

## 5. 中核機関の成果概要

### 5 - 1 タンパク 3000 プロジェクト 中核機関成果報告

グループ名 網羅的解析プログラム

機関名 理化学研究所

代表者名 横山 茂之

#### 1. 平成 17 年 10 月末におけるグループ全体の事業計画に対する事業の進捗状況の概要について

##### 1.1 目的および戦略

タンパク 3000 プロジェクトの目的にそって、生物学的・医学的に重要なタンパク質の構造・機能の解明をめざす。特に、疾患など医学的な観点から重要となるヒト（およびマウス）のタンパク質の研究に重点を置く。加えて、生物に共通する生命現象の基本的に重要なタンパク質の研究のため、細菌ならびに古細菌（真核細胞のモデル）にも取り組む。立体構造を活用して、タンパク質機能を制御する化合物の探索など、応用的な研究も推進し、成果の産業移転もめざす。構造的基盤に立ったタンパク質機能の理解を通じてライフサイエンスに寄与すること、また、そのための様々な技術を開発して研究体制を構築することが目的である。

多くの重要なヒト・マウスなどのタンパク質に取り組むことを可能にするため、機能ドメインによるファミリー分類に基づいて、ドメインの立体構造（「基本構造」）という視点を戦略の軸とする。実験的に決定された立体構造に基づき、同一ファミリーに属する他のタンパク質についても、ホモロジー・モデリングの手法によって構造情報を展開することができる。このように、できるだけ多くの重要タンパク質について構造的基盤に立った機能研究を可能にすることを目指す（新しいタイプの立体構造の発見は研究の結果であって目的ではない）。タンパク 3000 プロジェクトにおける網羅的研究拠点としての割り当てに従って、5 年間に 2,500 の基本構造を明らかにするという目標を設定している。

タンパク質の立体構造と分子機能の研究を生物機能の理解へと発展するためには、タンパク質によって構成されるネットワーク、システムとしての視点が極めて重要である。シグナル伝達や遺伝子発現の制御など、重要なメカニズムにおいては、様々な相互作用（タンパク質・タンパク質、タンパク質・核酸、タンパク質・脂質、タンパク質・低分子化合物等）により、機能的パスウェイが構築される。そこで、それらの相互作用の構造基盤を明らかにするため、複合体としての構造解析を進める必要がある。マルチドメイン構造は、分子内でのドメイン間相互作用、ドメイン・モチーフ相互作用に基づく、活性制御機構の必須の基盤である。基本構造の情報をフルに活用して、複雑なタンパク質システムの研究に発展させる。

構造解析に適した試料の調製は、現在、構造解析の最も深刻なボトルネックとなっており、全体のスループットや費用対効果に決定的に影響するステップである。タンパク質は極めて多様な性質を示すので、万能な方法は存在せず、タンパク質ごとに最適化していくことが必須である。そのため、新規の細胞系および無細胞系のタンパク質発現技術を開発し、適した方法を選択することを可能にするとともに、発現検討のプロセスを自動化・機械化し、大規模で論理的な条件検討を行うことを可能にしている。さらに、立体構造解析のための試料調製のみならず、機能解析など幅広い応用に向けた新規技術体系として発展させる。

立体構造の決定には、X線結晶構造解析とNMR解析を効果的に組み合わせる。全長で 200 残基以上のタンパク質、複合体試料については、X線結晶構造解析による決定を目指す。全長で 200 残基以下のタンパク質、および、ドメインとして 200 残基以下のフラグメントに分割できるものについては、まず<sup>1</sup>H-<sup>15</sup>N NMRスペクトルのスクリーニングを行い、立体構造決定の適否を判断する。良好な試料について、単量体は、<sup>13</sup>C,<sup>15</sup>N 二重標識多重共鳴NMRによって立体構造を決定し、多量体化しているものは、X線結晶構造解析により立体構造を決定する。ドメイン構造の情報は、マルチドメイン試料、複合体の構造解析の基礎として活用する。

これらの構造解析を実現するために、構造決定技術の高度化、新規技術の開発を継続して行っている。試料調製、データ測定から構造計算までの要素技術の著しい高度化に裏付けられた自動構造解析技術は、研究者の負担を著しく軽減するとともに、構造解析の質的な向上をも可能にする。他方、現在では構造解析が難しいターゲットに取り組むための次世代技術の開発も必須である。巨大複合体や膜タンパク質は、医学・生物学的に重要なタンパク質に多くみられ、今後の研究の焦点となる。そこで、試料調製技術、結晶化技術、NMR の高磁場化、放射光ビームラインのマイクロフォーカス化などの基礎研究を積み重ねている。

構築した研究体制を最大限に活用する、すなわち、ライフサイエンス、医学におけるニーズを的確に捉え、創薬などの応用的価値に確実につなげるため、アカデミアおよび産業界との有効な共同研究、協力体制を構築する必要がある。そのための制度設計、幅広い調査研究など、様々な取り組みを展開している。

##### 1.2 解析ターゲット

###### 1.2.1 ヒトおよびマウスのタンパク質

医学・生物学的に重要なタンパク質、創薬につながるタンパク質をターゲットとする。プロジェクト開始時に、ヒトおよびマウスから、重要な疾患関連タンパク質を探索し、癌、神経疾患、糖尿病などの重要疾患、

シグナル伝達などの重要分子機能に関連するタンパク質を数千種リストアップして、研究を進めた。その後、年度ごとに新たな情報を取り入れてターゲットタンパク質の精査を続け、ヒトの疾患の原因に関係があるもの、創薬ターゲットの候補になるものとして、現在、合計 1,263 のヒトタンパク質に絞り込んでいる。その疾患別の内訳は以下のとおりである。

癌・白血病	370	循環器疾患	81	呼吸器・消化器疾患	1
免疫・造血疾患	284	感染症	37	その他	84
神経疾患	192	内分泌疾患	34		
糖尿病等栄養疾患	174	肝臓・胆道・腎疾患	6	合計	1263

これらのターゲットタンパク質を、シグナル伝達や転写制御などのタンパク質ネットワークの分析を行って、分子間相互作用や制御機構の視点から整理した。さらに、ドメインレベルでのアミノ酸配列ファミリー（Pfam など）に基づいて整理し、他のファミリーメンバーやホモログも考慮に入れた。このようにして、機能と構造の両面から優先順位付けし、戦略的な研究計画を立案した。さらに、理化学研究所免疫・アレルギー科学総合研究センターの谷口克センター長、小原収グループディレクターらにより、免疫学において重要なタンパク質 2,697 種（上記の疾患関連タンパク質との重なりも多い）がリストアップされた。平成 17 年度は、その中でも緊急の重要ターゲット約 100 種を選定して、構造および機能の解析の共同研究を開始した。他方、理研疾病専門委員会を立ち上げ、医学薬学研究者による調査研究もサポートした（§1.8）。

### 1.2.2 植物のタンパク質

植物特異的タンパク質の立体構造解析は、ヒト、マウスなどのタンパク質の解析と相補的な意義がある。生物種としては、完全長 cDNA の集積されたシロイヌナズナ *Arabidopsis thaliana* を主な対象とした。機能の観点からは、環境ストレスに対応する転写制御因子を中心に、適応機構等の理解に寄与するターゲットを選定した。これにより、環境ストレス耐性植物の育種などに貢献することを目指す。現在までに 30 を超える立体構造を決定しており、植物タンパク質を専門とする海外のプロジェクトと比較しても、最大規模である。

### 1.2.3 微生物のタンパク質

生物体の必須な遺伝子は約 1,000 個で、基本的で重要な酵素タンパク質は微生物からヒトまで高い相同性を持つ。そこで、細菌および古細菌由来タンパク質の高次構造を明らかにし、その機能を理解して、広くヒトまで保存された生命現象の解明を目指す。他方、細菌タンパク質は、感染症研究にも重要で、転写・翻訳に関わる因子のように必須な酵素は、抗菌剤のターゲットとなる可能性が高い。古細菌タンパク質は、細菌型とは異なる真核生物型の特徴をもつため、病原菌とヒトの間での構造上の差異を推定するのに有用である。

