極神経活性相関の解析において、独立変数として  $\log P$  の 1 次と 2 次項を併せて入れるといつも相関関係が有意に改善された。そしていつも 2 次項の係数は正であった。素直に考えると殺虫活性にとっては  $\log P$  の値に「最悪値」が存在することになる。それを正当化するために、作用点とは異なるところに化合物が蓄積する部位があって、そこでの蓄積効果と作用点での濃度を高める効果が拮抗するためだろうと解釈してきた。正しいかもしれないしそうでないかもしれない。筆者には今でもこのことについて釈然としないものが残っている。

そのようなとき、筆者はピレスロイドや DDT の作用機構研究に近づくため、外部電極法や内部電極法のような初歩的な方法による神経活性の測定より、神経効果としてもっと本質的な作用として知られている Na 電流についての測定を行える膜電位固定法を取り入れたいと考えた。自ら実験をするためには、入手の容易な実験材料を用いて素人でも採用できる方法を見つける必要がある。ある大学でそれをさせていただけることがわかって、学生にその測定法を学んで実践してもらった。しかしどうも進捗が芳しくない。実験方法に問題があると考えた。彼の場合には学位がかかっているのでその実験はそのまま続け、筆者自身で行える他の道を探すことにした。藤田先生のご理解を得て、再度 3 ヶ月間訪米することになった。その時のことにもいろいろの感動談があるがこの拙文の本題ではないので割愛する。そこで学んだ方法を帰国後に取り入れて、その結果について QSAR を行った。いろんなおもしろいこともわかった。もっと本質に触れられる方法はパッチクランプ法だと思って、それを他の学生にやってもらった。しかしそれまでの電気生理学的実験の場合には自分で学び自分で実践した方法であったので、学生指導も自分の目線で行えた。しかしパッチクランプ法については耳学問であったので指導は順調ではなかった。そうこうしているうちに私も現職に就くことになり、装置を譲り受けて現在の勤務先へ運び込んだ。しかし如何せん、それを作動させるには筆者の手に負えず、そのうえ指導する学生もほとんどいなくなってそのまま放置する結果になった。

現職に就く数年前に藤田先生を通じて入手したネオニコチノイドの殺虫活性と、一種の外部電極法による神経活性試験をワモンゴキブリの神経を用いて測定し QSAR を行った。その後も学外から依頼を受けたこの系列の化合物について上記と同じ方法による測定を自ら行ったり、依頼先から派遣された学生による測定の手ほどきをしたりしている。その結果について QSAR 解析を行うこともある。このようなことを昔取った杵柄をとると言うのでしょうか。振り返って、筆者の QSAR についての実践があまりにも皮相であったことを今さらのように思いつつ筆を置くことにする。

## //// 研究紹介 ////

## 能動学習法を利用したペプチドワクチン開発

日本電気株式会社 市場開発推進本部 バイオ IT 事業推進センター 宮川 知也

## 1. まえがき

バイオテクノロジーの多くの分野、例えば新薬開発などにおいては、薬（化合物）の合成と試験という実験による探索が必要だが、一般に探索すべき対象の範囲は極めて広く、多数の実験候補からいかに効率的に有用な候補物質をスクリーニングできるかが、研究・開発現場の重要な課題となっている。

一つの新薬開発には、平均して約 300 億円の費用と約 10 年の期間が必要と言われており、その高コスト及び長期間が必要である原因の一つは、膨大な候補の中から有用な候補物質を探索するスクリーニングにあると言われており。従って、極めて多数の実験が必要な場合などに、ごく少数の実験結果から残りの実験結果を予測することができれば、実験コストや時間の大幅な節約が可能になる。

本稿では、コンピュータにおける機械学習法の一つである「能動学習法」を利用することにより、医薬品の一種であるペプチドワクチンの開発を効率化する方法を紹介し、具体的にどのような疾患への適用が可能であるか、将来的な見通しについても概観する。

## 2. 「能動学習法」によるペプチド結合実験の効率化

## 2.1 ペプチドワクチンとは

ウイルスや細菌などの病原体に生体がさらされると、病原体由来のたんぱく質は生体の細胞内で分解されてペプチド断片となり、抗原提示分子（ヒトの場合には、HLA (Human Leukocyte Antigen) がその役割を担う）に結合し、細胞外に抗原として提示される。血液中に存在し免疫機能を調節している T リンパ球細胞には、様々な抗原に対応すべく多様な特異性を持った細胞群があるが、HLA 分子に結合した病原体由来のペプチドを偶然に認識することができれば、この抗原を認識する細胞障害性 T リンパ球細胞の増殖を始めとする免疫機能が亢進し、再度の病原体進入時には、迅速に病原体を排除することができる。

ワクチンとは、こうした生体の機能を活用したものであるが、あらかじめ HLA 分子に結合する病原体由来のペプチド断片を予測することができれば、極めて効率的で副作用が殆どない、ペプチドワクチンとしての有望な候補物質となる（図 1）。ヒトの HLA 分子に結合するペプチドのアミノ酸配列を調べるために、人工合成ペプチドなどを用いた様々な種類のペプチド結合実験が行われている。

HLA 分子には百種類以上にも及ぶ型が知られているが、そのうち、HLA-A 型に結合するペプチド断片はアミノ酸が 9 個連続したものであり、その順列組み合わせは約 5000 億通りにもなる。一方、実際の結合実験は 1 年間で数百個程度が限度であり、10 年をかけても約 1 億分の 1 程度の実験しかできないことになる。従って、いかに少ない回数の実験から、求める特性を持ったペプチド配列を見つけるかが極めて重要となる。

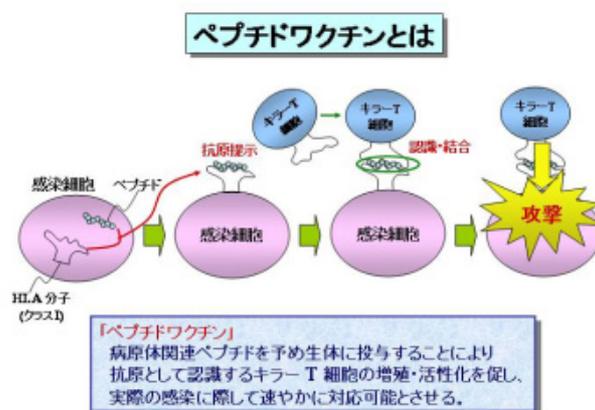


図1 ペプチドワクチンとは

## 2.2 「能動学習法」とは

「能動学習法」とは、少数（数十個）の実験結果を訓練データとして、実験の条件と結果との間にあるルールを複数の学習プログラムで学習し予測させる、機械学習法の一つである。個々の学習プログラムが出した予想結果の中から、予想の割れたものを「質問」として実験者に返し、測定実験をしてもらう。その実験結果を受けて、再度、複数の学習プログラムで学習をして予測させる。本研究開発では、学習機械の実装アルゴリズムとして、隠れマルコフモデル(HMM)を使用した。

実験者と同学習システムの間でこのやりとりを数回繰り返すことにより、システムは、少数のデータから非常に高精度な予測をすることができるようになる（図2）<sup>1,2,3</sup>。

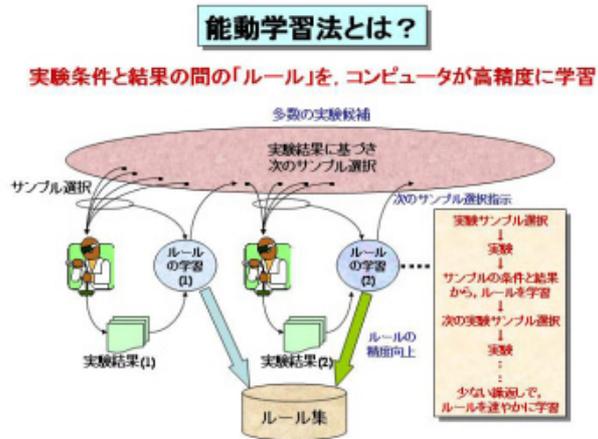


図2 「能動学習法」とは

## 2.3 結合ペプチド予測への「能動学習法」の適用

ヒト HLA 分子に結合するペプチド予測をするに先立ち、マウスの抗原提示分子である MHC (Major Histocompatibility Complex) 分子に結合するペプチド予測への「能動学習法」の適用を行ったが、これは、京都大学理学部宇高助教授との共同研究により進められた<sup>4</sup>。

具体的には、ある特定のマウス MHC 分子を対象を選び、まず最初に、学術文献などでその MHC 分子に結合することがわかっている、アミノ酸が 9 個連なったペプチド約 180 個を人工的に合成し、マウスの特殊な培養細胞を用いたペプチド結合実験アッセイ系を用いて、MHC 分子に対する結合定数を測定した。

