

# 仮想スクリーニング統合システム ～薬師“Xsi”ver.1.0～ の開発と応用

住友製薬(株) 研究本部  
化学研究所 山崎 一人  
ゲノム科学研究所 金岡 昌治

## Development and Applications of Virtual Screening System –Xsi ver. 1.0–

Sumitomo Pharmaceuticals Co., Ltd.  
Research Division  
Chemistry Research Laboratories  
Kazuto YAMAZAKI  
Genomic Science Laboratories  
Masaharu KANAOKA

Virtual screening, a computational method to identify bioactive compounds among a vast number of chemicals, is in use with a great expectation of saving time and cost necessary for actual screening of bioactive compounds. Virtual screening techniques can be categorized in two, namely Ligand-based Drug Design (LBDD) and Structure-based Drug Design (SBDD). Conceptually LBDD and SBDD can be complementary with each other, but have been developed and applied independently. In order to improve the current methods, we developed 'Multiple Docking' method which utilized LBDD and SBDD complementarily. In this review we describe this new method, along with a computer software system 'Xsi' developed to realize the method.

### はじめに

創薬研究では、薬理作用を有する化合物を見出すため、膨大な数の化合物が薬効スクリーニングに供される。コンビナトリアル合成技術やハイスループットスクリーニング技術等の進展により、化合物の供給や薬理評価試験の作業効率は飛躍的に向上したが、それでもなお目的の薬物分子を見出すには多大な時間とコストを要する。そこで薬効スクリーニングを計算機上で実行する仮想スクリーニングに、大きな期待が寄せられている。

仮想スクリーニングが使われ始めた当初は、実験的に得られる構造活性相関データに基づく手法が主流であった。この手法は、同じ薬理作用を有する化合物には共通する構造および物理化学的特性があるとの経験則に基づいている。古くはQSAR解析や重ね合わせ解析として知られ、近年ではケモインフォマティクスとして発展してきた。これらは薬理活性化合物すなわちリガンドの情報に基づく手法であることから、LBDD (Ligand-based Drug Design) と呼

ばれる。LBDDは、ここ10年間に飛躍的に技術が向上し、創薬研究に不可欠なツールとして定着している。この手法は構造活性相関情報に基づくため、薬理活性評価実験と密接に連携して用いられ、実験データの蓄積に伴って予測能は向上する。LBDDの活用は、目的の薬物分子を見出すまでの労力と時間を軽減し、結果的に成功確率の向上に寄与する。しかしながら、先験情報が不足している場合、すなわち構造活性相関が未知の領域を予測できないという根本的な問題を抱えている。

一方、近年の本格的なポストゲノム研究の進展と共に、医薬化合物の標的である蛋白質の立体構造が数多く解明されつつある。その結果、蛋白質の立体構造に基づく仮想スクリーニングが強く指向されるようになった。この手法は、相互作用する蛋白質と化合物は相補的な関係 (= 「鍵と鍵穴」の関係) にあるとの知見に基づいている。具体的には、ドッキングスタディーや分子シミュレーション等があり、総称してSBDD (Structure-based Drug Design) と呼ばれる。SBDDは、ここ数年間の精力的な研究によ

り、ようやく実用段階に到達してきている。本手法は、LBDD と異なり先験情報を必要としないことから、原理的には未知なる部分の予測が可能になるものと大きな期待が寄せられている。ただし、あたかも実験を代替できるかのような議論がなされがちであるが、実際はそのようなレベルにあるとは到底言えない。また、先験情報を前提としないということは、逆に構造活性相関や X 線構造解析等の実験データが蓄積しても、LBDD のように予測性能を向上できないことを意味する。従って、実際の創薬研究において、薬理実験等と連携して、LBDD と SBDD を如何に適切に利用していくかが課題となっている。

エストロゲン受容体に結合する 3 つのリガンドについて、X 線構造解析から得られた座標を重ね合わせて Fig. 1 に表示した<sup>1), 2)</sup>。各リガンドは受容体に対して相補的であると共に、リガンド同士は相互に類似していることが見て取れる。つまり、LBDD および SBDD はそれぞれ独立に技術向上が図られてきたが、本来は同じ現象を異なる観点から解析しているに過ぎない。また、上述のように両手法の特徴および問題点は完全に相補的と言える。従って、いずれもが実用レベルに達した現状において、両手法を統合した仮想スクリーニングが指向される段階に入ったと言える。近年、特定の蛋白質に対し複数の化合物に

ついてドッキングスタディーを行い、得られた結合モデルに基づき 3 次元 QSAR 解析を行う事例が多数報告されている<sup>3) ~ 6)</sup>。この解析手順は、確かに両手法を組み合わせたものであるが、単にそれらを段階的に適用したに過ぎない。真に求められる統合解析手法は、両者の問題点を克服し得るものであるべきだが、実際この方法では、LBDD 手法の問題点である構造活性相関が未知なる部分の予測も、SBDD 手法の問題点であるドッキングスタディーの精度向上も、共に改善され得ない。

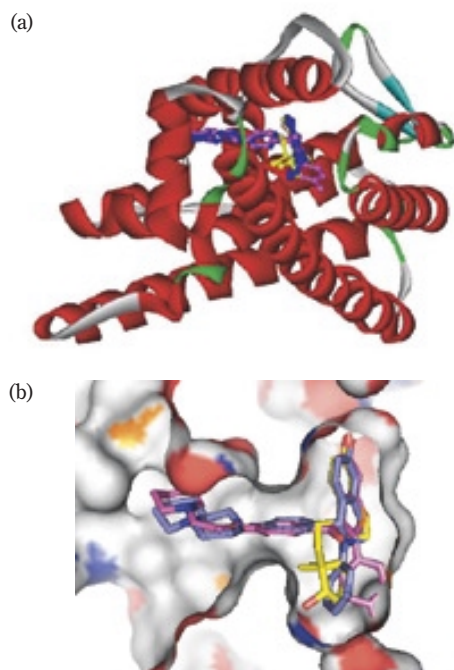
著者らは、LBDD と SBDD の各手法をシームレスに連携した解析を実行し得る仮想スクリーニング統合システム "薬師 ~ Xsi ~ ver.1.0" をみずほ情報総研(旧富士総合研究所)と共同で開発した<sup>7), 8)</sup>。また、LBDD と SBDD の各手法の問題点を克服する解析手法として、マルチブルドッキング解析法を考案した<sup>9), 10)</sup>。その有用性について、実際の創薬研究を想定した具体的な事例により検証した。

