

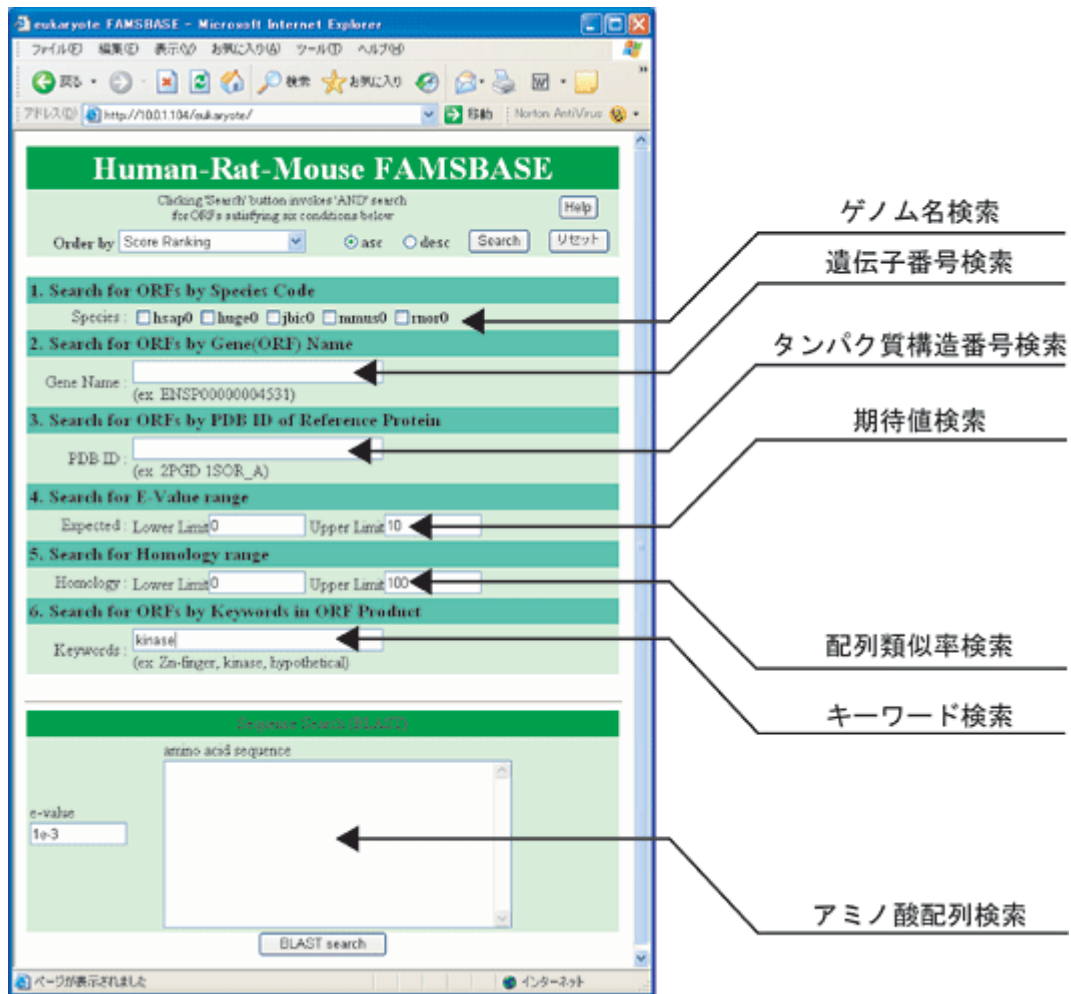
2006年9月28日  
独立行政法人 理化学研究所

## タンパク質立体構造予測データベースを全世界に公開

- 創薬研究の活性化や新規機能性生体物質の創製に向けて -

タンパク質は、様々なはたらきによって生命活動を担っています。このタンパク質のはたらき（機能）は、タンパク質の立体的なかたち（立体構造）との間に、密接な関係があります。多くの病気は、タンパク質が正しく機能しなくなることで生じるため、タンパク質の機能と形との関係を把握した上で、より良く効いて副作用の少ない薬を設計する方法が効率的です。そのためには、タンパク質の立体構造のデータが必要ですが、一般的に、これらの立体構造は、X線結晶構造解析法やNMR法などの方法を用いて、実験的に決定されます。しかし、実験は大変な作業で多くの手間と時間がかかる上、さまざまな要因によって、簡単には立体構造が調べられないタンパク質もあり、病気に関わる多くのタンパク質の立体構造を決定するのは容易ではありません。そこで、同じようなアミノ酸の並びを持ち、既に立体構造がわかっているタンパク質の立体構造から、構造未知のタンパク質の立体構造をコンピュータで予測する方法が開発されました。