我々は、細菌である高度好熱菌 *Thermus thermophilus* HB8 を中心として研究を開始し、特徴である熱安定性から結晶構造解析の高い成功率を得てきた。倉光の提唱した「丸ごと一匹」プロジェクトでは、*T. thermophilus* の 2,239 の遺伝子にコードされた全てのタンパク質を物理化学的に理解することを目指す。さらに、古細菌としては、超好熱菌 *Pyrococcus horikoshii*（遺伝子数 2,246）を中心に、タンパク質の熱安定性を利用して結晶構造解析を進めてきた。現在までに、細菌、古細菌の合計で 600 近い結晶構造を決定することに成功した。解析が進行したため、現在、構造未知のタンパク質は、この 2 種では解析が困難なものが多くなってきたため、他の微生物のオーソログに範囲を広げた〔細菌は、*Bacillus subtilis* (4,106), *Geobacillus kaustrophilus* (3,540), *Escherichia coli* (5,259), *Symbiobacterium toebii* (3,284), *Thermotoga maritima* (1,875), *Aquifex aeolicus* (1,559) など、古細菌は、*Aeropyrum pernix* (2,695), *Sulfolobus tokodaii* (2,826) など、括弧内は全遺伝子数〕。今後は、病原菌タンパク質にも直接に取り組む。

## 1.3 試料調製

### 1.3.1 ヒトおよびマウス由来 cDNA クローンの整備

ヒトについては、小原収博士（かずさ DNA 研究所）および菅野純夫博士（東京大学）との共同研究として、それぞれ、5,286 クローン（脳由来）および 4,063 クローンのヒト完全長 cDNA を得た。さらに、別途 26,620 クローンのヒト cDNA（各種臓器由来）を得た。これらの合計 35,969 クローンのヒト cDNA について、必要に応じて 2,450 クローンの塩基配列を解析し、タンパク質発現に利用可能に整備した。マウスについては、FANTOM の完全長 cDNA 52,260 クローン（FANTOM1、FANTOM2、FANTOM3）を中心として、1 万以上のクローンの塩基配列を解析し、合計 54,487 クローンを備えた。これは、タンパク質発現のための cDNA コレクションとして世界最大規模である。

### 1.3.2 無細胞タンパク質合成システムによるハイスループット試料調製

試料調製のボトルネックを解消するため、通常の方法に加えて、無細胞タンパク質合成技術の開発を行った。多種類のタンパク質試料をハイスループットに調製するブレイクスルーとなる無細胞タンパク質合成の

自動化システムを構築し、多数のヒト・マウス由来タンパク質の cDNA からの発現に成功した。

このシステムにおけるタンパク質試料調製過程では、まず、微量スケールの発現スクリーニングによって良好な試料・条件を選定する。発現タンパク質のC末端にGFPを連結し、フォールディングの蛍光レポーターとして利用している。mgスケールの<sup>15</sup>N標識試料を自動化発現・タグ精製を行う（ハイスループットの専用機を開発）。次に、これらに対する<sup>15</sup>N-HSQC NMRスペクトル観測により、発現と特性（溶解性等）の良好なコンストラクトを選定し、その中から、解析に好適なタンパク質を選定する。選定されたタンパク質は無細胞タンパク質合成によって<sup>15</sup>N/<sup>13</sup>C標識試料（NMR）または、SeMet置換試料（結晶解析）として大量合成され、最終的にハイスループット自動化クロマトグラフィー・システムで精製されるという流れを辿る。以上のように複雑で時間と人手を要し、また繊細な構造解析過程に供する試料調製というデリケートな過程を自動化する意義は大きく、国際的に見ても例がない成果である。

平成 17 年 10 月末までに、15,533 ドメイン領域(182,192 クローン)について発現および可溶性のスクリーニングを行った結果、5,276 ドメイン領域(16,579 クローン)の可溶性・高発現量のドメインを得た。これらの<sup>15</sup>Nラベル体を作成し、NMR Screeningによりサンプルの性質を検証した結果、1,378 箇所の立体構造解析可能と思われるドメイン領域を見いだした。NMR解析用試料調製数は約 1460 種類である。同様に、ヒト・マウスタンパク質などのX線構造解析には、約 900 種類のNativeとSeMetなどの試料を大量調製した。微生物（高度好熱菌、超好熱菌、大腸菌、病原菌など）タンパク質のX線構造解析には、約 3600 種類のNativeとSeMetなどの試料を大量調製した。

他方、現在は開発の焦点を移し、より高難度の合成困難なタンパク質（巨大複合体や貫通型膜タンパク質など：特許申請中）の無細胞合成の基本技術、非天然型アミノ酸をタンパク質に部位特異的に導入する基本技術などの開発を進めている（詳しくは§5に後述）。

#### 1.4 立体構造決定

##### 1.4.1 NMR 解析によるタンパク質構造決定

主に、高等生物（ヒト、マウス、シロイヌナズナ）由来タンパク質のうち、分子量が2万を超えない機能ドメインを対象としてNMR構造解析を実施した。平成17年10月末までに、1,193ドメイン領域についてダブルラベル体を調製した結果、これまでに932個の立体構造解析可能と思われるサンプルを得ている。このうち659個については立体構造解析を完了し、PDB登録済あるいはPDB登録可能な状態にある。90個についてはグローバルフォールドまで決定し、現在精密化を行っている。188個については現在NMR測定中もしくはシグナル帰属中である。また、上記とは別に23個のドメインについてはスクリーニングを介さずにNMR解析を行い、立体構造を決定した。この結果、プロジェクト開始から平成17年10月末までに、682構造（ヒト・マウス634構造、高度好熱菌等微生物9構造、その他39構造）の構造決定を行った（解析を行った個々のタンパク質の概要は別紙に記載）。

多次元多重共鳴NMRスペクトルの測定は、理研横浜研究所に設置された約40台のNMR装置（Bruker, Varian, 日本電子）を用いて、1試料あたり約3週間をかけて測定した。測定システムを整備し、これらの装置の利用可能時間（保守、故障、待機時間などを除く）の90%以上という、極めて高い利用率を達成した。一部のデータは、独立行政法人物質・材料研究機構を中心として開発した920MHz NMR装置でも行った。NMR法構造解析における技術の高度化として、NMRスペクトルデータの解析、シグナルの帰属、構造情報の抽出（特にNOEの帰属）構造計算など、特に人手と時間を要する過程の効率化・自動化を進めるソフトウェアKUJIRAを開発・改良した（特許出願）。CYANAは、多数のNOEデータの自動解析と立体構造計算を行うソフトウェアで、誤差を減らしながら計算速度を上げることに成功した。現在までに、CYANA version2.0を完成させ、外部に公開した。また、タンパク質の凝集・アミロイド形成に対して、高圧NMR法を適用した。

##### 1.4.2 X線結晶構造解析によるタンパク質構造決定

プロジェクト開始時から平成17年10月末までに、665構造（高度好熱菌等微生物581構造、ヒト・マウス等57構造、その他39構造）を決定した。（解析を行った個々のタンパク質の概要は別紙に記載）。平成16年度には、SPring-8の台風被害というアクシデントがあったものの、計画の遂行を果たすことができた。

