



構造活性相関部会・ニュースレター <1 October 2006>

SAR News No.11

「目次」

/// Perspective/Retrospective ///

結晶解析の結果を上手に活用するために 平山 令明 … 2

/// Cutting Edge ///

CALTA 理論と統合データベースが可能にする ADME と活性の両立

— “Early ADME 導入後” を担う創薬データマイニング—

小林 好真 … 7

ドッキングポーズ評価のためのコンセンサスコア戦略

小田 彰史
山乙 教之
広野 修一 … 13

/// Activities ///

<報告>

- ・構造活性フォーラム 2006 高橋 由雅 … 18
- ・SAR Promotion Award 平成 18 年度受賞者 赤松 美紀 … 19
- ・情報化学入門講座 (第 4 回) 高木 達也 … 20

<会告>

- ・第 34 回 構造活性相関シンポジウム プログラム … 21

//// Perspective/Retrospective ////

結晶解析の結果を上手に活用するために

東海大学医学部医学科基礎医学系 平山令明

1. はじめに

コンピュータ科学の進歩により、かなり複雑かつ大きな分子系についての計算を行うことも現在では可能である。さらに、ソフトウェアの使い勝手の向上により、分子構造についての知識が少ない研究者・技術者でも比較的容易にこのような計算を実行できるようになって来た。このこと自体は非常に歓迎すべきことであるが、ともすると物理化学的な観点から意味の無い計算をしたり、求められた結果を疑いもなく受け入れてしまう危惧もはらんでいる。

創薬を最終目的として行われる種々のシミュレーションでは、分子構造が最も重要な基礎データである。ほとんど全てのシミュレーションでは、X線結晶解析で得られた分子構造を基準あるいは参照構造として採用している。しかし、シミュレーションの基礎となる分子構造が不正確であれば、シミュレーションも正確には行えない。X線結晶解析の原理は極めて明快で、その原理を理解することは決して困難ではないが、シミュレーションを行う多くの研究者・技術者がほとんどこの方法論について学んだことがなく、その原理限界を知らない場合が決して少なくない。そこで本稿では、そうした研究者・技術者のために、X線解析結果を活用する上での注意点を中心に述べたいと思う。

2. 低分子 X 線解析結果の取扱い

2.1. X 線解析で求められるもの

X線は主に電子によって散乱される。従って、X線解析で求められるのは、結晶内の電子密度の分布である。結晶は基本的に等価な単位胞が3次的に規則正しく配列したものであり、結晶中にある全ての単位胞を平均して、代表であるひとつの単位胞中の電子密度を求める。電子密度の局所的な極大点に原子核があると考え、原子の位置を決め、その点の電子密度の高さに基づきその原子種を帰属する。十分に電子の数が異なる原子同士であれば、原子種の帰属は困難ではないが、CとN原子のように電子数の差が1である場合には、それらの電子状態によっては両者の原子種を判別することは難しい。例えば、4級のN原子と4級のC原子の区別は難しい。通常、X線解析では先験的な知識なしに構造決定ができると信じられているが、このような場合、元素分析やスペクトル・データがないと判定に誤りが入る可能性がある。

電子密度が局所的に極大になる点の座標で各原子核の位置を代表する。しかしH原子の場合、結合した相手の原子に電子が引き付けられている場合がほとんどであり、極大点はH原子の原子核位置より、結合した相手の原子側に寄っている。従って、X線解析で得られたH原子の位置は原子核の位置とは有意に異なる。通常X線解析ではC-H結合距離は1Åより短くなるが、原子核位置を求めることのできる中性子線回折によると、この結合は1.08Å程度になる。従って、X線構造を分子力場計算に用いる場合には、X線解析で求められたH原子を全て一旦削除した後で、全てのH原子位置を幾何学的計算で求めることが必須である。無論、X線解析で得られた位置は、幾何学的計算の結果を評価するうえで非常に重要な情報である。

原子は結晶の単位胞の中で凍結されているのではなく振動している。この効果を考慮するために各原子に温度因子というパラメータを与える。原子の動きを表現するパラメータという意味で結晶学領域では、温度因子という言葉より原子変位パラメータ (atomic displacement parameter) という言葉が最近では推奨されているが、本稿では温度因子という言葉を使う。温度因子が大きい原子は結晶内での動きが大きいということになる。柔軟な直鎖状分子内の原子では、一般的に温度因子は高くなり、インドール環のように比較的固い部分の原子では温度因子は低くなる。もしインドール環全体が動いている時には、環内の全原子の温度因子は大きくなる。振動を表現するために、振動の球状近似と回転楕円体近似の両方が使用されるが、低分子X線解析では後者が専ら用いられる。この近似を使うと、運動の方向性がある程度分かる。

全ての単位胞は等価であると仮定したが、実際には単位胞中の内容が若干異なる場合がある。よくあることは、その分子は水和物になっているが、全ての分子が水和物になっている訳ではないという場合である。50%の分子が水和物になっている場合、それを1単位胞内での存在量で表すと0.5となる。この値を占有率と呼ぶ。水和した分子が結晶の表面に多いのか、それとも少ないのかは当然結晶解析では分からない。また占有率が0.45なのか0.55なのかを結晶学的に決定することも困難である。系統誤差等の影響を受けた電子密度図では、偽の電子密度もよく現れるので、占有率が小さい原子の存在証明はかなり困難である。元素分析や熱分析を併用しても、二桁目の精度について保証することは一般的に容易ではない。

ある程度の自由度が分子にあり、複数の安定な立体配座をこの分子が取ることができると、複数の安定な立体配座を結晶中の異なる単位胞中で取り、それでも結晶を与えることになる。この状況は直鎖状の分子だけでなく、環構造においても起こる。この場合も、複数の立体配座に関係する原子を一つの単位胞中の座標で表現するために、占有率を用いる。例えばA及びBという二つの立体配座が50%ずつ存在すれば、その立体配座に関係する座標Aと座標Bの占有率は各々0.5ということになる。 α アノマーと β アノマーが1:1で結晶中に存在する糖の場合など、実際にこういう状態が起こることは少なくない。構造解析の論文の中には占有率が0.65と0.35のように報告されているものがあるが、これらの値は両原子の温度因子が同じになるように占有率を最適化した値である。しかし、この最適化には大きな仮定がある。つまり占有率の大小に関わらず、それらの原子が等しい温度因子を持つという仮定である。この仮定はいつでも成立するとは限らない。そして、パラメータの性質上、占有率と温度因子に相関があることにも注意すべきである。占有率を大きくすれば、温度因子は大きくなる。仮に占有率が1.0になっている原子でも、同じ分子内の原子に比較して異様に温度因子が大きいなら、その占有率は1.0より小さいことを疑うべきである。

2.2. 人為的なミス

低分子有機化合物のX線解析は現在ではかなり容易に行える状況になっている。しかし、このことは結晶学に精通していない研究者・技術者による構造解析例がかなり増えていることも意味する。その弊害は、多くの誤りが解析結果に含まれることである。最近では、多くの学術雑誌で低分子有機化合物の構造解析データをCambridge Structural Database (CSD)に供託することが義務付けられているので、毎年非常に多くの構造がCSDに付加されている。従って、何らかの参照構造が必要な時、CSDは非常に重要な拠り所となる。CSDでは解析の信頼性について疑わしいものについては、コメントを出している。しかしこのチェック機能は種々の理由から十分ではない。また当然のことではあるが、CSD側はその構造解析の真贋を保証するものではない。従って、利用者がその分子構造を利用する前に、構造解析の妥当性を検証する必要がある。

例えば、最近筆者が遭遇した例では、 R 因子は0.07であるが、結合距離に通常値から標準偏差の3倍以上ずれている構造がCSDの中に登録されていた。その構造解析に関するデータを検討してみると、空間群に誤りのある可能性が出てきた。そこで実験誤差内で格子定数が一致する結晶を再調整して構造解析をし直した。その結果、CSDに登録されていた空間群 $P2_1$ は誤りで、 $P2_12_1$ であることが明らかになり、正しい空間群を採用すると、全ての結合距離は妥当な値になった。この例は、低分子有機化合物の結晶解析で、それなりの R 因子を与えていても、その正しさを100%保証するものではないことを示している。先の理由から、このような例は決して例外的なものではない。データベースからデータを抽出して利用する場合全てに当てはまることであるが、公共あるいはそれに類するデータベースの大半は玉石混交であることを前提にすべきである。上記の例のような場合には、分子構造のチェックで怪しいと睨むことも出来るが、結晶解析結果を機械的に判断するソフトウェア(例えばcheckCIF)を一旦通してから活用することも一つの方法である。上記の例では、checkCIFを通すと文字通り山のように警告が出てくることが分かった。

2.3. 良質な構造の選択

医薬分子設計の種々の過程で、特定の化学構造の特徴や分子間相互作用について、定量的な情報を得たい場合、結晶解析は極めて有用な情報を与えてくれる。筆者の研究室では、特定の部分化学構造についての知見を得たい時には、必ずCSDを用いてまずチェックすることになっている。しかし、特に複数の解析結果がデータベース中にある場合に、どのデータを最も信頼できる構造として採用するかが問題になることが多い。X線結晶解析では、実際に結晶から回折されたX線強

度 ($|Fo|^2$) とモデル構造に基づいて計算したX線強度 ($|Fc|^2$) の差が小さくなるように、モデル構造を修正する。 F はX線の振幅を表し、添え字のoとcは観測値と計算値をそれぞれ意味する。最終的にそれ以上改良できなかった段階での次式 R (この式では F の絶対値を用いている) が論文やデータベースに登録されている。 Σ は観測した全ての回折X線について和を取ることを

$$R = \frac{\sum (|Fo| - |Fc|)}{\sum |Fo|} \quad (1)$$

意味する。 R 因子とは、モデル構造がどれだけ観測データを説明できるか、その程度を表している。2005年11月版 (V.5.72) のCSDには340,342件の結晶構造が収録されており、その内 R 因子が3、5、7および10%以下の構造は各々8、52、82、および95%である。

SやCIより原子番号の大きい原子を構造 (塩も含めて) に含まない場合3%以下の構造は精密解析に属し、非常に厳密に構造が求められていることを示す。X線解析で得られる位置としてのH原子位置も求められている。このカテゴリーの構造は十分に信頼できる。しかし、もしこのカテゴリーの構造中に異常な結合距離がある場合には、計算に使用されている回折X線の数と最大回折角 θ をチェックすべきである。(1)式から判るように、 R の値は使用するデータ数によって大きく変わる。最近データ捏造が工業界だけでなく科学界でも騒がれているが、 Fo と Fc の一致度の悪いデータを削除してしまえば、 R 因子はぐっと低くなる。 R 因子のみを機械的に論文採択の基準に利用している論文誌があるとすれば、このような操作をすれば潜り抜けることができる。

同様の条件で、 R 因子が5%以下の構造は、ほとんど問題がない構造であると判断できる。筆者の研究室ではまずこの条件で該当する構造のみを最初に見ることにしている。現在では R 因子が5%以下の構造は半分以上を占めるので、これで大抵の用は足りる。 R 因子が7%は微妙なところで、2.2の例に見るように、かなり大きな問題を含んでいても、この程度に下がることがある。 R 因子が7-10%の構造は原子の並び方には間違いがないが、原子種の判定が微妙であることがある。筆者の研究室では、固有の構造を見る場合にはこのカテゴリーに属する構造は考慮しないことを原則としている。

