

PNE
PROTEIN,
NUCLEIC ACID
AND ENZYME

蛋白質 核酸 酵素

別刷

「蛋白質 核酸 酵素」編集部

共立出版株式会社

〒112-8700 東京都文京区小日向 4-6-19

Tel.03-3947-2515 FAX 03-3944-8182

E-mail : pne@kyoritsu-pub.co.jp

<http://www.kyoritsu-pub.co.jp/>

抗SARS薬開発におけるウイルス蛋白質の立体構造予測

竹田-志鷹真由子・梅山秀明

創薬において、標的蛋白質の機能を解析し、効率よくドラッグデザインを行なうには、ゲノム情報のみならず蛋白質の立体構造情報が必要不可欠である。しかしながら、蛋白質の多くは立体構造がわかっていないのが現状である。そのような場合、コンピュータを用いたホモロジーモデリングという手法で、蛋白質の立体構造を予測することができる。蛋白質立体構造予測の国際コンテスト (CASP) が隔年に開催されていることから、その手法の開発および応用への国際的な関心の高さがうかがえる。本稿では、蛋白質立体構造予測について概説し、重症急性呼吸器症候群 (SARS) の治療薬開発においてホモロジーモデリングの手法が用いられ、注目されたことについて紹介する。

▶▶KEY WORDS : SARS コロナウイルス 蛋白質立体構造予測 ホモロジーモデリング

■はじめに■

2002 年末から 2003 年前半にかけて、重症急性呼吸器症候群 (severe acute respiratory syndrome ; SARS) がアジアを中心に猛威をふるった。世界保健機構 (WHO) の報告によると、半年ほどの間に 8,000 人以上の感染者を出し、そのうち 800 人近くが死亡した。SARS の原因ウイルスは新種のコロナウイルス (SARS-CoV) であった。SARS 発生から 4 年以上が経ったいま、世界中の研究者の努力により SARS-CoV の受容体や細胞侵入機構が解明されつつある。すなわち、SARS-CoV の受容体が ACE2 (angiotensin converting enzyme 2) であること¹⁾、あるいは細胞表面のスパイク構造を形成する S 蛋白質の細胞膜融合能活性化がゲノムの細胞侵入に重要な役割を果たしていること²⁾ などが明らかになり、治療薬やワクチン開発の標的となる分子の選択肢は広がっている。しかしながら、2003 年に SARS-CoV の全ゲノム配列が決定された直後、抗 SARS 薬開発の標的分子として真っ先に注目されたのは、SARS-CoV main protease (M^{pro}) であった。というのも、当然のことながら、ウイルス蛋白質の立体構造は当時まだ実験的に解かれていなかったが、SARS-CoV M^{pro} に関してはコンピュータを用いたホモロジーモデリ

ングという手法で立体構造予測が可能であり、予測した構造情報を基に治療薬開発が大いに加速されることが期待されたからである。

ここで、筆者らの専門である蛋白質立体構造予測の創薬研究における位置づけについて簡単に説明したい。創薬において、ターゲットとなる蛋白質の探索や新規化合物の探索などのさまざまな段階において *in silico* 技術が重要な役割を果たしているが、コンピュータを用いた蛋白質立体構造予測法もその重要な技術の 1 つである。ターゲットとなる蛋白質の機能を解析し、効率よくドラッグデザインを行なうには、遺伝子情報のみならず蛋白質の立体構造情報が必要不可欠である。しかしながら莫大な遺伝子情報に比べて、実験的に立体構造が決定されている蛋白質の数は少なく、アミノ酸配列がわかっているにもかかわらず立体構造はわからないターゲット蛋白質が多数存在しているのが現状である。そのような場合、コンピュータを用いた手法でその立体構造を予測することができる。実験的な制約などにより実験構造を得るのがむずかしい場合、予測構造を用いることにより創薬研究が加速されるのである。蛋白質立体構造予測の国際コンテスト (Critical Assessment of Techniques for Protein

Mayuko Takeda-Shitaka, Hideaki Umeyama, 北里大学薬学部 E-mail : shitakam@pharm.kitasato-u.ac.jp, umeyamah@pharm.kitasato-u.ac.jp

Three-dimensional structure prediction of the viral protein in the development of anti SARS drugs

Structure Prediction ; CASP) が隔年に開催されていることから、その手法の開発および応用への国際的な関心の高さがうかがえる (<http://predictioncenter.org/>)。

