

バイオインフォマティクスとドッキング研究の接点

白井宏樹,* 小堀正人

Interface between Bioinformatics and Docking Study

Hiroki SHIRAI,* and Masato KOBORI

*Applied Genomics, Molecular Medicine Laboratories, Astellas Pharma Co., Ltd.,
21 Miyukigaoka, Tsukuba City 305-8585, Japan*

(Received July 31, 2006)

We describe the prospects of bioinformatics for drug discovery and discuss the current status, problems, and future direction of the interface between bioinformatics and docking studies. We also describe our recent work on sequence and structure analysis using the guanidino-modifying enzymes superfamily as a good example.

Key words—bioinformatics; docking study; guanidino group-modifying enzymes superfamily

1. はじめに

バイオインフォマティクス (情報生物学) とドッキング研究 (計算化学) とは, ともに創薬において活用されているインシリコ技術という共通性を持つにもかかわらず, 意外なほどに迂遠な関係にある。創薬は生物学と化学が密接に絡み合った学際研究の典型であることから, バイオインフォマティクスとドッキング研究をより効率的に組み合わせ、創薬インフォマティクスとも呼ぶべき総合的なソリューションを模索することが望まれつつある。そこで本稿ではバイオインフォマティクスとドッキング研究両者の接点を探った。まずは, 創薬におけるバイオインフォマティクスを分類し, 主に物質学としてのバイオインフォマティクスが, ドッキング研究との接点となることを解説する。物質学としてのバイオインフォマティクス研究には, 蛋白質の立体構造情報 (以下, 単に構造情報と略す) を考慮した高次の配列解析や立体構造情報自体の解析が含まれるが, あるスーパーファミリーを対象にしたこれらの研究を紹介し, 生物学上どのような知見が得られ, また創薬においていかなる有用な情報が抽出できるかを解説する。これを踏まえて, バイオインフォマ

ティクスとドッキング研究を相互に活用することで, より効率的な研究が展開できる可能性について考察する。

2. バイオインフォマティクスとドッキング研究

創薬は, 新規な医薬品標的遺伝子の発見に始まり, その機能を増幅又は制御する化合物や生物分子を発明 (創製) する研究である。Figure 1に, バイオインフォマティクスとドッキング研究の対比を示した。バイオインフォマティクスは, その技術や扱う対象, 目的において極めて広範な分野であり, かつ常に著しい変貌を遂げている分野である。このため, 専門家でさえその全貌を捉えることが困難な分野といえよう。このため, 次の節においては, 創薬という切り口でその概観を掴むことにしているが, 取りあえずここでは, バイオインフォマティクスを「配列や構造, 発現や機能といった各種情報を整理・解析することで, 遺伝子や蛋白質の機能や進化, あるいは生命現象といった生物学上の重要なイベントに関する新規知見を発見する分野」と定義し議論を進めることにする。バイオインフォマティクスのみによって何らかの発見につながるというよりは, 多くの場合仮説を立てるのに直接的, あるいは間接的に役立つことで, 次にどのような実験を行えばよいかの指針を与え, その実験による検証によって発見が成立したり, また仮説が正しくなかった場合には実験結果を踏まえ, また別の生物情報を加味

アステラス製薬株式会社分子医学研究所ゲノム情報研究室 (〒305-8585 筑波市御幸が丘 21)

*e-mail: hiroki.shirai@jp.astellas.com

本総説は, 日本薬学会第 126 年会シンポジウム S31 で発表したものを中心に記述したものである。

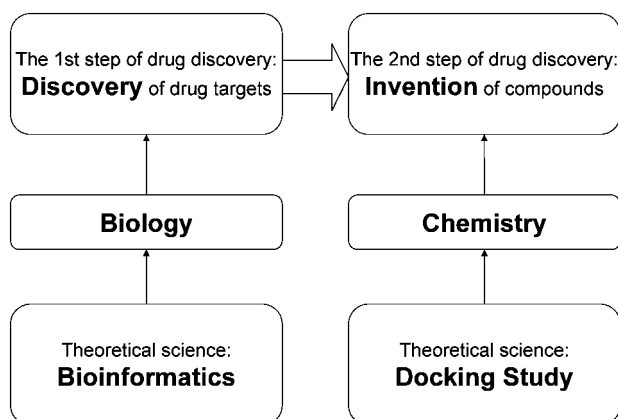


Fig. 1. Comparison of Bioinformatics and Docking Study

することで新たな仮説を作成するというサイクルに貢献することになる。創薬においては、昨今様々な活用がバイオインフォマティクスに求められつつあるが、創薬の最上流のステップである新規標的の発見研究における活用が最も求められている。

ドッキング研究は、標的とする蛋白質の活性部位の凹面に対して、それに相補的な化合物を多数の化合物候補より選抜する研究であり、それによって見出された化合物を合成展開することで、標的蛋白質の機能を十分に制御できる化合物を創製（発明）するという大目的の一部として機能している。

以上の対比を念頭においた上で、本稿の目的であるバイオインフォマティクスとドッキング研究の接点を模索するならば、それは通常発明を目的としているドッキング研究を、発見に利用する試みであり、逆に通常発見を目的としているバイオインフォマティクスを発明に利用する試みであるといえる。