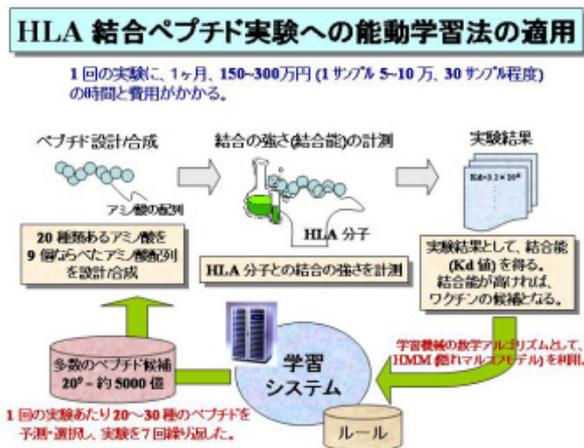


図3 ペプチド結合実験への「能動学習法」の適用

各ペプチドのアミノ酸配列とその結合定数の情報を最初の学習データとし、それよりランダム選択したデータセットを 50 個作成し、50 個の学習プログラム学習データとし、各学習プログラムで実行する。実験すべき候補選択の為にランダムにペプチド配列を作成し、50 個の学習プログラムを用いて結合定数を予測する。50 個の学習プログラムの結果で最も予測結果が割れたペプチド配列(50 個)選択し、次の実際のペプチド結合実験を行う。この結合実験の結果を次の訓練データとして積み重ね、再度学習をし、その次に実験をすべき 50 個のペプチド配列を選択する。

このやりとりを 7 回繰り返すことによって、予測精度の向上を図り、最終的にマウス MHC 分子に対する結合ペプチド予測プログラムを完成した（図3）。

## 2.4 結合ペプチドの高精度予測

このようにして作成した予測プログラムを用いて、ランダムなアミノ酸配列を持った 100 万個のペプチドに対して結合能の予測を行い、結合能が高いと予測された上位 1% の中から、人工合

成がしやすい任意の 41 個のペプチドを選抜した。これら 41 個の合成ペプチドを用いて、マウス MHC 分子への結合実験を行った<sup>4</sup>。

一般に、MHC 分子に対するペプチドの結合実験では、結合定数  $K_d$  値が $-6$ よりも小さい場合には、マウスでの T 細胞の増殖および活性化などの免疫機能の向上を伴うぐらいまで、MHC 分子に対する抗原ペプチドの結合強度が大きくなると予測されている。結合実験の結果、予測された 41 個のペプチドのうち、約 70~80% のペプチドが $-6$ 以下の  $K_d$  値を示した<sup>4</sup>。

従来、MHC 分子への結合ペプチドを予測するプログラムについては、結合に重要とされるアミノ酸部位（モチーフともいう）に注目したアンカー法などに基づいた、いくつかのプログラムが公開されているが<sup>5,6,7</sup>、その予測的中率は数%~30% 程度にとどまり、この「能動学習法」を利用した予測プログラムの的中効率の高さは特筆される。

### 3. ヒト HLA 抗原提示分子結合ペプチドの予測

続いて、ヒトの抗原提示分子であり、マウス MHC 分子と類似の機能を持つヒト HLA 分子に結合するペプチドを予測するプログラムの作成に着手した。

本予測プログラムの開発は、高知大学医学部免疫学教室の宇高教授との共同研究の形で進んでいる。

#### 3.1 ヒトを対象とした結合ペプチド予測の意義

ヒト HLA 分子に結合するペプチドを効率よく予測することができれば、それはすぐに、ヒトの疾患治療を目的としたペプチドワクチンの候補としての開発を進めていくことが可能となる。

従来用いられてきたワクチンでは、例えば感染症に対する予防効果を持ったワクチンを開発する場合、十分に T 細胞が認識できるような抗原を提供することが困難であるため、無毒化した生きた病原体やその構造体の一部などを材料として、ワクチンを作成することが中心となっている。しかし、無毒化したはずの毒性作用が復活してしまう例や、病原体の構造体の一部などが、人体に対して思わぬ副作用を引き起こすなど、しばしば重篤な副作用が報告されてきた。

更に、例えばインフルエンザワクチン開発の場合などでは、毎年、次の年に流行するインフルエンザの型を予想して、次年度分のワクチン開発を行っているが、この予測がしばしば外れてしまうなどの問題がある。

ワクチンを人工合成ペプチドという構造体で作成することが可能になれば、ペプチド以外の病原体由来の構造体を全く含まないことから、毒性の復活や思わぬ副作用の発生を防ぐことができるばかりか、製造コストが格段に安くなる上、予測結果からすぐに臨床試験に入れるなど、タイミングよく迅速なワクチン開発が可能になると考えられる。

#### 3.2 ヒト HLA 結合ペプチド実験と「能動学習法」の適用

ヒト HLA 分子に結合するペプチドを予測するプログラムの作成を目的として、この度 2003 年 7 月より、高知大学医学部免疫学教室の宇高教授との共同研究を開始した。

高知大学医学部の側で、ヒトの培養細胞などを用いたペプチド結合実験を実施し、NEC 側ではペプチド結合実験のデータを受けたペプチド予測プログラムの訓練と、次に結合実験をするべきペプチド配列と結合定数のリストを返答するという作業を繰り返している。本予測プログラムは、2004 年 6 月ごろには完成の予定である。

#### 3.3 ペプチドワクチンとしての開発可能性

ヒト HLA 分子には、A 型、B 型、C 型、DR 型、DQ 型、DP 型といった大きな分類も含めて百数十種類にも及ぶ型が存在し、個人個人が持つ HLA の型の違いが、例えば臓器移植の際の拒絶反応の原因とされている。

その中でも抗原提示に重要な働きを持つことが示されている HLA-A の型には、所属する人種に基づいた発現分布の偏りがあり、例えば日本人では、HLA-A24 型を持つ人が全体のほぼ 50% を占めるのに対して、欧米人では HLA-A2 型を持つ人が一番多いことが確認されている（図 4）。

そこで今回のヒト HLA 分子結合ペプチド予測プログラムの作成にあたっては、まず、この HLA-A24 型および HLA-A2 型に対する予測プログラムの作成から着手している。その完成後、順

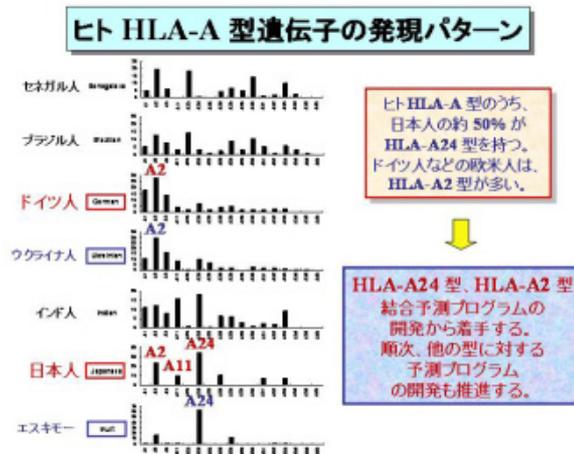


図 4 ヒト HLA-A 型遺伝子の発現パターン

また、免疫機能の活性化が確認されたとしても、実際に感染症の感染を予防できるか、がん組織などの縮退を引き起こすことができるかどうかなどについては、実際の患者などに投与してその治療効果を確認する臨床試験が不可欠である。更に、思いがけない副作用などが発現しないかどうかについても、臨床試験などで慎重に確認していく必要がある。

### 3.4 期待されるペプチドワクチンの適応疾患

#### 1) 感染症の感染予防効果を持ったワクチン

ウイルス・バクテリアなどの病原体に人体がさらされると、病原体由来のたんぱく質はヒトの細胞内で分解されて HLA 分子に結合し、細胞外に抗原として提示される。血液中の T リンパ球細胞が HLA 分子に結合した病原体由来のペプチドを認識すると、細胞障害性 T リンパ球細胞の増殖を始めとする免疫機能が亢進し、再度の病原体進入時には、迅速に病原体を排除することができる。

従って、あらかじめ HLA 分子に結合する病原体由来のペプチド断片を予測することができれば、極めて効率的で副作用が殆どない、ペプチドワクチンとしての有望な候補物質となる。また、一般の SARS（重症急性呼吸器症候群）、鳥インフルエンザウイルスなどの新興感染症に対しても、そのゲノム配列が発表された段階で極めて迅速にワクチン候補ペプチドを予測し、提案することが可能になる（図 5）。

#### 2) がん細胞の免疫治療を目的としたワクチン

がん細胞に対して、本来免疫系は、がん細胞特有のたんぱく質断片を認識して細胞障害性 T 細胞を始めとする免疫機能が亢進することにより、がん細胞を排除する働きを持っている。

従って、あらかじめ HLA 分子に結合するがん細胞固有のペプチド断片などを予測することができれば、極めて効率的で副作用が殆どない、抗がん剤としての有望な候補物質となる。

実際、白血病や乳がんなどに対しては、がん抗原ペプチド投与によるがん免疫療

次他の HLA 型に対する予測プログラムの作成を進めていき、将来的には主要な HLA 型の予測プログラムを、一通り完成していく予定である。

様々な HLA 型に対応した結合ペプチド予測プログラムを作成することができれば、将来的には、患者個人の HLA 型をあらかじめ別の方法などで検査することにより、各人の疾患に最適な複数のペプチドの組み合わせを選択するなど、テーラーメイド医療の実現も可能になる。

しかし、ヒト HLA 分子に結合することが予想されるペプチド配列が決まっただけでは、ワクチンとして実際に開発することはできない。まず、動物実験などにより、動物へのペプチド投与が、実際に免疫機能の活性化を引き起こすかどうかを確認する必要がある。