本稿では、蛋白質立体構造予測の手法のなかでも、現在最も精度がよいとされているホモロジーモデリングに焦点をあて、まずはじめに手法の概略、および CASP6 (2004 年開催)、CASP7 (2006 年開催) について紹介する。次に抗 SARS 薬開発における SARS-CoV M^{pro} のホモロジーモデリングについて紹介する。

I. 蛋白質立体構造予測

1. ホモロジーモデリング

ホモロジーモデリングとは、立体構造未知の蛋白質 (目的蛋白質) の立体構造を、類似の配列をもつ立体構造既知の蛋白質 (参照蛋白質) を参照して予測する手法であり、コンピュータを用いた蛋白質立体構造予測法のなかで、現在最も精度がよいとされている。ホモロジーモデリングは通常 2 つのステップよりなっている。

(1) 参照蛋白質の選定および目的蛋白質と参照蛋白質間のアライメント作成

(2) アライメントに基づいたモデル構造の構築

モデルの信頼性は、(1) および (2) の両方のステップに依存するため、どちらのステップも重要である。まず、(1) のステップで現在使われている代表的なプログラムは、PSI-BLAST³⁾ などの類似性検索プログラムである。PSI-BLAST により、参照蛋白質の候補および目的蛋白質とそれら候補間のアライメントが得られる。PSI-BLAST から出力されたアライメントは、精度を上げるために実験情報などを基に微調整されることもある。(2) のステップに対しては、世界中で各種ホモロジーモデリングソフトが開発されている。たとえば、筆者らはホモロジーモデリングソフト CHIMERA⁴⁻⁸⁾ と FAMS⁵⁻⁹⁾ を開発した。CHIMERA は目的に応じて対話的に研究者の工夫を入れながらモデリングできるソフトであり、公に得られる情報のみならず研究者がもつ独自の情報、知識などをモデルに反映させることにより、精度の高いアライメントやモデル構造を構築することができる。しかしながら得られるモデルの信頼性は、研究者の知識や経験も反映されるため、蛋白質立体構造に関する知識やモデリングに対する経験が必要である。一方 FAMS は、CHIMERA

の手法を自動化することを目指して開発された全自動ホモロジーモデリングソフトであり、アライメントを入力として全工程全自動でモデル構造を構築する。ゲノム上にコードされている蛋白質の網羅的な (大規模な) モデリングを行なう場合などにも適している。筆者らは、CHIMERA と FAMS の開発において、主鎖構造のみならず、側鎖構造も精度よく予測できることを目指した。蛋白質の機能、リガンドとの相互作用などを考える際に、側鎖のコンフォメーションは非常に重要なものであり、モデルを活用できるかどうかは側鎖の精度に大きく依存すると考えられるからである。

2. 蛋白質立体構造予測の国際コンテスト (CASP6, CASP7)

コンピュータで予測されたモデルを実際に研究で用いる場合には、モデルの精度を知っておくことは重要である。一般的には、モデルの精度はおもに目的蛋白質と参照蛋白質間のホモロジーに依存し、ホモロジーが高いほど精度がよくなる。ホモロジーが低い場合には、参照蛋白質の選定やアライメントのエラー、モデル構築のエラーなどが増えるため、モデルの精度が下がる傾向にある。また、同じ蛋白質をモデリングするにしても、用いる方法によって算出されるモデル構造が異なるため、方法の精度を知ることは重要である。精度の高いモデルを構築するうえでは、たとえば隔年に開催される蛋白質立体構造予測の国際コンテスト CASP で優秀な成績を収めている参加者のモデリング手法が大変参考になる。CASP は 1994 年に CASP1 として始まり、最近では 2006 年に CASP7 が開催された。CASP では、実験的に立体構造が解明 (公開) される直前の蛋白質のアミノ酸配列がそのホームページ上に出題され、それに対し参加者は蛋白質立体構造を予測し、指定された期間内にインターネットを介してそれぞれの予測構造を提出する。つまり正解となる立体構造が完全に伏せられた状態で行なう蛋白質立体構造予測のブラインドコンテストである。CASP において世界トップレベルの成績を収めたグループは、主催者から招待を受け、結果発表会において招待講演を行なうことになっている。

CASP への参加登録には server と human predictor の 2 区分ある。server として参加する場合は、人間の介入のない完全に自動化された方法でモデリングを行なわなくてはならないという制約があり、モデリングソフトその

