

# 標的受容体構造に基づく内分泌攪乱物質の三次元構造活性相関解析

富岡 伸夫、板井 昭子

株式会社医薬分子設計研究所

講演に先立ち、本シンポジウムでお話しをさせていただく機会をくださった主催者と菅野博士にお礼申し上げます。

私どもの会社である株式会社医薬分子設計研究所(IMMD)は、コンピュータを利用した医薬分子設計の研究と応用を目的に1995年に創立されました。本日は、コンピュータを用いた技術について最初にお話しし、続いて、内分泌攪乱物質研究へのその技術の応用についてお話しします。

このスライドは、医薬分子の設計と開発の一般的な流れを示したものです。何らかの生物活性を有するタンパク質を標的として選び、そのタンパク質の活性を変化させることのできる化学物質群をハイ・スループット・スクリーニング(HTS)による化学物質ライブラリーから選び出します。こうした物質のことを「リード物質」と呼びます。リード物質を、その活性に応じてさらに選別し、医薬分子の候補として優れた物質をいくつか選び出して、臨床試験を行ないます。

一般的な例では、これらの段階には多くの実験と長い時間が必要になることがあります。私どもの会社の使命は、これらの諸段階を、コンピュータ技術を用いること促進できる技術を提供することにあります。コンピュータはシリコンでできていますので、この種の技術のことを“*in silico*”技術と呼んでいます。

まず最初に、我々の *in silico* 技術の主要部分である「ドッキングシミュレーション」についてお話しします。

医薬分子は、体内の特定のタンパク質に結合することで、その活性を発揮するようになります。結合した分子は標的タンパク質と共に、これに示すような安定した複合体を形成します。この図では、タンパク質構造の骨格をワイヤで表示してあります。医薬分子が、タンパク質の形の合ったポケットに結合しています。このタンパク質-リガンド複合体の構造は、タンパク質結晶解析で解明されたものです。

ドッキングシミュレーションは、こうした複合体の構造を、結晶解析実験をせずに予測することを目的にしています。

ドッキングシミュレーションは、簡単に解けるような問題ではありません。リガンドの回転やタンパク質の変形、さらには結合回転によるリガンドの立体構造変化などの莫大な数の自由度を考えなければなりません。リガンドがタンパク質に結合したときの立体構造変化は「活性コンフォメーション」と呼ばれます。リガンドの活性コンフォメーションは、リガンドが単体でいるときの安定コンフォメーションとは同じではありません。

ドッキング問題を解決するために、私たちは自動化ドッキングプログラムすなわち“ADAM”を開発しました。

ADAM は、水素結合パターンの一致に基づき、ドッキングモデルを段階的に最良化することで、リガンドの相対的な定位と立体構造を系統的かつ効率よく探索します。ドッキングモデルの評価は、エネルギースコアに基づいて行ないます。ADAMは、高い精度と高速性が利点となっています。

理想的なドッキングシミュレーションでは、最上位にランクされるドッキングモデルの構造が結晶の実際の構造と一致する必要があります。

このスライドは、2種類のタンパク質のADAMドッキングの結果を示したものです。結晶で観察されたリガンドの構造は緑で、ドッキングモデルの予想構造を黄色で示します。ここに示されたリガンドは柔軟な立体構造であるにもかかわらず、ADAMはリガンドの立体構造と位置を正確に予測できています。リガンドの原子の、実験で得られた位置からのずれである平均二乗偏差は、きわめて小さい値です。

ADAMドッキングの精度は、多くのタンパク質-リガンド系で検討されました。その多くの場合で、結合したリガンドの位置と立体構造をきわめて正確に予測することができました。

ADAMドッキングは高い精度と高速性が特長の自動方法なので、それを *in silico* スクリーニングの目的に使用することができます。

このスライドは、ADAMを利用した *in silico* スクリーニングの典型的な流れを示したものです。

市場に流通している物質もしくは実存しない物質の 3 次元データベースを準備し、それぞれ化学物質を、ADAM でもって標的タンパク質にドッキングさせます。うまくドッキングできた化学物質は、もっとも安定な複合体の構造に基づいてスコア化されます。そのスコアをはじめとするいくつかの基準に基づいて、少数の適合する化学物質を選別します。