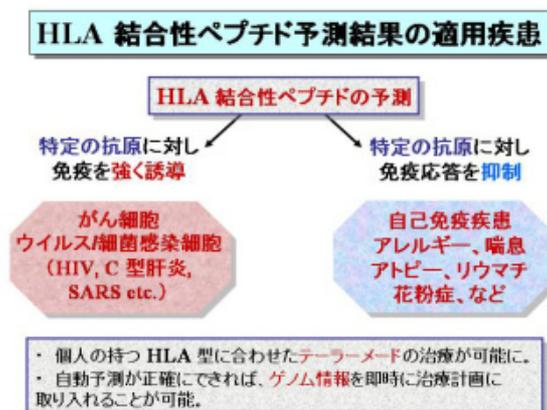


図 5 期待されるペプチドワクチンの適用疾患

法が試みられており、一部では好成績が得られている<sup>8</sup> (図5)。

### 3) アレルギー疾患の免疫治療を目的としたワクチン

花粉症やアトピー、喘息、リウマチなどに代表されるアレルギー疾患では、花粉などのアレルゲン (抗原) に対する免疫機能が異常に亢進することにより、その症状が現れる。しかし、HLA に結合したアレルゲン由来のペプチドをうまく選択することにより、免疫機能全体をそのアレルゲンに対する免疫寛容 (アナジー) 状態に誘導することが可能であると言われている。

例えば、そのアレルゲン由来のペプチド断片であって、ある程度 T リンパ球細胞にその存在は認識されるものの、HLA 分子への結合があまり強くないため T リンパ球細胞による認識が持続しない抗原ペプチドを選択することができれば、そのアレルゲンに対する免疫寛容 (アナジー) 状態に誘導することが可能であると考えられている。

したがって、あらかじめ HLA 分子に結合するアレルゲン由来のペプチド断片を予測することができれば、極めて効率的で副作用が殆どない、抗アレルギー製剤としての有望な候補物質となる (図5)。

### むすび

これまでに、マウス MHC 分子に結合するペプチドを予測する実験系において、「能動学習法」が極めて効率的な予測方法であることが実証された。今後、ヒト HLA 分子に結合するペプチドを効率よく予測するプログラムを順次作成することにより、感染症、癌、アレルギー疾患などを治療する目的を持ったペプチドワクチンのペプチド候補配列を提供する予定である。

最後に、本ニュースレターへの執筆の機会を与えて下さいました編集委員の皆様に深く感謝いたします。

### 参考文献

- 1) 馬見塚、安部 ; 「集団能動学習ーデータマイニング・バイオインフォマティクスへの展開ー」電子情報通信学会論文誌 (2002) Vol. J85-D-II, pp.717-724.
- 2) 麻生川ほか ; 「能動学習法を利用した創薬スクリーニング」NEC 技法 (2003) Vol.56, No.10, 33-37.
- 3) Mamitsuka H. "Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models." (1998) *Proteins*, 33, 460-474.
- 4) Udaka K. et. al. "Empirical evaluation of a dynamic experiment design method for prediction of MHC class I-binding peptides." (2002) *J. Immunol.*, 169, 5744-5753.
- 5) Parker K. et. al. "Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains." (1994) *J. Immunol.*, 152, 163-175.
- 6) Rammensee H. et. al. "SYFPEITHI: database for MHC ligands and peptide motifs." (1999) *Immunogenetics*, 50, 213-219.
- 7) Udaka k. et. al. "An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries." (2000) *Immunogenetics*, 51, 816-828.
- 8) Oka Y. et. al. "Wilms tumor gene peptide-based immunotherapy for patients with overt leukemia from myelodysplastic syndrome (MDS) or MDS with myelofibrosis." (2003) *Int. J. Hematology*, 78, 56-61.

## //// 研究紹介 ////

## 機械学習による薬物化学構造からの活性クラス識別

豊橋技術科学大学・知識情報工学系  
高橋 由雅†, 錦織 克己, 藤島 悟志

## 1. はじめに

現在、医薬や農薬などの開発研究の場では既存薬物のデータを背景に、薬化合物の化学構造と生理活性など種々の作用（または特性）との間の関係を積極的に見出し、これらの情報を活用することによって新規有用物質の合理的な開発を進めようとする様々な試みが行われている。また、多大な努力によって生み出される新薬の登場の一方で、例えば大きな期待を背負って登場した医薬品に対して重篤な副作用が報告されたり、農薬においてはそのヒト健康影響や残留性による環境影響などが社会的な問題となっている。

筆者らは、こうした化学物質のリスク評価の問題に関連し、薬物の化学構造情報のみからその活性（毒性）等の作用の有無を識別・予測することをねらいとし、化学構造の TFS(Topological Fragment Spectra)表現にもとづく構造類似性評価や機械学習を基礎とした活性予測やリスク推定の可能性について検討を進めている[1,2]。本稿では、分類学習モデルとして近年注目を集めているサポートベクターマシン（Support Vector Machine, 以下 SVM）を用いた薬物活性クラス分類における試みについて紹介する。

## 2. サポートベクターマシン (SVM)

計算機処理能力の急速な進展と大容量記憶媒体の低廉化に伴い、膨大な情報の収集が容易に行えるようになった。こうした現状を背景に、最近では様々な分野においてデータマイニングあるいはチャンス発見と呼ばれる、大量のデータから有用な知識を発掘するための新しい研究分野が誕生し、大きな関心を集めている。また、その知識発見の基盤となる機械学習の分野において、サポートベクターマシン（Support Vector Machine; 以下 SVM）[3]は現在最も優れた分類学習モデルの一つとして注目されている。SVMは、パーセプトロン型学習モデルである。SVMの基本的な構造は単純な線形識別関数であるにも関わらず、カーネル関数とマージン最大化といった工夫を加えることにより、チューニングを施した複雑な

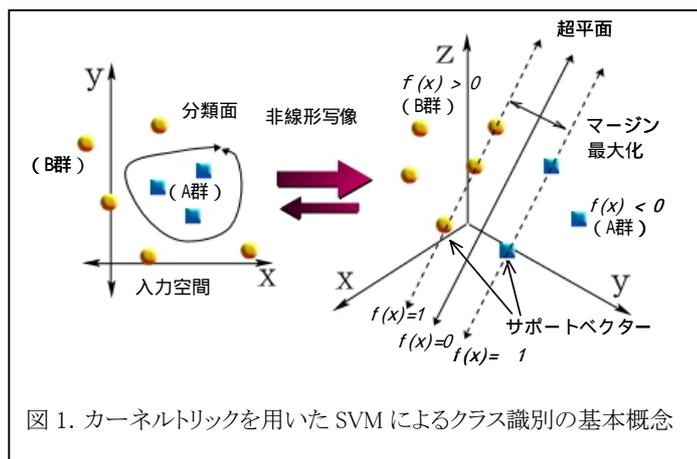


図1. カーネルトリックを用いたSVMによるクラス識別の基本概念

†連絡先: 〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1 豊橋技術  
科学大学 知識情報工学系(分子情報システム研究室) Tel:  
0532-44-6878, Email: taka@mis.tutkie.tut.ac.jp

多層パーセプトロンに劣らない性能を有することが知られている。

その基本概念は、ある与えられた線形分離可能な訓練集合  $\mathbf{X}(\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n)$ , ( $\mathbf{x}_i$  はクラスラベル  $y_i \in \{-1, 1\}, i=1, \dots, n$  をもつ) に対し、各パタン  $\mathbf{x}$  を事前に用意された非線形写像を通して高次元特徴空間  $\mathbf{z}$  に写像し、そこでのマージン最大の最適な識別面を得ることにある(図 1)。SVM による入力パタンベクトル  $\mathbf{x}=(x_1, \dots, x_d)$  の識別関数は、次のように表される。

$$f(\mathbf{x}_i) = (\mathbf{w} \cdot \mathbf{x}_i) + b \quad (1)$$

ここで  $\mathbf{w}$  は重みベクトル、 $b$  はバイアス項を表す。ここでの識別面は  $f(\mathbf{x}_i) = 0$  で表すことができる。このとき、マージン最大の識別面は次式の  $L(\mathbf{w})$  を下に示す条件下で最小化することによって得られる。

$$L(\mathbf{w}) = \|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w} = \sum_{i=1}^d w_i^2 \quad (2)$$

$$\text{制約条件: } y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad (i=1, \dots, n)$$

クラス分類のための決定関数は次式によって表すことができる。

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \quad (3)$$

ここで、 $\text{sign}$  評価関数の値が正であれば+1、負であれば-1 を返す単純な符号化関数である。この基本概念は線形分離不可能な場合にもスラック変数  $\xi$  を導入し、与えられた制約条件の下で次式を最小化することによって容易に一般化できる。

$$\frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^n \xi_i \quad (4)$$

$$\text{制約条件: } y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0$$

この制約条件下での 2 次最適化問題はラグランジュ乗数  $\alpha$  の導入によって次のように表すことができる。

$$W(\alpha) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j - \sum_{i=1}^n \alpha_i \quad (5)$$

$$\text{制約条件: } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^n \alpha_i y_i = 0$$

このように訓練パタン  $\mathbf{x}_i$  は内積を通してのみ最終解に現れることから、非線形写像やカーネル関数の概念を利用することによって一般的な非線形関数にも拡張できる。すなわち、ある写像  $\mathbf{x} \rightarrow \phi(\mathbf{x})$  が与えられるとき、その最終的な特徴空間での内積はカーネル関数  $K(\mathbf{x}, \mathbf{x}')$  によって置き換えることができる。

$$f(\mathbf{x}) = g(\phi(\mathbf{x})) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (6)$$