### 仮想スクリーニング統合システムの開発

既存の代表的な統合ソフトウェアとして、SYBYL<sup>®</sup> (TRIPOS 社)<sup>11)</sup> や MOE<sup>™</sup> (CCG 社)<sup>12)</sup> 等が市販されている。SYBYL<sup>®</sup> は低分子用の分子設計統合ソフトウェアとして古い歴史をもち、随時新しいモジュールが追加され、現在では高分子に関する計算機能も含む多機能な統合ソフトウェアへと発展している。一方、MOE<sup>™</sup> は、分子計算に適したプログラミング環境の提供を目的に開発された製品で、様々な計算機能はその環境下で稼動するプログラムとして付加的に提供される。両者とも GUI (Graphical User Interface) 環境が整備されており、計算の実行および解析を直感的に行える。また専用のプログラム言語が準備されており、複雑な連続計算処理も可能である。

しかしながら、両ソフトウェアは利便性と拡張性において一長一短である。SYBYL<sup>®</sup> のようにある程度纏まった機能がモジュールとして提供されることはユーザーの利便性が高い。すなわち、ユーザーはモジュール自体を開発する必要がなく、既存の機能をいかに組み合わせる適切な解析を行うかに注力できる。一方、MOE<sup>™</sup> のようにモジュール自体を開発できる環境は、仕様の変更や機能の追加といった拡張性に優れている。後述するマルチブルドッキング解析を実現するには、多数のモジュールを複雑に組み合わせるだけでなく、各モジュールに不足する機能を補う必要がある。よって、利便性と拡張性を両立した統合ソフトウェアが望まれることから、我々は独自に開発することとした。

統合ソフトウェアの開発は、国内のソフトウェア



(a) X-ray structure of the ligand binding domain of estrogen receptor  $\alpha$  with Tetrahydroisoquinoline (IXQC, blue), Raloxifene (1ERR, pink) and 17 $\beta$ -Estradiol (1ERE, yellow)  
 (b) Ligand binding site of estrogen receptor  $\alpha$  with the three ligands

**Fig. 1** Structures of three ligands bound to estrogen receptor  $\alpha$

開発企業の「みずほ情報総研 株(旧富士総合研究所)」と共同で行った。また、*in house*での利用に留めるのではなく、製品として販売することを前提とした。ソフトウェア会社と組むことは、先方のプログラム資産を流用できることに加え、専門家による高度なIT技術の恩恵を受けることができる。また、運用開始後に想定されるソフトウェアの保守や、仕様変更および機能拡張等の開発を委託することができる。国内企業であることは、円滑なコミュニケーションはもとより、メガファーマの意向が優先されがちな外国企業よりも柔軟かつ迅速な対応が期待できる。一方、製品として販売されることは、ソフトウェアの信頼性を確保する上で重要である。この統合ソフトウェアは、「薬師 (Xsi) ver.1.0」として2004年1月末にリリースされた<sup>7),8)</sup>。

## 1. 薬師 (Xsi) の概要

Xsiは、SYBYL®と同様に複数のモジュールで構成されている。各モジュールには、情報を格納するひとつのクラスと、そのクラスに対してデータ操作や計算処理を指示する複数の関数が準備されている。Xsiの実行は、専用スクリプトを介してCUI (Character User Interface) ベースで行う。専用スクリプトの記述方法は極めて単純であり、基本的には1対のクラスと関数を1行ごとに羅列するだけでよい。必要に応じて、条件文や繰返し文、各種演算子を利用できる。また、すべてのクラスは多次元配列として扱うことも可能である。Fig. 2には、SDファイルから読み込んだ複数の分子について分子力学計算により連続的

```

MoleculeSetFileSD sdfFile;
Array<Molecule> am;
MolecularMechanics mm;
Integer i;
//
sdfFile.setFileName ("test.sdf");
am = sdfFile.getMolecule();
//
i = 0;
while (i < am.size()) {¥
mm.setMolecule(am[ i ]);¥
mm.minimize();¥
mm.clear();¥
i = i + 1;¥
};
//
sdfFile.clear();
sdfFile.setFileName("test_mm.sdf");
sdfFile.setMolecule(am);
sdfFile.output();
quit;
    
```

} Declaration of class and array  
 } Import molecules from an SD file  
 } Energy minimization by molecular mechanics  
 } Output molecules to an SD file

**Fig. 2** Xsi script for energy minimization by molecular mechanics

に構造を最適化し、結果をSDファイルに出力するスクリプト例を示した。スクリプトを読み易くする為に加えコメント行を除くと、20行にも満たないスクリプトにより記述できる。

スクリプトを介したCUIベースでの実行は、解析手順を試行錯誤する計算化学者と、既定の計算を実行するだけの計算化学を専門としない研究者の双方にとって都合がよい。前者は、一度作成したスクリプトについて部分的な変更を繰り返し最適な条件あるいは手順を見出したり、他の用途に転用したりできる為、高い生産性を実現する。一方、CUIベースでの実行であればWeb等でのサービス提供が容易であり、後者には使い慣れない専用GUIよりも却って好ましい。

本プログラムは、C++言語で記述されており、Linux上で稼動する。こうした汎用技術の使用により、飛躍的な性能向上を続ける計算機環境へ容易に対応することができる。また、一部のクラスは並列計算に対応していることから、複数のCPUで構成されたPCクラスター上で稼動し、大規模な計算が可能である。最近では、グリッド計算環境への対応も完了している。当社では、ひとつのマスターノードと4つのスレーブノードで構成されたメモリー分散型のPCクラスター上で使用している。各ノードには、Intel社 Xeon2.4GHzのCPUと2GByteのRAMが搭載されており、SuSE Linux 7.3により稼動している。

## 2. モジュール構成

創薬研究に利用される計算化学手法は極めて多岐にわたり、無償および商用ソフトウェアが数多く存在する。その大半は、特定の解析を目的としており、関連する必要十分な機能を持たせることで利便性を高めている。その結果、解析目的の異なるソフトウェア間に重複する機能が生じている。例えば、3次元構造記述子に基づくQSAR解析と複数化合物の重ね合せ解析を目的とするソフトウェアでは、両者とも立体配座解析機能を必要とする。汎用的な解析を目的とする統合ソフトウェアでは、こうした機能の重複はソフトウェアの開発と保守の手間を増やすだけでなく、仕様の違いが解析結果に影響しかねない。Xsiでは必要な機能を網羅した上で相互に重複しないようにオブジェクト化し、それらの組み合わせを変えることで種々の解析目的に対応できるようなクラスおよび関数の構成とした。