3. 創薬におけるバイオインフォマティクスの分類と化学との接点

3-1. 創薬におけるバイオインフォマティクスの特徴と分類 バイオインフォマティクスは、その技術や扱う対象、目的において極めて広範な分野であり、かつ常に著しい変貌を遂げている分野である。研究対象である生物学(特に“オミクス生物学”)の質的・量的情報量の増大への対応が必要であり、一方で手段としての情報通信手段、ハードウェア、ソフトウェアの著しい進展に対応し活用することが望まれている。

バイオインフォマティクスにおけるこういった特徴に加えて、創薬におけるバイオインフォマティク

スへの期待も拡大を続けている。当初は創薬上流の新規標的探索への貢献が行われていたが、創薬の比較的下流に位置する毒性や副作用の予見や解析といった目的にも利用されるに至っている。さらには医薬品開発の段階での個人差の予見性や解析などへの活用も求められるようになってきている。

バイオインフォマティクスとドッキング研究との接点を模索するという本稿の試みを行うには、その前にまず、このようなバイオインフォマティクスの特殊事情を把握した上で、バイオインフォマティクスのどのような切り口がドッキング研究に対してその接点となるのか、整理しておく必要がある。このような鳥瞰なくしては、能力を超えた期待感を与えてしまうか、逆に不必要に消極的な展望を与えてしまう恐れがあるからでもある。

このために、まずは創薬におけるバイオインフォマティクスの役割を分類し、バイオインフォマティクス側からみた化学（化合物）研究、特にドッキング研究との接点を模索する（Fig. 2）。

バイオインフォマティクスの創薬への活用を考える際、インフォメーションテクノロジー（IT）としてのバイオインフォマティクスと理論科学としてのバイオインフォマティクスに大きく分類できよう。前節でバイオインフォマティクスは生物学上の新規知見の発見に貢献すると記したが、ITとしてのバイオインフォマティクスは、主として他のユーザーの利便性を向上させるなど実験研究者を介した広く間接的な貢献であり、理論科学としてのバイオインフォマティクスは、個別の課題に特化した、直接的な貢献を役割としている。

3-2. ITとしてのバイオインフォマティクス

ITとしてのバイオインフォマティクスはすべての基盤となるインフラ整備の作業と実験研究者が利用するデータベースの整備の作業に分けられる。



白井宏樹

アステラス製薬・分子医学研究所・主管研究員、東京医科歯科大学大学院・生命情報科学教育部・連携助教授、日本バイオインフォマティクス学会評議員幹事、創薬インフォマティクス研究会主査、薬学博士。1964年2月生まれ。大阪大学薬学部卒、同大学院修了後、田辺製薬研究員（この間生物分子工学研究所客員研究員、英国ケンブリッジ大学客員研究員）、理化学研究所・ゲノム科学総合研究所研究員を経て現職。

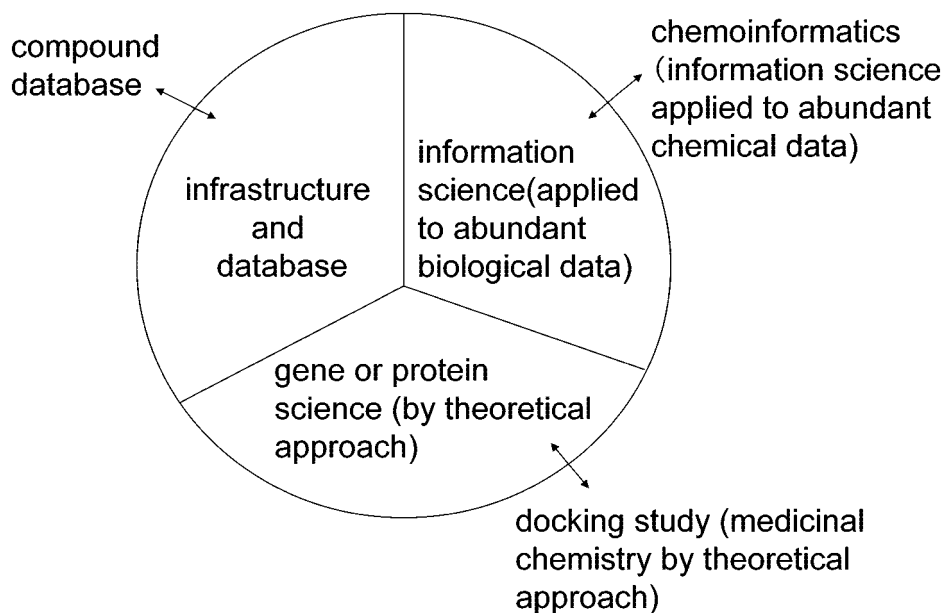


Fig. 2. Interface between Biology and Chemistry Based on *in silico* Technology

インフラ整備の作業は、メモリーの豊富なサーバーや計算能力の高い高速計算機の導入、ネットワークの構築、各種ソフトウェアの導入、有償無償で入手できるデータをダウンロードし、活用するお膳立てをする作業などである。各種トラブルの対応やトラブルを未然に防ぐ手立てを講じることも重要な役割である。データベース構築の作業は、各種生物情報を有効に活用するために必須の要素である。これには実験研究者が利用するデータベース、ソフトウェアの開発やユーザー・フレンドリーなグラフィカル・ユーザー・インターフェイス (GUI) の構築などの作業も含まれる。というのは、比較的簡単な解析や調査は、わざわざバイオインフォマティクスの専門家が行わなくとも実験研究者自身が使えこともあり、そのようなニーズに応える必要があるからである。そのために、実験研究者が正確に適切な情報にアクセスできるよう、またクリックする回数を減らしてストレスなく利用できるような仕組みを構築する作業である。