理研ゲノム科学総合研究センターのタンパク質構造・機能研究グループは、文部科学省が推進している「タンパク3000プロジェクト」において、タンパク質の立体構造に関する研究を行っていますが、今回、その成果として、タンパク質の立体構造を予測したデータをまとめたデータベースをウェブ上で公開します。このデータベースは、タンパク質だけでなく、薬を設計する上で役立つ様々な情報なども含まれています。公開はウェブサイト (<http://famshelp.gsc.riken.jp/famsbase/>) にて行われ、誰でも手軽に利用することができます。これにより、創薬研究はもとより、新規機能性生体物質の創製など様々な分野における応用が期待されます。



(図) ウェブ上で公開されるデータベース「RIKEN FAMSBASE」の一画面

2006年9月28日  
独立行政法人 理化学研究所

## タンパク質立体構造予測データベースを全世界に公開

- 創薬研究の活性化や新規機能性生体物質の創製に向けて -

### ◇ポイント◇

- 遺伝子情報から予測した約 600 万種類のタンパク質立体構造をウェブサイトで公開
- タンパク 3000 の成果の効率的な応用拡大
- タンパク質の立体構造予測を創薬研究などに応用可能

独立行政法人理化学研究所（野依良治理事長）は、医学、薬学、生物学、理工学などのライフサイエンス分野をはじめ、創薬やタンパク質機能改変を目指す研究者などが手軽に利用できるタンパク質立体構造予測（モデリング）データベースを 9 月 28 日に全世界に向け公開します。これは、理研ゲノム科学総合研究センター（榎佳之センター長）タンパク質構造・機能研究グループ（横山茂之プロジェクトディレクター）の梅山秀明客員主管研究員（北里大学薬学部教授）らの研究成果で、文部科学省で推進している「タンパク 3000 プロジェクト」の一環として行ったものです。

今回の研究では、創薬研究などに最も重要で公的に入手可能なヒトや実験動物（ラット・マウス）などの全遺伝子に対してホモロジーモデリング法\*1によるタンパク質のモデリングを行い、それらのモデリングデータとともに、モデリングされたタンパク質ごとのモデル評価値や、他の化合物と相互作用可能なタンパク質上の場所の情報、タンパク質の説明などを充実させて構築したデータベース「RIKEN FAMSBASE(リケン ファムスベース)」を公開します。このデータベースは、約 600 万種類のタンパク質モデリングデータが格納されています。さらには、継続的なデータの更新を可能にするための自動更新システムを開発し、本データベースが高品質のデータを提供できるようにしました。

研究グループは本年 1 月に、インフルエンザウイルスのノイラミニダーゼ（タンパク質）のモデリングデータをデータベース化して公開してきましたが、今回公開する包括的な立体構造予測データベースは、それを包含し、質、量ともにはるかに上回るものです。このようなデータベースは世界初であり、これまでにない、詳細かつ応用性の広いデータベース公開となります。この成果により、創薬研究の活性化が期待されます。また、タンパク質機能の改変による新規機能性生体物質の創製なども期待されます。

公開はウェブサイト (<http://famshelp.gsc.riken.jp/famsbase/>) で行います。

## 1. 背景

ヒトを始め様々な生物種のゲノムプロジェクトが進展し、現在では莫大な遺伝子情報が蓄積してきています。創薬研究などにおいて、標的となるタンパク質の機能を解析し、効率よくドラッグデザインを行うには、遺伝子情報のみでなくタンパク質の立体構造情報が必要不可欠です。しかしながら莫大な遺伝子情報に比べると、

X線結晶構造解析法や核磁気共鳴（NMR）法などを用いて実験的に立体構造が決定されているタンパク質の数は少なく（Protein Data Bank<sup>\*2</sup>に登録されている実験的に決定された構造の数は、2006年9月現在で約3万9千個）、アミノ酸配列がわかっているにもかかわらず、立体構造のわからないタンパク質が多数存在しているのが現状です。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、立体構造が未知であるタンパク質の立体構造予測（モデリング）をすることができます。梅山秀明客員主管研究員らは、既に277生物種のタンパク質モデリングデータ（約140万個）を収めた「FAMSBASE」を長浜バイオ大学から公開していますが、今回新たに、ヒト遺伝子、実験動物（ラット・マウス）遺伝子などを含んだ公的に入手可能な全遺伝子に対してホモロジーモデリングによるタンパク質立体構造予測を行い、それらのモデリングデータを加えた「RIKEN FAMSBASE」を理研より公開することになりました。