一方、個々の機能については、無償もしくは比較的安価な商用ソフトウェアの中にも、極めて優れたものが少なからず存在する。こうしたソフトウェアは一般的に広く普及しており、多くの研究者が操作に慣れ親しんでいる。著者らは、こうした既存のソ

ソフトウェアも積極的に活用することとした。

Xsi は、Table 1 に示す 27 個のクラスと 374 個の関数で構成されている。ファイル入出力機能には、4 つのクラスが準備されている。有機低分子には MOL および SDF 形式、蛋白質には PDB 形式、そしてテキストデータには CSV 形式を採用した。いずれの形式も各用途における標準的なファイル形式である。分子情報を格納するクラスには、Molecule と Universe が設定されている。前者にはひとつの分子、後者には複数の分子および蛋白質のアミノ酸配列情報等が格納される。また、種々の分子計算における制限情報を格納するクラスが、双方に準備されている。分子計算には、10 種類のクラスが設定されている。この中には、エネルギー計算を行うものとして、分子力学計算、モンテカルロ計算およびドッキング解析があげられる。これらの計算では、MMFF94s 力場パラメーター<sup>13)</sup>、溶媒効果として距離依存誘電率や GB/SA モデル<sup>14)</sup>を利用できる。また、ドッキング解析は、通常のエネルギー計算に基づく方法に加え、ポテンシャルフィールドを用いた高速な計算も可能である。一方、分子間の比較には、ファーマコフォアや形状等の特徴量をグリッド格子点上に投影した

座標系依存の特性値や、それらを主成分分析した座標系に依存しない特性値 (WHIM 記述子<sup>15)</sup>) 等が用いられる。Xsi が算出できる分子の特性値は主に 3 次元構造に基づくものであり、LogP やトポロジカルインデックス等の平面構造から算出する構造記述子は外部プログラムの利用を想定している。

Xsi には、これら以外に数値や文字列に対するクラスが準備されている。これらのクラスは、スクリプトにより四則演算等を行うことができる。また、多次元配列として柔軟に扱うことができ、その配列操作を簡便にする Utility クラスも準備されている。Xsi では、分子計算による結果や分子の特性値を CSV ファイルとして出力し、外部プログラムにより統計解析やインフォマティクス解析を行うことを想定している。多くの計算化学者にとって、使い慣れた汎用の統計解析ソフトウェアや最新の情報科学手法を利用できるメリットは大きい。こうした外部プログラムの多くは CSV 形式に対応しており、解析結果を Xsi に取り込み元の分子情報に反映させることは容易である。分子計算と連動して頻繁に用いられる一部の統計および情報科学手法は Xsi に実装されている。線型回帰分析、線型判別分析、類似性計算、クラスタ

**Table 1** Xsi function modules

Category	Xsi_Class	Xsi_Function	Description
File I/O	MoleculeFileMol	8	I/O for MOL format file (for single small molecule)
	MoleculeSetFileSD	9	I/O for SDF format file (for multiple small molecules)
	MoleculeSetFilePDB	11	I/O for PDB format file (for protein)
	CSVFile	22	I/O for CSV format file (for text data)
Molecular Information	Molecule	31	Information of small molecule
	Universe	9	Information of protein or multiple molecules
	Constraints	12	Constraints information for small molecule
	UniverseConstraints	20	Constraints information for protein or multiple molecules
Molecular Calculation	MolecularMechanics	19	Molecular Mechanics calculation
	MolecularOrbitalCNDO	2	Molecular Orbital calculation
	MonteCarlo	48	Monte Carlo simulation
	Docking	38	Docking simulation
	LigandAlignment	53	Ligand alignment
	RotationAngleReservoir	2	Storage of dihedral angles
	Descriptor	5	Structural descriptor calculation
	Field	5	Projection of molecular properties onto grid cross sections
Miscellaneous	PotentialField	8	Projection of potential energy onto grid cross sections
	RMSMinimizer	7	RMS calculation
	Integer	6	Integer variable
	Real	5	Real number variable
	String	6	Character string variable
	RandomNumberGenerator	5	Generation of random numbers
	ClusterAnalyzer	10	Cluster analysis
	Similarity	5	Similarity search
	Clique	6	Clique search
	Statistics	10	Regression and discriminant analyses, etc.
	Utility	12	Operation of array etc.

一分析およびクリーク探索等がそれに該当する。

### 3. 汎用的な分子計算機能

著者らは、創薬研究において頻繁に用いられる計算手法を実現するスクリプトを作成した (Table 2)。

分子の立体構造を生成したり、また実際に取り得る配座を解析することは、一般的に分子設計における最初の過程である。これらは、多数の分子を対象とすることから、連続的かつ高速な処理が期待される。よって、すべてのスクリプトは複数の分子を連続的に処理できるように記述した。また、標準の結合長および結合角をもつ立体構造を生成するだけの機能から、極小構造を網羅する詳細な配座解析までをカバーしており、用途に応じて使い分けできる。さらに、特定の置換基に限定した配座解析やコンビナトリアル的に構造を生成する機能も実現している。

LBDD 関連のスクリプトについては、分子特性値の算出、ケモインフォマティクス、ファーマコフォア解析、QSAR 解析およびリガンドアライメント等の多岐にわたる機能を実現している。ファーマコフォアについては、MOL 形式による出力が可能であり、PyMol<sup>16)</sup>等の分子構造ブラウザを用いて分子構造と対応させた表示が可能である。また、分子特性値を CSV ファイルを介して種々の情報科学的解析プログ