X線結晶構造解析における技術の高度化として、自動結晶化観察ロボット導入と結晶化スクリーニング条件セット改良による、結晶化の高効率化や、ビームライン自動結晶操作システム対応のラボ用ロボットシステム開発による結晶スクリーニングの高効率化、また、結晶写真自動評価システム開発や、結晶の回折能を効率的でバイアスの少ないものとするためのプレート上回折評価システムの実用化、さらには、設置したPCクラスター計算機による結晶解析計算の高速化と結晶解析計算の標準スクリプト開発などを行った。一方で、理研構造ゲノムビームラインBL26B1、B2では、膨大な数の結晶試料の迅速回折強度測定の自動化（ビームラインを一括制御するソフトウェア、結晶自動センタリングソフトウェアの開発などを中心とした回折データ収集システムなどの構築）・ロボット化（結晶自動マウントロボットの開発など）による効率的な連続運転の実現や、自動構造解析標準スクリプト構築などを行った。

また、結晶凍結技術開発（クライオ X、Z）や重原子試薬との共結晶化・重原子誘導体調製技術開発についても研究を進めた。さらに高難度タンパク質に対応するためのマイクロ結晶対応自動マイクロ光学系開発

や、高速 X 線検知システムの開発などの、微小結晶に焦点を合わせた技術開発も行った。

## 1.5 決定した立体構造のバイオインフォマティクス

### 1.5.1 International Structural Genomics Organization (ISGO)

我々は、International Structural Genomics Organization (ISGO) による国際的な取り決めにしたがって、ターゲット選択 (アミノ酸配列) 試料の調製、X 線結晶構造解析または NMR 解析による構造解析のプロセスの進捗状況を XML で公開している (企業との共同研究を除く) また、決定した構造座標などは PDB (Protein Data Bank) に登録を行っている。プロジェクト開始時から平成 17 年 10 月末までの PDB 登録数は 1,104 (そのうち低分子化合物との複合体: 70) であり、企業との共同研究または共同研究検討のため、PDB 登録を控えている数は 243 である。さらに、ISGO の目標である基本構造の網羅的な理解に向けて、決定した構造に基づいて新たに構造モデリングが可能になるタンパク質数 (アミノ酸配列データベース中で、配列同一性が、PDB に既存のタンパク質に対しては 30%未満で、我々のタンパク質に対して新たに 30%以上となったタンパク質数) を調べた。平成 14 年度では 28,635、15 年度では 45,212、16 年度では 273,271、プロジェクト開始時から 17 年 10 月末まででは 386,832 の構造モデルを構築可能としたことになる。また、1 構造あたりのモデル可能数は平均 287 で、国際水準以上であり、ISGO の活動に大いに寄与することができた。

### 1.5.2 ドメイン構造 (基本構造)

NMR 解析に進めたドメインを Pfam に基づいて分類すると、これまでに 175 ファミリー (676 個) について 1 個以上の立体構造を決定している。加えて、Pfam ではドメインとして検出するのが困難な新規性の高いドメインを 63 個見だし、そのうち 29 個については立体構造解析を完了している。ヒトタンパク質のドメイン、およびヒトタンパク質のドメインをモデル可能にする他生物種のドメインだけに注目すると、これまでに 163 ファミリーについて 1 個以上の立体構造を決定している。ヒトに於ける各 Pfam ファミリーに該当するドメインのうち、我々が最初に立体構造を決定ないしはモデル可能にしたドメイン数の占める割合をもってそのドメインに対する我々の貢献度とすると、我々が現時点で貢献をしたファミリーは 163 ファミリー中 132 ファミリー (約 81%) ある。そのうち、貢献度が 50% を超えるファミリーは 31 ファミリー (約 19%) であり、特に 11 ファミリーについては我々の貢献度が 100% である。163 ファミリーでの貢献は平均 25.2% である。また、我々の成果により他の成果を用いた場合よりも精度の高いモデルが組めるケースも多数あるが、それらも含めると 163 ファミリー中 160 ファミリー (約 98%) については我々の成果が貢献し、貢献度の平均値は 43.3% となる。現在、スクリーニングによって新たに得られた 186 個のドメインの大量調製の準備を進めている。これらには未着手の 59 Pfam ファミリー、および Pfam では検出できない 15 個の新規ドメインが含まれており、今後さらに多くの基本構造が明らかになると期待される。

X 線結晶解析のヒトタンパク質に対する成果としては、PDB 登録済 496 個には 306 種類 (469 個) の Pfam ファミリーが含まれ、このうちヒトタンパク質に見いだされるドメインは 224 ファミリーある。我々はこのうち 39 ファミリーについて新規に立体構造を決定、もしくは立体構造モデルを作成可能にした。特に 8 ファミリーは、我々が世界で最初にヒトタンパク質での立体構造を決定、もしくは立体構造モデル作成を可能にしたファミリーである。また、これとは別に 54 ファミリーについては、これまで知られていたホモログを用いた場合よりも高精度の立体構造モデルを作成可能にした。このうちでも 31 ファミリーについてはこれまでヒトでの立体構造決定は無く、モデル精度向上の意義は大きい。なお、ヒト・マウスのタンパク質構造のみでなく、細菌と古細菌のタンパク質構造も、ヒトのタンパク質構造のモデリングに寄与している。

### 1.5.3 マルチドメイン構造のドメイン分割計算

平成 17 年 10 月末までに我々が決定した構造座標に対して、構造ドメイン定義自動化プログラム (SCOP、CATH と比較できる構造を用いて良く対応することを確認) を適用して、マルチドメイン構造のドメイン分割を行い、ドメイン構造数を調べたところ、平均で 2 ~ 3 ドメインに分割された。このドメイン数に、シングルドメイン構造として決定した数とあわせて、この時期で既に約 2 千のドメイン構造を決定していることがわかった。したがって、目的とした 2,500 のドメイン構造 (基本構造) という目標に迫っている。

### 1.5.4 機能範囲 (functional coverage)

RSGIによる構造決定対象が、どのような生物学的機能に関連しているかを、生物学的機能の定義として Gene Ontology (GO; <http://www.geneontology.org/>) による生物学的機能カテゴリ (“Biological Process”) を用いて評価したところ、3,065 の機能カテゴリが関連付けられた。これは、全カテゴリ数 10,052 の 30.5% に相当する広範囲の機能をカバーしており、RSGIの対象タンパク質がヒトなどの高等真核生物から細菌・古細菌に及んでいることを反映している。他方、米国 NIH の PSI プロジェクトで構造解析されたものは、主として細菌あるいは古細菌のタンパク質であり、1,150 カテゴリ (全カテゴリの 11.4%) が関連付けられた。

主要な古細菌 (4 種) および細菌 (9 種類) のタンパク質に対し、全構造既知およびモデル可能なタンパク質に占める我々の成果の割合を評価した結果、細菌では 3.8%、古細菌では 6.6% を我々の成果が占めた。本プロ

ジェクト開始時点で既に PDB に約 20,000 エントリーの立体構造の蓄積があり、その後も約 10,000 エントリーが我々とは別に登録されたことを考慮すると、その約 2%の解析数でこれらの成果が得られている点は評価に値する。また、上記の細菌および古細菌のタンパク質を機能別に 11 カテゴリに分類し、それぞれの区分毎に我々の占める成果の割合を評価すると、特に古細菌のタンパク質合成、転写、RNA プロセッシング、アミノ酸合成、機能未知タンパク質、および細菌の RNA プロセッシングでの我々の成果は、それぞれ 9%~10% に達し、際立っている。加えて、我々の成果を用いることで以前よりも高精度のモデルが作れるタンパク質が多数あり(平均 9%; 合成や代謝関連のタンパク質では約 13-26%)、機能解析への応用が期待できる。