典型的なリード物質発見のためのプロジェクトでは、我々がこの段階において約 100 種類の化学物質を選択し、その物質を購入するなり合成するなりしてから、その物質の活性を検証するための実験を行います。もっとも良い場合には、選択した物質のうちおよそ 40%に活性が見られ、次の最適化の段階に進みます。

このスライドは、実際の医薬分子開発プロジェクトの中で我々が行なった *in silico* スクリーニングの結果を示したものです。選択した適合物質における活性物質の比である「成功率」は、標的タンパク質の難しさや仮定の置き方によって、数%から 40%までのばらつきがあります。

このスライドは、*in silico* スクリーニングをハイ・スループット・スクリーニングと比較した場合の利点についてまとめたものです。

まず、品切れの物質であっても、実在しない物質であっても扱うことができます。物質サンプルが必要になるのは、スクリーニングが終わった後です。選択する物質の数が少ないので、HTS には適さない手の込んだ試験法を用いることができます。ドッキングによってタンパク質-リガンド複合体の構造が判りますので、ぴったり合う化学物質の活性を向上させるための構造改変のヒントが得られます。さまざまなタンパク質の結晶構造の解明が迅速に蓄積されますので、*in silico* スクリーニングが適用できる標的物質の数も迅速に増えていきます。

先ほどのスライドで、*in silico* 技術に関して簡単にお話ししました。

この後は、内分泌攪乱物質の構造活性相関の研究における、我々のこの技術のエストロゲン受容体への応用についてお話しします。

エストロゲン受容体は、多くの内分泌攪乱物質の典型的な標的タンパク質です。

エストロゲン受容体は、N 末端ドメイン、DNA 結合ドメイン、そして C 末端のリガンド結合ドメインから成ります。リガンド結合ドメインの結晶構造は、いろいろな化学物質をリガンドとして解明されています。

この図は、DES と複合体を形成したリガンド結合ドメインの結晶構造を示したものです。ここでタンパク質の中に埋もれているポケットは、DES 結合部位です。エストロゲン受容体を変化させる内分泌攪乱物質は、このポケットに結合するものと考えられます。

このスライドは、既知の内分泌攪乱化学物質の構造活性相関を解析するためのコンピュータ計算法を示したものです。

内分泌攪乱物質の化学構造を、3 次構造に変換し、ADAM プログラムによってエストロゲン受容体の結合部位にドッキングさせました。タンパク質-リガンド複合体の構造を、エネルギー最小化プログラムによってさらに最適化し、結合自由エネルギーの点数を GenB プログラムによってスコア化します。最後に、これらのスコアを、実験で得られた相対結合親和性と比較します。

リガンドの結合親和性は、相対的結合自由エネルギー、すなわち“ $\Delta$ -G-結合”で得られます。

GenB プログラムでは、タンパク質-リガンド複合体の“ $\Delta$ -G-結合”を、これら方程式を用いて、立体構造の項、水和作用の項、相互作用の項を考え合わせることで、算出します。

GenB プログラムは、実験による結合親和性を再現するために、何種類かのタンパク質ファミリーのタンパク質-リガンド複合体を用いてパラメトリック化されています。このリガンドでは、実験による結合親和性が、1 桁程度の平均二乗誤差によって予測できています。

このスライドは、今回の研究で用いた結晶構造をまとめたものです。

エストロゲン受容体は、結合するリガンドの種類によって、立体構造が変化する部位がそれぞれ異なります。こうした立体構造変化を考慮に入れる目的で、我々は、アゴニストおよびアンタゴニストを含めたさまざまなリガンドと合成された複数の結晶構造を用いました。また、結合部位の近傍にあるいくつかのアミノ残基の立体構造変化についても考慮に入れました。

我々は、それぞれのタンパク質構造について、既知の 110 種類の内分泌攪乱化学物質を ADAM プログラムによってドッキングさせました。

続いて、それぞれの化学物質を GenB プログラムによってスコア化し、回帰分析によって物質ごとにもっとも大きい値のスコアを選びました。この実験では、110 種類の化学物質のうち 82 種類がスコア化できました。