SやCIより原子番号の大きい原子 (重原子) を含む場合には、回折データの質が良くても、 R 因子が下がらない場合がある。電子数の多い原子では、dやf軌道に電子が入るために、原子核周りの電子の広がりが球状ではない。しかし一般的には、原子のX線散乱能を球状関数で近似するので、どうしてもその差が埋まらない。従って、特に精密構造解析を行わないと、 R 因子は高めに出る。探索を目的とする化合物が重原子を含む場合は、検索に使用する R 値は若干大きくしても良い。

分子構造が硬い場合には、 R 因子は小さくなる傾向にある。これに対して、分子構造が柔軟である場合には R 因子は高くなる傾向にある。これは、後者では結晶内での分子同士の詰まり方 (パッキング) の規則性が減少するからである。自由度の高い分子の解析データを検索する場合には、予め R 因子を高めを設定しておくが良い。後述するように、分子の運動性が高いところの原子座標の不確定さは勿論大きい。

数が非常に少ないのが残念であるが、中性子線回折に基づく結晶解析の結果もCSDに登録されている。中性子線回折では水素原子位置も他の原子と同様の精度で決定できる。目的とする分子ないし部分構造を含む中性子線解析があれば、利用価値は高い。

3. タンパク質X線解析結果の取扱い

3.1. タンパク質結晶解析の特徴

タンパク質結晶解析も低分子結晶解析と全く同じ原理で行われるので、2章で述べたことは全く同様に当てはまる。タンパク質結晶解析の大きな特徴は、次の3点に要約される。第1は、結晶中にたくさんの水分子が含まれ、その大部分は明確なX線回折に寄与しないことである。第2は、バルクの水に接する分子表面を中心に分子の運動性が高いことである。第3は、これらの条件により、X線回折の条件が緊密にパッキングした低分子結晶に比較して格段に悪いということである。理想的なX線回折の条件は、単位胞中に十分な強さで分子が充填し、全ての単位胞中での原子位置が等価であることであるが、タンパク質結晶がこの条件を充分満たすことは稀である。

タンパク質と低分子間の相互作用を原子レベルで理解し、それに基づき分子設計を展開するためには、かなり質の高いタンパク質立体構造を利用する必要がある。これは通常の構造解析者が構造に求める質以上のものである。当然であるが、不正確な構造からは有用な知見は生まれない。タンパク質結晶解析の質を計るための指数として良く使われているのが分解能 (resolution) と R 因子である。分解能は、識別できる最小の 2 点間の距離と考えて良い。2 Å 分解能では、基本的に 2 Å 以上離れた点のみを識別できる。例えば一つの単位格子の長さが 50 Å で、その方向で観測できる回折 X 線が 25 次までであれば、分解能は 2 Å となる。しかしこの数字は観測できたデータの上限を意味しているため、それら上限のデータの質を決定するものではない。つまり分解能が良くても、その数字に相当する構造が求められていない場合があり得ることを示す。タンパク質結晶の性質上、分解能の近くで急速にデータの質が落ちることが一般的である。回折データは 3 次元的に出るので、球で近似すると、半径の 3 乗でデータ数が増えるので、分解能限界付近でデータの質が落ちると、質の悪いデータが占める割合が高くなる。最近の構造解析の論文では、データの質を保証するために、幾つか分割した分解能範囲におけるデータの質を示している。データの質の判定は最終的には構造解析の結果によって判断されるが、等価なデータの一貫度 (R_{merge}) で表されることが一般的に行われている。しかしこれは満遍なく均質にデータが取れたことを保証するものであるため、系統的な誤差が少なかったことを意味するものである。筆者らはドッキング等の原子レベルでの解析を目指す場合、分解能 < 2 Å を目安にしている。低分子有機化合物の X 線解析での分解能は通常 1.0 Å 程度である。

最近のタンパク質 X 線構造解析では R_{free} という値を上記の R 因子の代わりに使うことが多い。(1) 式の Σ では観測された全データについての和をとっているが、 R_{free} では全く異なる。 R_{free} を計算するには、予め観測データの中から少数のデータをランダムに除いておく。残りの大多数のデータを用いて評価すべきモデルが確定した時、そのモデルとこの少数のデータに基づいて計算されるのが R_{free} である。 R_{free} の計算に使われるデータはモデル構築用に使われていないので、モデルにその影響の入っていないデータと考えることができ (厳密には異なるが)、 R_{free} は全てのデータを含めて計算される R より、構造の確かさを示す、信頼性の高い指数とされている。 R_{free} が 0.20 以下を与える構造は概ね信頼できる構造を与えるものと、一般的には考えられている。ドッキング等の原子レベルでの解析を目指す場合、 $R_{\text{free}} < 0.20$ を筆者は目安にしている。しかし上記の事情を考慮することは勿論である。

3.2. 化学構造の曖昧さ

タンパク質結晶の特性により、分解能は高くない。このことは全ての非水素原子の座標を独立に決定できる十分な観測データが得られないということの意味する。仮に、2 Å 分解能での最高の電子密度が出て、タンパク質構造に対する予備知識がなければ、解釈は到底不可能である。つまりその電子密度分布に妥当な分子像をあてはめることはできない。勿論、C と O 原子の区別も客観的に行うことはできない。実際には我々が既にタンパク質の構造 (アミノ酸の構造) についての知識があるので、この知識を援用して分解能不足の問題を解消している。トリプトファンのような大きなアミノ酸残基の位置を頼りにアミノ酸配列を参考にして、最も電子密度と一致する残基の配置を求めるのである。低分子 X 線解析では各原子の原子パラメータ (座標と温度因子) 一つ一つの標準偏差を求めることが出来るので、結合距離、結合角、ねじれ角、分子の平面性などの分子内の幾何学的な値を定量的に議論できるだけでなく、水素結合、ファン・デア・ワールス接触などの非結合原子間相互作用の有無や強さなどを定量的に比較できる。しかしタンパク質 X 線解析では、各原子座標に対する標準偏差を見積もることは出来ない。従って、予め化学構造が分かっているリガンド分子の構造を電子密度から決定できないのは勿論のこと、二重結合か単結合か、酸が解離しているかどうか、水素結合は本当にあるかどうか、というような化学構造に対する定量的な回答はほとんど出来ない。結晶が良好で例外的に標準偏差を求めることが出来たタンパク質の構造解析例はあるが、非常に少なく、今後もそれが可能な系は稀であると推測される。

タンパク質の構造ファミリーの確認や美しいリボン構造の描画が出来ても、原子レベルでの分子構造や分子間相互作用を議論する上で、現状のタンパク質結晶解析から得られる原子パラメータの精度はかなり低いことをまず認識する必要がある。

3.3. それではどうするか

3.1 で述べた筆者による基準に合格するタンパク質構造解析でも大きな問題を持っているものは少なくない。問題をタンパク質とリガンド側の問題に分けて考えることにする。

タンパク質側の問題点は、おもに原子種の帰属、温度因子および占有率である。原子種の帰属が明瞭にできないので、アスパラギンおよびグルタミンの側鎖のOとN原子の区別はできない。またヒスチジンのN原子の位置も確定できない。水素原子の位置が推定できれば、予想される水素結合様式から、これらのヘテロ原子の帰属を行うことが可能であるが、タンパク質結晶解析ではそれができない。この問題に対して、Richardsonらは、タンパク質構造中のそのアミノ酸残基周囲の状況から最も妥当なヘテロ原子の帰属を提案するプログラムを報告している。筆者らの経験では、かなり精度良く決定されている構造であっても、その構造が妥当でないことがこのプログラムで指摘される。勿論、このプログラムが提案する構造が逆に必ずしも正しくない場合もある。つまり、原子種の判別は、タンパク質X線解析の限界外の問題であるので、可能であればX線以外の知識も総動員して、もっとも妥当な化学構造を推定するしか現状では方法がない。温度因子と占有率は非常に難しい問題である。筆者らは占有率が1.0以下のアミノ酸でその温度因子が 30 \AA^2 以上の場合には、その位置は疑わしいと判断している。また占有率が0.5以下ないし温度因子が 50 \AA^2 以上のアミノ酸の位置は、X線の構造はあくまで参考とし、むしろ分子力学的に妥当な側鎖構造を優先することにしてしている。これらの判断も実はかなり主観的である。原子レベルでの議論が必要な場合は、この基準でも甘いかも知れない。

リガンド側の大きな問題は、化学構造と解離状態である。Protein Data Bank (PDB)の書誌的データに記載されている化学構造や原著論文に記載されている化学構造が妥当でない場合が少なくない。分子設計の参考として複合体の中のリガンド分子の化学構造を用いる場合には十分に注意が必要である。解離状態についても、タンパク質X線解析は直接的な証拠を出せない。そこで、筆者らは、 -1 から $+1$ までの電荷を持った解離状態を発生させ、その中で最も妥当なものを考慮するようにしている。非水素原子の占有率が0.5以下あるいは温度因子が 50 \AA^2 以上の場合には基本的にそのリガンドの結合様式は信頼できないと筆者は考えることにしている。原子レベルでの比較対照をする場合には、原子の占有率が1.0で温度因子が 30 \AA^2 以下のリガンドを用いることにしている。

3.3. 誤差を見積もる

分解能も R_{free} も、その構造解析の質を保証しないものであるなら、何を根拠に解析された構造の質を評価すれば良いのか。これは非常に重要な問題であるが、余り熱心に議論されない問題でもある。タンパク質構造解析の論文中には、最終構造を評価する量として R_{free} 以外の種々な量が掲載される。第一は、Ramachandran Plotである。これは主鎖の ϕ と ψ 角をX-Y座標上にプロットしたもので、典型的な α ヘリックスや β ストランドなどの二次構造からのずれを表示したものであるが、どの程度ずれたら悪いのかという基準がない。第二は、結合距離や結合角などの幾何学的諸量の標準的な値からのずれである。しかし、これら標準的な値になるようなポテンシャル関数を用いて最適化する限り、これらのずれも構造の妥当性を100%保証するものではない。第三はLuzzati plotと呼ばれる座標精度の評価である。この方法は現在でも使われているが、Cruickshankにより、このような応用には不適切であることが指摘されている。Cruickshankはこれに代わるものとして、実験で得られる回折データの良好性と十分さを考慮して、原子座標の誤差を見積もる量としてdiffraction-component precision index (DPI)を提案している。DPIは、Blowにより変形され、より使いやすいものになっている。筆者は、ドッキングの正しさを客観的に見積もる必要がある時に、DPIを便利に使用している。DPIは非常に有用であるが、結晶学領域では余り活用されていない。しかし、構造解析の結果を活用する分子設計の領域ではもっと活用すべきであると筆者は考えている。

4. さいごに

紙面の都合で、十分な説明ができなかったが、結晶解析の結果を活用する上で予め知っておくべきことについては凡そカバーした積りである。本稿で述べたことをもう少し深く理解するには、やはり結晶解析の原理について学んで頂くのが、最良と考える。