さて、このような IT としてのバイオインフォマティクスにおいて、化学との融合を考えると、生物学と化学の両方の情報を相互活用できるシステム上あるいはデータベース上の仕組みを構築する作業ということになる。昨今創業統合データベースなるものが多く創られてきているが、これはデータベースによる生物情報と化学情報の融合といえる。例え

ば、遺伝子のアノテーションに関して、以前のバイオインフォマティクス・データベースにおいては、個々の遺伝子に関連する生物情報（遺伝子多型の存在、ゲノム上位置、家系解析などで知られた疾患との関連性、分子機能、発現組織、ホモログ情報など）に限られていたが、最近ではその遺伝子の機能を制御する化合物の存在やアッセイ結果の数値なども含まれるようになりつつある。このような情報が充実してくると、今まで主に生物情報だけを参考にして組み立てていた生物実験に変更をもたらすことができよう。昨今、ハイスループット・スクリーニングの普及などによって、一社内で蓄えているアッセイデータも膨大になっている。折角行われたアッセイ実験のデータは、いずれ別の研究において参考になることもあるから、これら情報を検索し易い仕組みの構築が必要である。ただし、入力されるデータに間違いが多いと、参考になるどころか負のバイアスをかけてしまうため、この仕組みが実際に有効活用されるためには、データの質にこだわる必要がある。さらにはそのために実験やその材料の Quality Control も含めた包括的なソリューションとして捉えておく必要がある。

3-3. 情報学としてのバイオインフォマティクス
理論科学としてのバイオインフォマティクスは情報学としてのバイオインフォマティクスと物質学あるいは生物物理学としてのバイオインフォマティクス

がある。

情報学としてのバイオインフォマティクスとは、すなわち情報学あるいは情報処理学の技術を大量の生物情報へ活用することである。マイクロアレイのデータ解析に代表されるが、エクセルなど実験研究者でも容易に活用できるソフトウェアでは対処しきれない大量のデータ解析を行うのが1つの役割である。少量の場合は実験研究者自身がデータをじっと睨んで何らかの解釈をすることも可能であるが、みるだけで何日も掛かってしまう大量データについては、情報処理技術を利用してそれらの解析の手助けをする必要がある。また、データ・マイニングと総称される知識抽出アルゴリズムを用いて、新たに重要な情報を抽出できる可能性がある。

一方、情報学を大量の化合物データの解析に用いる分野をケモインフォマティクスと呼んでいる。すなわち情報学の技術を使えば、対象が遺伝子であれ、マイクロアレイでの変動であれ、化合物の構造式であれ、“情報”として捉え、数学の定理に基づいて分類、解析していくことができる。創薬は、生物学と化学が絡み合った学際研究分野であることから、バイオインフォマティクスとケモインフォマティクスを組み合わせる創薬インフォマティクスとも呼ぶべき総合的なソリューションの模索が行われているが、情報学の立場でみると、そもそもバイオ、ケモに分類する必要性がないほど本質的に融合しているのである。

3-4. 物質学としてのバイオインフォマティクス

一方、物質学としてのバイオインフォマティクスは、遺伝子や蛋白質を単なる情報ではなく、物理的実体を持った存在として捉え、その物質としての遺伝子や蛋白質を対象にした課題を、手段は実験であれインシリコであれ活用して解決するという分野である。主に立体構造を考慮した高次の配列解析研究と、構造情報自体を対象とした研究に分けられる。遺伝子や蛋白質は物理的実体を持った存在であるので、物理化学の原則に則って解析することで、単に配列“情報”として扱っている場合より詳しい情報の抽出が期待できる。

情報学としてのバイオインフォマティクスが、数学の持つ演繹性を武器にして、生物研究の特性である帰納法的な解析を補完しているのに対して、物質学としてのバイオインフォマティクスは、不変の原

理である物理化学の演繹性を武器にして、やはり純粋に経験論だけからでは得られない情報の抽出を試みる分野といえる。

さて、現在も相同性検出の強化は依然としてバイオインフォマティクスにおける重要な研究課題である。ヒトゲノム配列が決定されたとはいえ、半数を超える遺伝子は機能が未知であり、また機能既知のものであっても、未同定の別途機能を有している可能性がある。機能や構造が未知の蛋白質が、別の機能や構造が既知の蛋白質と相同性の関係にあることを発見すれば、それらと同様の機能や構造を有するものと考えられ、新規な知見の抽出が可能となるのである。また機能が既知のもの同士であっても、その相同性関係が検出されないために、相互の情報が有効に活かせず、よってそれより一歩進んだ解析ができないケースも多い。

近い相同性関係は簡単に検出されて、多くの場合既知情報として整理されてしまっているが、遠い進化的類縁関係については、Fold 認識手法など蛋白質の構造情報を活用した検出方法が感度が高く、有効である。Fold 認識手法とは、ある蛋白質の配列やそのファミリーの配列プロファイルが、立体構造既知のファミリーのプロファイルと類似していることを検出することで遠い相同性関係にあることを見出す手法である。また、立体構造を重ね合わせることで、個別の構造だけからでは得られない重要な残基の情報を抽出することができる。

