

2009 年 7 月 23 日
独立行政法人理化学研究所

世界最大級のタンパク質結晶構造解析実験データベースを公開

—90 万件の結晶化条件などを整備した実験データ群が利用可能に—

本研究成果のポイント

○SPRING-8 の高輝度 X 線を使ったタンパク質結晶構造解析の膨大な実験データを集積

○実験データをタンパク質研究の参考情報や、新たな方法論開発に有効活用

○データの再利用や自動処理化が容易なセマンティックウェブ形式で共有化（当分野で世界初）

独立行政法人理化学研究所（野依良治理事長）は、微生物由来タンパク質の X 線結晶構造解析実験データを主体にした、生命科学研究に役立つタンパク質実験データベースをウェブ上で公開します。理研放射光科学総合研究センター（石川哲也センター長）タンパク質結晶構造解析研究グループの浅田征彦リサーチアソシエイトおよび国島直樹グループ副ディレクターらによる研究成果です。本データベースは、理研生命情報基盤研究部門（豊田哲郎部門長）の公開基盤「理研サイネス^{*1}」上で 7 月 23 日からダウンロードが可能となります（<http://database.riken.jp>）。

生命現象を原子レベルで理解し、医学・産業応用につなげるために、X 線結晶構造解析などによりタンパク質の立体構造を決定することは必要不可欠です。近年、世界中で行われている構造ゲノムプロジェクト^{*2}推進の結果、生命科学研究者が手軽にタンパク質の X 線結晶構造解析を実施できる研究基盤が整備されてきました。研究グループは、構造ゲノムプロジェクトの成果を社会還元する試みとして、大型放射光施設 SPRING-8^{*3}の高輝度 X 線を利用して集積した、微生物由来タンパク質、変異体タンパク質^{*4}、重原子標識タンパク質という 3 つの、系統的で詳細な結晶構造解析実験データベース（データサイズ 5.0TB=5×10¹²B、ファイル数 9700 万）を公開します。

目的タンパク質の結晶構造を得るためには、試料調製から始まり、結晶化、X 線回折実験を経て、計算による構造決定に至る一連の作業が必要で、各段階の作業から膨大な実験データが発生します。微生物由来タンパク質の結晶構造解析実験データベースでは、これらの実験データを研究のために自分で再編集して利用することが可能な形で一挙公開し、タンパク質の効率的な構造決定のためのソフトウェア開発などを促進します。また、生命科学者のタンパク質研究を、このデータベースに登録された世界最大規模の類似タンパク質情報により支援します。

変異体タンパク質の結晶構造解析実験データベースでは、研究グループがタンパク質結晶構造解析の方法論開発のために行った、多数の変異体タンパク質に関する実験データを公開します。これらの変異体タンパク質は、均一な条件で結晶化されるなど、相互比較を行う上で国内外に例のない有利な特長があります。このため、これらの実験データを再利用可能な形で公開することが、バイオインフォマティクス^{*5}分野での活用を加速し、創薬などへの貢献につながります。

タンパク質結晶構造解析では、白金などの重原子を含む試薬で標識したタンパク質結晶の作製が必要な場合があります。研究グループは、生命科学者へのタンパク質結晶構造解析のさらなる普及のため、目的タンパク質のアミノ酸配列や溶液条件に基づいて最適な重原子試薬が簡単にウェブ検索できる、当分野で世界最大のソフトウェア「*HATODAS*」を開発しました。重原子標識タンパク質の結晶構造解析実験データベースでは、*HATODAS* の検索基盤データを再利用可能な形で公開することにより、タンパク質工学^{*6}分野などへの応用を促進します。

本公開事業は、文部科学省「統合データベースプロジェクト^{*7}」の一環として実施しました。

1. 背景

世界中で行われている構造ゲノムプロジェクトにより、膨大な数のタンパク質立体構造が決定されてきました。理研放射光科学総合研究センターは文部科学省「タンパク 3000 プロジェクト（2002 年度～2006 年度）」に参画し、タンパク質結晶構造解析研究グループと放射光システム生物学研究グループ（倉光成紀グループディレクター）が中心となって、主に微生物由来タンパク質の結晶構造を SPring-8 の世界最高輝度を誇る X 線を用いて集中的に決定してきました。特に、高度好熱菌 *Thermus thermophilus* HB8^{*8} では、約 2,200 種の全構成タンパク質のうち国内外の成果を含めて約 2 割まで立体構造決定が進み、今やこの菌が最も構造解析の進んだ生物となりました。さらに、タンパク質結晶構造解析研究グループは、結晶化ロボット、構造解析支援ソフトウェア、タンパク質結晶工学技術^{*9} など、効率のよい結晶構造決定のための研究基盤整備を行いました。今回、これらの研究成果を社会還元する試みとして、放射光科学総合研究センターに存在するタンパク 3000 プロジェクト関連の膨大なタンパク質実験データ（技術開発関連データを含む）を、ウェブ上で公開しました。