**Table 2** General purpose functions for drug design programmed by Xsi scripts

	Tertiary structure generator
	Residual conformation optimizer
LBDD/	Conformation search for diverse set
SBDD	Conformation search for multiple minimum set
	Combinatorial conformation generator
	Substructure conformation search
	Structural descriptors calculation
	Molecular surface calculation (polar/non-polar/total)
	Similarity or k-nearest neighbor search
	Similarity matrix calculation
	Cluster analysis and diversity extraction
LBDD	Pharmacophore mapping with output of MOL format file
	Multiple ligand alignment based on pharmacophore similarity
	Regression and discriminant analysis with selection of explanatory variable
	Grid-projected descriptors calculation
	3d-QSAR (CoMFA/CoMSIA)
	Ligand alignment with conformation adjustment
	Random structure generator in active site of protein
	Water molecule mapping in active site of protein
	Potential field generator with output of PDB format file
SBDD	Grid-based Docking
	Cartesian-based Docking
	Substructure conformation search in active site of protein
	Binding energy calculation
	Molecular surface calculation for ligand/protein complex

ラムに与えることで、高度な解析が可能である。

SBDD 関連のスクリプトも同様に、多くの機能をカバーしている。ドッキング解析は、ポテンシャルフィールドを用いた高速な解析と、蛋白質の挙動を含む正確なエネルギー計算に基づく詳細な解析の両方を実現した。また、特定の置換基だけを可変としたドッキング解析も可能である。さらに、ドッキング解析から得られた結合モデルに基づき、水和自由エネルギーを加味した結合エネルギーの算出や、結合エネルギーとの相関性が指摘されている各種表面積の計算も行える。

### 4. 既存ソフトウェアの相補的利用

Xsi と相補的に用いる既存ソフトウェアを Table 3 に示した。これらのソフトウェアとは、SDF、PDB および CSV 形式といった汎用的な標準ファイル形式によりデータをやり取りする。その他のファイル形式については、BABEL 等の変換プログラムの利用を想定している。

Xsi と連携させた既存ソフトウェアの高度な利用方法として、標的蛋白質に対する結合状態が既知もしくは推定されている化合物に基づき、特定の置換基だけを種々変換した誘導体について連続的にドッキングスタディーを行う場合がある。この際、共通する部分構造と変換する置換基セットを準備する必要がある。前者は元の結合状態を保持したまま分子構造を改変する必要があり、DS Viewer Pro<sup>17)</sup>等を利用する。一方、後者には Accord for Excel<sup>17)</sup>の “ R Group Table ” 機能を利用すると便利である。通常、特定の位置に導入される分子フラグメントは、合成ルートに基づき適切な官能基を有した試薬から選択される。よって、試薬データベースから抽出した複数の化合物を Accord for Excel に取り込み、“ R Group Table ” 機能において該当する官能基を指定すれば、

**Table 3** External softwares used with Xsi

Category	Function	Software
Molecular Viewer		PyMol
	2D	ISIS Draw, Chem Draw
	3D	DS ViewerPro
Molecular Editor	R-substitution	Accord for Excel
	Database	ISIS Base, ChemFinder
	Protein	PDB Viewer
File Format Converter		BABEL
	Statistics	SPSS, S-PLUS
Analysis	Infomatics	RandomForest, LibSVM, NEUROSIM, BayesiaLab
	Molecular Properties	Dragon, WSKOW
Molecular Calculation	Molecular Dynamics	TINKER
	Molecular Orbital	MOPAC

導入したいフラグメントと結合位置がすべての分子に対して自動的に設定される。SDF ファイルを介して先の共通部分構造の座標データと共に Xsi に与えれば、適切な分子を構築できるだけでなく、導入したフラグメント部分だけを可変とする配座解析やドッキングを自動的かつ連続的に実行する。

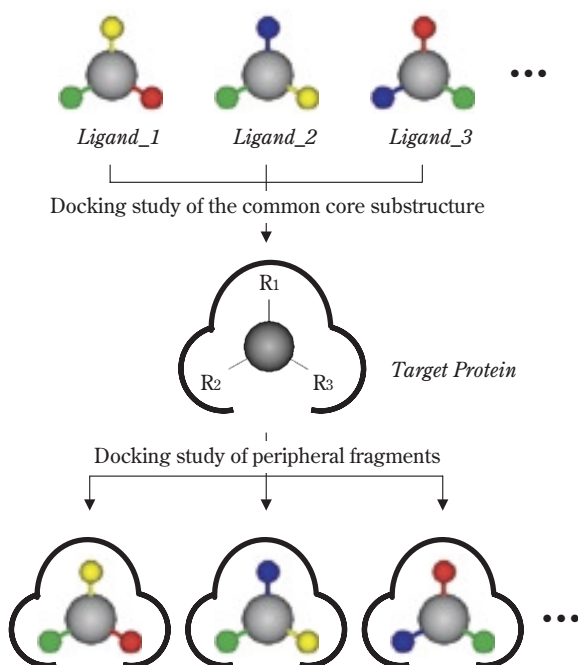
### マルチプルドッキング解析

著者らは、LBDD 手法の問題点である構造活性相関が未知な部分の予測と、SBDD 手法の問題点であるドッキングスタディーの精度向上を目的として、マルチプルドッキング解析法を考案した<sup>9), 10)</sup>。この解析手法は、複数の既知リガンドについて結合モデルを同時に得ることを特徴とする。また、得られた複数分子の結合モデルに基づき、ドッキングスタディーと薬理活性予測に対する各評価関数を算出する。そして、薬理活性が未知の化合物について、これらの評価関数を用いた仮想スクリーニングを行うことができる。

#### 1. 複数分子のドッキングスタディー

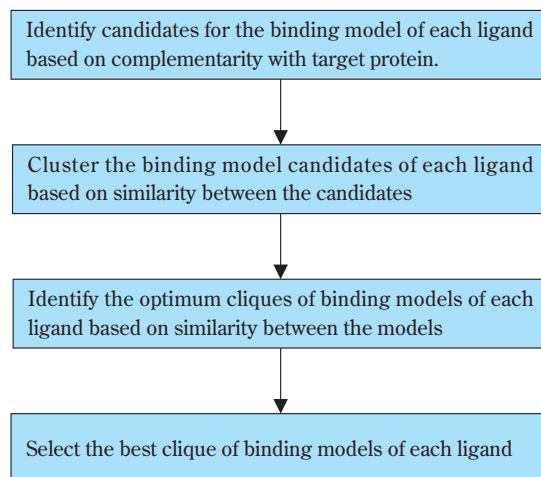
多数の事例から、共通の部分構造を有する誘導体は、同一の標的蛋白質に対して類似の結合様式を示す傾向にあることが知られている (Fig. 1)。この知見を利用して、マルチプルドッキング解析では、共通部分構造を有する複数の既知リガンドについて、結合モデルを同時に得ることから始める。