ドッキング研究は、物理化学の法則に基づいて蛋白質と化合物の相補性を解析する研究であることから、蛋白質においても、化学においても、その物質としての物理化学的性質を把握しておく必要がある。したがって化学との接点という観点では、この分類項目にあるバイオインフォマティクス、すなわち物質学としてのバイオインフォマティクスが、本稿の主題である、ドッキング研究との接点となるのである。

以上創薬におけるバイオインフォマティクス利用の分類と、化学との接点に関して記述した。ただし進展が著しいバイオインフォマティクス分野の、創薬への活用策を鳥瞰することは容易ではなく、ここで述べた分類も、かならずしも永久に成立するものではなかろう。また現時点でさえもっと別の切り口での分類もありえよう。この点に留意しながら、次

節において物質学としてのバイオインフォマティクスの研究例を紹介する。グアニジノ基修飾酵素 (GME) スーパーファミリーを対象にした配列解析, 及び構造情報解析の研究であり, それによってどのような生物学上発見を行えたのかを解説する。またこのような情報抽出が創薬においてどのような有用性があるかについても整理する。

4. GME Superfamily を対象にした配列解析, 構造情報解析研究

4-1. 配列解析による GME スーパーファミリーの発見 物質学としてのバイオインフォマティクスは, 特に蛋白質を対象にした場合, 高次の配列解析と構造情報解析が含まれる。この節では, GME スーパーファミリーを対象にした配列解析, 及び構造情報解析の研究を紹介する。¹⁾

GME は, グアニジノ基若しくはメチル化されたグアニジノ基の修飾を触媒する酵素をメンバーとしたスーパーファミリーであり, 共通の α/β propeller 構造を有する一方, 配列の多様性は極めて高い。構造における類似性や, 機能における類似性から, これらが互いに進化的類縁関係にあることは自

明に近いが, 旧来的な配列解析手法 (構造情報を用いない, 配列相同性検出法) によっては, かならずしも相同性関係が検出できず, また信頼あるアラインメントを作成することができない。なお GME に属する酵素は, 生物学上重要な機能を持ったものが多く, 興味深い研究対象でもある。

アミジノ基転移酵素 (amidino transferase, AT EC2.1.4.1) は, L-アルギニンのグアニジノ基内のアミジノ基を別の基質に転移する酵素であり (Fig. 3), GME の中で最も早く立体構造決定が行われたファミリーである。²⁾ 特にここで紹介する配列解析研究を行った時点では, 唯一立体構造が決定されていた。 α/β propeller 構造は, 五回回転対称性を持ったバレル構造であり, 基質であるアルギニンは, このバレルに嵌まり込んで認識される。立体構造を元に, 触媒残基の同定やメカニズムの提唱も既に行われていた。基質先端のグアニジノ基が酵素のバレルの底に位置する Cys-His-Asp からなる catalytic triad と, 別途 2 つの酸性残基によって認識される。AT において提唱されている触媒メカニズムは 2 つの協調的な求核反応が順次起こるというものであ