2. 公開内容

タンパク質結晶構造解析研究グループは、以下の 3 種類の結晶構造解析実験データベースを公開しました。

(1) 微生物由来タンパク質 (<http://database.riken.jp/item/rib220>)

研究対象のタンパク質について実験方針を検討する際に、当該タンパク質と類似しているタンパク質がデータベースに収録されていれば、その類似タンパク質の実験情報が参考になります。特に、*T. thermophilus* HB8 由来タンパク質の場合、必要であれば理研バイオリソースセンターから遺伝子を取り寄せて類似タンパク質を生産し、当該タンパク質と構造や機能を比較検討することが可能となります。

目的タンパク質の結晶構造を得るためには①目的遺伝子の発現、②発現タンパク質試料の精製、③精製試料の結晶化、④タンパク質結晶の X 線回折データ収集、⑤回折データに基づく計算による構造決定、という一連の作業が必要で、各段階の作業から膨大な実験データが発生します（図 1）。理研の複数の部署は、これらの膨大な実験データを詳細に系統立って保存してきました。しかし、この膨大なデータを公開するにあたって、データの保存形式が部署によって異なっていることや、手入力が多いためデータ整合性の確認作業が必要なことが問題となっていました。そこで研究グループは、放射光システム生物学研究グループや生命分子システム基盤研究領域（横山茂之領域長）の協力を得て、放射光科学総合研究センターに存在する構造解析実験データを 1 年半かけて編集・確認し、1 つのデータベースに統一する作業を行いました。

今回、微生物由来タンパク質にかかわる試料調製（発現プラスミド構築実験 10,000 件、培養実験 5,000 件、精製実験 3,000 件）、結晶化実験データ（結晶化条件 90 万件、観察画像 1,000 万件）と回折実験データ（200 件）を公開し、類似タンパク質の構造決定などを支援していきます。

(2) 変異体タンパク質 (<http://database.riken.jp/item/rib220> の一部)

研究グループは、変異導入によりタンパク質の安定性、結晶性、重原子との結合しやすさを改善する新しい方法を開発する目的で、2種類のモデルタンパク質を用い、多数の変異導入実験とその変異体の結晶構造解析を行ってきました。モデルタンパク質として用いたのは、研究グループが立体構造を決定した高度好熱菌 *T. thermophilus* HB8 由来 TTHB049 と超好熱古細菌 *Pyrococcus horikoshii* OT3^{*10} 由来 PH0725 です。

TTHB049 は、177 アミノ酸残基の小さな単量体タンパク質で、大量の試料を安定して調製できる利点があります。また、熱変性の温度が 79.2℃と、*T. thermophilus* 由来タンパク質としては低いので、変異導入により目的タンパク質の安定性がどのように影響を受けるかを調べるのに格好の材料といえます。また、もう 1 つのモデルタンパク質である PH0725 は、265 アミノ酸残基の単量体が 2 つ会合した二量体で、タンパク質として中程度の大きさを持ちます。野生型由来の PH0725 結晶の X 線回折で得られる分解能も、2.1 Å (1 Å は 100 億分の 1 m) と改善の余地を残す中程度の精度でした。さらに、大量の試料を安定して調製でき、変異導入によって容易に結晶型が変化しないため、変異導入によりどのように結晶の質が影響を受けるかを調べるのに適しています。この PH0725 を用いた一連の変異導入実験により、研究グループは、タンパク質結晶の品質が変異導入によって改善可能であることを示しました (図 2)。

今回、これら 2 種類のモデルタンパク質について、変異体の結晶構造解析に関する実験データ 150 件を公開し、タンパク質設計などを支援します。変異体タンパク質の構造解析は、専用の規格化した手順で行われているため、ほかの微生物由来タンパク質と比較して、より均一で詳細な実験データとして提供ができます。