ものの性能が試される。予測構造の提出の締め切りは、出題されてから48時間以内である。一方、human predictorとして参加する場合は、server区分のような方法の制約はなく、立体構造未知の蛋白質に対して立体構造をどれだけ正確に予測することができるか、いわば研究者の蛋白質立体構造予測に対する総合力を競うものであり、さまざまな手法を駆使して参加者は予測を行なう。human predictor区分の予測構造の提出の締め切りはserver区分よりも遅く設定されている。

筆者らのグループは、2004年に開催されたCASP6 (<http://predictioncenter.org/casp6/>) および2006年に開催されたCASP7 (<http://predictioncenter.org/casp7/>)に参加し、2回連続で招待講演を行なうことができ、筆者らの開発した手法の精度が世界トップレベルであることが実証された。まずCASP6についてであるが、予測期間は3カ月、出題は90ドメイン、参加グループは250をこえた。CASP6の結果については、学術誌 *Proteins* にCASP6特集号が組まれ、主催者側の審査報告および招待講演を行なったグループの参加報告が掲載されている¹⁰⁾。筆者らは、独自に開発した2つのシステムCHIMERAとFAMSを統合したウェブユーザーインターフェースシステムSKE-CHIMERAを用いて、CASP6のhuman predictor区分に参加した(グループ名:CHIMERA)。SKE-CHIMERAは、モデリングの際に参考になるさまざまな情報を自動構築してデータベース化し、それらの情報に必要に応じて研究者の工夫を取り込んでアライメントを構築(CHIMERA)、そのアライメントを基に全自動でモデルを構築(FAMS)、そのモデルを基にさらにアライメントを修正(CHIMERA)、それらをくり返しながら精度のよいモデルを構築するシステムである。筆者らは、イタリアで開かれたCASP6の結果発表会のHomology Based Modeling部門において上位入賞し、招待講演を行なうことができ(講演者:竹田志鷹真由子)、本方法の有用性が実証された(http://predictioncenter.org/casp6/meeting/presentations/CASP6_Program.doc)⁶⁾。

次にCASP7についてであるが、予測期間は3カ月、出題は124ドメイン、参加グループは300をこえた。筆者らは、CASP6後に新たに独自開発した蛋白質立体構造評価プログラムCIRCLE¹¹⁾を用いてCASP7のhuman predictor区分に参加した(グループ名:FAMS-ACE)。CIRCLEとは、埋没面積率と極性面積率によって決定されるアミノ酸残基の側鎖の環境を評価する関数(3D1Dscore)

と、2次構造予測と予測構造の2次構造の一致度をスコア化した関数(SSscore)により構造を評価するものである。CASP7の詳細な結果については学術誌 *Proteins* にCASP7特集号が組まれる予定である。FAMS-ACEは今回新たに開発したメタサーバー(メタセクター)である。CASPにおいては、前述したように出題してから48時間後にserver区分で参加しているグループの予測構造の提出が締め切られるが、それらの予測構造は即座にホームページで公開される。メタサーバーというのは、それらの公開された多数の予測構造情報を使って立体構造を予測する全自動サーバーのことである。FAMS-ACEはCIRCLEおよびConsensus¹²⁾のスコアを指標として公開された予測構造情報を基にFAMSで再モデリングした多数のモデル候補のなかから最高スコアのモデルを選び出す全自動メタサーバーである。アメリカで行なわれたCASP7の結果発表会では、Template Based Modeling部門において上位入賞し招待講演を行なうことができ(http://predictioncenter.org/casp7/meeting_docs/meeting_program.html)、CIRCLEとConsensusの組合せが構造評価法として有用であることが示された(講演者:梅山秀明)。

実際の創薬研究において求められる精度の高いモデルを構築するためには、最近のCASP7の結果から考察すると、現時点では、FAMS-ACEなどのメタサーバーを用いることにより多くのサーバーの情報を効果的に取り込み、必要に応じてSKE-CHIMERAなどの方法で研究者独自の情報や公に得られる情報などを反映させてモデルを構築することで、よりよいモデルが構築できると可能性が示唆される。筆者らは、人間が介入して行っていた手法を基に自動化できる部分はさらなる自動化を進め、迅速で精度のよい全自動モデリング法の開発を続けたいと考えている。