## 1.6 高難度ターゲットへの挑戦

### 1.6.1 巨大複合体

巨大複合体や膜タンパク質の構造生物学には、通常のタンパク質と比較して、未だ多くの困難があり、様々な技術開発が必須である。巨大複合体については、構成タンパク質の同定に始まり、活性状態の複合体を調製する技術、結晶化の技術、構造決定の技術などが特に高難度である。我々は、構成タンパク質の同定は、質量分析による解析、個々のタンパク質のリコンビナント発現と複合体の再構成を進めた。例えば、ヒトの翻訳開始因子複合体 eIF3 (600kDa)について 11 個のサブユニット (28~170kDa) を全て同定し、リコンビナントタンパク質から機能性の複合体を再構成することに成功し、現在、結晶化を試みている。リボソームは巨大複合体の代表例であるが、我々は、高度好熱菌の 30S サブユニットと種々の翻訳因子、リボソーム結合タンパク質、RNA などとの複合体の結晶化に成功し、一部については、高分解能の構造決定を終えている。出芽酵母の 80S リボソーム、60S および 40S サブユニットについて、ひとつのリボソームタンパク質に His タグを付与することで、大量に高純度で精製することに成功した。40S サブユニットの結晶化に成功して、現在、回折データの測定中である。また、タンパク質・DNA 複合体についても、特異的な構成成分をもつヌクレオソームのリコンビナントタンパク質からの再構成に成功した。高度好熱菌の RNA ポリメラーゼについては、最近、プロモータ DNA とのオープンプロモータ複合体、転写終結シグナル RNA との複合体などの結晶化に成功した。このような *in vitro* および *in vivo* の巨大複合体の大量調製には、各サブユニットの量比や加える順序、シャペロンの利用などで、複合体ごとの差異が大きく、今後のさらなる技術開発が必要である。

### 1.6.2 膜タンパク質

膜タンパク質、特に、膜貫通領域の多い Integral Membrane Proteins については、試料の大量調製、結晶化は、非常にチャレンジングな課題である。構造解析としては、既に、腸内細菌の V 型 Na<sup>+</sup> ATPase の膜貫通サブユニットと阻害剤 (DCCD) との複合体、光合成細菌の photosystem II などの結晶構造解析に成功している。さらに、ヒトの膜タンパク質をリコンビナントで大量調製するため、昆虫細胞や酵母における過剰発現などの *in vivo* 法に加えて、大腸菌、小麦胚芽、ヒト培養細胞などから作成した無細胞のタンパク質合成系を用いる *in vitro* 法を開発した。合成された膜タンパク質を、界面活性剤で可溶化された状態で蓄積し、精製、活性測定、結晶化へと進む。また、フォールディングのため、可溶化した膜タンパク質をリボソームに再構成し、再び可溶化することも行う。約 300 種類のヒト G タンパク質共役型リセプター (GPCR) を中心として、リセプター、トランスポーターなどの発現を試み、数種類の GPCR で、十分な合成量と純度を達成した。

また、細胞外領域には、多くの糖鎖が付加される場合が多い。結晶化には、不均一な糖鎖は不利である。我々は、CHO 細胞の糖転移酵素欠損株などを用いてタンパク質を調製し、さらに、エンドグリコシダーゼによって可能な限り除去している。これに加えて、RNAi 法により、糖鎖修飾を制御することを試みている。また、ヒト培養細胞由来の無細胞タンパク質合成系を用いて、糖タンパク質の合成に成功している。

## 1.7 立体構造情報利用による制御化合物の探索

### 1.7.1 機能解析

機能解析として、タンパク質の分子間相互作用や制御活性 (転写・翻訳の制御、細胞間・細胞内シグナル伝達系など) の解析を進めた。分子間相互作用解析については、two-hybrid 法、pull-down 法、MALDI/TOF 及び ESI 型の質量分析法などを用いた。また、制御活性解析については、半自動細胞周期解析法、RNA 干渉法等を導入した。これらの実験的な手法に加えて、分子動力学計算による機能解析を検討し、平成 16 年に世界最高速の LSI 「MDGRAPE-3 チップ」の開発に成功した。この技術開発により、原子レベルでのタンパク質の機構を優れた精度で且つ短期間に解明する事が可能となった。さらに、医薬品候補物質とタンパク質の結合を高精度で見積もる事が可能になった。

### 1.7.2 *in silico* スクリーニング

タンパク質構造を創薬に応用するため、決定した立体構造からリガンド結合部位を予測し、*in silico* リガンドスクリーニングを進めた。SARS ウイルスのプロテアーゼの立体構造 (結晶構造、ホモロジーモデル) に基づき、SARS ウイルス増殖阻害活性を有する化合物の取得に成功し、動物実験によって有効性と安全性

を確認した（特許申請済）。現在、薬剤とタンパク質の複合体の X 線結晶構造解析を進め、さらなる最適化を目指している。さらに、細菌の転写・翻訳など必須の酵素群の立体構造を用いて、*in silico* リガンドスクリーニングを行っている。既に、バリル tRNA 合成酵素を阻害する複数の化合物が得られ、黄色ブドウ球菌などに対する抗菌作用と動物細胞に対する無毒性を確認した。現在、安全性等の試験を進めている。また、強い抗カビ剤の候補化合物や、新規抗癌剤の候補化合物を得て、研究を続けている。さらに、C 型肝炎ウイルス、AIDS ウイルス、鳥インフルエンザウイルスなどの抗ウイルス剤の研究も開始した。

### 1.7.3 産業連携

創薬のために、産業界との緊密な共同研究と速やかな成果移転を行うための「パートナー制度」をスタートした。パートナー制度とは、まず、基本的な共同研究契約を締結し、ターゲット情報を共有する。次に、発現に成功したタンパク質試料、決定された構造情報、結合化合物に関する特許などをタイムリーに成果移転する。この制度によって、理研から提供された試料や構造を利用し、企業保有の化合物ライブラリーを用いる研究から生まれる発明は、企業から特許出願する。以上の流れを持つパートナー制度は、産業界との連携や成果移転による研究活性化などに極めて有用である。事例としては、癌関連タンパク質、免疫・炎症関連タンパク質、遺伝性疾患関連タンパク質などが挙げられ、これらについての共同研究を開始した。

## 1.8 理研疾病専門委員会による調査研究

### 1.8.1 医学・薬学分野での重要タンパク質および研究者の調査

タンパク質構造研究に対する医学・生物学研究者からのニーズを調査し、研究テーマの選定に反映し、的確な共同研究を遂行することが極めて重要である。平成 17 年度、タンパク 3000 プロジェクトの委託により、理研疾病専門委員会（委員長 笹月健彦）を設置した。その元で重要疾病ごとの 6 つの分野別委員会（タンパク 3000 プロジェクトの 9 拠点からの推薦に基づいて、全国の大学・研究機関から 7~10 名ずつの委員を選定）をスタートし、今後 5~10 年の疾病関連の医学薬学研究において重要なタンパク質および創薬ターゲットと考えられるタンパク質の調査・検討を行った。現在までにリストアップされた重要タンパク質の総数は 2,381 となり、膜タンパク質、巨大複合体、糖タンパク質も多数含まれている（各 372、77、662）。なお、この中には、我々が独自に調査してターゲットとしている 1,263 のヒト疾病関連タンパク質の大半が含まれていた。さらに、それらの重要タンパク質の機能研究に関して著名な我が国の医薬研究者についても調査を行い、300 名を超えるリストを作成した。他方、情報共有と連携の基盤となる情報管理システムやデータベースの検討を行う「情報プラットフォーム委員会」を設置して、今後の研究ネットワークの構築に備えた。

	分野別委員会名 [主査名(敬称略) 委員数]	ターゲット数	著名な研究者数
A1	脳/神経/アポトーシス委員会 [高橋良輔、10]	621	75
A2	メタボリックシンドローム委員会 [門脇 孝、7]	185	19
A3	免疫/造血/アレルギー委員会 [谷口 克、7]	772	39
A4	癌/白血病委員会 [垣添忠生、8]	316	123
A5	感染症委員会 [北 潔、10]	580	29
A6	消化器・循環器・呼吸器・骨・難病等重要疾病委員会 [笹月健彦、8]	135	57
	ターゲット数合計(非重複)	2381	-
B1	情報プラットフォーム委員会 [五條掘 孝、8]	-	-