複数分子のドッキングスタディーは、共通部分構

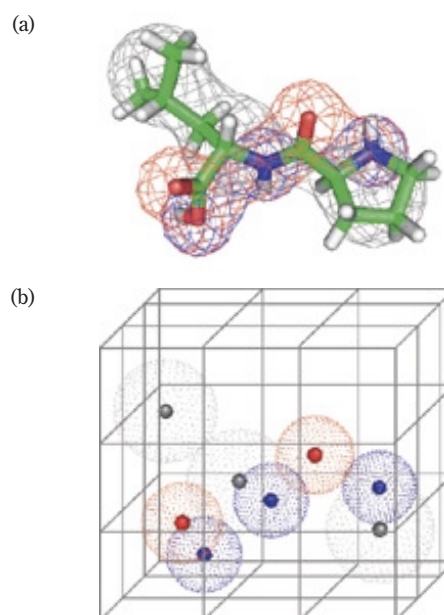


**Fig. 3** Scheme for the stepwise docking study

造と周辺置換基の各結合モデルを段階的に探索する (Fig. 3)。各段階の計算手順を Fig. 4 に示したが、両段階は同じ計算手順を踏む。つまり、各分子について標的蛋白質との相補性から結合モデル候補を列挙し、それらのファーマコフォアや形状等の相同性から全分子の最適な結合モデルの組み合わせを見出す。後者には、クリーク探索と呼ばれるグラフ理論の手法を用い、より多数の分子がより相互に類似している結合モデル候補の組み合わせを探索する。この探



**Fig. 4** Procedure of docking study for both common core substructure and peripheral fragments



Hydrogen bond acceptor (red), Hydrogen bond donor (blue) and Hydrophobic center (gray).

(a) Molecular properties mapped on a model compound.

(b) Molecular properties projected onto grid cross sections

**Fig. 5** Examples of molecular properties

索を容易にする為、事前に各分子について多様な結合モデル候補を抽出しておく必要がある。なお、ファーマコフォアや形状等の類似性は、各結合モデルの水素結合能、電荷および疎水性といった特性値を共通のグリッド格子点上に投影した情報に基づいて算出する (Fig. 5)。

このように、共通部分構造と周辺置換基の探索は同じ計算手順により行うが、各手順の方法は双方で若干異なる。まず、標的蛋白質との相補性から結合モデル候補を探索する手順では、前者は活性部位近傍の広範囲を粗く探索するのに対し、後者は限定された領域を詳細に探索する。よって、前者では精度こそ落ちるものの高速な計算が可能なポテンシャルフィールドに基づく探索<sup>18)</sup>を、後者では低速ながら精度の高い計算が可能なCartesian座標系に基づく探索を行う。なお、後者では必要に応じて蛋白質の構造変化を考慮する。また、各分子ごとに多様な結合モデル候補を抽出する手順では、前者がクラスタ分析から得られた各クラスタの中心を採用するのに対し、後者は最小エネルギーの結合モデル候補を採用する。さらに、最適な結合モデル候補の組み合わせを決める方法も、前者では共通部分構造の類似性を指標とするのに対し、後者では後述する結合エネルギーと薬理活性値との相関性等の指標に基づく。

以上の解析から得られた複数分子の結合モデルは、ひとつひとつが蛋白質に対して相補的であると同時に、結合モデル同士が良好な対応関係にある。すなわち、通常のドッキング手法における評価関数と最適解探索の不十分さを、結合モデル間の相同性により補う方法と言える。また、共通部分構造と周辺置換基の段階的な探索や、各手順における目的に応じた計算方法の選択により、広範かつ詳細な探索を効率的に行うことができる。

## 2. 評価関数の作成と仮想スクリーニング

複数分子のドッキングスタディーから得られた結合モデルに基づき、薬理活性が未知である任意の化合物について仮想スクリーニングを行う為の評価関数を作成する。仮想スクリーニングは、標的蛋白質に対するドッキングと、得られた結合モデルに基づく薬理活性値の予測の2段階で行う (Fig. 6)。従って、両段階に対する評価関数を作成する。

ドッキングスタディーでは、標的蛋白質との相補性と、既存の結合モデルに対する類似性とを段階的もしくは同時に評価する。前者の役割は結合モデル候補を列挙することにある為、汎用的評価関数をそのまま利用する。対象分子の最適な結合モデルの選択および標的蛋白質に結合し得ないとの判別は、後

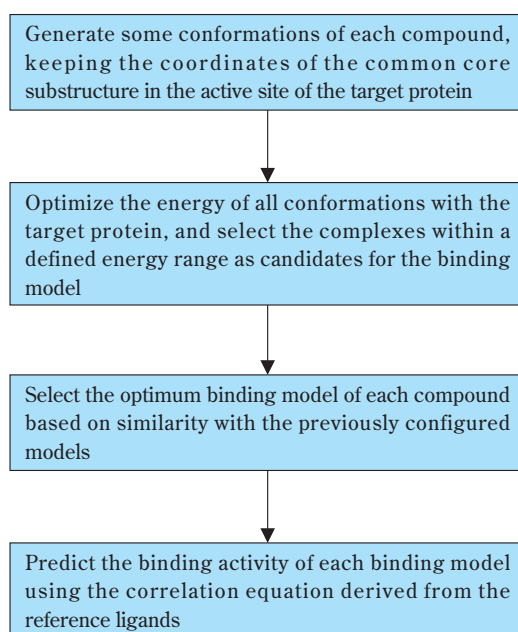


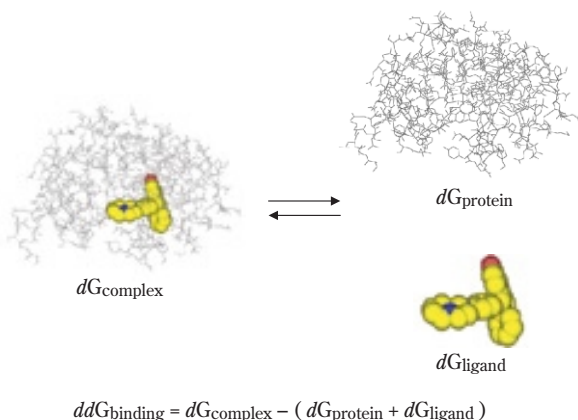
Fig. 6 Procedure of virtual screening based on the multiple docking model