### 3. SVM による薬物活性クラス分類

筆者らは、上位機アルゴリズムに従い、薬物活性クラス分類のための汎用 SVM ツールを作成した。実装に際しては、Dong らの提案した SMO(Sequential Minimal Optimization)改良アルゴリズム[4]を用いるとともに、TFS を入力ベクトルとした SVM を実装し、実験を行った。非線形写像のカーネル関数には次式で示す Gaussian RBF(Radial Basis Function)カーネル関数を用いた。

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{\sigma^2}\right) \quad (7)$$

上記 SVM の学習に際しては調整パラメータ C,  $\sigma$  について対象とする問題に適切な値を定める必要がある。これらの値は予備試験による Trial and Error で決定した。ここでは, C=100,  $\sigma=40$  を用いた。また, SVM は基本的には 2 クラス分類モデルであり, 複数クラスに関する分類のためには SVM を組み合わせる必要がある。ここでは, k クラス分類問題(k $\geq$ 3)を解くための一般的な組み合わせ法である one-against-the-rest を利用した。

治験薬構造データベース MDDR[5]より抽出した 4 種の異なるドーパミン受容体(D1, D2, D3, D4)に作用するアンタゴニスト 1,364 種を対象に, SVM を利用した活性クラス識別を試みた。SVM への入力シグナルには各化合物構造の TFS(Topological Fragment Spectra)表現を用いた。実験に際しては対象データを事前に訓練集合 (1,227 化合物 ; D1:155, D2:356, D3:216, D4:500) と予測集合 (137 化合物 ; D1:18, D2:39, D3:24, D4:56) に分割し, 学習及び予測実験を行った。学習の結果, 訓練集合 1,227 化合物全ての活性クラスを全て正しく認識することができた。

つぎに, 得られたモデルを活性クラスが未知の予測集合に適用し, 予測実験を試みた。その結果, 137 化合物中, 123 化合物 (89.8%) の活性クラスを正しく予測することができた(表 1)。このことは化学構造情報からの薬物活性クラス識別における構造記述子としての TFS の有効性を改めて強く示唆するとともに, 薬物活性クラス分類における SVM の有用性を示すものである。

表 1 SVM によるドーパミン受容体 D<sub>n</sub> アンタゴニスト (1364 化合物) の活性クラス分類と予測

Class	Training		Prediction	
	Data	%correct	Data	%correct
All	1227	100	123/137	89.8
D1	155	100	15/18	83.3
D2	356	100	31/39	79.5
D3	216	100	22/24	91.7
D4	500	100	55/56	98.2

#### 4. 人工ニューラルネットワーク (ANN) との比較

筆者らは先に, TFS を入力シグナルとした人工ニューラルネットワークが薬物の活性クラス識別に対して良好な結果を与えることを示した[2]。そこで, 上記 TFS 表現を基礎とした SVM の薬物活性クラス識別における分類・予測性能を ANN のそれと比較検討した。比較に際しては前述のドーパミン受容体アンタゴニスト 1,364 種を対象に, その 90%を訓練集合, 10%を予測集合とする ten-fold cross validation 法を用いて実験を行った。その結果, SVM は ANN に比べて全ての活性クラスでより良好な予測結果を与えた(表 2)。これにより, 薬物活性クラス識別における TFS を特徴記述ベクトルとしたサポートベクターマシン

表 2 SVM と ANN による識別率・予測率の比較

Clas s	SVM		ANN	
	Training %correct	Prediction %correct	Training %correct	Prediction %correct
All	100	90.6	87.6	81.0
D1	100	87.5	76.5	73.2
D2	100	86.1	80.8	72.2
D3	100	88.3	91.1	88.9
D4	100	95.5	94.4	90.5

\* ten-fold cross validation test による。

は、人工ニューラルネットワークによるそれに比べ、より優れた予測安定性を有することが示された。

#### 4. おわりに

上記の ANN および AVM を用いた活性クラス分類においては、何れの場合も識別対象とする特定の活性を有する化合物群のみを用いた結果であり、どの活性クラスにも属さないノイズ化合物群を含まないものであることに注意すべきである。一般に、特定の薬物活性を有する化合物は膨大な既存化合物群の中のごく一部であり、それ以外のはノイズデータと考えることができる。従って、実用的な観点からは、大量のノイズデータ存在下で特定の活性クラスを識別することが要求される。このことから、現在、ノイズデータ存在下での同様なアプローチの有効性について引き続き検討を進めているところである。

#### 文献

- 1) 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.
- 2) S. Fujishima, Y. Takahashi, Classification of Pharmacological Activity of Drugs using TFS-Based Artificial Neural Network, *J. Chem. Inf. Comput. Sci.*, in press.
- 3) V. N. Vapnik : *The Nature of Statistical Learning Theory*, Springer, 1995.
- 4) J. Dong, A. Krzyzak, and C. Y. Suen. A fast SVM Training Algorithm. In *Proceedings of First International Workshop, SVM 2002*, pages 53-67 Niagara Falls, Canada, 2002.
- 5) MDL Information Systems, Inc., *MDL Drug Data Report*, Ver. 2001.1 (2001).

## //// 報告 ////

## 第31回構造活性相関シンポジウム実施報告

第31回構造活性相関シンポジウム実行委員  
星薬科大学 高橋典子

第31回構造活性相関シンポジウムは、日本薬学会構造活性相関部会の主催、日本化学会ほか関係諸学会の共催で、平成15年11月18日(火)、19日(水)の2日間、市川紘先生を実行委員長として、星薬科大学を会場に開催されました。本年度シンポジウムでは特別講演2件と、口頭発表(20件)とポスター発表(30件)合わせて50件の一般講演が行われ、活発な議論が交わされました。

第1日目は午前と夕方に、配座探索や蛋白質-低分子相互作用の解析、物性予測といった分子設計の基盤的研究と農薬分野への応用についての一般講演が行われました。さらに、午後には二つの特別講演とポスターセッションが、情報化学討論会と合同で開催され、2時間にわたって熱心な議論が交わされました。東京大学大学院総合文化研究科の浅島誠先生による特別講演「構造と機能からみた器官形成と形づくり」では、最近話題のES細胞からの形態形成について、低分子化合物の構造とその生理活性という観点から、わかりやすくご講演頂きました。また、京都大学大学院農学研究科/慶應大学先端生命研究所の西岡孝明先生による特別講演「ゲノム情報から細胞内化学反応ネットワークを再構築するーバイオインフォマティクスから化学分析までー」では、メタボローム解析によって生命現象を化合物の代謝・構造変換の視点で解釈し、それに基づく新規な疾病治療薬の開発の可能性が示されました。第2日目は分子設計の具体的事例などの一般講演とともに、ポスターセッションが行われました。

今回からの新しい試みとして、優秀なポスター発表に対する「ポスター賞」を創設し、構造活性相関部会の常任幹事による投票の結果、次の3演題が第1回の受賞者に選ばれました。

- KP14 分子重ね合わせに基づく活性化合物解析システムの開発 (医薬分子設計研・野中はるみ, 富岡伸夫, 板井昭子)
- KP19 標的蛋白質の Induced Fit を考慮したリガンドドッキング(1): ブラウン動力学法を用いた鍵穴サンプリング (北里大薬・山乙教之, 広野修一)
- KP28 SOM フィルターを用いた QSAR モデリング (豊橋技科大・秋元紗恵, 佐々木英史, 高橋由雅)

ポスター賞受賞講演については、SAR News の本号に内容をご紹介頂いております。

以上のとおり、併催した第26回情報化学討論会とあわせて約350名の参加者を得て、盛会のうちに無事終えることができました。次回は藤沢薬品工業の田中明人先生のお世話で、つくばでの開催が決定しています。最後に、本シンポジウムの開催に当たり、多数のご助言・ご助力を頂きました日本薬学会の関係各位に深くお礼を申し上げます。

## //// 第 31 回構造活性相関シンポジウム・ポスター賞 ////

## KP14 分子重ね合わせに基づく活性化合物解析システムの開発

(医薬分子設計研) 野中 はるみ、富岡 伸夫、板井 昭子

## はじめに

標的蛋白の立体構造が未知の場合に、複数の活性化合物から活性発現に必要な構造や物性の条件を抽出する方法として「分子重ね合わせ法」がある。従来の分子重ね合わせ法は、原子位置や分子骨格の類似性に基づいたものであったが、同一結合部位に結合するために原子位置が一致する必要はなく、また骨格の異なる化合物の重ね合わせに比べて得られる情報が少ない、という問題があった。これに対して我々は、骨格の異なる分子の三次元構造を同一空間中で重ね合わせ、標的蛋白の結合部位の形状・性質、共通の相互作用サイト、活性化合物中の活性に必須な官能基などを推定する分子重ね合わせ法 RECEPTS を開発してきた[1-3]。分子重ね合わせ計算には、対象とする分子のコンホメーション、分子同士の相対配置(回転・並進の自由度)、重ね合わせモデルの評価などの解決すべき問題があるが、RECEPTS では結合部位に想定される水素結合部位の対応付けに基づく最小自乗法計算と、分子間の物理化学的性質の一致度の三次元格子点を用いた評価法により、これらの問題の解決を図ってきた。