(3) 重原子標識タンパク質 (<http://database.riken.jp/item/rib108>)

研究グループは、タンパク質構造解析支援ソフトウェアの 1 つとして、重原子データベース「*HATODAS*」を日立ソフトウェアエンジニアリング株式会社と共同で開発しました (図 3)。目的タンパク質の結晶構造を決定するために、白金や水銀などの重原子を含む試薬で標識したタンパク質結晶を作製する場合があります。その場合、どの重原子試薬が適しているかを実験的なスクリーニングで検討する必要があり、解析経験のない研究者には気軽に取り組みづらいものでした。*HATODAS* は、既知の重原子標識タンパク質をデータベース化することにより、使用すべき重原子試薬が簡単に検索できるようにすることを目的としたシステムで、統合データベースプロジェクトに参画する以前から、ウェブ上で公開していました。*HATODAS* に目的タンパク質のアミノ酸配列を入力すると、標識物質として使える可能性の高い重原子試薬を出力します。さらに、溶液条件を入力することで、最適な重原子試薬の絞り込みができます。

今回、これら重原子標識タンパク質に関する実験データ 500 件を、統合データベースの一部として、再利用 (研究のための再編集) や一括ダウンロードが容易な形にして提供し、タンパク質工学分野の新たな方法論開発などを促進します。

3. 今後の期待

公開する 3 種類のデータベース各々への今後の期待は、以下のとおりです。

1 つ目の微生物由来タンパク質の実験データベースでは、結晶構造解析の膨大な実験データが詳細かつ系統的に整理されています。従って、本実験データの傾向分析などに基づ

き、タンパク質の効率的な構造決定のためのソフトウェアなどを開発することが可能です。さらに、例えば代表的なベンチマークセット（評価基準データ）として採用されるなど、生命科学分野において多様なデータを比較検討するための基準となる情報基盤として、広範な社会貢献が期待されます。2つ目の変異体タンパク質の実験データベースは、ホモロジーモデリング^{*11}の高精度化などに活用できます。現在のホモロジーモデリング技術では、例えば、ヒト由来タンパク質の立体構造を微生物由来類似タンパク質の既知立体構造から高精度で予測し、創薬につなげることは非常に困難です。本データベースでは、すでに変異体の系統立った詳細な立体構造を提供しているため、解答に合うようにアルゴリズムを改良することが可能です。こうして得た高精度ホモロジーモデリング技術は、類似タンパク質の立体構造予測に大きな威力を発揮すると期待できます。3つ目の重原子標識タンパク質の実験データベースでは、タンパク質工学分野への展開が注目されます。研究グループはこれまでに、重原子を結合するモチーフ配列を多数見つけています（図 3）。従って、これらのモチーフをタンパク質工学的に目的タンパク質に導入することで、計画的に重原子標識をすることが可能となり、さまざまな応用が期待できます。

今回のタンパク質実験データベースは、理研生命情報基盤研究部門が開発している公開基盤「理研サイネス」(<http://database.riken.jp>)から公開しました（図 4）。この公開基盤上では、各データベースがセマンティックウェブと呼ばれる国際標準形式で再構築されています。この形式の最大の特長として、各データ間の関係について意味付けがなされているため、データの再利用や自動処理化が容易なことが挙げられます。また、今回公開した3種類のデータベースは、どれも、未加工の実験データまでさかのぼって利用できるという特長を持っています。従って、データベースを一括ダウンロードし、ほかのデータと組み合わせるなどして新たなデータベースを構築することが可能になり、大規模で予想外の展開が期待できることとなります。今後は、今回公開したデータについて国際的な合意を得ながら、生命科学者が利用しやすいように、さらに整備を進めていく予定です。2010年度末までに、統合データベースの一部として生命科学分野の幅広い研究者に役立つ体制を確立していきます。

<報道担当・問い合わせ先>

(問い合わせ先)

独立行政法人理化学研究所

放射光科学総合研究センター

タンパク質結晶構造解析研究グループ

グループ副ディレクター 国島 直樹 (くにしま なおき)

TEL : 0791-58-2937 FAX : 0791-58-2917

播磨研究所 研究推進部 企画課

TEL : 0791-58-0900 FAX : 0791-58-0800

(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

(SPring-8 に関すること)