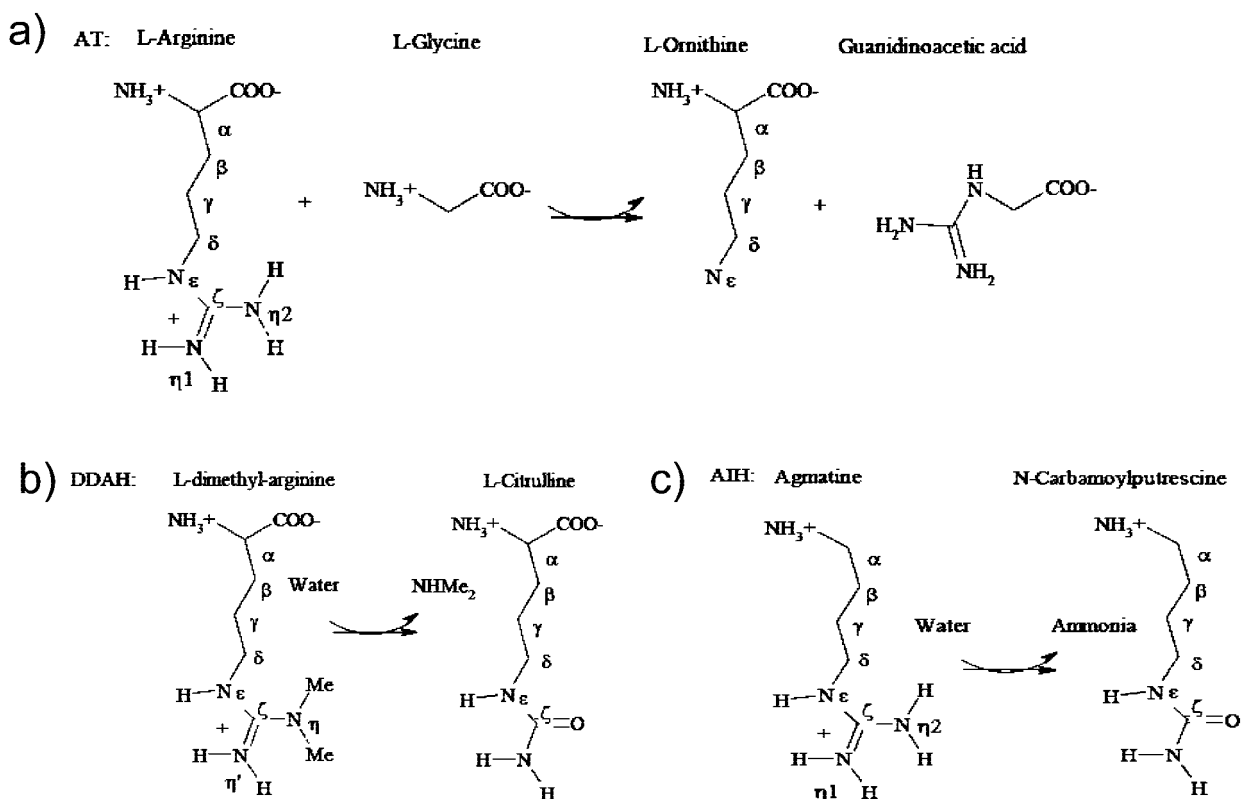


Fig. 3. Molecular Function of a) AT, b) DDAH, and c) AIH

る。哺乳類 AT はアミジノ基がグリシンに受け渡されてクレアチン生合成に利用され、*Streptomyces* においてはリン酸化イノサミンに受け渡されて抗生物質産生に利用される。哺乳類と *Streptomyces* だけに知られているユニークなファミリーであるが、結晶構造からは同じ触媒反応メカニズムを有していることが推察されている。

さて、われわれは Fold 認識手法の 1 つである FUGUE³⁾ (詳細は後述する) を用いた配列解析により、当時構造未知であった 4 種類の酵素が、AT と相同性関係にあり、かつ類似した触媒機構を有していると予測した。これら 4 つとは、アルギニン脱イミノ基酵素 (arginine deiminase, ADI; EC 3.5.3.6), NG,NG-ジメチルアルギニンジメチルアミノ基加水分解酵素 (NG,NG-dimethylarginine dimethylaminohydrolase, DDAH, EC 3.5.3.18), *Porphyromonas gingivalis* 由来ペプチジルアルギニン脱イミノ基酵素 (*Porphyromonas gingivalis* peptidyl-arginine deiminase, PPAD) とその弱い配列相同性を示したホモログ (PPAD homologue, PPADH), 及びサクシニルアルギニン脱水素酵素 (succinylarginine dihydrolase, AstB; EC 3,-,-,-) である。

ADI は、アミノ酸の L-アルギニンのグアニジノ基の脱イミノ基反応を触媒する。病原性微生物において必須としている酵素である一方、ヒトには存在しないので、有望な抗菌薬標的として知られている。また ADI 蛋白自体が抗癌活性を有していることも知られており、バイオ医薬品としての期待も高まっている。

DDAH は、NG,NG-dimethylarginine dimethyl L-arginine (asymmetric dimethyl L-arginine, ADMA) を加水分解する酵素である。ADMA は一酸化窒素合成酵素 (nitric oxide synthase; NOS) の競合的な阻害剤であり、DDAH は高血圧の標的分子となっている。DDAH と ADI の配列は弱い相同性関係が以前より検出されていた。

PPAD はペプチドの C 末端のアルギニン残基のグアニジノ基から脱イミノ基反応を触媒する酵素であり、アンモニアを発生する。発生するアンモニアにより、口腔内の酸クレンジング・サイクルから中和により菌の生育を守る役割を果たしていると推測されている。すなわち歯周病菌の病原因子と考えられており、創薬ターゲットである。なお 556 残基か

らなる PPAD の N 末 380 残基からなるドメインと弱い相同性がみられる機能未知のファミリーが存在していたため、これを PPADH と呼ぶことにした。

AstB は病原性株を含む大腸菌の主たるアルギニン代謝経路 (arginine succinyl transferase (AST) 経路) 上の分子であり、病原性大腸菌感染症の治療薬標的として有望視されている分子である。

ADI と DDAH, PPAD と PPADH はそれぞれ互いに弱い相同性関係が検出されていたが、それ以外の GME 間では全く相同性関係が見出されていなかった。単にバイオインフォマティクスとしての配列解析ツールが検出できなかったという点だけでなく、いずれもアルギニンか修飾したアルギニンを基質とするという機能上の類似性がある (Fig. 3) にも係わらず、見出せなかったという点が重要であろう。ともあれ、FUGUE を用いて、これら構造未知の酵素の配列が、AT と同様の構造を取っている可能性が高いことを検出し、GME スーパーファミリーの発見となった。

PSI-BLAST は、BLAST により得た多重配列からプロファイルを作成し、そのプロファイルを問い合わせにして再度相同性検索を行うというステップを繰り返すことにより雪達磨式に相同配列を収集する手法であるが、FUGUE は、この PSI-BLAST を利用し、かつ立体構造情報を利用することで検出能を高めている。FUGUE では、ファミリー毎に立体構造情報を考慮したプロファイルが準備されている。そして問い合わせ配列に対し、PSI-BLAST で収集した相同配列から作成された多重整列データを、各ファミリーのプロファイルと比較し、適合度を計算する。適合度は異なる特徴や性質を持つ数値を比較するときによく利用される、Z スコアと呼ばれる統計的数値で評価される。この値が大きいくほど平均値からのずれが大きいく、すなわち偶然ではない相同性検出が行われた可能性が高いと解釈される。あらかじめ立体構造既知の配列を利用して試行された結果から、信憑性ある検出であることを示す Z スコアの閾値を求めている。FUGUE により、ADI, DDAH, PPAD, PPADH, AstB がすべて AT と同様の構造を有し、互いに遠い相同性関係にあると予測した。すなわち、同一スーパーファミリーに帰属される関係にあり、当時 PPADH を除いて機能が既知であったが、いずれもグアニジノ基を修飾