分子力学や分子軌道法の計算により分子構造が簡単に求められるようになり、実験的に求められた構造データを軽視する向きがある。しかし、理論科学はまだ未熟で、暫くの間は実験科学による軌道修正がないと、ゲームの世界の話に落ち込んでしまう危険性を多分にはらんでいる。

//// Cutting Edge ////

CALTA 理論と統合データベースが可能にする ADME と活性の両立

— “Early ADME 導入後” を担う創薬データマイニング —

第一製薬（株）創剤代謝研究所 小林好真

1.1. Early ADMEで「振るい落とす」だけでは、創薬成功確率は上がらない。

「臨床試験に失敗する最大の原因は体内動態」¹⁾。このように言われてから約10年がたった。今や製薬業界では、創薬初期段階からのeADME (early ADME)評価によりADME/物性が望ましくない候補品を避けるのが当然となった。その結果、体内動態が原因で臨床試験に失敗する薬剤は確かに世界的に激減した²⁾。筆者が勤務する第一製薬でも研究部門全体を挙げて新しい創薬システムを導入し、たとえば最近数年間で臨床開発段階に上げた経口剤は全て良好な経口吸収性と体内動態を有している³⁾。

しかしそれでもなお、近年日米欧で承認される新薬数は減少傾向にあるうえ開発コストは上昇の一途である。いっそうの創薬効率化の必要に迫られている。ADMEに関しても当然、化合物をeADMEスクリーニングでただ「振るい落とす」だけでは非効率極まりない。とくに資金力に劣る日本企業は、ADME、活性、毒性、その他全てに十分な資質を有する化合物を得るために、リード探索時から創造力をフル回転させる必要がある。

1.2. ADME(T)と活性の“非両立性”・・・二兎を追うのが難しい？

とはいえ、ADME(T)と有効性両方の条件を満たす化合物の創製は容易ではない。「活性が高い化合物を作ってみたらADME(T)/物性が悪くなり、ADME(T)がよい化合物は活性がなくなった」という事例が各社で頻発していると聞いている。

この「ADME(T)と活性の“非両立性”」はどんなときに起きるのか？なぜそうになってしまうのか？本当に、これらの“二兎”を同時に追う方法はないのか？つまり、この非両立性を回避してADME(T)と活性、両方が優れている化合物を創製する方法はないのか？さらに言えば、本当に製薬各社の大量のeADME(T)データは十分に構造との相関研究に生かされているのだろうか？文字通り「スクリーニング(振るい落とす)」だけに終わっていないだろうか？薬物の構造-ADME(T)相関ならびに構造活性相関をどのように精査し、どのようにリード骨格を選んでいくべきだろうか？

1.3. 二兎を追うための解決策

筆者らは、ADME(T)と活性の“非両立性”の原因の一つを解析し、物性評価とデータマイニング手法を駆使して、ADME(T)と活性の双方に優れた化合物を得るための方法論を確立した。今回この稿では、第一製薬社内で活用されており、他社の創薬現場でも取り入れられ始めた構造-活性-ADME(T)戦略のコア、CALTA理論および統合データベースKameDBを中心に事例紹介をしたい。

2.1. Log D-活性相関図— CALTA-map

手始めにまず、Fig. 1を見てみよう。これは、第一製薬内での創薬探索プロジェクトのいくつかについて、log DをX軸、log (“活性”)をY軸に取ったグラフに、探索化合物をプロットしたものである。第一製薬ではこれを「CALTA-map」と名づけて活用している（命名の由来は後述）。なお、log Dはオクタノール/緩衝液（緩衝液のpHは主に7.4）の分配係数であり、全て実測値である。プロジェクトによっては酸性化合物と塩基性化合物が混在する場合があるが、Fig.1にはどちらかだけをプロットしたものを提示した。活性についての詳細は公開できないが、基本的にはIC₅₀値のようなin vitroでの実測データである。

さてそれぞれのCALTA-mapを見ると、2パターンに分かれることが一目瞭然である。きれいに逆三角形型に分布する場合と全体に“均等散布”する場合とである。逆三角形はみなCALTA-mapの“右上”に位置しており、三角形を大きく外れるプロットは皆無である。興味深いことに、プ

プロジェクトによって全く異なる種類の活性評価系を使っているのにも関わらず、三角形の“斜辺”の傾きは必ず約-1となっているのである。もう一つ興味深いのは、標的が膜近傍のタンパクの場合は全て“逆三角形型”のCALTA-mapになるという事実である。（“近傍”という表現を用いたのは薬物結合部位が同定されていないケースも含まれているため）。一方、“均等散布”型はいずれも標的が膜とは無関係の位置にあると考えられている創薬プロジェクトで確認された。

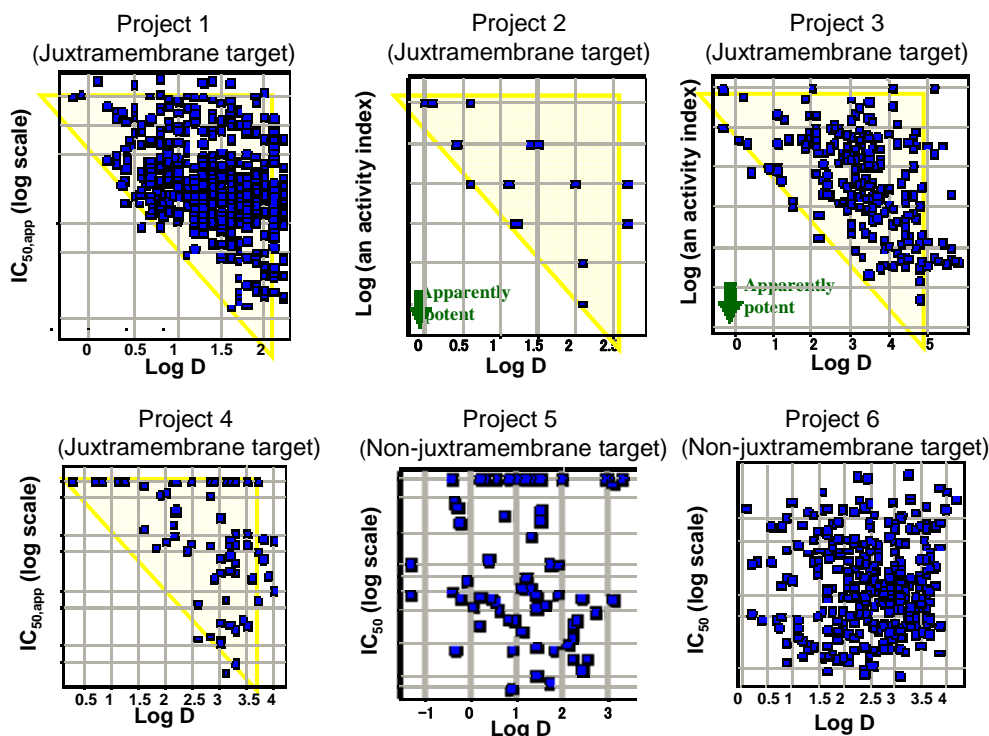


Fig. 1. CALTA map analysis for some drug discovery projects in Daiichi Pharm.

Project 1-4 (juxtramembrane-target) : The plots were distributed in an “inverted triangle” shape.

Project 5,6 (non-juxtramembrane-target) : The plots distributed “evenly”.

2.2. CALTA 理論（第一段階） ～「真の活性」の概念～

Fig.1 に示した現象は、実は我々が以下の理論からあらかじめ予測していたとおりのものであった。

膜タンパクが標的のとき、まず一つのモデルとしてFig.2 のような相互作用様式を考える。ここで、平衡関係を考えれば、*in vitro* の評価データとして観測される ($K_{i,app}$) は、標的との特異的な親和性 ($K_{i,int}$) と膜への分配 (P) の積となる (式1)。活性として IC_{50} を取れば、式2 が成立する。対数を取れば式3 となる。

古くから知られるように、生体膜への分配は octanol/緩衝液分配係数と平行とみなせる場合が多いので、式3 は $\log D$ を用いて式4 のように表すことができるものとする。

当然、ドラッグデザインにおいては脂溶性に依存しない「真の活性」、つまり標的との特異的な親和性 ($K_{i,int}$) の向上が必要である。 P を上昇させても非特異的に「見かけの」観測活性が上昇するだけである。

ところが標的が膜近傍蛋白のとき、考慮を怠るとリード骨格を見誤る危険性がある。例えば薬物-標的相互作用を模式図 (Fig.3) のように単純化した場合、標的結合と関係しない部位 (Portion B) の変換であっても、脂溶性の部分構造を入れれば P が上昇し、観測活性が向上する。観測活性と構造式だけを眺めていると Portion B を活性発現に必須なリード骨格と誤認してしまう恐れがある。当然、本来は真の活性に重要な Portion A についての構造活性相関を解明し、最適化された部分構造をリード骨格とすべきである。

しかし通常は、膜近傍蛋白標的については $K_{i,int}$ を大量に直接実測するスクリーニング系の構築は困難である場合が多い。そこで、脂溶性と見かけの活性とを実測し、それらの相関から真の活性

を求めていく必要が生じる。これがCALTA (Correlation Analysis for Lipophilicity and Target Affinity) 理論である。

2.3. CALTA 理論 (第二段階) ～CALTA map～

式4を、 $\log D$ をX軸に $\log IC_{50,int}$ をY軸に取ったグラフに図示するとFig.4 のようになる。真の活性 ($IC_{50,int}$) が等しい化合物は傾き-1 の同一直線上にのり、そのY切片が真の活性に相当する。また、真の活性を高める構造変換を行えば、その直線は“左下方向”に移動する。

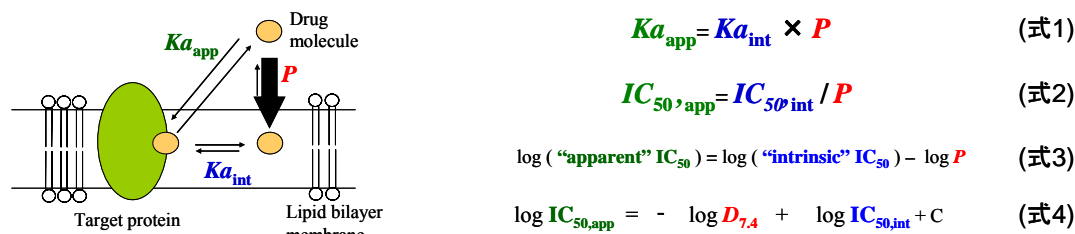


Fig. 2. Illustration of CALTA theory

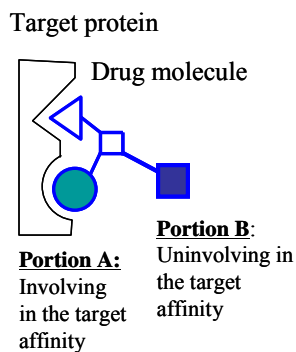


Fig. 3. An interaction model between the drug and the protein

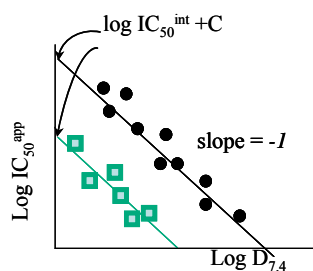


Fig. 4. CALTA map and lead discovery

つまり、Fig.1 に示した“逆三角形型”は、この直線の集合の結果得られたものとみなすことができよう。“逆三角形”の斜辺が-1となる説明もつく。なお、Portion Bに位置する部分構造が立体障害などの理由により活性を減じる事例もあるが、その場合は直線の上方にプロットが外れ、下方にずれることはない。