II. SARS-CoV 蛋白質の立体構造予測

SARSは2002年末にアジアで出現した。世界的な注目の中、SARS-CoVの全ゲノム配列は異例のスピードで解析・同定され、2003年5月1日には2つの研究グループがそれぞれ結果を学術誌 *Science* (電子版)に発表した^{13,14)}。ゲノム解読が終わった段階で、まず抗SARS薬の標的分子として注目されたのはSARS-CoVの複製を制御しているSARS-CoV M^{pro}であった。一般的に、ウイルスの蛋白質分解酵素は、ウイルスゲノムから翻訳された

前駆体蛋白質を個々の機能蛋白質に分けるプロセッシングに重要であり、ウイルスの複製過程において不可欠なものである。したがって、複数のウイルスにおいて蛋白質分解酵素は抗ウイルス薬の有力な標的になっている。抗SARS薬開発を急ぐためにもその立体構造は必要不可欠であったが、当然のことながら実験構造はまだ解かれていなかった。SARS-CoV M^{pro}の場合、Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/>)¹⁵⁾に登録されているTGEV (transmissible gastroenteritis coronavirus) M^{pro}のX線構造 (PDB ID: 1LVO)を参照蛋白質として用いることにより、ホモロジーモデリングで立体構造を予測することが可能であったため (SARS-CoV M^{pro}とTGEV M^{pro}の間のホモロジーは44.5%)、おそらく多くの研究グループがSARS-CoVのゲノム情報とTGEV M^{pro}のX線構造情報を基にSARS-CoV M^{pro}のモデルを構築し、自らの抗SARS薬開発に役立てたに違いない。研究グループのなかには、予測したSARS-CoV M^{pro}の立体構造モデルを抗SARS薬開発に広く活用してもらうために、座標を公開するグループも複数みられ、ホモロジーモデリングという手法が注目を集めた。座標公開の日付の早かった順に2グループ紹介する。

まず筆者らのグループであるが、参照蛋白質であるTGEV M^{pro}がホモダイマーの構造をとっていることから、独自に開発した蛋白質複合体モデリングソフトFAMS Complex¹⁶⁾を用いてSARS-CoV M^{pro}のホモダイマーの構造を予測した。FAMS Complexは立体構造既知の蛋白質複合体構造 (参照蛋白質)を基に、立体構造未知の蛋白質複合体構造 (目的蛋白質)をホモロジーモデリングするソフトである。FAMS Complexは単体蛋白質モデリングソフトFAMSを基に開発されたもので、FAMSの精度の高さを保持したまま複合体をモデリングできるように開発されている。モデルの座標公開は2003年5月9日であったが (<http://www.pd-fams.com/>)^{7,17)}、筆者の知る限りでは世界で最も早い公開であった。多くの研究者が、実験構造が解かれるのを待たずに、筆者らのモデルを用いて抗SARS薬開発を開始した。次に座標を公開したのはAnandらのグループである。彼らは、参照蛋白質であるTGEV M^{pro}のX線解析を行なったグループである。彼らは、新たにHCoV 229E M^{pro} (human coronavirus strain 229E M^{pro})のX線構造を解析し、TGEV M^{pro}とHCoV 229E M^{pro}の両方のX線構造を参照してモノマーの構造を予測した。彼らのモデルは2003年5月13日に学術誌

Science (電子版)に掲載され¹⁸⁾、座標は2003年5月20日にPDBにて公開された (PDB ID: 1P9T)。

世界中でホモロジーモデルを用いた抗SARS薬開発が進む中、2003年7月29日に、SARS-CoV M^{pro}のX線構造がPDBより公開された (PDB ID: 1Q2W)。そこで、モデルの精度を

表1 SARS-CoV M^{pro}のモデルとX線構造の重ね合わせにおけるRMSD

重ね合わせた残基	RMSD (Å)			
	Cα原子		全原子	
	筆者らのモデル	1P9T	筆者らのモデル	1P9T
活性部位	0.78	1.24	1.42	1.66
活性部位ドメイン (ドメインIとII)	1.16	1.99	1.89	2.55
αヘリックスドメイン (ドメインIII)	1.68	2.75	2.67	3.60
全残基 (モノマー)	2.54	3.48	3.10	3.98
全残基 (ホモダイマー)	2.82	—	3.31	—