財団法人高輝度光科学研究センター 広報室

TEL : 0791-58-2785 FAX : 0791-58-2786

<補足説明>

※1 理研サイネス

理研サイネス (RIKEN SciNeS : RIKEN Life Science Networking System) は、ライフサイエンスを主体にしたデータベースの構築基盤システムを研究機関内で一元化し、「セマンティックウェブ形式」と呼ばれる国際標準規格に準拠したデータ公開を大規模に実施するために、理研の生命情報基盤研究部門で運営されている情報公開基盤である。理研サイネスを利用することにより、研究機関内の個々の研究者はウェブサーバーを維持する必要がなくなり、各自のデータベースを研究成果物としてスムーズに外部へ向け発信・発表することができる。日本の研究者が国際連携研究を進めるための情報基盤として役立つ。

参考資料 :

<http://omicspace.riken.jp/publications/toyoda2009ja.pdf>

http://www.riken.jp/r-world/info/release/press/2009/090331_2/detail.html

※2 構造ゲノムプロジェクト

タンパク質の立体構造は、生物を原子レベルで理解するために必要不可欠な情報である。生物の全遺伝子 (ゲノム) を多種類の生物について解読するゲノミクスが実現した後、それらのゲノムがコードするタンパク質の立体構造を網羅的に決定し、立体構造に基づく研究開発の基盤として役立てようとする大規模な研究プログラムが世界的に進められている。この学問分野を構造ゲノミクスという。日本では 2002 年度から 2006 年度までの 5 年間に「タンパク 3000 プロジェクト」と呼ばれる構造ゲノムプロジェクトを国家プロジェクトとして実施し、多数の構造決定を行うとともに、効率的な構造解析のための研究基盤を整備した。

※3 大型放射光施設 SPring-8

兵庫県の播磨科学公園都市にある世界最高輝度の放射光を生み出す理研の施設。SPring-8 の名前は **Super Photon ring-8GeV** に由来。放射光とは、電子を光とほぼ等しい速度まで加速し、電磁石によって進行方向を曲げた時に発生する、細く強力な電磁波のこと。SPring-8 では、この放射光を用いて、ナノテクノロジー、バイオテクノロジーや産業利用まで幅広い研究が行われている。

※4 変異体タンパク質

遺伝子操作により、目的タンパク質の特定のアミノ酸残基を、任意の別の種類のアミノ酸に置換することができる（部位特異的変異導入）。この技術で野生型タンパク質に突然変異を導入したものを変異体タンパク質という。

※5 バイオインフォマティクス

応用数学、情報学、統計学、計算機科学などの技術応用によって生物学の問題を解こうとする学問。「生命情報学」「生物情報学」などと訳される。近年、多くの生物を対象に実施されているゲノムプロジェクトや構造ゲノムプロジェクトによって、大量のバイオ関連情報が得られるようになり、それらの情報をタンパク質の系統解析、構造予測、相互作用予測など有用なバイオインフォマティクス技術につなげることが求められている。

※6 タンパク質工学

タンパク質は生命を維持するためのさまざまな機能を発揮し、生体内で多種多様な働きを持つ。タンパク質工学は、天然のタンパク質を人為的に改変し、望みの機能を持つ人工タンパク質を設計することを目指している。タンパク質は各々特定のアミノ酸配列を持ち、特異的な立体構造を形成することにより特定の機能を発揮する。従って、タンパク質工学の手法としては、遺伝子操作による変異体タンパク質の作製が一般的に用いられる。

※7 統合データベースプロジェクト

ライフサイエンス研究を支える基盤として、生命情報データベースの統合化を推進する文部科学省のプロジェクトである。わが国のライフサイエンス関係データベースの利便性向上を図るため、ライフサイエンス関係データベース整備戦略の立案・評価支援、データベース統合化の基盤技術開発、ポータルサイトの整備などを行い、統合化を推進する。理研は 2007 年度から 4 年計画で行われている「統合データベースプロジェクト補完課題」に課題名「植物オミックス情報および蛋白質構造情報」で参画し、実験データなどを同プロジェクトに提供する。

※8 高度好熱菌 *Thermus thermophilus* HB8

静岡県伊豆半島にある峰温泉から発見された、75℃の高温環境に生育する細菌（バクテリア）。約 2,200 種類の遺伝子から成る比較的小型のゲノムを持つ。また、その構成タンパク質は熱に強く、実験を行う上で都合がよい。理研放射光システム生物学研究グループは、*T. thermophilus* HB8 をモデル生物としたシステム生物学（細胞内のあらゆる生命現象を、その構成物質の構造と機能に基づくシステムの観点から理解することを目指す学問）を推進している。