したがって、構造変換を行ったときCALTA-map上で傾き-1の直線に沿って果てしなくプロットが動いていくような部位は真の活性に重要とは言えない。逆に良いリード骨格とは、map上でより左下の直線を為す骨格であると言える。研究者は、CALTA-map上でプロットを見ながら、直感的に骨格を判断し、構造変換を着想できる。

2.4. CALTA 理論創薬と ADME データマイニング。～二兎を追うために～

“逆三角形型”のCALTA-mapの場合、活性を最大限に上げようとするれば、Portion Bで最大限に脂溶性を上げて非特異的な「見かけの活性」を上げればよい。ところが、高すぎる脂溶性は薬物のADME(T)特性を望まぬ方向に向かわせる。従って、活性とADME(T)の“非両立性”が生じてくる。

実際、Fig.1で“逆三角形型”に分類された4プロジェクトについてのeADME(T)評価結果は、LogDが高くなるほど全て“悪化”の傾向にあり、いずれの化合物も当時社内ですべて設けていた振るい落とし基準の何かには引っかかってしまっていた。

よく知られるように数十年来のQSAR研究の結果、非常に多くの毒性項目への脂溶性の関与が世界中で報告されている。つまり高脂溶性の高活性化化合物に関しては、たとえ既存の非臨床試験

をクリアしたとしても、予想もしない毒性による臨床試験の中止や、上市後に新たに見出された副作用による市場撤退の危険性が、常に付きまとうのである。

したがって、非両立性を避けるためにも、脂溶性に依存しない「真の活性」を可能な限り高め、不必要に脂溶性を上げないことが重要である。

2.5. CALTA 理論でのリード探索・・・Key Based Clustering と 3D-QSAR 解析

以下、in silico アプローチを用いて、CALTA 理論を活用した「真の活性」に重要なリード探索を行った一例を紹介しよう。

まずFig. 1 に示したあるプロジェクトの化合物群について、構造式を表すMolKey (MDL)をもとにクラスター分類 (DecisionSite™ (Spotfire), Tanimotoの距離およびWard法を使用) を実施した。

得られた各クラスターそれぞれから化合物を選び、Catalyst™ (Accelrys, HypoGen機能を活用) を用いてpharmacophoreを生成・評価した。この際のポイントは、入力する活性値として、式4から算出した $IC_{50,int}$ を用いた点である。

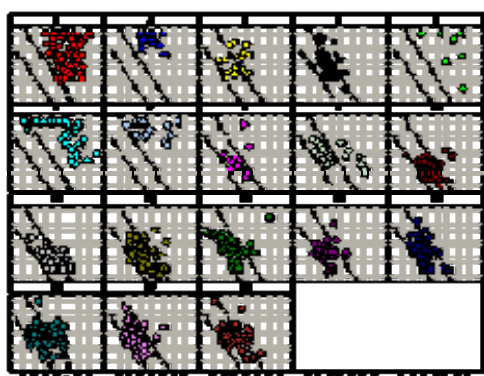


Fig. 5. CALTA-map for each of the clusters that were generated by a hierarchical cluster classification of the compounds of Project 1 based on their structural features.

It shows that the clusters ranged from right-upper to left-lower on the map. This indicates that the compounds' substructures affected their intrinsic activity.

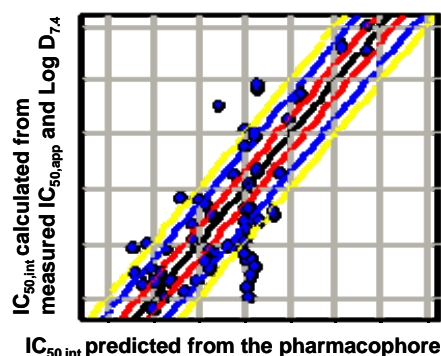


Fig. 6. Predictivity of one of the pharmacophores generated using $IC_{50,int}$ that was estimated based on CALTA theory

結果をFig 5,6 に示す。Fig 5はクラスター数を18にした例である。構造式情報のみで分類しただけであるが、それでもCALTA mapの右上～左下の方向に化合物が分類されている傾向が見取れる。化合物の構造と密接な関係があるのは“真の活性”であることが反映されている結果だろう。Fig 6は、化合物の構造式と $IC_{50,int}$ とを元に得られたあるpharmacophoreの予測精度を評価したものである。X軸はpharmacophoreに新規化合物の構造式を当てはめて得られた予測 $IC_{50,int}$ である。Y軸は化合物の実測 $IC_{50,app}$ および実測log Dとから、式4を用いて計算された $IC_{50,int}$ である。それなりに良い予測精度であった。ここで得られたpharmacophoreがリード骨格となり得る。一方、 $IC_{50,app}$ を活性値として入力したときのpharmacophoreは非常に低い予測精度であった(結果図示せず)。

なお、構造式の表現/クラスタリング方法は日々進歩を遂げているものの、現実にはまだ、化学者が目で見て構造を分類するほうがより正確な場合が多い。Fig 5の創薬プロジェクトについても、目で見て構造を分類することで、明確な直線群や真の活性に重要な部位を見つけることが出来た。もう一つ付け加えると、近年 clog D の予測精度も向上しているが、CALTA 理論を十分活用するには $\pm 0.1 \sim 0.2$ 程度の精度が必要であるため、clog D 値のみでは困難なことが多い。精度が十分な実測値の蓄積が重要となる。

3.1. リード最適化, 適切な ADME(T)特性を追求するための注意点

CALTA 理論を駆使してリード骨格を見出したら、次のステップでは ADME(T)特性を考慮し、適切な脂溶性を決めるリード最適化作業が必要となる。もちろん現実には、ADME(T)は化合物の

脂溶性のみで決まるわけではない。以下のような点もリード最適化の際に注意しなくてはならない。

第一に、ADME(T)には表面電荷分布、酸解離定数、分子体積、等さまざまな物性が関わる。さらに、近年非常に重要視されている反応性代謝物に由来する毒性や非可逆的酵素阻害に関しては構造特異的要素の寄与も非常に大きい（余談であるがこれらは今後構造活性相関研究者たちのホットな研究対象となっていくことであろう）。このように、無駄な脂溶性を無くすことはまず大前提であるが、その他にも考慮しなければならないことは多い。

第二に、薬物の体内動態を考えたとき、自分たちの開発しているプロジェクトの望ましい体内動態像が分からない場合が少なくない。こうした場合にどうやって化合物の構造を最適化していけば良いのだろうか。創薬探索段階で eADME の各データに統一基準を設けて振り落としするのは、良い化合物を落としリスクがあまりに高い。

これらの問題点に対応するために、現在我々が取り組んでいるのが次に述べる「臨床 PK/eADME/in silico 統合データベースに基づくデータマイニング戦略」である。

3.2. 臨床 PK/eADME/in silico 統合データベースに基づくデータマイニング戦略

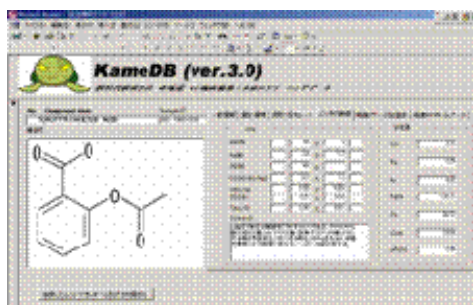


Fig.7. A Clinical PK/eADME/in silico Integrated database (Kame DB)

動態の目標値が具体的に示せない初期段階でも有効なアプローチとして、「データマイニングによる市販薬との比較」が挙げられる。第一製薬での事例では、様々な領域の市販薬原薬を 500 化合物ほど買い集め、自社のeADME評価に全て通し、各種in silicoパラメータを算出し、臨床でのPKパラメータ、剤形、用法用量、上市年、等々の付属情報も併せて全て統合データベースに登録する作業を行っている。（なお、このときデータを一つ一つ専門家が見て意味あるデータにそろえなおすcuration作業が重要かつ非常に大変である）。それが、Fig.7にしめす“Kame DB”である。これらのデータに統計評価を施すことで、in silicoパラメータやeADMEデータと、臨床PKとの相関解析が可能となる。そのほかにも、非常に多くの情報

をこのデータベースから得ることができる。

3.3. 臨床 PK/eADME/in silico 統合データベースに基づくデータマイニングの一例

上記のデータマイニングの一例として、eADME データの変化と臨床での“成功率”とを比較した結果を紹介する。

ここで“成功率”を定義するに当たり、「開発ステージ phase II a において開発継続の判断がなされた薬物には、ヒトにおける吸収性及び対象疾患に対する薬理活性は認められており、基本的な“薬としての性質(Drugability)”は付与されている」と仮定した。つまり Phase II b 以降に進んだ医薬品を、“成功”とみなした。一方、当社探索研究の結果で臨床に進まなかった化合物を“失敗”とみなした。いささか荒い定義であるが、今回目的とするマイニングには十分と考えた。統計手法としては、logistic 単回帰分析を実施し、各パラメータごとに odds 比を求めた。その結果を Table 1 に示す。つまり、たとえば、log D が 0.5 大きくなったとき成功確率は 0.79 倍に、分子量が 50 増えると成功確率は 0.33 倍に、という統計的な推測が下せる。ただ実際にはこれらのパラメータ間相互の相関関係が高いため、単回帰ではなく特別な重回帰分析が必要である。また領域ごとに最適な体内動態があるため、領域ごとに目指す eADME の値も当然変わる。これらの点も踏まえると複雑な統計的研究が必要となるが、その詳細は、また稿を改めて紹介したい。

このように、リード骨格を最適化する過程では、log D やそのほか多数の eADME データの動きに注意しながら、統計的な成功確率を 1 つの目標値にしていく手法が有効と考えられる。化合物の順位付けにも役立つので実用的である。今回、薬効や毒性に関する統計については触れなかったが、今後同様の考えを拡張していく価値はあるだろう。

Table 1. Drugability represented by odds ratios

Parameters M.W.	JP2 solub. (µg/mL)	Log D _{7.4}	CYP3A4 %inhibition	Microsome Stability	Protein Binding%	
Unit	50	50	0.5	10	10	
Odds ratio	0.33	1.02	0.79	0.74	1.28	0.77

4. まとめ

以上、創薬探索において ADME(T)特性と活性とを両立させるための創薬データマイニング手法を2点紹介した。(1)膜近傍タンパク標的プロジェクトにおいては、観測活性 $-\log D$ をプロットした CALTA-map 上で“真の活性”に重要な分子部分を見出し、リード骨格を選定する。(2)自社の系での eADME スクリーニングで多数の市販薬を評価し、臨床 PK/eADME/in silico 統合データベースを作成。統計的手法により自社探索化合物との比較を行い目標となる eADME プロファイルを見出す。

なお、当然だがこうしたデータマイニングのためには、必要十分な精度とスループットを有する評価系の確立と、その適切なデータベース化が前提条件である。さらに、複数の専門分野の協同も不可欠である。今回の我々の研究も、社内の構造活性相関、統計学、物理化学、分析化学、計算化学、そしてもちろん薬物動態学・・・等の専門家の共同チームによる成果である。

最後になりましたが、これら同分野異分野の方々大勢のご指導、特に第一製薬創剤代謝研究所 須藤賢一所長、東京大学大学院理学系研究科橘和夫教授のご指導に、心から感謝申し上げます。

5. References.

- 1) Kennedy T., *Drug Discovery Today* **2**, 436-444 (1997).
- 2) Kola I., Landis J., *Nature Reviews Drug Discovery* **3**, 711-715 (2004).
- 3) 鈴木忠生, 薬剤学 **66**, 106-109 (2006).