者により行う。よって、ドッキングの評価関数は、既存の結合モデルに対する類似性を評価するものを意味する。なお、この評価関数に、標的蛋白質に対する相補性スコアを加味することも可能である。

既存の結合モデルに対する類似性の評価には、分子特性を共通のグリッド格子点上に投影した情報を用いる。比較的単純な方法には、既存の各結合モデルに対するK-Nearest Neighbor法<sup>19)</sup>や、各格子点ごとに複数分子の特性値を平均化した情報に対する類似性を評価する方法があげられる。より高度な評価法としては、One-Class SVM法<sup>20)</sup>やSOM法<sup>21)</sup>等の情報科学的解析手法の利用が想定される。いずれにせよ、既存の結合モデルを学習データとし、より擬陰性の少ない方法を採用する。また、真の結合状態とは異なる結合モデル候補や既知不活性分子の結合モデル候補を学習データに加えることで、擬陽性を低く抑える試みも有効と考えられる。こうした方法は、構造活性相関情報が蓄積されるにつれ学習モデルが充実し、結果的に予測精度の向上が期待できる。また、対象分子について標的蛋白質との複合体構造が実験的に明らかにされれば、該当する分子の結合モデルを実測データに置き換えた上で、他の分子の結合モデルを修正することも可能である。

一方、結合モデルに基づく薬理活性値の予測はチャレンジングな課題である。標的蛋白質に対するリガンドの結合活性は、可逆的な平衡状態においては結合と非結合の各状態間の自由エネルギー差によって決まる (Fig. 7)。しかしながら、溶媒分子を含む複雑な系について、エントロピーの寄与や静電相互

作用を正確に見積もるには相当の計算量を要する。近年、拡張アンサンブル分子動力学シミュレーションによる自由エネルギーの算出<sup>22)</sup>や、分子軌道計算による正確な相互作用エネルギーの見積<sup>23)</sup>等の試みがなされているが、仮想スクリーニングに要求されるスループットには遠く及ばない。著者らは、現実的な対応として、連続体モデルによる水和の自由エネルギーと力場関数に基づくポテンシャルエネルギーから結合エネルギーの近似値を算出する方法を採用している。その際、蛋白質とリガンド間の相互作用エネルギーに相当する項と、水和エネルギー差に相当する項に分離し、既知の薬理活性値を指標とした回帰式として算出する。一方、力場関数は静電相互作用を正確に見積もれない場合が多い。それを回避するひとつの方法として、極性・非極性の溶媒露出表面積による回帰式が薬理活性値を良好に予測できることが知られている<sup>24)</sup>。これら2つの方法を試み、既存の構造活性相関をより正確に予測できる方を採用する。



**Fig. 7** Free energy of binding between ligand and target protein

### 3. マルチプルドッキング解析の有用性

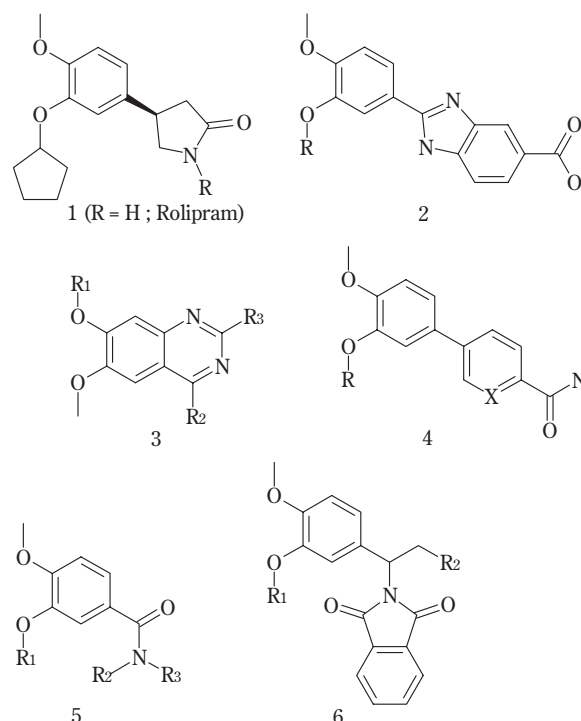
以上述べた仮想スクリーニング手法は、標的蛋白質と化合物の相補性、および複数化合物間の類似性を適切に組み合わせる解析する。従って、SBDDおよびLBDDの各機能をシームレスに連携した解析を実行し得るXsiのような統合ソフトウェアにより、初めて可能になった。

一方、本手法は、一般のドッキングスタディーとは異なり、先験情報として周辺誘導体の構造活性相関情報を若干ながら必要とする。通常、HTSにおける複数のヒット化合物に共通して見られる部分構造、あるいは既知化合物の部分的改変から見出された重要な部分構造を起点に、周辺誘導体の探索合成は始められる。従って、本手法に要求される先験情報を

利用できないケースは稀であり、実用上問題となることは殆どないと言える。当然ながら、マルチプルドッキング解析では、標的蛋白質の立体構造が必要である。実験的に得られた座標データを使えない場合は、ホモロジーモデリング等により構築されるが、通常ドッキングスタディーではその構造の精度を問わずに利用される。また、結合するリガンドに依存した蛋白質の構造変化についても十分に考慮されていない。一方、本手法は既知の構造活性相関情報を用いた評価関数の構築と、その評価関数を利用した仮想スクリーニングとを段階的に行う為、前者において蛋白質構造の妥当性を間接的に検証できる。また、通常ドッキングスタディーとは異なり、構造活性相関情報の蓄積や、X線構造解析等による複合体の座標データの解明により、随時予測精度を向上させることができる。

### Phosphodiesterase-4 阻害剤への適用

マルチプルドッキング解析の有用性を検証する為、実際の創薬探索研究を想定したモデル実験を行なった。モデル実験の対象には、喘息等の疾患に対する治療薬として期待され、グローバルに研究開発が進められているPhosphodiesterase-4 (PDE-4) 阻害剤を選択した。PDE-4を阻害する化合物としてRolipramが古くから知られており (Fig. 8)、複数のグループから代表的な市販SBDDソフトウェア (Dock、FlexX、AutoDock) によるドッキングスタディーの結果が