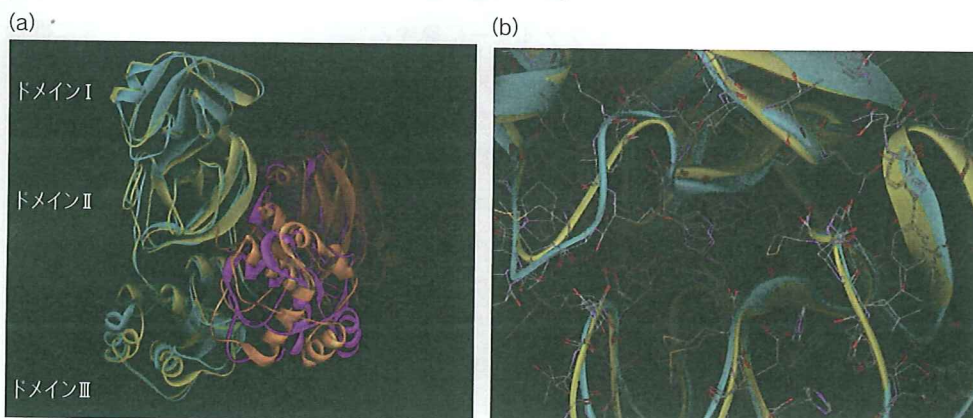


図1 SARS-CoV M^{pro}のモデルとX線構造の重ね合わせ¹⁷⁾

モデル構造: 水色と赤紫, X線構造: 黄色とオレンジ. (a) SARS-CoV M^{pro}のホモダイマー構造. (b) SARS-CoV M^{pro}の活性部位近傍.

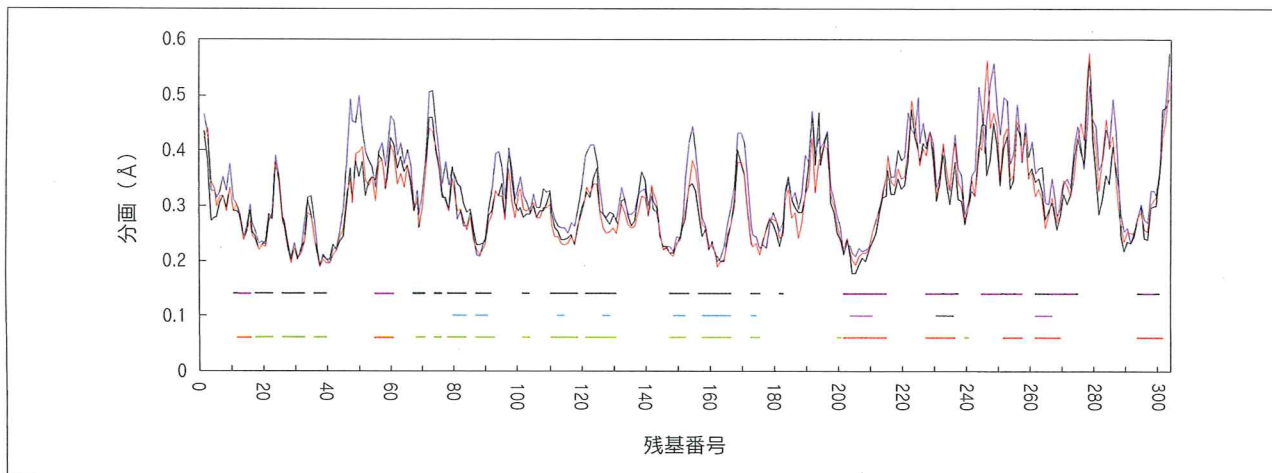


図2 SARS-CoV M^{pro} のモデルと X 線構造の C α 原子のゆらぎ¹⁷⁾

筆者らのモデル構造：赤，1P9T：青，X 線構造：黒。横棒は 2 次構造を形成している箇所を示している。筆者らのモデル構造の α ヘリックス：赤紫，1P9T の α ヘリックス：薄紫，X 線構造の α ヘリックス：オレンジ，筆者らのモデル構造の β シート：緑，1P9T の β シート：水色，X 線構造の β シート：黄緑。

検証するために、筆者らのモデル構造と X 線構造の比較を行なった¹⁷⁾。まず、X 線構造にモデル構造を重ね合わせて、構造の類似性を評価したが、図 1 と表 1 に示すとおり非常によく予測できていたことが実証された。予測どおりに SARS-CoV M^{pro} の X 線構造はホモダイマーを形成しており、それぞれのモノマー分子は 3 つのドメインからなっていた (図 1)。ドメイン I、II は β バレルドメイン (活性部位ドメイン) であり、活性部位はドメイン I と II の間に位置している。ドメイン III は α ヘリックスドメインである。表 1 に X 線構造とモデル構造の間の RMSD (root mean square deviation) をまとめてあるが、とくに活性部位では小さな値となっており、創薬に最も重要な活性部位が精度よく予測されていたことが実証された。比較のため X 線構造と 1P9T の間の RMSD も表 1 に示したが、筆者らのモデルのほうが精度よく予測できていたことが示された。