//// Cutting Edge ////

ドッキングポーズ評価のためのコンセンサススコア戦略

東北薬科大学 小田彰史、北里大学薬学部 山乙教之、広野修一

1. はじめに

薬物標的となる生体高分子の立体構造を元にドラッグデザインを行う Structure-Based Drug Design (SBDD) において、標的タンパク質と低分子化合物との結合様式を推測することは非常に重要である¹。これまで生体高分子-リガンド複合体構造を計算機的に予測する、すなわち計算機によるドッキングのためのプログラムは多数開発されている。これらのプログラムは一般的に一つの系に対して多数のモデル候補（「ポーズ」と呼ばれることが多い）を予測するが、これら多数のポーズのうちのどれが複合体構造としてもっともらしいかを評価する必要がある。そのためには結合自由エネルギーを計算しなければならないが、煩雑な統計力学的操作を行うことなく結合自由エネルギーを見積もるための手法として、これまでにいくつものスコア関数が提唱されてきた。スコア関数は経験的スコア関数、知識に基づいたスコア関数、力場に基づいたスコア関数の三つのグループに分けられる¹。これらのスコア関数はポーズの選別以外にも次に示すような役割を担っており、計算機によるドッキングにおいて重要な要素となっている。

計算機によるドッキングにおいて、スコア関数は三種類の役割を果たす²。まず1つ目は、ドッキング計算の最中にドッキングポーズのスコアを計算し、妥当なポーズ群を与えることである。すなわち、作成されたポーズの評価だけではなく、ポーズの作成そのものにもスコア関数が必要である。2つ目は、多数作成されたポーズのうち、どれをピックアップすべきかを評価するという用途である。これについては上で既に述べた。3つ目はある一つの標的タンパク質に対して多数の低分子化合物をドッキングさせ、ヴァーチャルスクリーニングを行う場合である。このとき、多数の化合物のうちのどれがその標的タンパク質に最も適合しているか、スコア関数を用いて評価する。この場合、スコア関数は結合親和性を予測する役割を果たす。しかしヴァーチャルスクリーニングを効率的に行うためにはまず結合様式を正しく予測する必要があるという報告があるため¹、スコア関数の評価を行う場合は、まずポーズの選別能力について議論する必要がある。

このようにスコア関数はSBDDにおいて非常に重要であり、これまで数多くのスコア関数が提唱されている。しかし、これまでに提唱されてきたスコア関数にはそれぞれ一長一短あり、決定的なスコア関数というものは存在しない。これは計算する系によって、あるいは計算条件によって用いるスコア関数を選ばなければならない可能性を示唆しており、複合体構造未知の系の計算ではそのような検討が困難な場合がある。そこで近年、ポーズの選別やヴァーチャルスクリーニングのためのスコア関数として、コンセンサススコアの考え方が提唱されている^{3,4}。これは、複数のスコア関数を同時に使用してポーズ選別・ヴァーチャルスクリーニングを行うという方法であり、個々のスコア関数の持つ弱点を補完しあうことで、よりよいスコアリングを可能にしようというアイデアである。ポーズ選別に対しては、スコア関数を単独で用いる場合と比較して、コンセンサススコアは実験的に得られた複合体構造を再現するようなポーズを選別する能力が高いことが報告されている。しかし一方で、いくつかのコンセンサススコアの能力を、実際の利用を見越して比較した研究は少ない。ヴァーチャルスクリーニングに対するコンセンサススコアの能力比較については行われているが⁴、ポーズ選別のためにコンセンサススコアを使った場合に対する網羅的な研究は存在しない。さらにコンセンサススコアの使用法について、通常スコア関数同様スコア値として評価するか、スコア値を順位付けして順位として評価するか、全ポーズ中での百分率で評価するかについて、統一的に比較を行った例はまだない。

そこで本稿では、計算機によるドッキングプログラムの一つFlexXにおけるポーズ選別のために、どのようなコンセンサススコアを使用すれば効率的であるかについて述べる。その際、コンセンサススコアを作り出すための戦略として9種類の戦略を比較した。また、スコア関数としては9種類を用意し、その511通り(2⁹-1通り)の組み合わせすべてを網羅的に検討した。この結果を

用いることで、より効率的に複合体モデルを与えるようにFlexXを改良することが可能であると期待している。また、これはコンセンサスコアをポーズ選別に使用する場合の網羅的な検討として他に類を見ない研究であり、FlexX以外の計算手法を用いる場合においても有用であると思われる。

2. 方法

ここでは、計算手法の概略を述べる。詳細については論文⁵を参照のこと。

2.1 テストセットと使用したスコア関数

コンセンサスコアの能力を評価するためのテストセットとしては、220種類の立体構造既知の生体高分子-リガンド複合体を使用した。複合体の立体構造については RCSB Protein Databank (PDB) から入手した。その PDB データから、ドッキング計算に使用する標的タンパク質、リガンド低分子それぞれの構造を抽出した。

妥当なポーズを選び出すための母集団となるポーズ集団については、SYBYL 6.9 の FlexX (version 1.11.1) によって求めた。出力されるポーズ数の上限を 500 個までとし、それ以外のパラメータは FlexX の default の設定を使用してドッキング計算を行った結果、それぞれの系に対して 6 個から 500 個のポーズが得られた。複合体モデルが妥当であるか否かを決める判定基準には、リガンド分子の実験構造と計算構造の root mean squared distances (RMSD) を使用した。PDB に収録されている立体構造を「正解」として、その正解構造との RMSD (リガンド中の重原子のみ) が 2.0 Å 以下の解を「妥当な解」と定義した。いかなるドッキングプログラム・スコア関数でも厳密に「正解」を求めることは極めて困難であるため、多数のポーズの中からどのくらい効率的に「妥当な解」が得られたかを一つの指標とした。

コンセンサスコアを形成するスコア関数としては、FlexX score, G_score, D_score, PMF score, ChemScore, DrugScore, PLP, ScreenScore, Xscoreの 9 種類を使用した。コンセンサスコアではスコア関数を組み合わせて使用するが、これら 9 種類のスコア関数からなる $2^9 - 1 = 511$ 通りの組み合わせすべてを扱った。すなわち、あるスコア関数を単独で使用した場合から、9 つすべてを使用した場合まで、すべての組み合わせについてテストを行ってベストの組み合わせを求めた。

2.2 ランキングとスケールリング

スコア関数は、それぞれ異なった基準で複合体モデルに対するスコアリングを行っている。そのため各スコア値からそのままコンセンサスコアを作成した場合、不適切な偏りが生じる可能性がある。その問題の解決法として、スコアをランクに変換する操作と、スコアのスケールリングの操作とが考えられる。

まずランキングについては、各ポーズに対して得られたスコア値に従ってポーズを並べ替え、順に 1 位、2 位…とランク付けする。そしてこのランクをあたかもスコアであるかのように使用して、コンセンサスコア操作を行う。たとえば 1 つの系に対して 500 個のポーズが得られたとすると、それぞれのポーズに対して 1 位から 500 位までのランクが付く。これを 9 種類のスコア関数すべてに対して実行すると、スコア関数の評価基準によらず常に 1 から 500 までの範囲内に収まる数値に変換できる。この数値を用いることで、異なった評価基準に基づくスコア関数を組み合わせてコンセンサスコアに用いることができる。文献 4 の “rank-by-rank” はこのランク化を使用したコンセンサスコアの例である。

スケールリングについてはいくつかの方法が考えられるが、本研究では auto-scaling を行った。これは式(1)のように、それぞれのポーズに対して得られたスコア値を、最大値 1、最小値 0 になるようにスケールリングする方法である。

$$x_{\text{scaled}}^{\text{A score}} = \left(x^{\text{A score}} - x_{\text{min}}^{\text{A score}} \right) / \left(x_{\text{max}}^{\text{A score}} - x_{\text{min}}^{\text{A score}} \right) \quad (1)$$

ここで、 $x_{\text{scaled}}^{\text{A score}}$ は A score (たとえば FlexX score, G_score, など) で求めたスコア値をスケールリングした結果得られた値、 $x^{\text{A score}}$ は A score で得られた生の値、 $x_{\text{min}}^{\text{A score}}$ はポーズ全体の中で最も小さい A score の値、 $x_{\text{max}}^{\text{A score}}$ はポーズ全体の中で最も大きい A score の値である。これにより、スコア関

数同士の評価基準の違いを克服して、コンセンサスコア操作を行うことができる。我々は既に auto-scaling されたスコア値の平均を取ることで文献⁴の“rank-by-number”を改良するという手法を提唱しており、これをAASS (Average of Auto-Scaled Scores) と名付けている⁶。

2.3 コンセンサスコア戦略

本稿では、これまで異なったコンセンサスコア戦略であるにもかかわらず同一の名称で呼ばれていた戦略について、それぞれ違いを明確にした名称を定義する。これまでもすでに、いくつかのコンセンサスコア戦略が提唱されてきた。文献⁴には、“rank-by-number”、“rank-by-rank”、“rank-by-vote”の三種類が提案されている。“rank-by-number”、“rank-by-rank”はそれぞれスコア値の平均、ランクの平均をそのままコンセンサスコアとして取る方法である。これら平均値によるコンセンサスコアをポーズの選別に使用する場合、「コンセンサスコア値が閾値 x 以下のポーズを採用する」という基準と「コンセンサスコア値の上位 y 位以上を採用する」という基準、さらに「コンセンサスコア値の上位 $z\%$ 以内を採用する」という三つの選別基準がある。これらのうちのどの基準を用いるかによってポーズを選び取る能力に差が出るのが予想されるが、これまでその比較検討はされていなかった。そこで今回、“rank-by-number”、“rank-by-rank”で得られたコンセンサスコアを閾値 x で判断する場合、上位 y 位まで取った場合、上位 $z\%$ まで取った場合のそれぞれのポーズ選別能力を比較した。これらはそれぞれ閾値で判断した場合を“number-by-”で、順位で判断した場合を“rank-by-”で、百分率で評価した場合を“percent-by-”で表現する。しかし“number-by-rank-by-number”といった表現は冗長すぎるため、“number-by-rank-by-number”の代わりに“number-by-number”と表記する。同様に、“number-by-rank”、“rank-by-number”、“rank-by-rank”、“percent-by-number”、“percent-by-rank”の6種類の表記を採用する。表1にこれらについてまとめてある。なお、“rank-by-number”と“rank-by-rank”については、文献⁴とは意味が違うので注意すること。また我々が文献⁶で提唱したAASSは“number-by-number”、“rank-by-number”、“percent-by-number”に該当する。本稿ではコンセンサスコアの比較に付随して、閾値 x, y, z の最適化も行った。