するのを触媒する酵素であったため、guanidino-modifying-enzymes (GME) スーパーファミリーと命名した。PPADH に関しては、他の GME メンバーの機能の類似性から、やはり同様の機能（すなわち、アルギニンかそれに類似した化合物を基質とし、アンモニアなどを発生させる酵素機能）を持つと予測した。

検出にあたっては、単に Z スコアが高いというだけではなく、次のいくつかの「構造上の観点」で矛盾がないことを根拠にこの検出に確信が持つに至った。まず、構造上の重要な残基（すなわち埋もれた疎水性残基、Arg169, His406, Thr409 といった構造上重要な水素結合を有する残基）が保存されている点、次に欠失や挿入はループ領域に集中している点、さらに指紋領域ともいえる C 末付近の 3 残基連続の Gly（構造上は蛋白質の中央を貫通するもので、小ささが要求されている）が保存されている点、そして最後に立体構造モデルが不具合なく構築できた、という 4 点である。すなわちこれら物理化学的な見地に立ったインシリコ上での検証を行い、確かな予測と断定した訳である。

さて、PPADH は、その後アグマチン加水分解酵素 (agmatine iminohydrolase; AIH) という、アルギニンに類似した化合物であるアグマチンを基質とした酵素であることが分かり、⁴⁾ 上記の機能予測が正しいことが検証された。

そして、DDAH, ADI, AIH, AstB の立体構造が決定され、その結果、これらが AT と同じ α/β propeller 構造を取るという予測の正しいことが検証された。^{5,6)}

4-2. GME スーパーファミリーの発見により得られたもの 次にこの発見によって、どのような生物学上の新規知見が抽出・発見できたのかを項目として整理した。そして各項目について、創薬における利点を記述した。

4-2-1. ホモログの存在 ホモログが存在していることを知ること自体が、重要な生物情報であり、特に分子進化を考える際には、最も基本的で重要な情報となる。

創薬における利点としては、ホモログが検出されていなかった場合に比べて、より実験データの解釈が容易になったり、また選択性をみるべき対象が明確にできたりする点が挙げられる。例えば、GME

においては、多くが創薬標的蛋白質として有望とされているが、AT のホモログであることから、副作用が懸念される。すなわち、創薬においては、単に標的蛋白質への親和性だけでなく、副作用を引き起こすであろう別の蛋白質との選択性も高い必要がある。もしホモログの関係にあること自体が分かっているならば、互いに選択性をみるべき対象として認識できない可能性がある。逆にホモログ分子が分かり、それとの選択性が必要であることが分かれば、最初から標的分子だけでなくそのホモログとの選択性を考慮したアッセイ系を組む必要があるし、また選択性を考慮した化合物の設計や合成が可能である。したがって、ホモログの検出は、創薬において生物、化学実験ともに大きく影響を与える重要な情報である。

4-2-2. 構造情報 構造情報については、AT は既知で、その他の PPAD, PPADH (AIH), ADI, DDAH, AstB では未知であったが、後者のすべてにおいて構造予測を行うことができるようになった。構造情報は、通常は実験的手法により比較的長い年月と高いコストを掛けて得られる貴重な情報である。インシリコで構造が予測できることは、構造決定に比べて、それに掛かるコスト的削減、期間短縮が行えることを意味する。当然ながら実験的手法との違いは正確さが担保されていないという極めて重大なものであるため、情報の質の限界をよく把握した上で活用する必要がある。この発見による創薬における利点としては、主題であるドッキング研究を行う際には標的蛋白質の構造情報がないと始まらないことから、ドッキング研究が不可能であるところを可能にした、という点が挙げられる。

4-2-3. 活性残基 活性残基については、AT は既知で、その他の PPAD, PPADH (AIH), ADI, DDAH, AstB では未知であったが、多重配列アラインメントを作成してみると AT の触媒残基 (Asp170 Asp254, His303, Asp305 and Cys407) は構造保存領域に存在しており、かつ基本的に一致することから、これらにおいても触媒残基が推定できた。一致しない箇所については、あとの節で紹介するが、それが分子機能の差異に繋がっていることが合理的に推測できた。創薬における利点としては、PPAD, ADI, DDAH, AstB などはいずれも重要な機能を持った分子であるが、その触媒残基を把握で

きたことで、実験を行う際に触媒残基の変異体を作成してその医薬品標的としての確かさを検証することなどが可能とすることができた。

4-2-4. 活性ドメイン 活性ドメインについては、ATは既知で、その他のPPAD, PPADH(AIH), ADI, DDAH, AstBでは未知であったが、今回、ATと同一構造を取る領域が同定されたので、その領域を活性ドメインと定義できた。創薬における利点としては、スクリーニング系構築などにおいて蛋白質を調製する際に、蛋白質全長をすんなりと精製できればよいが、かならずしもそうではなく、部分的に調製せざるを得ないときなどに、どこが機能上必須のドメインであるかを把握しておくことで、効果的な調製を可能とすることができた点である。