### 1.8.2 タンパク質機能研究の緊急課題

疾病別の特に緊急な重要テーマに関しては、分野別委員会の選定した医学・薬学研究者にタンパク質機能研究を再委託し（研究費を配分）、平成 18 年度にタンパク 3000 プロジェクトとの構造解析共同研究を行うための準備を進める。

## 1.9 機能別成果の例

遺伝情報系タンパク質を例として、構造・機能解析の成果を以下に述べる。

### 1.9.1 DNA 複製・修復

ゲノムは、様々な外的・内的要因による変化にさらされているにもかかわらず、常に安定に遺伝情報を保持し、次世代に伝えている。そのような染色体の安定維持に必須であり、生存に不可欠な「DNA 複製・組換え・修復」を担うタンパク質について研究を行っている。ここでは、相同組換えを担うタンパク質構造基盤を初めて解明した 2 つの結果を紹介する。

二重鎖 DNA 切断の相同組換えによる修復機構の中で、重要な相同的対合反応を触媒するヒト Rad52 タンパク質の結晶構造を 2.85 分解能で決定したところ、11 量体のリング構造であることがわかった。リング外周

にある塩基性アミノ酸が局在する溝が、単鎖DNAと二重鎖DNAの両方を結合させ、両者間で相対的対合反応を促す機能部位であった。( *Mol. Cell*, 2002 )

真核生物による減数分裂では相対的対合反応が行われるが、この相同DNA組換えは損傷したDNAの修復にも使用され、細胞が正常に機能するために不可欠である。この現象に関わるタンパク質・Dmc1の立体構造を3.2 で決定した。その結果、Dmc1は、16量体から成る巨大なダブルリング構造をしていることが明らかになった。さらに、中心の穴に二重鎖DNAが、リング側面の穴に単鎖DNAが結合し、ダブルリング構造の内部で相同対合反応が行われた後、中心の穴から組換えが終了したDNAが出てくるという複合体モデルを示した。( *Mol. Cell*, 2004 )

### 1.9.2 転写

ウイルスからヒトに至る多くの生物はDNAを持っていて、DNA上の遺伝情報はメッセンジャーRNA ( mRNA ) へと転写された後、タンパク質へと翻訳され、細胞あるいは生命体全体に伝達される。RNAポリメラーゼ(RNAP)による転写反応を高次構造レベルでたどるため、高度好熱菌RNAP( マルチサブユニット )、T7ファージRNAP ( シングルサブユニット ) などを用いて、転写開始、伸長、終結までの各ステップでの結晶構造解析を目指している。

高度好熱菌 RNAP に 因子が結合した複合体 ( ホロ酵素 ) の結晶構造解析を行った結果、RNA ポリメラーゼ全体がカニのはさみのような構造をとっており、 因子が DNA のプロモーターを認識し結合するのに都合の良い仕組みになっていることが判明した。また、DNA 結合部位の溝が 1 本鎖 DNA のみ結合できるような構造になっていることや、 因子が伸長過程への移行時に解離する機構などを解明した。( *Nature*, 2002a)

緊縮調節におけるRNAP活性制御機構を探るため、グアノシン4リン酸 ( ppGpp ) とRNAPの複合体の結晶構造解析を行った。ppGppがRNAポリメラーゼの活性中心近傍に結合することが初めて明らかになった。さらに結晶構造に基づく分子モデリングや生化学的解析からはppGppがRNAポリメラーゼの活性中心の立体配置に影響を及ぼしたり、DNAと直接相互作用したりすることによって、ポリメラーゼの活性を制御していると示唆された。( *Cell*, 2004b)

転写制御因子 DksA タンパク質の結晶構造解析により、DksA が ppGpp に依存する転写調節に重要な役割を果たしていることを解明した。DksA が RNAP の基質侵入孔内でマグネシウムイオンを介した結合をすることにより、ppGpp と RNAP 複合体の安定化に寄与していることを明らかにし、RNAP の基質侵入孔が複数の転写因子による制御作用のために重要であることを示した。( *Cell*, 2004c)

一方、T7ファージRNAPとDNA: RNAハイブリッドとの複合体の結晶構造解析を行い、転写伸長複合体の構造を明らかにし、開始複合体と比較した。18塩基対からなるDNA: RNAハイブリッドのうち上流の8塩基対の部分がRNAポリメラーゼの活性中心に位置することでRNAポリメラーゼの末端側130アミノ酸残基及び活性中心部分の立体構造が大きく変化していた。転写開始複合体には認められなかったmRNAの出口となるチャンネルが形成された。( *Nature*, 2002b)

さらに、このT7ファージRNAPとDNA: RNAハイブリッド複合体に、基質アナログAMPcPPを結合させて結晶構造解析を行った。では、AMPcPPは、DNA: RNAハイブリッド分子上のDNA鎖の相補ヌクレオチドと塩基対を形成し、非活性型である開いた構造をしたT7RNAポリメラーゼに結合していた。そのため、T7ファージRNAPは、活性型である閉じた構造へ変化する前に、Yヘリックスの上で正しいヌクレオチドを識別していると考えられる。( *Cell*, 2004a)

この他に、転写に関しては、様々な転写制御因子の構造解析を進めている。また、真核生物における転写制御とエピジェネティクスに関連して注目を集めているヒストンの修飾 ( リジンのアセチル化、メチル化など ) についても、構造生物学的な研究を進めている。例えば、ヒトBrd2のプロモドメインの結晶構造解析を行ったところ、他のプロモドメインとフォールドは同じであったが、ユニークな 2 量体を形成しており、10アミノ酸残基にわたってヒストンH4テールを認識し、Lys12がアセチル化されていてLys8がアセチル化されていない状態を認識することを見いだした。これは、ひとつのタンパク質がヒストン・コード ( ヒストン修飾の組み合わせによって制御が規定される ) を認識することの初めての発見となった。さらに、最近、ヒストンH3の脱メチル化酵素であることが判明したLSD1について、そのSWIRMドメイン構造のNMR解析により、H3テールを認識することを初めて明らかにした。(いずれも未発表)

### 1.9.3 RNAのプロセシングと修飾

tRNA は、他の RNA と同様に DNA を鋳型と鋳型として転写合成される。転写された直後の前駆体 tRNA にはその 5'末端や 3'末端に成熟 tRNA には見られない伸長配列が存在し、イントロンを持つものもある。そのような前駆体 tRNA が機能を持った成熟 tRNA になるためには様々なプロセシングやスプライシングを受ける必要がある。まず、3'末端の伸長配列を除去する RNase PH, tRNase Z, tRNA splicing における連続する 3 つの反応 ( 切断、連結、修繕 ) を行う endonuclease, ligase, phosphotransferase、また、CCA 付加酵素などの立体構造を決定した。また、tRNA の成熟過程では、様々な tRNA 機能を達成するために、位置特異的な塩基修飾を行う酵素群が働く。それらの修飾酵素についても精力的に構造解析を進めている。

そのような修飾酵素のひとつ、アーケオシン tRNA グアニントランスグリコシラーゼ ( ArcTGT ) はアク

セスが困難なコア部分の D アームにある G15 位を修飾する酵素であるが、tRNA との複合体の構造解析を行ったところ、酵素が L 字型から I 型へと構造変化を起こした tRNA と結合することを解明した。また、これにより I 型という第二の tRNA の形態を初めて明らかにし（ラムダ型と命名）さらに ArcTGT による配列非特異的かつ位置特異的な tRNA 15 位の認識機構を解明した。（*Cell*, 2003）

#### 1.9.4 翻訳

翻訳に関しても、タンパク質や RNA の立体構造に基づいて、全ての過程のメカニズムを解明することを目指して、まずアミノアシル tRNA 合成酵素、翻訳因子（多くは GTPase）、リボソームなどに注目して、構造解析を進めている。アミノアシル tRNA 合成酵素については、大半の酵素の構造を既に決定した。延長因子 EF-G などについては、初めて活性型の GTP 結合型の立体構造を決定し、リボソーム上での活性制御機構を明らかにした。リボソームについても、いくつかの結晶化に成功している。低分子量 GTPase である Era については、GTP 結合型の構造を明らかにして、リボソーム 30S サブユニットのアセンブリーにおける役割を明らかにした。