一方、“rank-by-vote”は多数決によるコンセンサスコア戦略である。この戦略では、スコア値がある投票基準を満たしていれば1票を与え、多くの票を集めたポーズが採用される。文献^{3,4}では「上位 $z\%$ 」を投票基準にしているが、たとえばSYBYLに実装されているCScoreなどでは「スコア値が閾値 x 以下（デフォルトでは $x=0.5$ ）」が投票基準である。また、「上位 y 位以内であれば1票」という戦略もありうる。今までのところ、“rank-by-vote”戦略における投票基準として百分率を用いた場合と閾値を用いた場合、順位を用いた場合についての比較もまた行われていない。そこで本稿では、投票基準として百分率を用いた場合を“vote-by-percent”、閾値を用いた場合を“vote-by-number”、順位を用いた場合を“vote-by-rank”として、この三者の比較を行った。つまり文献⁴の“rank-by-vote”は我々の用語では“vote-by-percent”に、CScoreは“vote-by-number”になる。なお“vote-by-”戦略に使用するパラメータとしては、投票のための閾値と得票数の閾値の二種類ある。たとえば“vote-by-percent”では、あるスコア関数で得られたスコア値が上位 $z\%$ 以内であれば1票を獲得し（ z : 投票のための閾値）、合計得票数が w 票以上であればそのポーズを「当選」とする（ w : 得票数の閾値）。本稿では、“vote-by-”戦略の両方の閾値についての最適化を行った。詳細については参考文献^{3,6}を参照のこと。

2.4 コンセンサスコアの能力の評価

これらのコンセンサスコアの評価法として、まず必ず妥当な解を得られる条件を求めた。こ

表1 コンセンサスコアの種類（文献5より許可を得て引用。Copyright 2006 American Chemical Society）

		コンセンサスコアの作成法	
		rank-by-number	rank-by-rank
ポーズ選択 の際の閾値	number-by-	number-by-number	number-by-rank
	rank-by-	rank-by-number	rank-by-rank
	percent-by-	percent-by-number	percent-by-rank

れは、たとえば“number-by-”戦略であれば「220個の系すべてにおいて、コンセンサスコア値が x 以下のポーズの中に、最低一つは妥当な解が含まれている」というような閾値 x を求めることである。同様に“rank-by-”戦略の y と“percent-by-”戦略の z を求めた。また同様に“vote-by-”戦略では「いかなる系でも、 w 票以上の得票数のあったポーズの中に最低一つは妥当な解が含まれている」というような w を求める。これらのそれぞれの条件下で、ポーズの数を全体の何%まで絞ることができたかをコンセンサスコアの能力として評価した。これを「圧縮率」($p_{compress}$)と呼ぶ。

またこれとは逆に、各閾値 x, y, z, w を様々に変化させたときに、220個のテスト系のうちいくつかの系で妥当な解が得られているかについても評価した。この評価は、すべての系で妥当な解を得るのではなく、できるだけ多くの系で妥当な解を得られるような条件を見出すために行う。220個すべての系で妥当な解を得ようとする圧縮率が非常に悪くなるため、この評価基準を用いて、いくつかの系を犠牲にすることで圧縮率を高めることができないかを検討した。この場合、圧縮率だけではなく220個の系のうちのいくつかで妥当な解を得たかという割合である「正答率」($p_{accurate}$)についても評価した。これらの評価基準を用いて、効率的にポーズを選別できるコンセンサスコア戦略、およびスコア関数の組み合わせを求めた。

3. 結果と考察

220個の系すべてで最低一つは妥当な解が得られるような閾値を求め、その閾値を用いたときの圧縮率を算出した。すなわちこれは、正答率が100%であるような条件下での圧縮率である。それぞれの戦略で511通りの組み合わせをすべて試しているが、本節では最良の圧縮率を与えたスコア関数の組み合わせ(“best combination”)について議論する。この“best combination”での結果を表2に記した。この表には閾値と圧縮率の両方を記してある。たとえばnumber-by-numberでは、best combinationはFlexX score, G_score, PMF score, D_score, PLP, ScreenScoreのセットであった。これらの6つのスコア値をauto-scalingした後に平均を取り、それが0.452以下だったポーズを採用すればいずれの系でも必ず妥当な解が得られることを意味している。この基準を採用することで、検討すべきポーズの数を約40%に削減することができる(圧縮率=0.399)。この表から、「必ず妥当な解を得る」という目的のためには、vote-by-number戦略が最も適していることがわかる。vote-by-number戦略でのbest combinationはFlexX score, G_score, PMF score, D_score, PLP, ScreenScore, Xscoreの組み合わせで、auto-scaling後のこれらの値が0.5以下であれば一票を投じ($x=0.5$)、全7つのスコア関数のうち5票が集まればそのポーズを採用する($w=5$)、という方法が最適であった。vote-by-戦略については、ヴァーチャルスクリーニングに対してこれまで文献で報告されてきたうちのいくつかはvote-by-percentを使用していたが³⁴、少なくともポーズ選別において「必ず妥当な解を得る」という目的にはCScoreのようにvote-by-numberを使用したほうがはるかに良好な結果を与えることもわかった。second bestはnumber-by-numberで、これは上述の通り6つのスコア関数の組がbest combinationであったが、number-by-rank, rank-by-number, rank-by-rankではいかなるコンセンサスコアよりもDrugScore単独で評価したほうが優れた結果となった。表1においてこれらのスコア関数の結果が同一なのは、単一スコアが最良の結果を示してしまったためである(スコア関数を単独で使用した場合は、すべてのrank-by-戦略、そして

表2 すべての系で妥当な解を与えることを求めた場合のコンセンサスコアの能力比較(文献5より許可を得て引用。Copyright 2006 American Chemical Society.)

(a) Number-by-, rank-by-, percent-by- 戦略					
	閾値	圧縮率		閾値	圧縮率
Number-by-number	$x = 0.452$	0.399	Number-by-rank	$x = 151$	0.519
Rank-by-number	$y = 151$	0.519	Rank-by-rank	$y = 151$	0.519
Percent-by-number	$z = 62.0\%$	0.617	Percent-by-rank	$z = 62.4\%$	0.623

(b) Vote-by- 戦略			
	投票のための閾値	得票数の閾値(w) / スコア関数の数	圧縮率
Vote-by-number	$x = 0.5$	5 votes / 7 voters	0.371
Vote-by-rank	$y = 150$	2 votes / 3 voters	0.510
Vote-by-percent	$z = 70\%$	3 votes / 3 voters	0.502

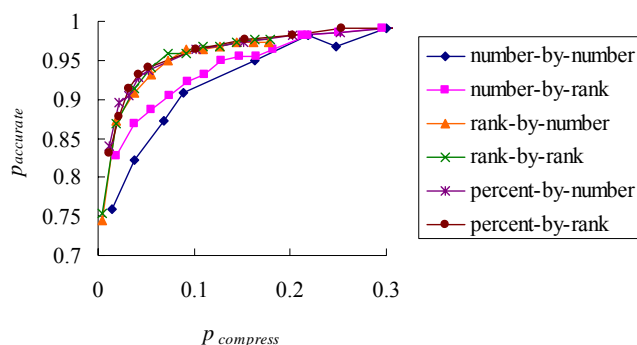


図1 閾値を変動させて圧縮率と正答率のバランスを見た場合のコンセンサスコアの能力比較 (文献5より許可を得て引用。Copyright 2006 American Chemical Society)

number-by-rank 戦略の結果が同じになる)。同様に、percent-by-rank もいかなるコンセンサスコアよりも PLP 単独で使用した結果のほうが優れていた。これらの結果は、これらの戦略が「必ず妥当な解を得る」という目的には適していないことを示唆している。また、percent-by-戦略はどちらも非常に悪い圧縮率しか与えることができなかった。これは得られたポーズ数が多い系も少ない系も同じ割合でポーズを pick up するため、すべての系で必ず妥当な解を与えるためには閾値を大きくしなければならなくなるためだと考えられる。たとえばポーズが5個しかない系で3個 pick up する必要がある場合、上位60%までを取らなければならなくなるが、同じ基準を使うとポーズが500個ある系では300個 pick up することになる。これは percent-by-戦略が本質的に持つ問題点であり、この戦略が「必ず妥当な解を得る」という目的には適していないことを示唆している。

上記の結果は、「すべての系に対して妥当な解を与える」、つまり「正答率100%」という条件下でのコンセンサスコアの能力についての議論であった。しかしこのような条件では、ほんの少数の例外的に評価が難しい系が存在するだけで不必要に閾値を高く設定する必要があるため、高い圧縮率は望めなくなってしまう。そこで引き続き、閾値をコントロールして、圧縮率を低く抑えつつできるだけ多くの系に対して妥当な解を与えるような条件を検討する。vote-by-戦略は閾値が2つあり、スコア関数の組み合わせも考慮したうえでこのような調整を行うのは困難であることから、ここでは number-by-戦略、rank-by-戦略、percent-by-戦略について検討する。

number-by-戦略、rank-by-戦略、percent-by-戦略のそれぞれに対して、様々な閾値を用いたときの正答率と圧縮率の関係について図1に示した。この図からわかるように、同等の正答率を与えるような条件同士で比較した場合、2つの number-by-戦略は rank-by-戦略、percent-by-戦略と比較してかなり圧縮率が悪い。この結果は「正答率100%」を必要とするか、「ある程度高い正答率と圧縮率を両立させる」を必要とするかによって異なった戦略を使用しなければならないという興味深い知見を与えている。従って、目的に応じてコンセンサスコア戦略を使い分けるか、あるいは両方の長所を生かすように組み合わせることが重要であろう。

以上、複数のコンセンサスコアについて系統的な能力比較を行った結果について述べてきた。これらのテストはコンセンサスコアのポーズ選別能力の最初の網羅的な評価であり、これによってコンセンサスコアを SBDD で使用する際の指針が得られたのではないかと考えている。

4. 参考文献

- 1 D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug. Discovery*, **3**, 935-949 (2004).
- 2 小田彰史, 広野修一, *機能材料*, **25**, 21-28 (2005).
- 3 P. S. Charifson, J. J. Corkery, M. A. Murcko, W. P. Walters, *J. Med. Chem.*, **42**, 5100-5109 (1999).
- 4 R. Wang, S. Wang, *J. Chem. Inf. Comput. Sci.*, **41**, 1422-1426 (2001).
- 5 A. Oda, K. Tsuchida, T. Takakura, N. Yamaotsu, S. Hirono, *J. Chem. Inf. Model.*, **46**, 380-391 (2006).
- 6 M. Katsuki, V. T. G. Chuang, K. Nishi, K. Kawahara, H. Nakayama, N. Yamaotsu, S. Hirono, M. Otagiri, *J. Biol. Chem.*, **280**, 1384-1391 (2005).