## 2. 研究手法

タンパク質のモデリングには、研究グループが開発した「全自動ホモロジーモデリングソフト:FAMS（Full Automatic Modeling System）」が用いられました。ホモロジーモデリング法とは、立体構造未知のタンパク質（目的タンパク質）の立体構造を、類似のアミノ酸配列を持つ立体構造既知のタンパク質（参照タンパク質）を参照して予測する手法であり、コンピュータを用いたタンパク質立体構造予測法の中で現在最もモデルの精度が高い手法とされています。FAMSは、このホモロジーモデリング法に基づいて開発されたソフトで、FAMSを用いて予測されたモデルの精度の高さは、タンパク質立体構造予測の国際コンテストにおいて実証されています。今回、公的に入手可能な全遺伝子に対して、FAMSを用いてタンパク質の立体構造を予測し、データベース化しました。

## 3. 研究成果

既に長浜バイオ大学から公開されている277生物種のタンパク質モデリングデータ（約140万個）を収めた「FAMSBASE」に対して、今回新たにヒトのタンパク質モデリングデータ（793,612個）、実験動物（ラット・マウス）のタンパク質モデリングデータ（505,628個）、全生物のタンパク質モデリングデータ<sup>\*3</sup>（3,368,709個）を加えてデータベース化し、「RIKEN FAMSBASE」としてウェブサイト上で公開することにより、格納された情報を誰でも自由にダウンロードすることができる形にまとめました（図1）。データベースには、予測した立体構造それぞれについて、モデル評価値や、他の化合物と相互作用可能なタンパク質上の場所の情報、タンパク質の説明なども格納され、公開されるタンパク質は、約600万種類にのぼります（図2）。さらには、継続的なデータの更新を可能にするための自動更新システムを開発し、本データベースが高品質のデータを提供できるようにしました。

（ウェブサイト:<http://famshelp.gsc.riken.jp/famsbase/>）

研究グループでは、すでにインフルエンザウイルスの増殖に必須なノイラミニダーゼのモデリングデータ（総数1,603個）を公開していますが（2006年1月20日プレスリリース:新型インフルエンザに対応するためのタンパク質立体構造データベースを全世界に公開）、今回は、ヒトのタンパク質、実験動物（ラット・マウス）

のタンパク質、277 生物種のタンパク質、全生物のタンパク質モデリングデータを包括的に公開することになります。このようなデータベースは世界初であり、これまででない、詳細かつ応用性の広いデータベース公開となります。

#### 4. 今後の期待

本研究によるタンパク質立体構造モデルの構築・データベース化・公開によって、創薬研究の活性化が期待されます。既に理研では、抗 SARS ウイルス剤（2004 年 9 月 8 日プレスリリース:抗 SARS ウイルス剤の候補化合物を発見）、抗インフルエンザウイルス剤、抗 HCV（C 型肝炎ウイルス）剤開発などの研究を様々な研究機関と共同で開始し、いくつかの重要な成果をあげています。また、タンパク質機能の改変による新規機能性生体物質の創製なども期待されます。

(問い合わせ先)

独立行政法人理化学研究所ゲノム科学総合研究センター  
タンパク質構造・機能研究グループ

プロジェクトディレクター 横山 茂之  
Tel : 045-503-9196 / Fax : 045-503-9195

北里大学薬学部製薬学科生物分子設計学教室 教授  
独立行政法人理化学研究所ゲノム科学総合研究センター  
タンパク質構造・機能研究グループ

客員主管研究員 梅山 秀明  
Tel : 03-5791-6330 / Fax : 03-3446-9553

独立行政法人理化学研究所ゲノム科学総合研究センター  
タンパク質構造・機能研究グループ

上級研究員 松本 武久  
Tel : 045-508-7471 / Fax : 045-508-7468

独立行政法人理化学研究所 横浜研究所

研究推進部 企画課 溝部 鈴  
Tel : 045-503-9117 / Fax : 045-503-9113

(報道担当)