この RECEPTS の概念に基づき、より幅広い化合物に対応できるようにアルゴリズムを改良し、さらに重ね合わせ計算結果に基づいて構造活性相関解析やデータベース検索計算を行うことができるようにした、より実践的なシステム KeyRecept を開発した。

## システムの構成と特長

KeyRecept システムは以下の特長を備えている。

- 1) 2～10個程度の骨格の異なる化合物を自動的に重ね合わせ可能
- 2) 標的蛋白側に想定される水素結合性官能基や疎水性グループの位置を表すダミー原子の対応関係に基づき、計算対象分子のコンホメーションと相対配置の可能性を網羅的に探索
- 3) コンホメーションの自由度が高い分子でも、重ね合わせの初期段階で結合軸回転角度を粗く刻み、その後精密化する手法を採用し、計算可能としている
- 4) 分子の物理化学的性質の三次元空間中での一致度を評価する種々の「指標値」(van der Waals 形状、分子表面形状、水素結合、静電相互作用など)を三次元格子点を用いて計算し、有望な分子重ね合わせモデルの選択に利用
- 5) 重ね合わせモデルに基づき、活性未知化合物の活性予測をする2種類の方法(SAR法、サイトモデル法)を備え、状況による使い分けが可能
- 6) コンホメーションの自由度の高い化合物でも高速なデータベース検索計算が可能

本システムによる計算の流れを図1に示す。最初に Superposer により標的蛋白の結合部位の特徴を反映する分子重ね合わせモデルを作成する。重ね合わせモデルに基づく解析・予測手法として、SAR法とサイトモデル法の2通りが用意されている。SAR法では、異なる活性値をもつ複数の化合物を重ね合わせモデルにさらに重ねて指標値を求め、活性値と指標値の間で重回帰分析を行い、活性の強弱の予測モデルが得られる。DBS-SAR 計算では、データベース中の活性未知化合物を順次重ね合わせて指標値を求め、SAR 計算で得られた重回帰式を利用して活性を予測する。SAR 法には十分な数の化合物の活性データが必要であり、適切な相関式が得られない場合は適用できないのに対し、そのような場合でもサイトモデル法は適用可能である。

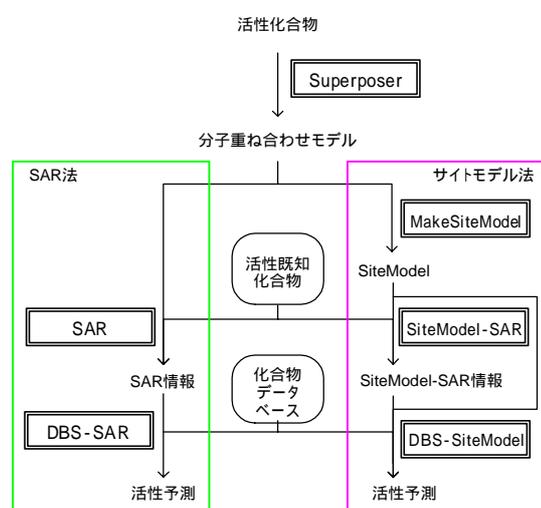


図 1 KeyRecept システム

サイトモデル法では、Superposer で得られた分子重ね合わせモデルから、活性発現に重要なサイトポイントとグリッド領域を抽出してサイトモデルを作成する。DBS-SiteModel 計算では、サイトモデルに対してデータベース中の活性未知化合物を順次重ねて指標を求め、各化合物が活性をもつ可能性を予測する。化合物の活性データがある場合は、重ね合わせ指標と活性値の重回帰分析を行い、それを活性予測に利用することもできる。手持ちの化合物データの量や質により SAR 法またはサイトモデル法を選択することや、比較的簡便なサイトモデル法で予備的なスクリーニングをした後に SAR 法を適用するなど、状況に応じた多彩な解析が可能である。

### 計算例

Angiotensin II Receptor (AT1) Antagonist は、Angiotensin converting enzyme により作られた Angiotensin II が AT1 受容体に結合するのを妨げ、Angiotensin II の血管収縮作用、血管壁肥厚作用、動脈硬化作用、心筋収縮力増強作用、心筋肥大作用を抑制するため、降圧薬、抗動脈硬化薬として用いられている。今回は、化合物の構造と活性データを文献[4-7]から入手し、計算対象とした。一連の計算に利用する化学構造を含む化合物データ (図1の「活性化化合物」

「活性既知化合物」に該当する) を ISIS により作成し、そこからエクスポートした RD ファイルからプログラム Key3D[8]による三次元化を行い、さらに属性情報の付加を行って解析用データベースを作成した。結合軸回転と各種ダミー原子の発生は、KeyRecep の自動割り振り機能により行った。

Superposer 計算対象分子を図2に示す。mol-1, mol-2 の n-Bu 基は両分子共通の構造であり、結合軸回転をしても指標値に有効に寄与しないことが予備計算から明らかになったので、根元の一つの結合軸回転だけを残し、その先の結合軸回転は停止させた。

Superposer 計算結果のうち最上位のモデルを図3に示す。3化合物に共通のイミダゾール環は一致しており、テトラゾール環とカルボン酸が共通の水素結合に関与すると予測された。全体的な分子形状もよく一致していた。分子重ね合わせ計算の結果が「正解」であるかどうかは、各分子と標的蛋白質との複合体結晶構造により検証できるが、今回対象とした Angiotensin II Receptor は結晶構造が未知なので、この分子重ね合わせ計算の結果が「正解」であるかどうかは検証できない。そこで、「正解」に近い重ね合わせモデルであれば、他の化合物の活性の強弱を正しく判別できるはずであること、また、実際の研究開発では活性未知化合物の活性予測が重要な意味をもつことから、SAR 法やサイトモデル法において、他の化合物の活性の強弱を説明できる重ね合わせモデルが得られているかどうかを検証した。

### SAR 法

Superposer 計算結果の最上位のモデルに対して SAR 計算を行った。IC<sub>50</sub> が 0.0013~150 μM の 13 化合物について、重ね合わせで得られた指標値と活性値の重回帰分析を行い、相関が一番良かった回帰式を示す。van der Waals 形状の一致がよいほど高活性が期待できることを示している。

$$\log(\text{IC}_{50}) = 7.630 + 10.907 \times \text{vdw} \quad R = 0.904 \quad n = 13$$

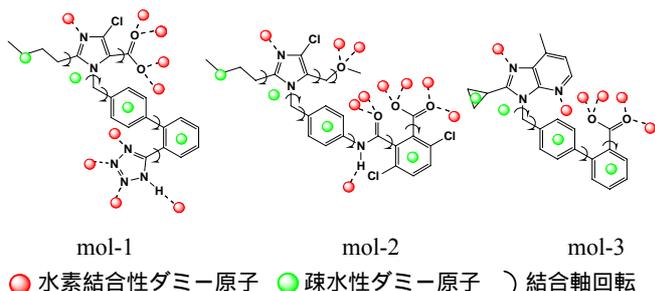
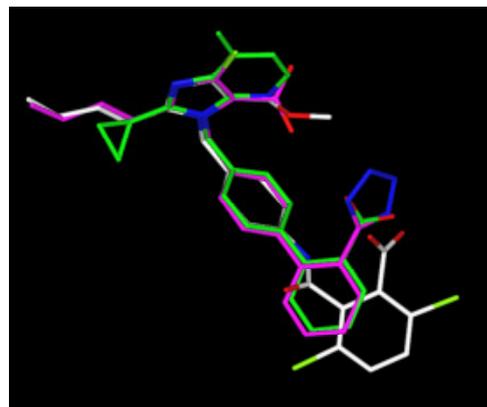


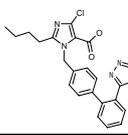
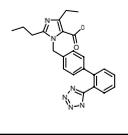
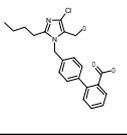
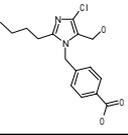
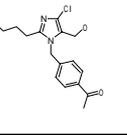
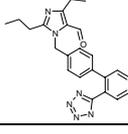
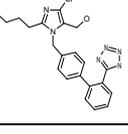
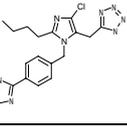
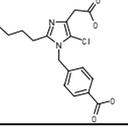
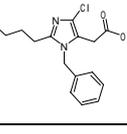
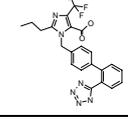
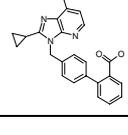
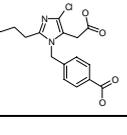
図2 Superposer 計算対象化合物



mol-1:ピンク mol-2:白 mol-3:緑

図3 Superposer 計算結果

表 1 SAR 計算・SiteModel-SAR 計算利用化合物

構造式	IC <sub>50</sub> ( $\mu$ M)	構造式	IC <sub>50</sub> ( $\mu$ M)	構造式	IC <sub>50</sub> ( $\mu$ M)	構造式	IC <sub>50</sub> ( $\mu$ M)	構造式	IC <sub>50</sub> ( $\mu$ M)
	0.0013		0.006		0.23		1.7		28
	0.002		0.019		1.2		19		150
	0.003		0.052		1.6				