4-2-5. 触媒機構 触媒機構については、ATが既に提唱されていたが、その他のPPAD, PPADH(AIH), ADI, DDAH, AstBでは提唱されていなかった。ATにおいて提唱されている触媒メカニズムは2つの協調的な求核反応が順次起こるというものである。分子機能の類似性から、同様な多段階の協調的な求核反応によると提唱した。創薬においては、化合物を最適化する段階でその触媒機構を参考にした効率的な設計が行われることから、化合物設計に用いられる有用な情報を獲得できたことになる。

4-2-6. 分子機能 分子機能については、AT, PPAD, ADI, DDAH, AstBは既知であったが、PPADH(AIH)は未知であった。しかし今回、AIHがGMEスーパーファミリーのメンバーであることが分かり、このことからアルギニンかそれに類似した化合物を基質とした酵素であるという次元の予測を行うことができた。もし加水分解するならばアンモニアが発生する可能性が高いことから、*Helicobacter pylori* (胃炎の原因となる細菌)由来のAIHは病原因子の可能性があると予測した。*Helicobacter pylori*の病原因子の1つとしてウレアーゼという酵素が有名であるが、この酵素は菌体内のウレアを分解してアンモニアを発生させ、それにより胃の強酸性を中和しコロニー生育を守るのに利用されている。AIHも胃に豊富に存在するアルギニンを分解しアンモニアを発生させることが可能であるため、ウレアーゼと同じ理由で新規創薬ターゲットであると推察した訳である。

事実AIHは、アグマチンを加水分解することでアンモニアを発生するので、この仮説は依然成立しており、実験による検証を待っている段階ではある。

以上のように、蛋白質の構造情報を活用した配列解析研究によって、新規知見を抽出でき、創薬標的候補の発見に加え、創薬プロセス全体にまたがって有用な情報を抽出できることが分かった。

4-3. 構造情報解析によるGMEスーパーファミリーの機能上重要な残基の抽出 以上のように、配列解析を用いることでGMEスーパーファミリーを発見し、またそれが実験により正しいと検証されたことを紹介した。一方上では述べなかったが、GMEにおいては基質はアルギニンかそれに類似した化合物であり、触媒部位の残基が微妙に異なることで、精密にGME内の化学反応の差異を出すことができていたことが分かった。

では次に、これらの立体構造が実験的に決定されてしまうと、それ以上にバイオインフォマティクスの技術を活用しても有用な情報を抽出できないのであろうか。その回答が、2つ目の研究としての構造情報解析にある。実験的に決定された構造を重ね合わせてみたところ、それぞれの構造決定された論文においては、さほど重要視されていなかった残基が新たに重要であることを見出すことができた。重ね合わせの結果、基質アルギニン中のカルボキシル基が、かならず酵素のアルギニン残基に認識されているという共通性があり、しかも触媒部位付近のアミノ酸の構成に加えて、そのアルギニンの位置に起因してGMEの酵素機能がさらに詳細に分類可能となった。すなわち、GMEの配列から、GMEの酵素機能を分類するルールを作成することができた。現在、GMEに属するとみられる配列は、明らかにATのメンバーであるものがADIなどと注釈されたりして混乱してしまっているが、この分類法を利用することで、GMEスーパーファミリー内の機能をより詳細に検出できることが分かった。

一方創薬における有用性としては、まずは上記のようなルールを見出したことで、新たな機能予測、そして新規標的発見へとつながる点である。また、ホモログとの選択性をみるのに重要な残基を抽出できた点である。今回、GMEにおいて、活性中心とその付近の酸性残基だけでなく、基質カルボキシル基を認識するアルギニンが選択的に重要であること

を見出したが、それによって選択性向上の指針を立てるのに有用な情報を提供することができた。

5. 創薬インフォマティクスへの展望

GME に関する一連の研究から、配列解析や構造情報解析により、生物学上の新規知見だけでなく、創薬における有用な情報も抽出できることを述べた。それではバイオインフォマティクスがドッキング研究にどう貢献できるかという観点で整理し考察してみた (Fig. 4)。まず、ホモログ情報の提供である。ドッキング研究、あるいはそれを含めた化合物創製という目的において、親和性だけでなく選択性の高い化合物を取得するのに利用価値が大きい。次に構造情報の提供である。ドッキング研究を行うには、まず蛋白質の構造情報が必要である。このため、配列解析によって構造予測が行え、それによってドッキング研究が成立することの重要性は自明といえる。第三に、親和性や選択性に重要な残基の情報の提供である。構造情報が既に入手できる状態であっても、それらを重ね合わせるなどして解析することで、単一の構造を眺めていただけではみられなかった重要な残基の役割が明確になることがある。そのような残基の重要性を認識した上でドッキング研究を行うことは精度向上に重要である。

今後の創薬を展望すると、これら2つの貢献に加えて、通常注目している活性中心の部位だけでなく、アロステリック部位 (活性中心ではないが化合

物が結合することで蛋白質の機能が亢進若しくは制御されるような部位) における結合化合物を見出すことで、創薬機会を向上させることが考えられる。そのために、バイオインフォマティクスによりアロステリック部位を見出すアルゴリズム開発なども重要になろう。