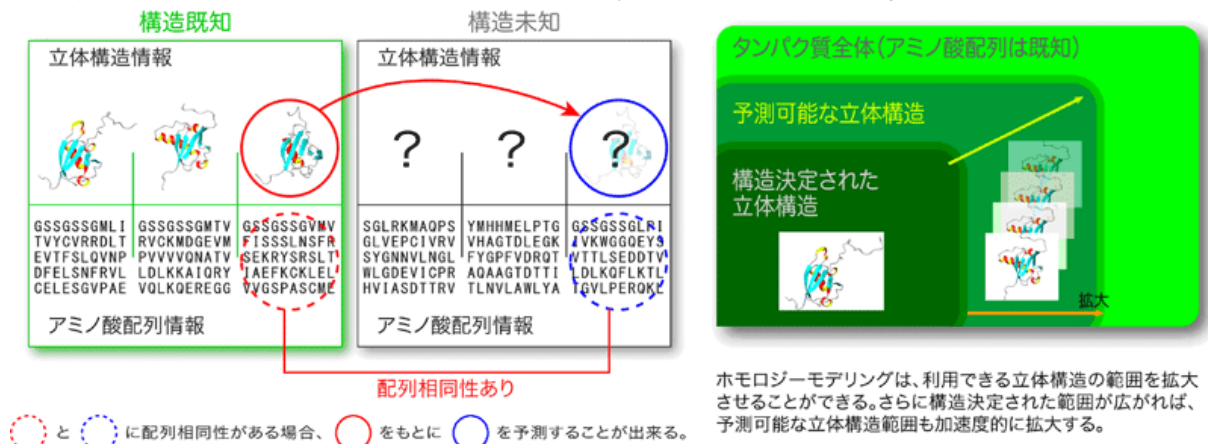
独立行政法人理化学研究所 広報室 報道担当

Tel : 048-467-9272 / Fax : 048-462-4715  
Mail : koho@riken.jp

## <補足説明>

### ※1 ホモロジーモデリング法

ホモロジーモデリング法とは、アミノ酸配列に相同性（生物学的に意味のある類似性）のある構造既知タンパク質の立体構造を鋳型として、構造未知タンパク質の立体構造をコンピュータで予測する手法のことである。これは生物進化において類縁関係のあるタンパク質どうしでは、アミノ酸配列に多少の変化が起きていても、それらの立体構造はよく保存されているという経験則に基づいている（下図左）。この手法は他の予測法と比較して予測精度や計算時間の面で優れており、構造データベースが充実するにしたがって予測可能なタンパク質数が加速度的に増加するため、さらに精度向上が見込まれる等の特徴がある（下図右）。



### ※2 Protein Data Bank

生体高分子の立体構造情報に関するデータベース。X線結晶回折やNMRなどの実験によって明らかにされたデータが収められている。ほとんどがタンパク質の立体構造で、一部核酸とタンパク質の複合体や核酸の構造も含まれる。

### ※3 全生物のタンパク質モデリングデータ

National Center for Biotechnology Information (NCBI) の「重複のないタンパク質配列データベース (nr, nonredundant protein sequence database)」に登録されている全生物のタンパク質のアミノ酸配列に対し、FAMS を用いて予測したモデリングデータ。