//// Activities ////

構造活性フォーラム 2006 「創薬研究のためのデータ科学：基礎と応用」 開催報告

構造活性フォーラム 2006
実行委員長 高橋 由雅

標記フォーラムは、構造活性関連部会の平成 18 年度行事の一環として、平成 18 年 6 月 30 日(金)、豊橋商工会議所（豊橋市花田町字石塚 42-1）9F ホールを会場に開催された。

今回のフォーラムでは構造活性関連および創薬研究に関連するデータ科学の先端技法に関する話題を取り上げ、特に計算機実験におけるバリデーション手法や機械学習とデータマイニングの基礎、さらには創薬研究における能動学習応用事例の解説を中心に、「創薬研究のためのデータ科学：基礎と応用」と題して企画した。普段は業務に追われ、大学卒業後は新しい方法論に関しての基礎的、理論的な話を聴く機会の少ない企業研究者を念頭に、1 日だけ学生に戻ったつもりで講義を受講する機会を提供できればとの趣旨で企画させていただいたものである。このことから、通常の講演会より長めの 1 コマ 75 分（質疑応答を含む）の講義形式とし、午前 2 件、午後 2 件の講演（講義）を組み、特に方法論については関連分野の研究に携わる人のみならず、これからこの分野の勉強を始めようと考えている人にも十分理解できるよう事前をお願いをした。講師の先生方には基礎的事項を中心に平易な解説に努めていただき、参加者からも好評を得ることができた。当日のプログラムは下記のとおりである。

- (1) 「計算機集約型手法による回帰・判別型分析法のバリデーション手法」
大阪大学大学院薬学系研究科 高木達也
- (2) 「データマイニングの基礎と創薬への適用事例」
関西学院大学理工学部情報科学科 岡田 孝
- (3) 「データ分類学習：サポートベクターマシン (SVM) の基礎と応用」
豊橋技術科学大学知識情報工学系 高橋由雅
- (4) 「GPCR リガンドスクリーニングへの能動学習の応用」
田辺製薬(株) 清水 良

本フォーラムの日程は 1 日のみの地方での開催であったこと、また、内容的にもこれまでの話題とは大きく異なったものであることから、正直なところ当初はどのくらいの参加申込が得られるかの心配もあった。最終的には募集人員 60 名に対し、一般の参加登録者は 55 名、講師を含め 59 名の参加を得ることができ、盛会裏に無事終了することができた。上記プログラム終了後は講師を囲んでミキサーを開催した。プログラム参加者の大半が出席し、引き続きにぎやかな交歓が行われたことを見届け、関連分野に携わる多数の参加者に有益な機会が提供できたものと胸を撫で下ろした次第である。

末筆ながら、本フォーラムの開催にあたりご支援ご協力を頂きました皆様方に心から御礼申し上げます。

//// Activities ////

SAR Promotion Award 平成 18 年度受賞者

庶務幹事 京都大学大学院 赤松美紀

構造活性相関部会では、常任幹事会での慎重な審議を経て、平成 17 年度より、構造活性相関研究の発展を促進するための事業として当該制度を設けた。本年度は、この奨励制度をホームページに掲載し、広い分野からの応募を奨励する制度として確立し、構造活性相関のますますの発展を計ることとした。

平成 18 年度は、6 月 30 日の常任幹事会において、受賞者を以下の 2 名に決定した。

- | | |
|-------|---|
| 氏名 | 合田浩明 (ごうだ ひろあき) |
| 所属 | 北里大学 薬学部 創薬物理化学教室 (助教授) |
| 参加学会名 | 2006 秋期アメリカ化学会国際会議 |
| 開催期日 | 2006 年 9 月 10 日 - 14 日 |
| 開催場所 | アメリカ合衆国, サンフランシスコ |
| 演題 | 計算化学手法を用いた新規な非ペプチド性キチナーゼ阻害剤の論理的分子設計
(ポスター発表) |
| 氏名 | 原田俊幸 (はらだ としゆき) |
| 所属 | 京都大学大学院農学研究科・応用生命科学専攻・生物調節化学分野 (修士 2 回生) |
| 参加学会名 | 第 16 回 エクダイソンワークショップ |
| 開催期日 | 2006 年 7 月 10 日 - 14 日 |
| 開催場所 | ベルギー, ゲンツ大学 |
| 演題 | ショウジョウバエ脱皮ホルモン受容体に対するステロイド型アゴニストの結合
(ポスター発表) |

2 名の受賞者の報告は、次号の SAR News に掲載される予定である。

//// Activities ////

情報化学入門講座（第4回）報告

阪大院薬・高木達也

2006年2月7日（火）、日本化学会館7階ホールにおいて情報化学入門講座（第4回）が、日本化学会情報化学部会主催、日本薬学会構造活性相関部会協賛にて開催されました。昨年の同時期に2日間の日程で行われた同講座（第3回）よりも日程が縮小したものの、29名（部会員・共催学会員10名、協賛学会員5名、非部会員10名、学生4名）の皆様にご参加いただきました。当日のプログラムは以下のようです。

1. 統計的検定の考え方（大阪大学 高木達也）
2. ケモインフォマティクス、バイオインフォマティクス、二度おいしいデータマイニング（奈良先端科学技術大学院大学 金谷重彦 教授）
3. 化合物の類似度とクラスタリング（筑波大学 中山伸一 教授）
4. ケモメトリックスソフトウェア Chemish を用いた演習（東京大学 船津公人 教授・荒川正幹 助手）

当日は、日本化学会情報化学部会長（当時）の細矢治夫・お茶の水女子大学名誉教授のご挨拶から始まり、最初の2題の座長を中山伸一教授が、後半2題の座長を産業技術総合研究所の内丸忠文博士が担当され、スムーズに講習会が進みました。

これらの講義および演習を通じて「情報化学の全体像を把握できた」というコメントや「今後の仕事の方向性が間違っていなかったことがわかり安心した」等の感想を参加者から頂くことができ、盛会のうちに終了することができました。なお、内容などの詳細は未定ですが、同入門講座は来年度も引き続き開催される予定と伺っております。



講演の様子



演習の風景

///// Activities /////

<会告>

第34回構造活性相関シンポジウム

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

会期 平成18年11月14日(火)～15日(水) [第29回情報化学討論会と併催]

会場 朱鷺メッセ(新潟市万代島 6-1) <http://www.tokimesse.com/>

参加予約申込締切 10月27日(金)

講演時間 特別講演 60分、招待講演 40分、奨励講演 30分、一般講演 25分(講演番号に*印)
または15分。詳細は下記ホームページをご覧ください。

第1日(11月14日)

座長 福島 千晶(10:20 - 10:45)

K01* Lead Generation of Novel Androgen Pure Antagonists Using Chemogenomics-like Approach (中外製薬) ○大田 雅照、橘 一生、吉野 仁、白石 拓也、今岡 郁博、加藤 伸明、中村 光昭、佐藤 晴彦、川田 洋充、永牟田 雅弘、小沼 悦郎

座長 助川 正之(10:45 - 11:15)

K02 脱皮ホルモン受容体に対するステロイド型アゴニストの結合予測(京大院・農) ○原田 俊幸、新井 宏和、渡辺 文太、赤松 美紀、中川 好秋、宮川 恒**K03** GPR54 アゴニストの構造活性相関研究(京大院・薬) ○富田 健嗣、大石 真也、大野 浩章、(京大院・農) 赤松 美紀、(京大院・薬) 藤井 信孝

座長 田中 明人(11:25 - 11:55)

K04 [奨励講演] ケミカルゲノミクス情報に基づく化合物探索(京大院・薬) 奥野 恭史

座長 船津 公人(13:15 - 14:15)

JS [特別講演] ナノ材料の知識の構造化(東大院・工) 山口 由岐夫

ポスターセッション(14:15 ~ 16:45)

KP01 ミニレクチンの設計、合成、糖鎖認識(野口研) ○戸潤 一孔、川上 宏子**KP02** 抗血管新生作用を有する chiral haloacetylcarbonyl-2-nitroimidazole の構造解析(名大院・生命農学) ○大倉 一人、(徳島大院・ソシオテクノサイエンス) 宇都 義浩、(岐阜薬大) 永澤 秀子、(徳島大院・ソシオテクノサイエンス) 堀 均**KP03** MD シミュレーションによる口蹄疫ウイルス GH ループの解析(北里大院・基礎生命) ○東 寛子、(北里大・理) 米田 茂隆**KP04** ADA 阻害剤の methyl 基導入に伴うダイナミクスと結合エネルギーの変化(京大院・薬) ○小杉 貴洋、仲西 功、北浦 和夫**KP05** ヒト免疫不全ウイルス I 型インテグラーゼ阻害剤の結合様式の決定と構造最適化(北里大・薬) ○藤本 拓、山乙 教之、合田 浩明、広野 修一**KP06** 取り消し**KP07** Inverse Agonist Activity of several 5-HT_{2A} Receptor Antagonists at the Constitutively Active Human 5-HT_{2A} receptor (新潟薬大・薬) ○Habib Abul Muntasir, Mohiuddin Ahmed Bhuiyan, Murad Hossain, Jun Kawakami、(新潟薬大・応用生命) Masaji Ishiguro、(新潟薬大・薬) Tadazumi Komiyama, Takafumi Nagatomo**KP08** ノシセプチン受容体のコンピュータモデリングによる立体構造予測と変異体活性測定(群馬大・工) ○阿久澤 夏世、武田 茂樹、(サントリー生有研) 石黒 正路**KP09** ヒト Organic Cation Transporter 1 (hOCT1) に対するリガンド化合物の三次元ファーマコフォアの推定と 3D-QSAR (北里大・薬) ○仲井 崇、曾根 大介、中込 泉、山乙 教之、合田 浩明、広野 修一**KP10** Protein-based alignment を用いた PPAR δ リガンドの三次元定量的構造活性相関解析(北里大・薬) ○中込 泉、山乙 教之、広野 修一、宮地 弘幸、春日 淳一**KP11** コンピュータによる非ペプチド性 β -セクレターゼ阻害剤の分子設計(北里大・薬) ○松下 泰雄、中込 泉、山乙 教之、広野 修一