※9 タンパク結晶工学技術

近年、世界中で行われている構造ゲノムプロジェクトの推進の結果、生命科学研究者が手軽に結晶構造解析を実施できる研究基盤が整備された。その一方で、膜タンパク質や超分子複合体をはじめとする、生物学的に重要であっても、現在の解析技術では構造決定が困難な「難解析性タンパク質」には、ほとんど手が付けられていない。理研タンパク結晶構造解析研究グループは、難解析性タンパク質の構造解析に革新をもたらす技術として、タンパク質分子を操作して解析可能な品質の結晶を得ることができる「タンパク結晶工学」の確立を目指す。例えば、変異導入によりタンパク結晶の品質を改善する技術（Mizutani *et al.* (2008) *Acta Cryst. D* 64, 1020-1033）や、結晶化基板によりタンパク結晶化を制御する技術（Sugahara *et al.*

(2008) *Acta Cryst. D* **64**, 686-695) を開発している。

※10 超好熱古細菌 *Pyrococcus horikoshii* OT3

ヒトを含む真核生物、バクテリアなどの原核生物と並ぶ、生物の第3のグループである古細菌の一種で、沖縄海溝内の海底熱水鉱床から発見された。98℃の極限環境に生育するため、その構成タンパク質は熱に非常に強く、実験を行う上で都合がよい。従って、タンパク質研究によく用いられる。

※11 ホモロジーモデリング

タンパク質の立体構造を効果的に予測する方法の1つ。アミノ酸配列の類似したタンパク質（アミノ酸同一残基率でおよそ30%以上）は立体構造も類似しているという原理に基づき、目的タンパク質のアミノ酸配列と、目的タンパク質に類似したタンパク質の立体構造から、目的タンパク質の立体構造を予測し、モデルを作成する。この方法で得た立体構造モデルをホモロジーモデルという。