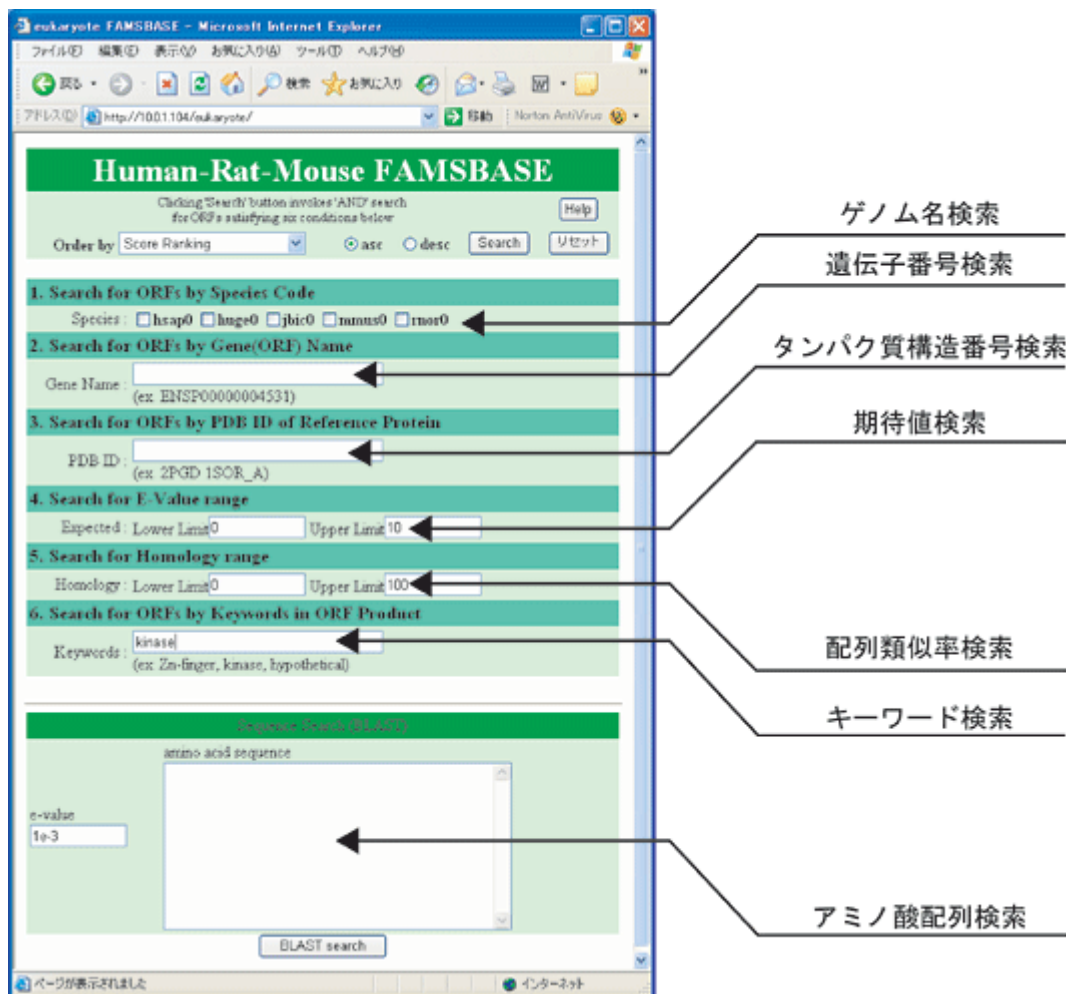


図1 「RIKEN FAMSBASE」の Human-Rat-Mouse FAMSBASE 画面

遺伝子番号やタンパク質構造の番号、遺伝子名、タンパク質のアミノ酸配列などから検索が可能となっている。

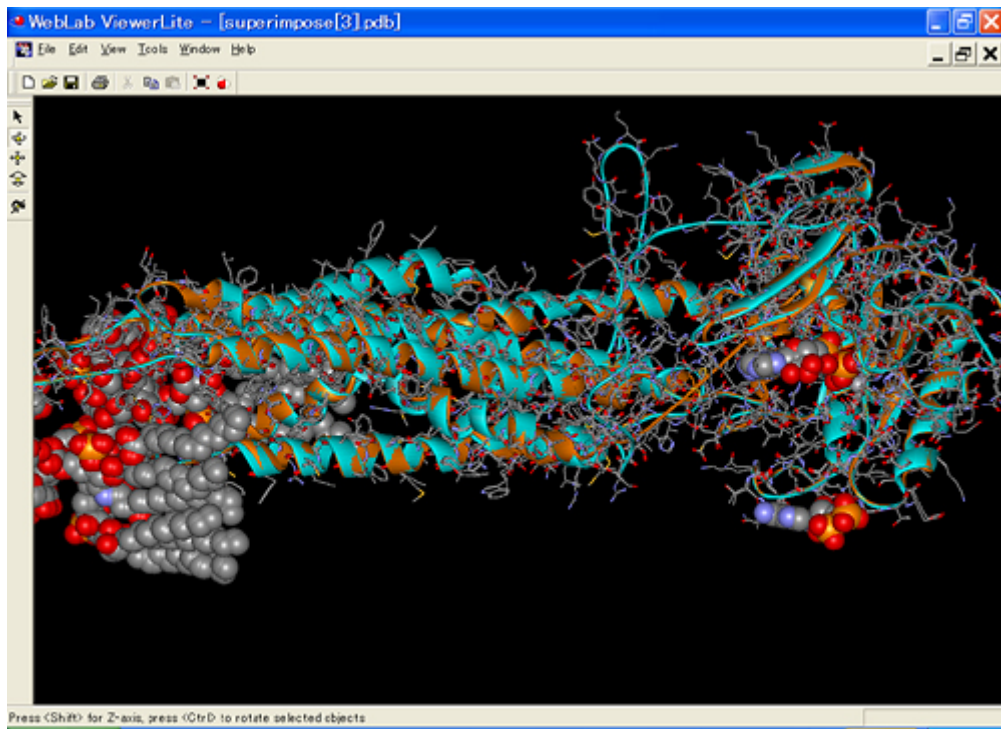


図 2 「RIKEN FAMSBASE」により公開されるタンパク質モデルの一例