Vasa はショウジョウバエ、DEAD-box RNA helicase であり、胚のパターン形成と生殖細胞の形成に必須である。発生中の胚において後極に局在し、そこで特定の mRNA の翻訳を促進する。我々は Vasa と RNA、ATP アナログとの三重複合体の結晶構造解析を行い、DEAD-box helicase について初めて ATP 結合型の構造を明らかにし、他の ATP 依存的モーターとは全く異なる RNA unwinding mechanism を明らかにした。（*Cell*, in press）

## 2 . グループにおけるタンパク質の構造解析について

	平成 17 年 4 月～平成 17 年 10 月末	(参考) 平成 14 年 4 月～平成 17 年 3 月末
(1)PDB登録数 <sup>1</sup>	158 (X 線: 88、NMR: 70)	946 (X 線: 470、NMR: 476)
(2)構造解析を終了したがPDB登録を保留しているタンパク質の数 <sup>1</sup>	115 (X 線: 28、NMR: 87)	128 (X 線: 126、NMR: 2)
(3)PDB登録の有無に関わらず構造解析を終了したタンパク質の数 <sup>1</sup>	157 (X 線: 64、NMR: 93)	1190 (X 線: 603、NMR: 587)
(4)ドメイン単位の基本構造の数 <sup>1</sup>	1007 (X 線: 707、NMR: 300)	989 (X 線: 599、NMR: 390)
(5)新たに構造モデリングが可能になった数 <sup>1</sup>	81130 (X 線:29930、NMR:51200)	305702 (X 線:79727、NMR:225975)

## 3 . 論文掲載数 <sup>2</sup>

	平成 17 年 4 月～平成 17 年 10 月末	(参考) 平成 14 年 4 月～平成 17 年 3 月末
・件数	78	347

## 4 . 成果の産業連携について <sup>4</sup>

	平成 17 年 4 月～平成 17 年 10 月末	(参考) 平成 14 年 4 月～平成 17 年 3 月末
(1)特許出願数（国内）		
理化学研究所が単独出願	13 件	35 件
共同研究で出願	4 件	14 件
（合計）	17 件	49 件
特許出願数（海外）		
理化学研究所が単独出願	1 件	12 件
共同研究で出願	0 件	10 件
（合計）	1 件	22 件
(2)成果の産業移転を目的とした共同研究の件数及び内容	<b>タンパク質の高次構造・機能解析に基づく医薬品候補物質の探索研究</b> (1) 個別創薬共同研究 ( 121 件 ) 医薬品候補物質探索のターゲットとなるタンパク質の高次構造及び機能解析とその情報に基づく医薬品候補物質の探索研究を、独占的な個別契約に基づいて実施している。(以下、契約済み件数) 大学・政府機関 83 件 (内 6 件契約満了、4 件中止) 民間企業 38 件 (内 16 件契約満了、5 件解約)	



	<p>個別創薬共同研究の実例</p> <ol style="list-style-type: none"> <li>1. 「タンパク質立体構造情報に基づく C 型肝炎ウイルス (HCV) 増殖阻害化合物の探索および設計」、山梨大学との共同研究：慢性肝炎、肝硬変、肝細胞癌を引き起こす HCV について、あらゆる型のウイルスに有効な治療薬の開発を目的とする。</li> <li>2. 「インシリコ技術による抗ウイルス分子標的薬の創薬研究」、株式会社インシリコサイエンスの共同研究：タンパク質立体構造情報に基づいた創薬 (SBDD) を実施するにあたり、立体構造を予測するモデリング技術を活用して、抗ウイルス分子標的薬の効率的な研究・開発を目的とする。</li> <li>3. 「高病原性トリインフルエンザウイルス由来タンパク質の立体構造解析および抗ウイルス剤の開発」、鳥取大学との共同研究：現在、大きな脅威となっている高病原性 (強毒型) トリインフルエンザウイルスをターゲットとして、SBDD によって抗ウイルス剤を開発することを目的とする。</li> <li>4. 「SARS (重症急性呼吸器症候群) コロナウイルス増殖阻害化合物の探索」東京医科歯科大学ならびに国立感染症研究所との共同研究：すでに in silico で、SARS ウイルスの増殖阻害の強い活性を示す化合物を得ている。化合物と標的酵素との複合体の X 線結晶構造解析を進め、最適化を行い、実用化を目指している。</li> </ol>
(3)成果の産業移転に関する具体的な例	<ol style="list-style-type: none"> <li>1. 製薬関連企業 タンパク 3000 の成果を効率良く産業移転する目的で、昨年度からパートナー制度を立ち上げた。データ配信は、平成 16 年 5 月下旬から平成 17 年 10 月下旬までに、864 クローンの発現成功タンパク質情報と、49 クローンの構造解析成功タンパク質情報 (ヒト由来 37 クローン、マウス由来 12 クローン) を実施した。これまでに 5 社と契約締結した (2 社は契約終了)。また、別途 2 社が契約内容の検討中である。本制度は、非独占契約を基本としており、独占契約を基本とする通常の個別共同研究 (12 件、内 6 件終了) と相補的な役割を果たしている。</li> <li>2. 技術関連特許 タンパク 3000 で開発した新規技術の内、特許化したものについては、これらの技術の実用化をさらに進め、国内外で広く使われるようにするために、企業へのライセンス活動を実施している。本年度は、非天然アミノ酸導入技術に関するもの 2 件、RNA 固定化技術に関するもの 1 件、新規試薬用酵素に関するもの 1 件、MS 関連技術に関するもの 1 件、NMR 関連技術に関するもの 2 件について、交渉あるいは交渉中である。すでに、安定同位体標識アミノ酸合成技術に関する特許の国内企業へのライセンスが決定している。 結晶化関連技術開発 TERA：竹田理化学工業との共同開発による全自動結晶化観察ロボット「TERA」を、民間企業であるプロテインウエーブが導入し、医薬メーカーからの受託研究のスピードアップに活用している。本機の導入で結晶化成功率が 2 倍以上、タンパク質の受託研究効率は 10 倍以上になった。 タンパク質試料調製技術開発 Cell-free EXPRESS：独自技術である全自動無細胞合成系タンパク質試料調製装置が基礎となり、Cell-free EXPRESS として (株) ワイ・イー・シーの製造により結実した (市販レベルに到達)。</li> </ol>
(4)出願した特許の具体的な例	<ol style="list-style-type: none"> <li>1. 癌細胞に対して特異的致死性を有する、細胞分裂阻害剤の特許を出願した。</li> <li>2. 安定同位体標識アミノ酸合成技術に関し、ライセンス契約締結済の企業とともに開発を行い、有用データに基づいて優先権を伴う出願を行っている。 (その他は § 5 を参照)</li> </ol> <p>平成 14 年度～平成 17 年度の特許出願：理研単独出願 (国内) 48 件、共同研究出願 (国内) 16 件、理研単独出願 (海外) 13 件、共同研究出願 (海外) 11 件</p>
5. 本プロジェクトの推進に係る技術開発に関する成果について <sup>4</sup>	
<p>5.1 タンパク質試料調製のための技術開発</p> <p>5.1.1 高等動植物 cDNA クローンのスクリーニングとタンパク質試料調製のための完全自動化・最適化技術</p> <p>ゲノムにコードされたタンパク質の構造や機能の網羅的解明、すなわち構造ゲノム科学研究には多数のタンパク質試料を迅速に調製できる、ハイスループットなタンパク質発現系が必要である。本プロジェクトでは無細胞タンパク質合成系を核として、cDNA ライブラリーにコードされた多数のタンパク質を同時並行的に、迅速に発現調製する系を構築した。また、さらなる新規手法や自動化開発などを進めている。NMR 法のための安定同位体標識タンパク質の調製や X 線結晶構造解析用の SeMet 導入タンパク質調製に成功した。</p>	