得られた回帰式を利用して、次に DBS-SAR 計算を実施した。計算対象にしたデータベース化合物は、市販化合物データベース ACD 中のテトラゾールまたはカルボン酸の部分構造を有する化合物からランダムに選択した 1036 個とした。計算の結果、最も予測活性値の良かった 2 化合物を図 4 に示す。これらは、既に Angiotensin II Receptor Antagonist として知られている candesartan と valsartan であった。図 4 に示した活性値は、SAR 計算で利用したデータと異なる実験系で測定されたものなので、数値そのものを比較することはできないが、これらが高活性の化合物として判定されたことは、本方法の有効性を示すものである。

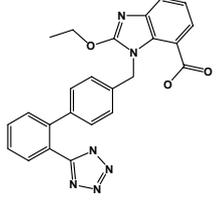
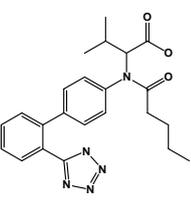
		
	candesartan	valsartan
IC <sub>50</sub> ( $\mu$ M)	0.028	0.0027
SAR 法	0.036	0.027
サイトモデル法	0.017	0.020

図 4 活性予測値最上位化合物

### サイトモデル法

Superposer 計算結果の最上位の重ね合わせモデルから、サイトモデル(図 5)を以下の条件で作成した。

**疎水性サイトポイント(白球、矢印)** 疎水性原子団に対応する疎水性ダミー原子に基づき発生

**水素結合性サイトポイント(色球、矢印)** 標的蛋白側に予想される水素結合性官能基の位置を示す水素結合性ダミー原子が、2 分子以上から同じ位置に発生される場合に設定

**vdW 領域(白い網)** 3 化合物の van der Waals 体積の和

**HB 領域(色付きの網)** 水素結合性サイトポイント周囲の領域

このサイトモデルが適切かどうか、SiteModel-SAR 計算を行い検証した。指標値と活性値との重回帰分析の結果、得られた相関式を以下に示す。SAR 法と同様に、van der Waals 形状が活性に重要であることが示された。

$$\log(\text{IC}_{50}) = 7.058 + 14.755 \times \text{vdw} \quad R = 0.931 \quad n = 13$$

SiteModel-SAR 計算の結果を利用して、データベース検索計算を行った。計算対象のデータベースは SAR 法と同一のものを用いた。計算の結果、SAR 法でも最も予測値が良かった candesartan と valsartan が最上位に予測された(図 4)。これらの化合物の予測活性が最上位になったことは、本方法が活性未知化合物の活性予測に有効であることを示している。

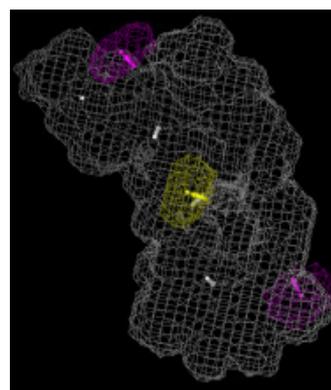


図 5 サイトモデル

## まとめ

KeyRecep システムにより、標的蛋白質立体構造未知の場合における分子重ね合わせ法と、その結果を利用した構造活性相関解析や活性未知化合物の活性予測によるバーチャルスクリーニングが可能となった。一連の計算は、インターフェース機能や結果解析機能などを備えた GUI から操作できるようになっている。本システムは、標的蛋白質の構造が未知の系での活性化化合物の構造活性相関解析に有用であり、特に HTS などのランダムスクリーニングから得られる多様な骨格を有する活性化化合物の解析に有効な方法となることが期待される。

本研究は、新エネルギー・産業技術総合開発機構 (NEDO) の「医用化合物スクリーニング支援システム」の研究開発の一環として行った。最後に、本ニュースレターに発表の機会をくださいました編集委員の皆様には深く感謝いたします。

## 参考文献

- [1] Kato, Y., Itai, A. and Iitaka, Y., *Tetrahedron*, **43** (1987) 5229
- [2] Itai, A., Kato, Y., Tomioka, N., Iitaka, Y., Endo, Y., Hasegawa, M., Shudo, K., Fujiki, H. and Sakai, S., *Proc. Natl. Acad. Sci. U.S.A.* **85** (1988) 3688
- [3] Kato, Y., Inoue, A., Yamada, M., Tomioka, N. and Itai, A., *J. Comput.-Aided Mol. Des.*, **6** (1992) 475
- [4] John V. Duncia et al., *J. Med. Chem.*, **33** (1990) 1312
- [5] David J. Carini et al., *J. Med. Chem.*, **33** (1990) 1330
- [6] John V. Duncia et al., *Med. Res. Rev.*, **12** (1992) 149
- [7] Ruth R. Wexler et al., *J. M. Chem.*, **39** (1996) 625
- [8] 中村建介, 富岡伸夫, 板井昭子, *JCPE Journal*, **12** (2000) 177

## KP19 標的蛋白質の Induced Fit を考慮したリガンドドッキング(1) :

### ブラウン動力学法を用いた複合体構造最適化

(北里大学薬学部) ○山乙 教之, 広野 修一

### 1. はじめに

タンパク質-リガンドドッキングにおいて、現在広く流通しているプログラムは、タンパク質を剛体として取り扱っている。そのため、リガンドとの複合体形成による標的タンパク質の induced fit は考慮されておらず、リガンドが結合していない状態の X 線構造や、アミノ酸配列からモデリングされたタンパク質立体構造に対して、ドッキングを行っても、十分な精度の複合体立体構造が得られないことが分かっている。また、今後発売予定の induced fit を考慮したと謳われているプログラムにおいても、自由度を持っているのは側鎖のみであり、主鎖レベルの大きな induced fit が起きるタンパク質には、対応できないと予想される。

このような induced fit に対応する手段の一つは、ドッキングで得られた不完全なタンパク質-リガンド複合体構造について、あらわな水分子を用いた水溶液中で分子動力学(MD)法を行い、平衡化することである。その場合、あらわな水分子を用いた水溶液中での分子動力学法では、非常に計算時間が掛かるという問題点がある。そこで、我々は、連続体モデル中で水の影響を粘性力とランダム力として取り扱う、ブラウン動力学(BD)法プログラムを開発し、ドッキング問題に対して適用を試みた。

### 2. 方法

溶液中でブラウン運動する粒子の運動は、ランジュバン方程式(1)式によって記述される。

$$m \frac{dv}{dt} = \mathbf{F} - m\zeta \mathbf{v} + \mathbf{F}^B \quad \dots (1)$$

ここで、 $m$  は、粒子の質量、 $\mathbf{v}$  は、速度、 $t$  は、時間、 $\mathbf{F}$  は、相互作用力、 $m\zeta$  は、摩擦係数、 $\mathbf{F}^B$  は、ランダム力である。 $\zeta$  は、ストークスの法則から、(2)式で導かれる。

$$\zeta = 6\pi\eta a / m \quad \dots (2)$$

ここで、 $\eta$  は、粘度、 $a$  は、粒子半径である。

我々は、BD 法のためのアルゴリズムとして、Ermak と Buckholz によって開発された(3)式を採用した。<sup>1,2)</sup>

$$\begin{aligned}
 \mathbf{r}(t+h) &= \mathbf{r}(t) + \frac{1}{\zeta} \mathbf{v}(t) (1 - e^{-\zeta h}) \\
 &+ \frac{1}{m\zeta} \mathbf{F}(t) \left\{ h - \frac{1}{\zeta} (1 - e^{-\zeta h}) \right\} \\
 &+ \mathbf{r}^B(t+h) \\
 \mathbf{v}(t+h) &= \mathbf{v}(t) e^{-\zeta h} + \frac{1}{m\zeta} \mathbf{F}(t) (1 - e^{-\zeta h}) \\
 &+ \frac{1}{m\zeta h} (\mathbf{F}(t+h) - \mathbf{F}(t)) \\
 &\times \left\{ h - \frac{1}{\zeta} (1 - e^{-\zeta h}) \right\} + \mathbf{v}^B(t+h) \quad \dots (3)
 \end{aligned}$$

ここで、 $\mathbf{r}$  は、位置、 $h$  は、タイムステップ、 $\mathbf{r}^B$  はランダム変位、 $\mathbf{v}^B$  はランダム変速である。

### 3. 結果と考察

Induced fit が起きるタンパク質に対するドッキングの検証系として、intestinal fatty acid binding protein にたいするパルミチン酸のドッキングを試みた。<sup>3-5)</sup> この系では、パルミチン酸が結合している状態の X 線構造(PDB code 2IFB)<sup>6)</sup> と結合していない状態の X 線構造(PDB code 1IFB)<sup>7)</sup> が共に得られている。主鎖の全体的な形は、非常に良く似ているが、図中の○で囲まれた部分で、主鎖の流れが少し異なっている (図 1)。

しかしながら、リガンドが結合していない X 線構造に対して、SYBYL FlexX を用いて、パルミチン酸をドッキングさせた結果、水素結合は、正しく結ばれているものの、パルミチン酸の形は、X 線構造 (図 2) との rmsd が 1.30Å と大きくゆがんだ構造となった (図 3)。

このドッキング構造に対して、ブラウン動力学法を適用し、構造を平衡化したところ、X 線構造との rmsd が 0.88Å のより X 線に近い立体構造が得られた (図 4)。