- KP12** リガンド結合部位に存在する水分子との競合を考慮した COMBINE 解析 (京大院・薬) ○横田 彰宏、中村 真也、仲西 功、北浦 和夫
- KP13** 等級活性予測を指向した Comparative Molecular Field Analysis (阪大院・薬、田辺製薬) ○大軽 貴典、(田辺製薬) 清水 良、(阪大院・薬) 岡本 晃典、(大阪大谷大・薬) 川瀬 雅也、(阪大院・薬) 白國 優子、(大阪大谷大・薬) 錦織 理華、(阪大院・薬、阪大微生物病研) 高木 達也
- KP14** アゾール系化合物の CYP2B/3A との結合および酵素活性阻害様式の解析 (徳島大院・薬) ○糸川 大祐、西岡 大貴、福島 淳治、安田 知世、山内 あい子、中馬 寛
- KP15** 分子軌道法による亜鉛含有タンパク質と阻害剤との相互作用解析 (徳島大院・薬) ○吉田 達貞、福島 淳治、中馬 寛
- KP16** 分子軌道法計算によるニトロ多環芳香族ケトンの構造-変異原性相関 (東邦大・理) ○小野里 磨優、大島 茂
- KP17** VDR と各種リガンドとの相互作用に関する理論的研究 (立教大・理) 常盤 広明、○大塚 智世、(徳島大・薬) 山岸 賢司、(東京医歯大・生体材料工学研) 山本 恵子、山田 幸子
- KP18** Ab initio FMO 法の SBDD への応用 (東大・生産技術研) ○甘利 真司、(立教大・理) 望月 祐志、(国立衛研) 中野 達也、(徳島大院・薬) 中馬 寛
- KP19** 構造活性相関モデル作成のための 28 日間反復投与毒性試験データの解析 (製品評価技術基盤機構) 佐藤 佐和子、○櫻谷 祐企、山田 隼、前川 昭彦
- KP20** 1-β-O-acyl glucuronides の親電子反応性に対する構造活性相関 (北海道薬大) ○馬場 暁子、吉岡 忠夫
- KP21** kinase-focused library の構築および方法論の検討 (第一製薬) ○島田 多堅
- KP22** 構造記述子の情報量に基づく類似活性化合物の探索手法 (阪大院・情報科学) ○木村 浩章、瀬尾 茂人、竹中 要一、松田 秀雄
- KP23** 化合物とチトクローム P450 との相互作用に関する構造活性相関解析 (京大院・薬) ○馮 春来、山下 富義、(お茶大・理) 伊藤 貴之、(京大院・薬) 橋田 充
- KP24** 化学反応コンパイラ「React」によるバイオパスウェイの動態解析 (徳島大院・薬) ○浅野 良治、木下 崇司、岡崎 紀明、中馬 寛
- KP25** 分子構造類似性に基く催奇形性予測ソフトウェア(SimScore) (セイラシステム) ○谷 真琴、A.Ammar Ghaibeh、佐々木 幹夫、(徳島大院・薬) 坂本 久美子、中馬 寛、山内 あい子
- KP26** 機能性食品設計のためのフラボノイドデータベース構築 (徳島大院・薬) ○木下 崇司、Zsolt Lepp、(徳島大院・栄養) 河合 慶親、寺尾 純二、(徳島大院・薬) 中馬 寛
- KP27** Support Vector Machine によるチトクローム P450 基質および阻害剤選択性の予測および解析 (徳島大院・薬) ○西岡 大貴、近藤 崇泰、糸川 大祐、木下 崇司、Zsolt Lepp、中馬 寛
- KP28** A method to identify potential lead-hopping among various therapeutic areas using SVM (徳島大院・薬) ○Zsolt Lepp、木下 崇司、中馬 寛
- KP29** A Novel 3D-QSAR Method Using Voronoi Grid and Support Vector Regression (VSVR- QSAR) (セイラシステム) ○A. Ammar Ghaibeh, Makoto Tani, Mikio Sasaki, (徳島大院・薬) Hiroshi Chuman
- KP30** SOM による高次元構造特徴空間の可視化と合理的な初期マップ生成 (豊橋技科大) ○渡邊 旬、藤島 悟志、高橋 由雅
- KP31** NTG の階層関係を利用した薬物構造データマイニング (豊橋技科大) ○栗林 滝、藤島 悟志、高橋 由雅
- KP32** 生体高分子ドッキング解析システム MIZUHO/ BioStation の開発 (みずほ情報総研) ○佐藤 智之、加藤 昭史、福澤 薫、(国立衛研) 中野 達也、(神戸大院) 田中 成典
- KP33** ドッキングソフトウェアの能力評価およびパラメータ最適化 (東北薬大) ○小田 彰史、岡安 愛、神山 由紀子、吉田 朋起、松崎 久夫
- KP34** 「並列創薬 (Parallel drug design)」と「コンビナトリアルケミストリー/合成」の融合による「インシリコンビ (In Sili-Combi)」の提案 (富士通) ○湯田 浩太郎、(富士通九州) Jose Martin Ciloy、北島 正人
- 座長 石黒 正路 (16:45 - 17:45)
- KS [特別講演]** 膜蛋白質結晶構造解析とその創薬への応用 (インペリアルカレッジロンドン) 岩田 想

第 2 日 (11 月 15 日)

座長 久保寺 英夫 (9:00 - 9:25)

K05* 分子シャペロンHSP47が認識するコラーゲンの構造 (新潟薬大・薬) ○西川 良美、浅田 真一、北川 幸己、小出 隆規

座長 高橋 由雅 (9:25 - 10:05)

K06* K-step Yard samples法の開発とADME-T予測への適用 (富士通) ○湯田 浩太郎

K07 能動学習法による創薬スクリーニング -PDE-5 のリガンド探索- (日本電気) ○襲田 勉、宮川 知也、山下 慶子、藤原 由希子、麻生川 稔

座長 丹羽 朋子 (10:05 - 10:45)

K08 構造類似度算出機能 (SimScore) を備えた薬物催奇形性情報システム (徳島大院・薬) ○坂本 久美子、(セイラシステム) 谷 真琴、A. Ammar Ghaibeh、佐々木 幹夫、(徳島大院・薬) 中馬 寛、山内 あい子

K09* 特徴的リード構造抽出のための部分構造精錬システムの開発 (豊橋技科大) ○藤島 悟志、(関西学院大・理工) 岡田 孝、(豊橋技科大) 高橋 由雅

座長 粕谷 敦 (11:00 - 11:30)

K10 ポイントフッ素化 γ -ラクトンにおける分子配座の支配要因 (鳥取大・工) ○早瀬 修一、川面 基、伊藤 敏幸

K11 ヒト血清アルブミン上の pH 依存性リガンド結合サイトに対するケトプロフェンの結合 (北里大・薬) ○山乙 教之、(熊本大院・薬) 金子 健一、小田切 優樹、(北里大・薬) 広野 修一

座長 仲西 功 (11:30 - 12:00)

K12 [奨励講演] 酵素反応の理論計算: QM/MM モデリング、全系量子計算による解析 (産総研・RICS) 石田 豊和

座長 広野 修一 (13:55 - 14:35)

K13 [招待講演] 有機水和物結晶の分子動力学シミュレーション (北里大・理) 米田 茂隆

座長 辻下 英樹 (14:35 - 15:05)

K14 フラグメント分子軌道法を用いた HIV-1 protease と阻害剤との相互作用解析 (徳島大院・薬、CREST-JST) ○山岸 賢司、(徳島大院・薬) 吉田 達貞、(立教大・理) 川和田 美里、常盤 広明、(徳島大院・薬) 中馬 寛

K15 ADA 阻害剤の methyl 基導入に伴うダイナミクスと結合エネルギーの変化 (京大院・薬) ○小杉 貴洋、仲西 功、北浦 和夫

座長 梅山 秀明 (15:15 - 16:05)

K16* VDR と各種リガンドとの相互作用に関する理論的研究 (立教大・理) ○常盤 広明、大塚 智世、(徳島大・薬) 山岸 賢司、(東京医歯大・生体材料科学研) 山本 恵子、山田 幸子

K17* Ab initio FMO法のSBDDへの応用 (東大・生産技術研) ○甘利 真司、(立教大・理) 望月 祐志、(国立衛研) 中野 達也、(徳島大院・薬) 中馬 寛

座長 藤原 巖 (16:05 - 16:20)

K18 2006年 CASP7 コンテストにおける自動サーバーによるタンパク質モデリング: 日本、米国、欧州の実力 (北里大・薬) ○岩館 満雄、寺師 玄記、加納 和彦、高谷 大輔、細井 亜紀男、大田 数広、竹田-志鷹 真由子、梅山 秀明

参加費 [一般] 予約 8000 円、当日 9000 円 [学生] 予約 3000 円、当日 4000 円

併催の第 29 回情報化学討論会に参加できます (要旨集含む)

懇親会 11 月 14 日 (火) 18:30、ホテル日航新潟 (情報化学討論会と合同) [一般] 予約 6000 円、当日 8000 円 [学生] 予約 3000 円、当日 4000 円

連絡先 〒956-8603 新潟市東島 265-1 新潟薬科大学応用生命科学部 石黒正路 米田照代

TEL 0250-25-5152 FAX 0250-25-5021 E-mail qsar@niigatayakudai.jp

ホームページ <http://www.geocities.jp/sar2006n/>

構造活性相関部会の沿革と趣旨

1970年代の前半、医農薬を含む生理活性物質の活性発現の分子機構、立体構造・電子構造の計算や活性データ処理に対するコンピュータの活用など、関連分野のめざましい発展にともなって、構造活性相関と分子設計に対する新しい方法論が世界的に台頭してきた。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、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のが本部会の前身の構造活性相関懇話会である。1975年5月京都において第1回の「懇話会」(シンポジウム)が旗揚げされ、1980年からは年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定期的に開催されるようになった。

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。構造活性相関懇話会は1994年にその名称を同研究会に改め、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すこととなった。2002年4月からは、日本薬学会の傘下組織の構造活性相関部会として再出発し、関連諸学会と密接な連携を保ちつつ、生理活性物質の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している。現在それぞれ年一回のシンポジウムとフォーラムを開催するとともに、部会誌のSAR Newsを年二回発行し、関係領域の最新の情勢に関する啓蒙と広報活動を行っている。

本部会の沿革と趣旨および最新の動向などの詳細に関してはホームページを参照頂きたい。
(http://bukai.pharm.or.jp/bukai_kozo/index.html)

編集後記

日本薬学会構造活性相関部会の部会誌SAR Newsの第11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、編集委員長が清水良から藤原巖に交代しました。また、山上知佐子先生が編集委員を退任され、新たに久保寺英夫が加わりました。

Perspective/Retrospectiveでは、平山令明先生(東海大学)にX線結晶解析の原理限界と得られる座標データの信頼性についてご執筆頂きました。分子設計に携わる研究者にとって昨今のSBDDや分子シミュレーションでX線解析結果を安易に用いることに対する警鐘であり、その対策の道標となることでしょう。Cutting Edgeでは、小林好真先生(第一製薬)にリード探索段階で薬理活性とADMEプロファイルを両立させるためにデータマイニングと可視化を利用する研究、小田彰史先生(東北薬科大学)には、ドッキングスタディにおいて複数のスコア関数を用いたコンセンサスの取り方が結果にどう影響するかを検討した研究を、それぞれ紹介していただきました。立場は異なりますが、いずれも最先端の研究で大いに読者の参考になったことと存じます。

このSAR Newsが単なる部会の連絡用機関紙に止まらず、今後とも構造活性相関研究のトレンドを把握する機会を会員にご提供できることを編集委員一同願っております。皆様の忌憚のないご意見をどうかお寄せください。

(編集委員会)

SAR News No.11 平成18年10月1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) 藤原 巖
清水 良
石黒正路
黒木保久
高橋由雅
福島千晶
久保寺英夫

*本誌の全ての記事、図表等の無断複写・転載を禁じます。