このスライドは、82 種類の化学物質について、ドッキングシミュレーションから得られたスコアと実測された相対結合親和性との比較をまとめたものです。ドッキングスコアは相対的結合自由エネルギーを表わしているものですので、この値がより小さい負の値であるほど、結合親和性が大きいことになります。相対結合親和性の対数が 2 であるものが、 $\beta$ -エストラジオールへの結合親和性を表わします。

この分析の対象となった化学物質の多様性にも関わらず、結合親和性の予測値と実測値との間の相関は、まずまず良好なものでした。過大評価ないし過小評価してしまう例が依然としてありますので、現在我々は、この種の誤差を減らすためにシミュレーションの手法やパラメータの改良を現在でも続けています。

その一つの例として、*in silico* スクリーニングの予備試験を行ないました。

この試験では、50 種類の既知の内分泌攪乱化学物質を、「入手可能な市販試薬データベース (Available Chemicals Directory)」から任意に選び出した大きな数の物質群に含めて、各物質をエストロゲン受容体にドッキングさせてみました。そして適合化学物質を前述の方法によってスコア化し、スコアでランク付けをしました。

「入手可能な市販試薬データベース」に記載されている物質のほとんどは内分泌攪乱物質ではありませんので、理想的には、既知の内分泌攪乱化学物質が適合物質の上位にランクされるはずですが。

このスライドは、任意に選んだ 24,000 種類の化学物質の中での、既知の内分泌攪乱化学物質のランクを示したものです。最下部にプロットされた点が、最上位のランクを表わします。

試験した 50 種類の内分泌攪乱物質のうち、42 種類は上位 1000 位の中にランクされており、その中の多くは 100 位以内でした。この結果によって、化学物質のデータベースから新規の内分泌攪乱物質候補を探索するのに我々の *in silico* 技術が有用であることがうかがえます。この試験においては、上位にランクし損なったり、標的タンパク質構造にドッキングできなかった物質が数種類ありました。こうした「偽陰性」の事例を減少させるためにも、我々の計算手法をさらに改良する必要があります。

最後に、今回お話した研究に関して IMMD の皆さんにお礼申し上げます。また、実験データと貴重なアドバイスをいただいた NIHS および CERI の皆さんにもお礼申し上げます。

ありがとうございました。

## 質疑応答

カブロック：ありがとうございました。質問はありますか？

質問：2つ質問があります。1つ目は、ADAMではタンパク質骨格にフレキシビリティを導入することはできますか？

富岡：できません。それで、私はERの結晶構造を複数用いています。

質問：2つ目の質問です。あなたがお調べになった化学物質のうちで、塩素や臭素といったハロゲンを持った化学物質にADAMは適用できるのでしょうか？それら化学物質をポケット内でモデル化できますか？

富岡：すみません。もう一度お願いします。

質問：塩素や臭素を持った化学物質の試験も行なえたのですか？

富岡：それは大丈夫です。塩素化や臭素化した化学物質にも使えます。しかし我々のプログラムにはまだ限界があります。化学物質は、タンパク質に少なくとも1個の水素が結合している必要があります。現在我々は、この制限を取り除くためにプログラムを改良しているところです。

質問：すみません、日本語でお願いします。

おっしゃったのかもしれないのですが、エストロゲン・レセプターのお話なのですが、これはエストロゲン様の活性自体の強さも、この構造からわかるのでしょうか。例えばDSとかE2とか、あとビスフェノールAなども、たぶん引かかってくると思うのですが、その強さなどもその構造から大体推測されるのでしょうか。

富岡：ええ。それがこの結果で、レラティブ・バインディング・アフィニティは、縦軸がエクスペリメンタルで横軸がプリディクションとなっています。

カブロック：他に質問はありませんか？

質問：質問してよろしいですか？24,000種もの化学物質をどのくらいの迅速さで扱えるのですか？スクリーニングに、実際どのくらい時間がかかったのでしょうか？

富岡：必要な計算時間は、化学物質1種につき、ドッキングに関して10~20秒、構造の最適化に関しておよそ1分、スコア化にもう1分ほどかかります。それほど速くはありません。2分または3分かかります。全体としては、我々は並列リナックスマシンを用いていますが、全部の分析を2日または3日で実行することができます。