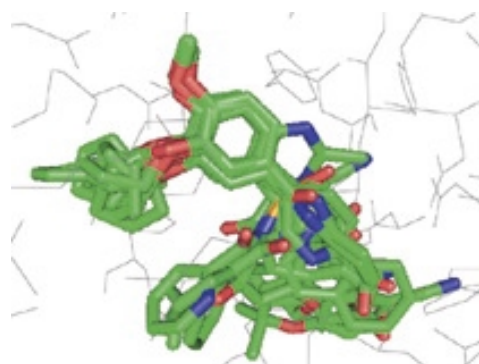
**Fig. 8** Analogs of PDE-4 inhibitors



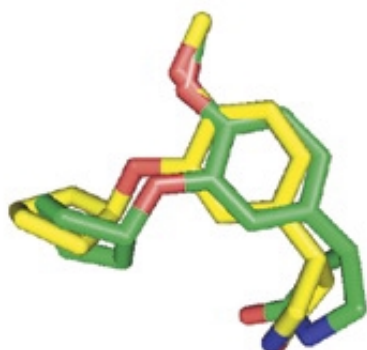
2002年に報告された<sup>25), 26)</sup>。その翌年にX線構造解析による複合体構造について報告されたが(PDB ; 1OYN) いずれの結合モデルも真の結合状態とは全く異なるものであった<sup>27)</sup>。つまり、計算により結合モデルを推定するのが極めて困難な対象系と言える。

モデル実験では、PDE-4阻害作用を示す6系統のカテコール誘導体78化合物を解析対象とした(Fig. 8)<sup>28) ~ 33)</sup>。そのうち、各系統から任意に2化合物づつを選択した計12化合物について、複数分子のドッキングスタディーを行った。文献報告と同様に、PDE-4の座標データにはX線構造解析から得られたリガンド非結合状態のものをを用いた(PDB ; 1F0J)<sup>34)</sup>。先述した手順(Fig. 4)に従い解析したところ、1化合物を除いて良好な結合モデルが得られた(Fig. 9)。この11化合物の中にはRolipramが含まれているので、X線構造解析から得られた複合体構造と比較した。その結果、計算から得られた結合モデルは真の結合状態を良好に再現するものであり、両者のヘテロ原子同士のRMS値は1.09であった(Fig. 10)。

11化合物の結合モデルに基づき、残りの66化合物(8個の不活性化化合物を含む)について仮想スクリーニングを行った。ドッキングスタディーにおける11



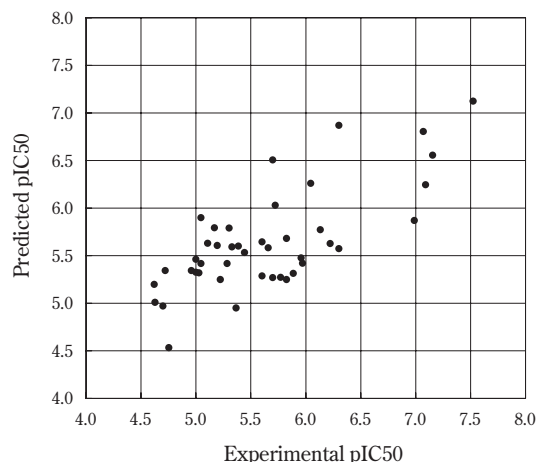
**Fig. 9** Binding models of 11 known PDE-4 inhibitors



**Fig. 10** The X-ray structure of Rolipram bound to PDE-4 and the binding model by multiple docking analysis

化合物の結合モデルとの類似性評価は、各グリッド格子点に投影された複数分子の特性値を平均化した情報に基づき行った。結合モデルの妥当性は、PDE-4との複合体エネルギーおよび、それを構成する相互作用エネルギーと水和エネルギーに基づいて判断した。すなわち、いずれかのエネルギー値が極端に大きいものは不活性と判断した。その結果、58個の活性化化合物のうち48化合物(83%)の結合モデルを得た。

この48個の結合モデルについて、薬理活性値の予測について検討した。まず、相互作用エネルギーと水和の自由エネルギーから算出した結合エネルギーについて検討したところ、薬理活性値との間に相関が見られたものの、負の相関であることから不適切と判断した。一方、相互作用エネルギーだけに着目した場合、薬理活性値との間に一定の正の相関が見られたものの、薬理活性値の予測には不十分なものであった。そこで、極性・非極性の溶媒露出表面積に基づく方法を試みたところ、比較的良好な結果が得られた。すなわち、薬理活性値を指標に極性および非極性の溶媒露出表面積差を説明変数とする線型回帰分析を行ったところ、式(1)に示す回帰式が得られた。この回帰式の相関係数(R)は0.733であり、5化合物をはずれ値として分析から除いた。算出された薬理活性の予測値と実験的に得られた値との関係をFig. 11に示した。これによると、多少のばらつきは見られるものの、3オーダー弱の範囲にわたりpIC50値を良好に予測し得ることが分かった。この範囲の下限はHTS等におけるヒット化合物に期待されるレベルの薬理活性強度であり、上限は開発候補化合物の対象となり得るレベルである。すなわち、探索合成を支援する目的に十分に合致する結果と言える。一方、仮想スクリーニングの対象に含まれる8個



**Fig. 11** Correlation between experimental and predicted pIC50's of PDE-4 inhibitors

の不活性化化合物のうち、結合モデルが得られた6化合物について式(1)から薬理活性値を求めた。その結果、2化合物こそ $pIC_{50}$ 値が6.5前後の比較的強い活性と予測されたが、残りの4化合物は5.5以下の弱い値を示した。これらの化合物の実験による $pIC_{50}$ 値は4ないし5以下であり、その近傍の弱い活性を有している可能性は否定できない。いずれにせよ、こうした仮想スクリーニングは活性向上を目的に利用される為、不活性もしくは弱い活性しか有さない化合物を識別するのに有効と言える。

$$pIC_{50} = 1.303 + 0.00441 * Polar\_ASA + 0.00494 * Non\ Polar\_ASA, n = 43, R = 0.733 \quad (1)$$