では逆にドッキング研究がバイオインフォマティクスに貢献できることはあるだろうか。ヒトゲノム配列が決定されたが、機能未知の配列が約半数存在しているため、それらの機能を予測し、解析することは極めて重要である。Fold 認識手法などを使って分子機能の種類までは予測できることが可能である (GME の例では PPADH (AIH) の分子機能として、アルギニンに類似した基質を加水分解してアンモニアを発生するだろう、という予測できたことがこれに該当する) が、リガンドの予測までは困難であった (PPADH (AIH) の基質がアグマチンであるという予測には至らなかった)。しかし、アグマチンはよく知られた細胞内に存在する化合物であり、既に生体内化合物データベース LIGAND⁷⁾ にも入っている。今後は LIGAND に登録されている化合物を対象にドッキングするなどして、生体内リガンドの予測を行うことも有効であろう。

また、薬効や副作用があるがその標的分子がみつからない医薬品や化合物が多く存在している。種々の実験的手法によって、その標的分子を見出す

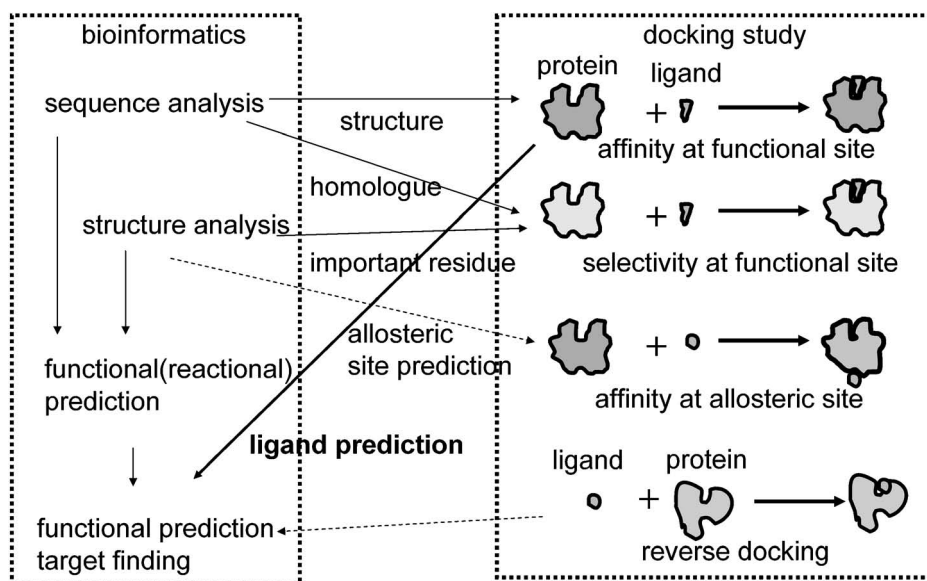


Fig. 4. Interface between Bioinformatics and Docking Study

試みが行われているが、このようなアプローチにおいて、リバースドッキング研究とでも言うべき、化合物の蛋白質結合部位を見出す研究も重要になってこよう。

このようにみると、生物研究における発見を目的として活用されているバイオインフォマティクスの発明研究への有効性が展望できたり、逆に化学研究における発明を目的として活用されているドッキング研究の発見研究への有効性が展望できた。またアロステリック部位予測やリバースドッキング研究などが実用化されれば、これらがさらに融合した総合的なソリューションとしての創薬インフォマティクスなるものが展望できてこよう。

謝辞 GME に関する一連の研究は、水口賢司博士（ケンブリッジ大・生化学部門）との共同研究である。創薬インフォマティクスへの展望に関しては、平山令明教授（東海大学医学部）、曾我真司氏（アステラス製薬）との日頃の議論を参考にまとめたものである。また IT としてのバイオインフォマティクスについては、水野克也氏（アステラス製薬）、寺地知帆氏（銜レバルファイブ）との日頃の

議論を参考にまとめたものである。さらに情報学としてのバイオインフォマティクスについては角山和久博士（アステラス製薬）との日頃の議論を参考にまとめたものである。この場で深謝申し上げたい。

REFERENCES

- 1) Shirai H., Blundell T. L., Mizuguchi K., *Trends Biochem. Sci.*, **26**, 465–468 (2001).
- 2) Humm A., Fritsche E., Steinbacher S., Huber R., *EMBO J.*, **16**, 3373–3385 (1997).
- 3) Shi J., Blundell T. L., Mizuguchi K., *J. Mol. Biol.*, **310**, 243–257 (2001).
- 4) Nakada Y., Itoh Y., *Microbiology*, **149**, 707–714 (2003).
- 5) Das K., Butler G. H., Kwiatkowski V., Clark Jr. A. D., Yadav P., Arnold E., *Structure*, **12**, 657–667 (2004).
- 6) Murray-Rust J., Leiper J., McAlister M., Phelan J., Tilley S., Santa-Maria J., Vallance P., McDonald N., *Nat. Struct. Biol.*, **8**, 679–683 (2001).
- 7) Goto S., Nishioka T., Kanehisa M., *Nucleic Acids Res.*, **28**, 380–382 (2000).