次に、基準振動解析により、モデル構造と X 線構造の動的構造の比較を行なった。図 2 に示すとおり、筆者らのモデルは X 線構造と似た動的構造を示しており、蛋白質のゆらぎを考慮した蛋白質-リガンド相互作用の解析などにも本モデルが十分有用であったことが実証された。比較のため 1P9T の動的構造も図 2 に示したが、2 次構造のくずれもありゆらぎが大きく、X 線構造のそれを再現できていない場所もあった。

以上の検証より、筆者らのモデルは X 線構造と比べて

も遜色なく創薬に用いることができる程度に精度よく予測できていたことが実証された。さらに、筆者らの手法は迅速に立体構造を予測することが可能であり、X 線構造が解かれる前において抗 SARS 薬開発を加速するために十分有用であったと考察される。

■おわりに■

創薬のターゲットとなる蛋白質の多くは立体構造がわかっていないのが現状であり、ホモロジーモデリングは創薬において欠かすことのできない手法である。本稿では SARS-CoV M^{pro} のモデリングを紹介したが、精度よくかつ迅速に予測構造を提供できる本手法は、たとえば新型インフルエンザなどの新しい感染症が出現した際になど、一刻も早く治療薬を開発するためにも有効であるものと思われる。ホモロジーモデリングは目的蛋白質と同じファミリーに属する参照蛋白質の立体構造を基に構造予測するものであるので、ホモロジーモデリングで立体構造予測を行なえるかどうかは参照蛋白質の有無により決まる。国際的な構造ゲノムプロジェクトによりあらゆる蛋白質ファミリーの代表構造が実験的に決定されつつあり、近いうちに多くの蛋白質の立体構造がホモロジーモデリングで予測できることになる。このような意味でも、ホモロジーモデリングはますます欠かすことのできない重要な手法となることは疑いない。

文 献

- 1) Li, W. *et al.* : *Nature*, 426, 450-454(2003)
- 2) Simmons, G. *et al.* : *Proc. Natl. Acad. Sci. USA*, 101, 4240-4245(2004)
- 3) Altschul, S. F. *et al.* : *Nucleic Acids Res.*, 25, 3389-3402(1997)
- 4) Yoneda, T., Komooka, H., Umeyama, H. : *J. Protein Chem.*, 16, 597-605(1997)
- 5) Takeda-Shitaka, M. *et al.* : *Curr. Med. Chem.*, 11, 551-558(2004)
- 6) Takeda-Shitaka, M. *et al.* : *Proteins*, Suppl 7, 122-127(2005)
- 7) 竹田-志鷹真由子・梅山秀明 : *薬剂学* 64, 168-171(2004)
- 8) 竹田-志鷹真由子・梅山秀明 : *ゲノム医学*, 5, 399-402(2005)
- 9) Ogata, K., Umeyama, H. : *J. Mol. Graphics Mod.*, 18, 258-272(2000)
- 10) *Proteins*, Supple 7(2005)
- 11) Terashi, G. *et al.* : *Proteins*, in press
- 12) Ginalski, K., Elofsson, A., Fischer, D., Rychlewski, L. : *Bioinformatics*, 19, 1015-1018(2003)
- 13) Rota, P. A. *et al.* : *Science*, 300, 1394-1399(2003) ; published online 1 May 2003 (10.1126/Science.1085952)
- 14) Marra, M. A. S. *et al.* : *Science*, 300, 1399-1404(2003) ; published online 1 May 2003 (10.1126/Science.1085953)
- 15) Berman, H. M. *et al.* : *Nucleic Acids Res.*, 28, 235-242(2000)
- 16) Takeda-Shitaka, M. *et al.* : *Med. Chem.*, 2, 191-201(2006)
- 17) Takeda-Shitaka, M. *et al.* : *Chem. Pharm. Bull.*, 52, 643-645(2004)
- 18) Anand, K. *et al.* : *Science*, 300, 1763-1767(2003) ; published online 13 May 2003 (10.1126/Science.1085658)