一方で、目的タンパク質の可溶性向上を目指し、様々な発現タグ（融合タンパク質）の検討を行い、劇的な可溶性向上効果や極めて高い精製度を実現した。これらの成果を反映し、ハイスループットな自動化技術として、96穴プレート8枚の同時処理により全工程を10時間で完了し、一日一台あたり約1,800種類のタンパク質を発現することができる自動合成装置を完成させた。これらの合成最適化、ならびに自動化技術開発成果により、NMRスクリーニングの高速化や効率化が達成され、その1週間あたりの最大キャパシティは First Stage : 2112 constructs/w、Second Stage : 384 constructs/w、Third Stage : 144 constructs/w、Large Scale Prep. : 12 samples/w という驚異的な数値となっている。

#### 5.1.2 合成困難なタンパク質などの合成技術の検討と詳細設計

無細胞タンパク質合成系を、創薬に直結した、より高難度タンパク質に応用させるため、膜タンパク質についての研究を重ねている。無細胞タンパク質合成系においては、疎水性の高い領域を有するタンパク質は不溶性の沈殿となる傾向がみられる。そこで、膜タンパク質を不溶化させることなく合成することを目的とし、数種類の界面活性剤を用いて、膜タンパクの合成量や可溶性に与える影響について検討した。その結果、Gタンパク質共役受容体の2アドレナリン受容体の発現に対するジギトニン添加によって大量のタンパク質精製を可能とした。

また、大腸菌由来以外の細胞を用いて、様々な合成困難なタンパク質調製を行うための研究も進行している。ほ乳動物細胞由来無細胞タンパク質合成系の研究においては、eIF4GとeIF2Bの組み換え体をHeLa抽出液に添加し、網状赤血球抽出液を越える活性を得ることに成功し、糖鎖を持つタンパク質を効率よく合成することに成功した。一方、培養細胞で多種のタンパク質を発現させる技術開発においては、*Drosophila* Schneider Line 2 (S2)細胞の適用により、バキュロウイルス系では発現が難しかったヒトおよびマウスタンパク質の細胞外ドメインの分泌発現を行い、活性を保持したタンパク質の迅速な調製に成功した。

21種類以上のアミノ酸からなる「超」タンパク質の作成や、それらのタンパク質どうしを結合させることで、癌や遺伝病のメカニズム解明に極めて有用とされる新規タンパク質合成技術開発を行った。このうち、動物細胞における非天然型アミノ酸のタンパク質への部位特異的導入法とその利用法の開発では、パラベンゾイルフェニルアラニンを動物細胞内においてタンパク質 Grb2 に組み込んだところ、Grb2は細胞内で、標的である上皮成長因子受容体と相互作用することが確認された。さらに、細胞への特定波長の光照射により、受容体との共有結合による架橋タンパク質の作成に成功した。この技術開発により、解析したいタンパク質の遺伝子を用意し、人工的アミノ酸とともに培地に加えるだけでタンパク質の指定の位置の置換が可能となり、多数のタンパク質が関与するネットワークの解明への活用が見込まれる。

### 5.2 NMR解析のための技術開発

#### 5.2.1 NMR解析技術の自動化ならびに高度化

タンパク質の立体構造を解析する多次元NMR法が発展し、分子量3万程度のタンパク質についても構造決定が可能になった。本プロジェクトではスペクトルデータの解析やシグナルの帰属、構造計算およびそれらデータの管理を行うソフトウェア開発とより高解像で高分子量対応のNMR法の技術開発を行っている。

KUJIRAは、NMRスペクトルデータを柔軟にかつ迅速に解析できるよう工夫されたNMR解析ツールである。主な機能は多くの2次元あるいは3次元スペクトルデータの統合的管理とシグナル帰属データの管理、他のソフトウェアで計算された立体構造の解析、他のソフトウェアで解析されたNOE帰属データの解析である。帰属速度の迅速化などの改良を重ね、Quick assign (QA)法と Simulated annealing (SA)法を開発し、これらを組み合わせることで帰属の正答率を向上させた。自動化されたNOE帰属・立体構造計算のためのCYANAソフトウェアパッケージは、大幅なバージョンアップ (CYANA version 2.0) を果たした。CYANAは大規模なRIKEN Super Combined Cluster (RSCC)システム上でプログラムを実行することによりCYANAの処理速度の向上を図っている。また、ケミカルシフト帰属の欠如が自動NOESY帰属を用いた構造計算の結果に及ぼす影響の減少、さらに、自動構造計算の過程で新たにケミカルシフト帰属を得るための拡張や、多量体を形成するタンパク質の構造を計算するアルゴリズム開発などを行った。

#### 5.2.2 NMR計測技術の高度化

NMR計測技術の高度化としては、高磁場化によるNMR法の発展を目指している。まず、920MHz NMRの4ケルビンプローブ開発の第1段階として、500MHzでの4ケルビンプローブの開発を進めた。次に、1GHz NMRスペクトロメータの開発に向けたNMR磁石の開発において、独立行政法人 物質・材料研究機構に協力し、1GHz NMR超伝導磁石の実証試験として、物質・材料機構にある大口径21T超伝導マグネットを使用し、1GHz NMR磁石の発生磁場23.5Tを目標とする超伝導磁石の製作を行った。さらに、NMR磁石の永久電流モード運転に不可欠な超伝導接続の開発を行った。今回は1GHz NMR磁石最内層コイルの有力候補である線材についての超伝導接続の新たな接続方法を開発した。結果として、接続部の電圧・電流測定において磁場中 (~0.4T)でも超伝導状態が保たれている事を確認した。

### 5.2.3 固体NMRと新規磁場配向技術開発、<sup>17</sup>Oの利用など

NMR法を用いた新規手法の開発も行った。酸素の固体NMR開発においては、ペプチドやタンパク質中の酸素を質量数17の同位体でエンリッチする方法を開発した。合成用保護基付きアミノ酸を原料にし、活性エステル型を使うことで温和な条件で酸素17水からアミノ酸に酸素原子を移すことに成功した。

膜結合型タンパク質を配向させるための新規磁場配向二重膜の開発においては、パイセルの短鎖部分を中性界面活性剤に置換した脂質二重膜で磁場配向実験を試し、その親水・疎水部に取り得る炭素鎖長を詳細に検討した。さらに、陰イオンとの相互作用についても配向試料の四重極核観測から遂行し、双極子モーメントの差が脂質二重膜の磁場配向度に強く影響することがわかった。

磁場配向膜への種々の脂質ドーピング系の開発と膜結合型タンパク質との相互作用解析においては、磁場配向膜にコレステロールをドーピングしたところ、脂質二重膜溶解ペプチド(メリチン)に対する安定性が著しく向上した。また、DMPGのような酸性脂質のドーピングにより、塩基性残基局在により脂質膜と相互作用する膜表在タンパク質との相互作用を溶液NMR法で観測する系を世界で始めて確立した。異方的構造情報を駆使したタンパク質立体解析法については、磁場配向膜が、水溶性タンパク質の立体構造決定にも有用であることを示し、糖タンパク質(50kDa)やマルチドメインタンパク質(30kDa)に適用した。

最適制御理論の適用による高分子のNMR法の感度増大については、コヒーレンス移動の効率を最適制御理論による緩和最適化パルスシーケンスを設計することにより最大化し、分子量800kDaのシャペロンタンパク質GroELのような、生体高分子の系にも導入できることを示した。この系においては交差相関緩和誘導分極移動(cross-correlated relaxation induced polarization transfer, CRIPT)を用いた場合に、<sup>1</sup>H-<sup>15</sup>Nコヒーレンス移動の効率が従来の方法と比較して平均して20-25%増加することが判明した。