## おわりに

LBDDとSBDDの各機能を統合したソフトウェアであるXsiの開発とその応用事例について概説した。ここで紹介した事例はあくまでモデル実験であるが、すでにXsiは様々な創薬研究課題に応用されており、着実に実績を上げている。また、本稿では触れなかったが、Xsiの用途は仮想スクリーニングに留まらず、ADMEおよび毒性の予測やリパーズプロテオミクス研究の支援等にも活用されている。Xsiは随時機能拡張が図られており、次版の開発および次々版の計画策定が進行中である。今後、ユーザーの増加によりプログラムの信頼性が向上すると共に、本ソフトウェアを利用したより高度な仮想スクリーニング手法の出現が期待される。

最後に、Xsiの共同開発先であるみずほ情報総研(株)バイオテクノロジー室の関係者に深くお礼を申し上げます。

## 引用文献

- 1) A. M. Brzozowski, A. C. W. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson, and M. Carlquist, *Nature*, 389, 753-758 (1997)
- 2) J. Renaud, S. F. Bischoff, T. Buhl, P. Floersheim, B. Fournier, M. Geiser, C. Halleux, J. Kallen, H. Keller, and P. Ramage, *J. Med. Chem.*, 48, 364-379 (2005)
- 3) C. L. Kuo, H. Assefa, S. Kamath, Z. Brzozowski, J. Slawinski, F. Saczewski, J. K. Buolamwini, and N. Neamati, *J. Med. Chem.*, 47, 385-399 (2004)
- 4) K. Jozwiak, S. Ravichandran, J. R. Collins, and I. W. Wainer, *J. Med. Chem.*, 47, 4008-4021 (2004)
- 5) A. A. Soderholm, P. T. Lehtovuori, and T. H. Nyronen, *J. Med. Chem.*, 48, 917-925 (2005)
- 6) R. Ragno, M. Artico, G. D. Martino, G. L. Regina, A. Coluccia, A. D. Pasquali, and R. Silvestri, *J. Med. Chem.*, 48, 213-223 (2005)
- 7) <http://www.mizuho-ir.co.jp/science/xsi/index.html>
- 8) Y. Inagaki, M. Hamada, K. Yamazaki, M. Kanaoka, and H. Chuman, CBI学会2004年大会(2004)
- 9) K. Yamazaki, and M. Kanaoka, PCT/JP02/11401 (住友製薬)
- 10) K. Yamazaki, M. Kanaoka, and Y. Inagaki, CBI学会2004年大会(2004)
- 11) <http://www.tripos.com/>
- 12) <http://www.chemcomp.com/>
- 13) T. A. Halgren, *J. Comp. Chem.*, 17, 490-641 (1996)
- 14) W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.*, 112, 6127-6129 (1990)
- 15) R. Todeschini, M. Lasagni, and E. Marengo, *J. Chemometrics*, 8, 263-272 (1994)
- 16) <http://pymol.sourceforge.net/>
- 17) <http://www.accelrys.com/>
- 18) I. D. Kunts, *Science*, 257, 1078-1082 (1992)
- 19) D. Chema, and A. Goldblum, *J. Chem. Inf. Comput. Sci.*, 43, 208-217 (2003)
- 20) E. Byvatov, and G. Schneider, *J. Chem. Inf. Comput. Sci.*, 44, 993-999 (2004)
- 21) Z. R. Yang, and K. C. Chou, *J. Chem. Inf. Comput. Sci.*, 43, 1748-1753 (2003)
- 22) A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers*, 60, 96-123 (2001)
- 23) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, *Chem. Phys. Lett.*, 313, 701-706 (1999)
- 24) J. S. Bardi, I. Luque, and E. Freire, *Biochemistry*, 36, 6588-6596 (1997)
- 25) O. Dym, I. Xenarios, H. Ke, and J. Colicelli, *Mol. Pharmacol.*, 61, 20-25 (2002)
- 26) P. Pospisil, T. Kuoni, L. Scapozza, and G. Folkers, *J. Recept. Sig. Trans.*, 22, 141-154 (2002)
- 27) Q. Huai, H. Wang, Y. Sun, H. Y. Kim, Y. Liu, and H. Ke, *Structure*, 11, 865-867 (2003)
- 28) I. L. Pinto, D. R. Buckle, S. A. Readshaw, and D. G. Smith, *Bioorg. Med. Chem. Lett.*, 3, 1743-1746 (1993)
- 29) J. B. Cheng, K. Cooper, A. J. Duplantier, J. F. Egger, K. G. Kraus, S. C. Marshall, A. Marfat, H. Masamune, J. T. Shirley, J. E. Tickner, and J. P. Umland, *Bioorg. Med. Chem. Lett.*, 5, 1969-1972 (1995)
- 30) R. J. Chambers, A. Marfat, J. B. Cheng, V. L. Cohan, D. B. Damon, A. J. Duplantier, T. A. Hibbs,

- T. H. Jenkinson, K. L. Johnson, K. G. Kraus, E. R. Pettipher, E. D. Salter, J. T. Shirley, and J. P. Umland, *Bioorg. Med. Chem. Lett.*, 7, 739-744 (1997)
- 31) J. G. Montana, G. M. Buckley, N. Cooper, H. J. Dyke, L. Gowers, J. P. Gregory, P. G. Hellewell, H. J. Kendall, C. Lowe, R. Maxey, J. Miotla, R. J. Naylor, K. A. Runcie, B. Tuladhar, and J. B. H. Warneck, *Bioorg. Med. Chem. Lett.*, 8, 2635-2640 (1998)
- 32) G. W. Muller, M. G. Shire, L. M. Wong, L. G. Corral, R. T. Patterson, Y. Chen, and D. I. Stirling, *Bioorg. Med. Chem. Lett.*, 8, 2669-2674 (1998)
- 33) B. Charpiot, J. Brun, I. Donze, R. Naef, M. Stefani, and T. Mueller, *Bioorg. Med. Chem. Lett.*, 8, 2891-2896 (1998)
- 34) R. X. Xu, A. M. Hassell, D. Vanderwall, M. H. Lambert, W. D. Holmes, M. A. Luther, W. J. Rocque, M. V. Milburn, Y. Zhao, H. Ke, R. T. Nolte, *Science*, 288, 1822-1825 (2000)

PROFILE



山崎 一人

Kazuto YAMAZAKI

住友製薬株式会社 研究本部  
化学研究所 分子設計研究グループ  
主任研究員



金岡 昌治

Masaharu KANAOKA

住友製薬株式会社 研究本部  
ゲノム科学研究所所長  
(元 分子設計研究グループ)  
農学博士