質問：通常ではない結合分子は見つかりましたか？従来の一般的な意味でのことですが、何か妙なリガンドとか、新しい分子とか、エストロゲン結合性物質には挙げられていなかったようなものがありましたか？

富岡：エストロゲン受容体については、この手法によって非ステロイド性の物質が1、2個見つかりました。我々の主な目的は、その他のタンパク質に関する薬物を見つけることであり、その他のタンパク質についてはすでに成功をいくつかおさめています。

質問：1つ質問があります。あなたにとっては不公平な質問かもしれませんが、米国EPAが今すぐあなたのモデルを使用するとしたならば、化学物質のおよそ2%または3%が詳しい分析が必要なものとして検出されるのでしょうか？

富岡：と言いますと・・・

質問：あなたはその試験で50種の陽性物質を得ました。それは上位500種の物質においてであり、24,000種のうちの500種ですので2%か3%になります。ですので、活性の目印が立つ数字がそれくらいになるのではないですか？

富岡：この領域に活性がある物質ということですね。そうした物質の中には偽陽性のものが多数あると思います。500種ないし1,000種の化学物質を試験したとすると、活性物質対不活性物質の比率はおそらく

30%か 40%になるでしょう。それ以外は偽陽性によるヒットです。

質問：1 つ質問があります。たぶん私がきちんと理解していないのだと思いますが、得られた結果についてハイ・エフィシエンシーで序列をつけていますが、その 500 種の中で、かなりの比率のものがハイ・プレディクタブルだったのですが、非常にスケールアウトしたのも、エストロゲンのバインディング・アビリティを持っていましたね。その意味合いがよくわからないのですが、それは先生の・・・。

富岡：これでしょうか。スケールアウト。

質問：ええ。そこには出ていないのですが。非常にエフィシエンシーが高かったのですが、いくつかのスケールアウトしたのはプログラムのせいなのか、それとも何か別の要素なのか。

富岡：いえ、やはりまだ評価法の問題だと考えております。ですので、どちらかというところ、それはフォールス・ネガティブに近い。本当はやはり上位に行ってほしいのですが、タンパク構造の課程がこの方法に非常に重要で、先程、菅野先生もおっしゃっていましたように、エストロゲン・レセプターはヘリックス 10 からヘリックス 12 が非常にフレキシブルで、それによって構造を変えます。ですからその過程によって、まだコンパウンド【化学物質】によっては評価が悪くなってしまっているケースがあると考えており、そういう例だと思います。

質問：もう一つそれに関連して、フォールス・ネガティブが非常に低くなるのはいいのですが、本当は無限に低くなってほしいわけですね。そうしないと、スクリーニングの意味がない。ファーストステップのスクリーニングでそういう量を設定したらということですが、そういうことについてはどのようにお考えですか。

富岡：これは、まだ 0 にするというのは正直言って難しいと思います。先程言いましたように水素結合がないような場合でもドッキングを可能にするとか、技術的な改良で、限りなく近づけていくことはできるのではないかと考えております。

質問：ありがとうございます。

菅野：すみません、手短にお願いします。

質問：では手短にということで日本語でやらせていただきます。

このドッキングモデルで、アンタゴニストとアゴニストのときのレセプターのホールディングが違います。このモデルの場合は、要するにアクティブ・フォームでデザインなさっているのか、それともC端、C末が中に入っていない状態で、アンタゴニストの場合はC末が入っていないですね。ですからアゴニストとアンタゴニストの場合は、おそらくかなり評価が変わってくると思いますが、そこはこのモデルではどのようにお考えでしょうか。

富岡：これは両方タンパクを使っており、実際には両方のタンパクでドッキングを試行します。そうしますと、アンタゴニスト・フォームのタンパク質にしか入らないコンパウンドも出てきますので、その結果はあとでマージするという扱いをしています。

今、予測できているのはやはりレラティブ・バインディング・アフィニティだけであり、アンタゴニストであるかアゴニストであるかという予測は、まだできていないと思います。

菅野：たいへんありがとうございました。