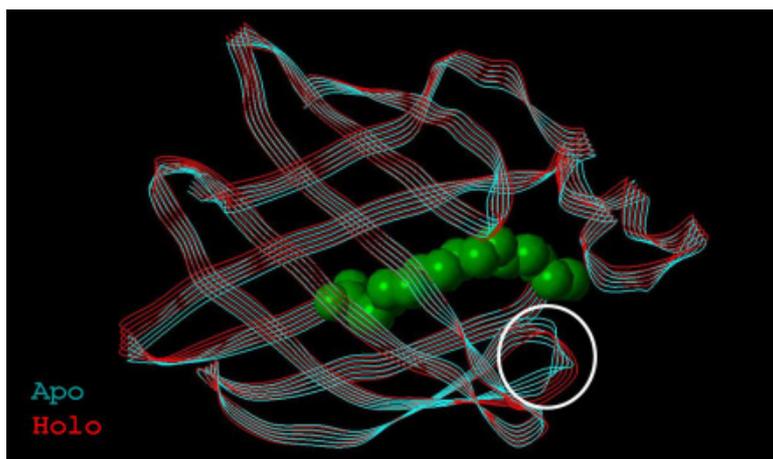


図 1 Intestinal fatty acid binding protein-パルミチン酸複合体の X 線構造 緑色：複合体中のパルミチン酸、赤色：複合体のときのタンパク質構造、水色：タンパク質単独のときのタンパク質構造、白丸：活性部位(N,C $\alpha$ ,C $\beta$ ,C,O)の rmsd=0.56Å

### 4. 結論

ブラウン動力学法により、ドッキングプログラムで得られた複合体立体構造を出発点として、より X 線構造に近い立体構造をサンプリングすることができた。従って、この手法は、タンパク

質-リガンド複合体立体構造のモデリングに有用であることが示された。

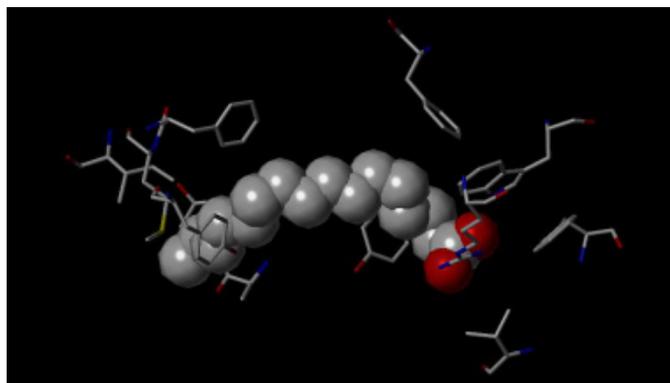


図2 複合体 X線構造中のパルミチン酸とその結合部位近傍の立体構造

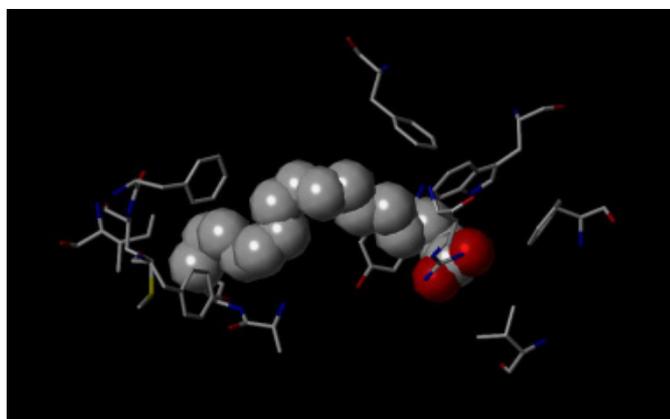


図3 FlexX で得られた複合体モデル パルミチン酸の rmsd=1.30Å

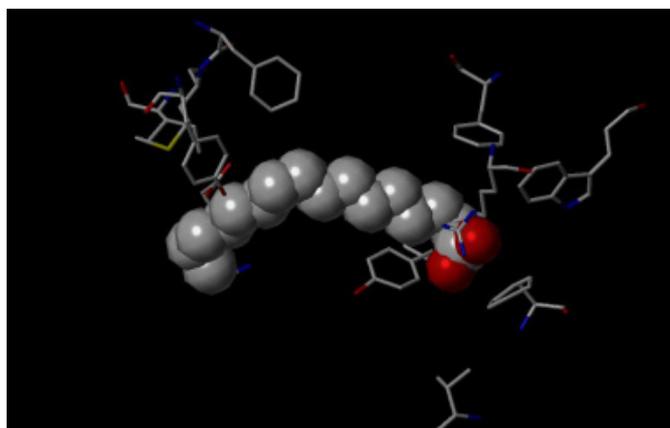


図4 ブラウン動力学法で平衡化された複合体モデル パルミチン酸の rmsd=0.88Å

#### 参考文献

- 1) Ermak, D.L.; Buckholz, H. *J. Comput. Phys.* **1980**, *35*, 169-182.
- 2) 神山新一; 佐藤明 *流体ミクロ・シミュレーション (分子シミュレーション講座3)*; 朝倉書店, **1997**.
- 3) 山乙教之; 広野修一 *第30回構造活性相関シンポジウム講演要旨集* **2002**, pp 27-30.
- 4) 山乙教之; 広野修一 *第31回構造活性相関シンポジウム講演要旨集* **2003**, pp 115-118.
- 5) Yamaotsu, N.; Hirono, S. *in preparation*.
- 6) Sacchettini, J.C.; Gordon, J.I.; Banaszak, L.J. *J. Mol. Biol.* **1989**, *208*, 327-339.
- 7) Sacchettini, J.C.; Gordon, J.I.; Banaszak, L.J. *Proc. Nat. Acad. Sci. USA* **1989**, *86*, 7736-7740.

**KP28 SOM フィルターを用いた QSAR モデリング†**

(豊橋技科大) ○秋元 紗恵, 佐々木 英史, 高橋 由雅

一般に、構造的に多様な物質を対象とした構造活性相関モデルの利用に際してはその予測性能が大きく低下するという問題がある。本研究ではこうしたケースでのより高い精度の予測モデルの獲得を目的とし、自己組織化特徴地図(Self-Organizing feature Map:SOM)を応用した QSAR モデルの生成を提案し、通常の一括モデリングとの予測性能を比較、検討した。

SOM は、ランダムな出発点から位相地図を組織化できる 2 層のネットワークである。この結果得られる地図は、ネットワークに与えられたパターン間の自然な関係を示す。アルゴリズムを計算機に実装し、実データがどのようにマッピングされるか、類似構造マッピングの観点から実験を行った。実験には水溶解度の値が既知である 283 化合物の化学構造データを用いた。SOM への入力には Ghose らの部分構造記述子を用いた。SOM による構造類似化合物のマッピングの結果、同一の(あるいは近傍の)競合層ニューロンには構造的に類似のものが集まっていることが確認できた。これにより、SOM を利用した構造類似化合物群のフィルタリング処理の可能性が示唆された。

次に、これらの化合物群に対し、SOM を用いて構造的に類似した化合物集合を獲得し、これらのデータを用いて部分空間ごとの局所的な個別モデルを作成することにより、より精度の高い予測を実現することをねらいとして検討を行った。まず始めに類似構造データの選択のための SOM フィルターを作成した。出力ユニット(競合層)は 3×3 (9 ユニット)からなる目の荒い SOM フィルターとした。ここでの水溶解度の推算には原子団寄与法による加成性モデルを用いた。原子団の定義には SOM フィルターへの入力同様、Ghose ら記述子を用い、モデルの作成には当研究室で別途開発された MOPET を用いた。SOM フィルターを利用した個別モデルの活用により、前述の 283 物質の水溶解度  $\log S(S:\text{mol/L})$  のデータ予測を試みた。実験に際しては 255 化合物 (90%) をランダムに選び訓練集合とし、残り 28 化合物 (10%) を検定用の予測集合(バリデーションセット)として用いた。これらの結果を、SOM フィルターを介さず、予測集合のすべてのサンプルを一括モデルで予測した場合の結果と比較したところ、SOM フィルターにもとづく個別モデルの利用による方が、より良好な予測結果を示すことが明らかとなった。

以上、本研究では SOM フィルターを用い、類似した構造ごとに個別モデルを作成・利用することでより精度の高いデータ予測を行うためのアプローチを提案し、実データを用いてその有用性を示した。しかし、個別モデリングのためのデータ集合の分割に伴い、個々のモデルに対するサンプル数の減少によるオーバーフィッティングや、モデル選択の妥当性評価等の課題も残る。引き続き、複数のデータを用いて実験を重ねながら実用性を検証していきたい。

†秋元, 佐々木, 高橋: SOM フィルターを用いた QSAR モデリング, 第 31 回構造活性相関シンポジウム, KP28 (2003).