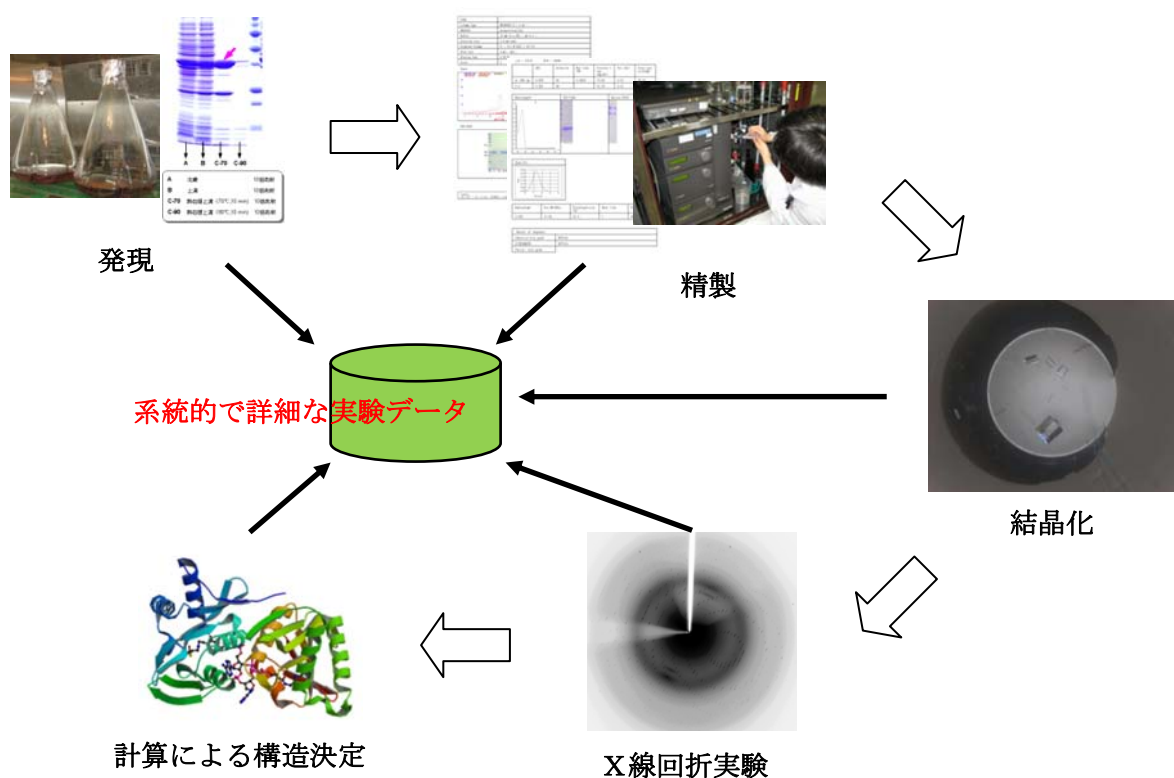


図1 タンパク質結晶構造解析の流れと実験データ

Wild-type (2.1 Å resolution)

T146R (1.6 Å resolution)

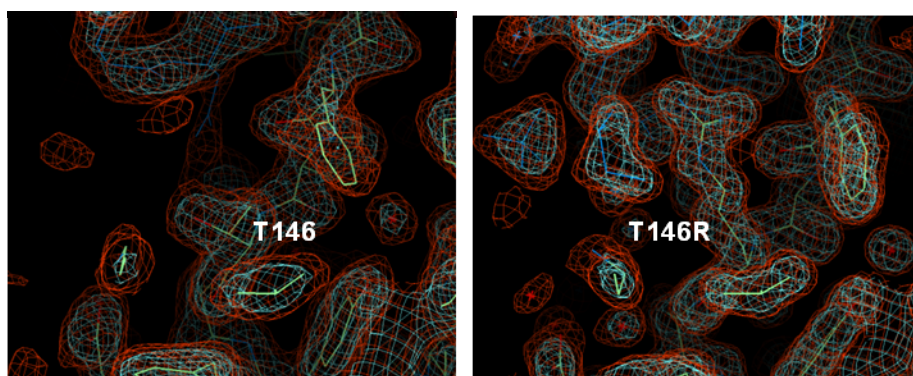


図 2 変異導入によるタンパク質結晶の品質改善

(Mizutani *et al.* (2008) *Acta Cryst. D* **64**, 1020-1033)

左：野生型 2.1 Å 分解能 右：T146R 変異型 1.6 Å 分解能

超好熱古細菌 *Pyrococcus horikoshii* OT3 由来 PH0725 タンパク質を用いて実験した。野生型と T146R 変異型のどちらも同じ空間群に属し、似た格子定数を持つ結晶である。変異を導入した結晶パッキング部位を同じ方向から見た結果、原子の位置を示すかご状の電子密度が、T146R 変異型でよりはっきり見えており、変異導入により、結晶の品質（X 線回折分解能）が改善することが分かった。ここで T146R 変異型とは、野生型 PH0725 タンパク質における 146 番目のスレオニン（3 文字表記で Thr、1 文字表記で T）残基をアルギニン（3 文字表記で Arg、1 文字表記で R）残基に置換してできる PH0725 の変異体タンパク質のことを指す。図中、白文字でラベルしたアミノ酸残基が、146 番目のスレオニンおよびアルギニンである。

HEAVY ATOM - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(H) ブックマーク(B) ツール(T) ヘルプ(H)

http://hatodas.harima.riken.go.jp/hatodas_v1/cgi-bin/frame.cgi

よく見るページ スマートブックマーク

Motif

motif No 18
motif atom: Pt
motif: RXXQM

motif No 20
motif atom: Pt
motif: LXXXM

motif No 33
motif atom: Sm
motif: EE

motif No 35
motif atom: Sm
motif: EXXGE

motif No 38
motif atom: Sm
motif: EXD

motif No 40
motif atom: Sm
motif: PXXXY

motif No 41
motif atom: Sm
motif: EXXG

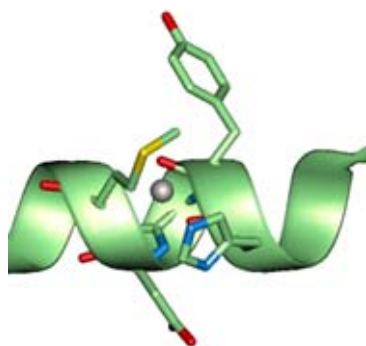
motif No 42
motif atom: Gd
motif: EE

Detail	PDB ID	Binding Site	Binding Secondary	Heavyatom Reagent	Heavyatom Name	Precipitant	Buffer	pH	Protein Name
detail	1L4U	YRALMAKRA	-	K ₂ PtCl ₄	Pt	PEG 4K	HEPES	7.5	-
detail	1QVR	EAAARLRMALES	alpha	-	Pt	PEG 4K	Acet.	4.1	-
detail	1QVR	IRIDMTEYMEKH	-	-	Pt	PEG 4K	Acet.	4.1	-
detail	1QVR	ARLRMALES	-	-	Pt	PEG 4K	Acet.	4.1	-
detail	1UAS	GRSVMERYT	-	-	Pt	2-propanol	Acet.	4.5	-