## //// お知らせ ////

## 構造活性フォーラム 2004

## 「創薬科学における ADME 予測 -基礎から応用まで-

日本薬学会構造活性相関部会では、毎年、構造活性相関の方法論の発展をはかるとともに、薬物と生体高分子との相互作用の分子機構、生体高分子の構造と機能、薬物の薬効・吸収・代謝・排泄・毒性、環境関連物質の構造と毒性・機能などに関する話題を取り上げ、創薬研究の新展開をはかるため構造活性フォーラムを開催しております。今年も、主として、近年話題となっております創薬科学における薬物動態特性（吸収・分布・代謝・排泄）の予測に焦点を絞り、その基礎となる分配係数の予測からトランスポーター、代謝にいたるまでの演題を中心に下記内容のフォーラムを企画しました。初心者にも分かり易いコースとして設定されておりますので、医薬・農薬に限らず、さまざまな分野において、このテーマに興味を持っておられる方々の参加をお待ちしております。

主催 日本薬学会構造活性相関部会 ([http://bukai.pharm.or.jp/bukai\\_kozo/index.html](http://bukai.pharm.or.jp/bukai_kozo/index.html))  
 協賛 日本化学会, 日本農芸化学会, 有機合成化学協会, 日本分析化学会, 日本農薬学会, 日本薬物動態学会, 近畿化学協会, Combinatorial Chemistry 研究会  
 会期 平成 16 年 6 月 25 日 (金) 10:00~17:40 18:00 よりミキサー  
 会場 京大会館 101 号室  
 (京都市左京区吉田河原町 15-9, Tel: 075-751-8311, <http://www.kyodaikaikan.jp/>)  
 [交通] JR-近鉄京都駅より市バス, 206 番 (交通事情により 40 分~1 時間), 四条京阪あ  
 るいは四条河原町より市バス 201, 31 番 (20~40 分), いずれも京大正門前下車南西へ  
 徒歩 5 分; 京阪電車鴨東線丸太町駅下車徒歩 10 分

1. 「分配係数から細胞透過性を予測できるか?」 赤松美紀 (京大院農)
  2. 「低分子化合物の log P の測定と予測 -水素結合性基の扱い方と問題点-」 山上知佐子 (神戸薬大)
  3. 「分配係数 log P の物理化学的解釈とその応用」 中馬 寛 (徳島大薬)
  4. 「生体異物と生体内脂質を輸送する ABC 蛋白質ファミリー」 植田和光 (京大院農)
  5. 「トランスポーターと非線形局所動態解析 (in loci) における最近の進歩」 山岡 清 (京大院薬)
  6. 特別講演 「創薬科学のための KEGG」 金久 實 (京大化研)
- 総合討論

フォーラム終了後、講師を囲んで簡単なミキサーを開催致します。(無料)

参加申込締切 5 月 21 日 (金) 定員 (100 名) になり次第締切

参加費 一般: 6,000 円, 学生: 3,000 円

参加申込方法 氏名, 所属, 連絡先 (住所, 電話・Fax 番号, 電子メール) を明記の上, 電子メール, Fax, または郵便にて御申込下さい。できるだけ, 電子メールの御利用をお願いいたします。

参加費は銀行振込 (みずほ銀行 百万遍支店 普通預金, 口座名: 構造活性フォーラム 2004, 口座番号: 2381090) により事前に御振込下さい。複数名の参加費をまとめて振込まれる場合は, 内訳を御連絡下さい。

申込先 〒606-8502 京都市左京区北白川追分町京都大学大学院  
 農学研究科地域環境科学専攻 赤松美紀

(Tel & Fax: 075-753-6489 E-mail: [akamatsu@kais.kyoto-u.ac.jp](mailto:akamatsu@kais.kyoto-u.ac.jp))

実行委員長 京都大学大学院薬学研究科 藤井信孝

実行委員 京都大学大学院農学研究科 赤松美紀

## //// お知らせ ////

## 第 32 回構造活性相関シンポジウム

- 日時** 平成 16 年 11 月 30 日 (火) ・ 12 月 1 日 (水)
- 会場** つくば国際会議場 (つくば市竹園 2-20-3)
- 交通** つくばバスセンターより徒歩 10 分 (つくばセンターまでは東京駅、羽田空港よりバス 約 1 時間)。詳細は <http://www.epochal.or.jp/> をご覧ください。
- 主催** 日本薬学会 構造活性相関部会
- 共催** 日本化学会, 日本農芸化学会, 日本分析化学会, 日本農薬学会
- 討論主題** ①QSAR 基本パラメータ・手法、②QSAR への情報数理的アプローチ、③分子モデリング、④医農薬等への応用、⑤吸収・分布・代謝・毒性・環境毒性と QSAR、⑥生理活性物質の活性評価、⑦バイオインフォマティクス、⑧コンビナトリアルケミストリーと創薬、⑨分子情報処理 (データベースを含む)、⑩その他
- 発表形式** 口頭 (講演 25 分または 15 分、討論 5 分を含む)、またはポスター (優秀な発表にはポスター賞授与)。
- 特別講演**
1. 「siRNA や miRNA に基づく創薬」 (東大、産総研) 多比良 和誠
  2. 「(仮) データマイニングーその発展と今後ー」 (阪大・産研) 鷲尾 隆
- 発表申込** 7 月 16 日 (金) 締切必着、E-mail により受付
- 講演要旨** 9 月 24 日 (金) A 4 版用紙を使用し、本文 (和文または英文) は 2 または 4 頁。要旨要約は英文半頁。執筆要項は URL に掲載します。
- 参加登録予約申込締切** 10 月 29 日 (金) 必着  
発表申込、参加登録及び送金方法などの詳細については URL に掲載します。
- 参加登録費** : 情報化学討論会と共通 (含要旨集)  
[一般] 予約 8000 円、当日 9,000 円 [学生] 予約 3,000 円、当日 4,000 円  
要旨集前送希望の場合は郵送料 1,000 円を別途申し受けます。なお、費用振込み後、参加取り消しによる返金には応じられません。
- 懇親会** (情報化学討論会と合同) : 11 月 30 日 (火) 18:30- 場所: 同会議場内多目的ホール (予定)  
[一般] 予約 6,000 円、当日 8,000 円 [学生] 予約 3,000 円、当日 4,000 円
- 連絡先** 〒292-0812 千葉県木更津市かずさ鎌足 2-6-7 (株) リバース・プロテオミクス研究所  
田中明人 TEL 0438-52-3990、FAX 0438-52-3986  
E-mail: [qsar32@reprori.jp](mailto:qsar32@reprori.jp)、[http://bukai.pharm.or.jp/bukai\\_kozo/index.html](http://bukai.pharm.or.jp/bukai_kozo/index.html)

## 構造活性相関部会の沿革と趣旨

本部会は構造活性相関懇話会として、1975年5月 京都において第1回シンポジウムを開いたのが始まりである。1975年度は2回のシンポジウムを開催し、以降1978年までは依頼講演4～5件、半日の簡素な形式であった。1980年より一般講演を募集し、年1回の構造活性相関シンポジウムが関係諸学会の共催の下で開かれるようになった。1993年より同シンポジウムは日本薬学会医薬化学部会の主催の下、関係学会の共催を得て行なわれることとなった。1994年より構造活性相関懇話会の名称を同研究会と改め、シンポジウム開催の実務担当グループとしての役割を果たしてきた。2002年4月からは、日本薬学会の支援を受けて構造活性相関部会として新しく組織化され、関連諸学会とも密接な連携を保ちつつ構造活性相関に関する学術・研究の振興と推進に向けて活動することとなった。

1975年当時、関係する領域における科学技術のめざましい発展にもなあって、医農薬を含む生理活性物質の構造活性相関と分子設計に対する新しい方法論が国内外に台頭してきた。このような情勢に呼応するとともに、研究者の交流と情報交換、海外諸国における研究の紹介、および国内における研究発表と方法論の普及の場を提供することを目的に設立された。以来、懇話会として**構造活性相関シンポジウム**の実行支援のほか、南江堂より、化学の領域 増刊122号：薬物の構造活性相関（ドラッグデザインと作用機作研究への指針）、および同増刊136号：同第二集（ドラッグデザインと作用機作研究の実際）をそれぞれ1979年と1982年に編集、出版するとともに、構造活性相関講習会を開催するなど設立の趣旨に応じた活動を進めている。本部会の沿革と趣旨および最近の動向などの詳細は、([http://bukai.pharm.or.jp/bukai\\_kozo/index.html](http://bukai.pharm.or.jp/bukai_kozo/index.html)) ホームページを参照願いたい。

## 編集後記

構造活性相関研究会の部会誌 SAR News の第6号をお届けいたします。ご多忙の中、ご執筆頂きました諸先生方に心よりお礼申し上げます。

本号より、編集委員長が黒木保久先生から清水に交代しました。不慣れな新米委員長のため、本号をお届けするのが遅くなってしまい、申し訳ございませんでした。また、中川好秋先生が編集委員から会計に替われ、山上知佐子先生と福島千晶先生に新たに編集委員会に加わって頂きました。

昨年11月には東京・星薬科大で第31回構造活性相関シンポジウムが開催され、たいへん盛況でした。今回から「ポスター賞」が新しく設けられ、3件の研究のご発表が第一回ポスター賞を受けられました。本号ではこれら3件のご発表について、演者の先生方にご紹介頂きました。このポスター賞が伝統ある構造活性相関シンポジウムのさらなる活性化につながると信じております。

今年は、6月には藤井信孝先生と赤松美紀先生のお世話で構造活性フォーラム2004（京都・京大会館）が、11月末にはリバース・プロテオミクス研究所の田中明人先生のお世話で第32回構造活性相関シンポジウムがつくば国際会議場で開催されます。皆様奮ってご参加下さいますようお願い申し上げます。  
(編集委員会)

SAR News No.6 平成16年3月31日

発行：構造活性相関部会（常任世話人代表：藤原 英明）

SAR News  
編集委員会  
(委員長) 清水 良  
石黒正路  
黒木保久  
高橋由雅  
福島千晶  
藤原 巖  
山上知佐子

\*本誌の全ての記事、図表等の無断複写・転載を禁じます。