Red Character is binding Residue with Heavy Atom

完了

検索結果の表示例

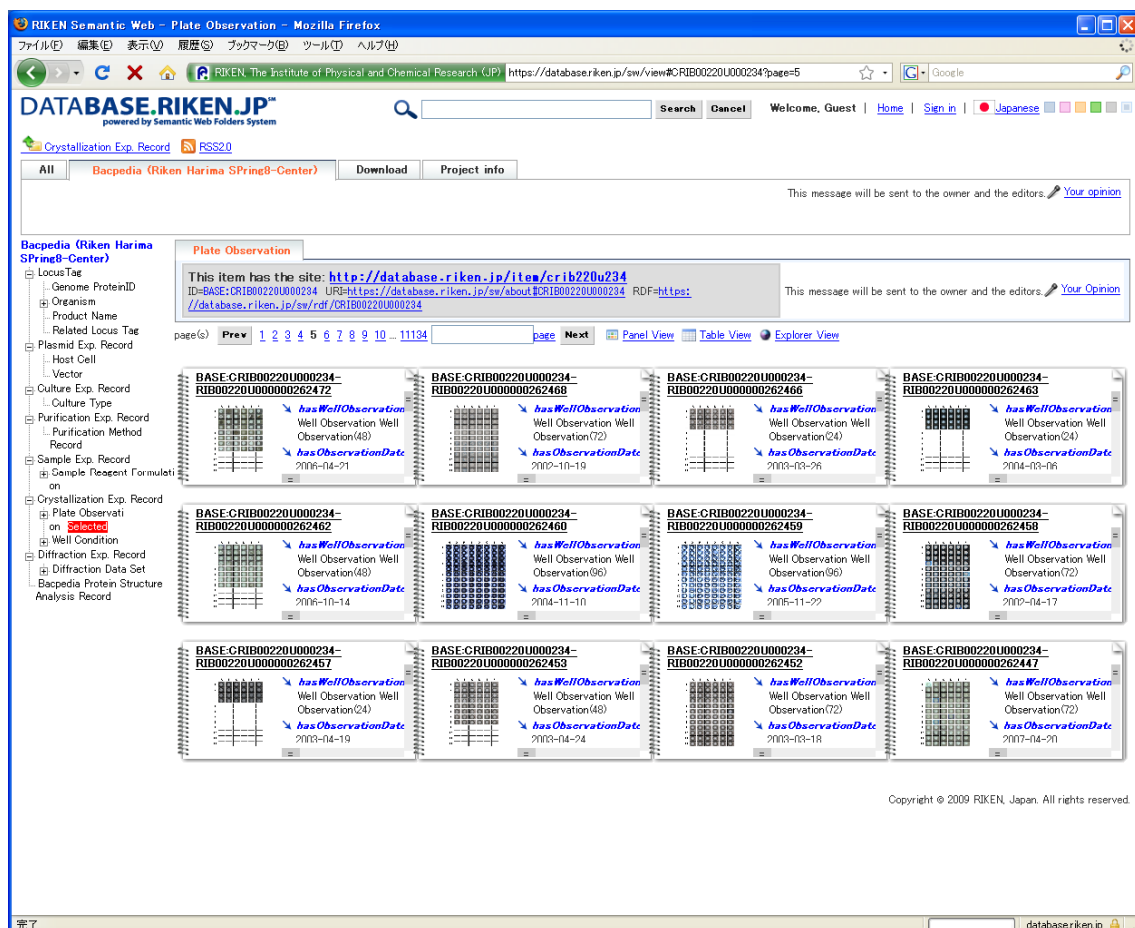


重原子結合モチーフ

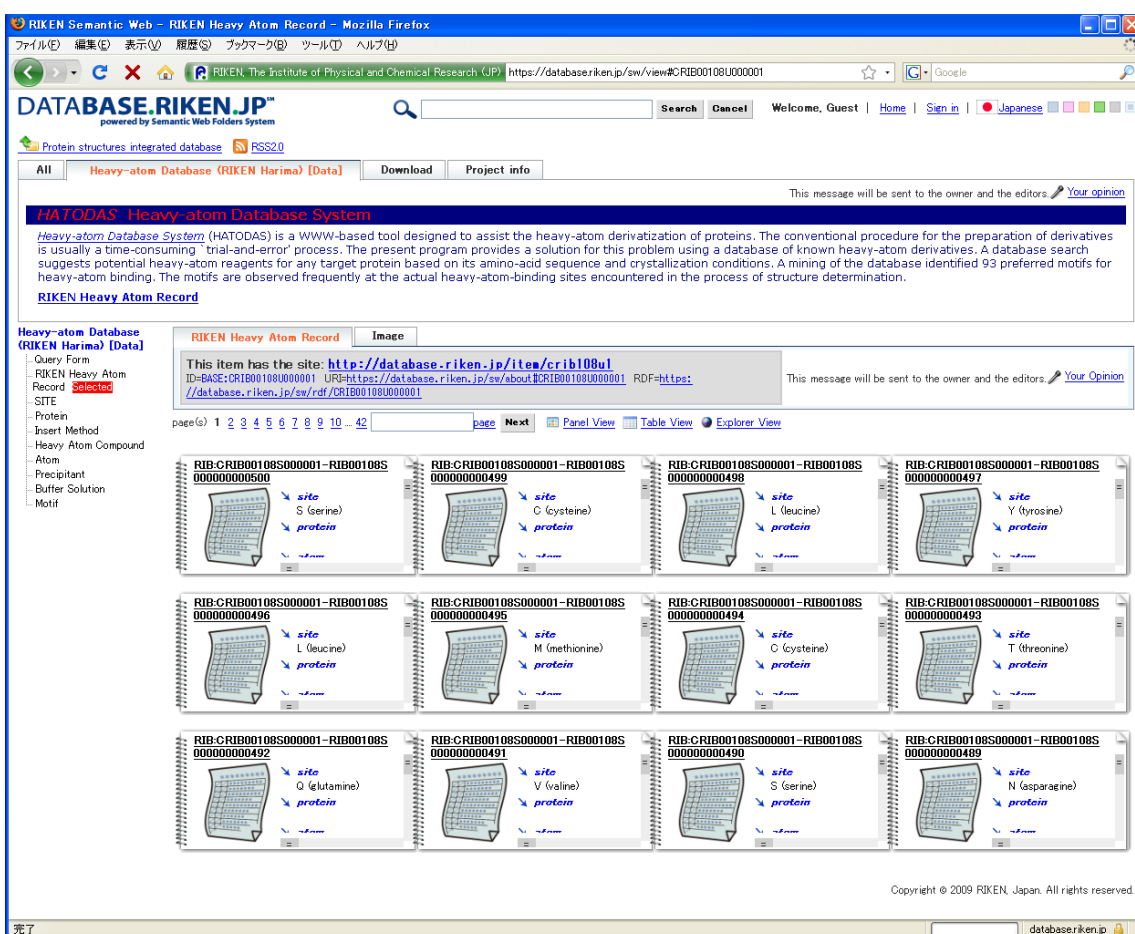
モチーフとは、複数種のタンパク質に共通して現れる局所的なアミノ酸配列のことであり、特定の共通機能（この場合重原子結合）が示唆される。なぜ重原子に対するモチーフが多数発見されたのか不明であるが、タンパク質の何らかの生理機能と関連する可能性もある。

図 3 重原子データベース HATODAS (<http://hatodas.harima.riken.jp>)

(Sugahara *et al.* (2005) *Acta Cryst. D* **61**, 1302-1305;
Sugahara *et al.* (2009) *J. Appl. Cryst.* **42**, 540-544)



結晶解析実験データ (<http://database.riken.jp/item/crib220u234>)



重原子実験データ (<http://database.riken.jp/item/crib108u1>)

図 4 公開基盤「理研サイネス」(<http://database.riken.jp>)上のデータベース画面