### 5.3 X線結晶構造解析のための技術開発

世界最大規模の第三世代大型放射光施設SPring-8における、理研構造ビームライン(BL26B1&B2)を有効に活用し、従来は煩雑な過程となっていたタンパク質結晶化や構造決定をハイスループット化するため、本プロジェクトでは、X線結晶構造解析の作業効率向上や自動化を含めた新しい技術開発を進行させ、構造ゲノム科学に対応している。

まず、タンパク質結晶化のプロセスはX線結晶構造解析の過程において最も手間のかかる作業とされるが、それは数百から数千条件を用いる結晶化スクリーニングや、その膨大な条件下での結晶化ドロップ観察という過程を、熟練した研究者が一つ一つこなさなければならなかったからである。TERAは本過程を全自動化するロボットで、オイルマイクロパッチ法を採用した結晶化分注ロボット、結晶化ウェル自動観察装置、2500枚の結晶化用プレートと125枚の試薬プレートの保管棚、各々の装置間を連携するロボットハンドを持つプレート搬送系ロボットからなるシステムである。本システム開発によって、タンパク質結晶化作業を24時間行えるようになり、作業効率と結晶化の再現性が大幅に向上した。さらに、セミマイクロ単位の微少結晶化技術開発では、セミマイクロ結晶化分注装置に加えて、セミマイクロ自動結晶化観察ロボットを導入し、その分注精度および再現性を検証するとともに50ミクロン以下の微小結晶の画像を鮮明に得ることを可能にした。

次に、得られた結晶は、構造解析に適しているか評価する。ここでは実際にX線を照射して評価を行うが、1つずつの結晶をその都度、顕微鏡で取り出すことなく、一括して結晶化プレート上処理可能な迅速X線結晶評価装置を開発し、作業の効率化を進めた。これにより良い結晶を素早く選別することが可能となった。

選別された結晶は、理研構造ゲノム科学ビームライン(BL26B1&B2)などにより解析されるが、本ビームラインは、凍結サンプルチェンジャーSPACE(SPring-8 Precise Automatic Cryo-sample Exchanger)の導入による自動運転が特徴である。分光器等ビームラインの光学機器はすべてSPring-8標準の装置を採用し、ビームライン制御ソフトBSS(Beamline Scheduling Software)を利用することでX線の波長変更から回折データ収集まで、一貫した自動運転を実現した。

また、ビームライン用サンプルの事前評価・準備システムとして、実験室系のX線回折装置にも凍結サンプルチェンジャーSPACEと同じサンプルトレイを使用する自動回折強度データ収集システムの開発導入を行い、実験室で一日に可能な回折チェック結晶数が従来の2.5倍に増加した。さらに、比較的良好な結果を得られた結晶を簡単に選択・収集可能であり、また、BL26B2においてデータセットを自動測定できるため、全体的に大幅な効率化に繋げることができた。さらに、数十ミクロンの微小結晶評価と解析を目標にした、偏向磁石ビームライン用の高輝度マイクロビーム光学系の技術開発や、X線検出器の高速化も進行し、X線CMOS検出器をベースに大面積高速フラットパネル検出器の開発による低ノイズ化を推し進めている。

一方で構造解析の成功のためには、多くの結晶解析ソフトウェアの組み合わせを含めた具体的実験経験が欠かせず、これらのノウハウをアルゴリズム化した自動化が大規模解析に果たす役割は大きい。自動構造解析の標準スクリプト構築においては、自動解析システム及び構造評価システムのソフトウェア開発を行った。自動解析システムは標準化された解析手順を自動化するもので、重原子による実験的位相付けに対応するため、データベース検索による重原子スクリーニング条件候補の提示、重原子データによる位相計算、自動モ

デル構築、構造精密化までを視野に入れている。また、構造評価システムは、解析された構造座標に基づき、主に構造比較と相互作用様式の標準データを自動作成してデータベースへ格納し、さらにこのデータをもとに論文等に直接利用できる資料の自動作成機能により、構造評価作業が効率化できるというものである。

#### 5.4 分子動力学専用計算機プロテインエクспロアの開発、タンパク質科学および創薬への応用研究

MDGRAPE-3 (プロテインエクспロア) の開発においては、分子動力学計算専用 LSI MDGRAPE-3 チップを開発し、さらに性能評価、量産を行った。当初の目標を上回る 300MHz 以上の周波数での動作が確認され、198Gflops 以上のピーク性能を達成した。冷却性能評価ボードを開発し、MDGRAPE-3 ボードの開発を行った。このボードを筐体に搭載して十分に冷却可能であることを確認した。高速シリアル通信ボード経由で PC に接続し、PC との間の転送速度が目標性能 600 ~ 800Mbyte/s に達していることを確認した。また、PC に直接装着・搭載可能な PCI-X 規格準拠ボードの製造も行った。さらに、分子動力学計算パッケージの専用計算機への移植においては、専用計算機を利用するための基本的なライブラリーを作成し、分子動力学計算パッケージ Amber-8 の基本部分の移植を行った。孤立系で 13,000 原子の系で比較したところ、PC 一台に PCI-X 版 MDGRAPE-3 カードを一枚挿入することにより、150 倍計算速度を向上させることができた。さらにリアルタイムの対話型分子動力学シミュレーションの改良を行なった。このシステムは、薬剤分子とタンパク質との相互作用を体感することができるため、将来薬剤設計に応用できる可能性がある。

### 6 . 平成 16 年度の評価に対する反映方法について

平成 16 年度の評価に対応して、よりいっそう高難度のターゲットに対する取り組みを強化した。また、評価意見に従い、医学・薬学の研究者との強い連携体制を構築することとし、疾病研究において重要なタンパク質の優先付けに反映させるための調査研究を行い、併せてタンパク質機能の研究者の調査も行うことにより、共同研究ネットワークの基礎作りを行った。産業界との個別共同研究については、定期打合せなどに基づき、解析ターゲットの優先順位付けに反映させた。その結果、共同研究による構造・機能の総合的な理解を目指す研究体制の確立に向けて大きく進展した。今後は、広範な生命科学分野の研究者との連携についても、同様の取り組みを行っていききたい。さらに、評価結果を反映して、高度なタンパク質試料調製技術の成果移転として、立体構造解析試料としてだけでなく、機能解析や抗体作成などのための試料の生命学者に対する供給を試験的に開始し、十分な手応えを得ている。また、技術開発をはじめとする若手研究者の活躍を推進し、ソフトウェアの開発などの充実が強化された。

### 7 . 中核機関としての独自の目標 ( 解析数、特許出願数等 ) に対する達成度、定期的な見直し体制等について

現在のところ、順調に目標を達成している。次年度の計画についても、§1 においても記述したように、広範な研究者と共同して、弾力的に、最適な研究体制の構築に向け努力している。

### 8 . 中核機関として、外部への広報をどのように実現しているか <sup>5</sup>

パンフレット、ホームページ ( <http://www.rsgi.riken.jp>; <http://protein.gsc.riken.go.jp> ) による研究体制、研究成果などの紹介や、国際シンポジウム開催 ( 構造ゲノム科学国際会議や、日英シンポジウムなど ) による、若手からシニアの研究者の成果発表・交流の場を積極的に設けた。また、解析したタンパク質構造情報については、Protein Data Bank (PDB) への登録と、ISGO の国際ルールである Target DB での解析進捗状況の開示に努め、またその構造情報を構造予測の技術錬磨のためのオリンピックと呼ばれる Critical Assessment of techniques for protein Structure Prediction (CASP) に提供、および、ISGO の代表としての国際機関の先導や、そのホームページ ( <http://www.isgo.org/> ) 立ち上げなど、国際貢献にも広く尽力した。また、無細胞タンパク質合成技術や、NMR ソフトウェアを中心とするハンズオンワークショップを定期的に行い、世界各国からの参加者への技術普及を行った。

9 . 各年度の委託費 ( 百万円 )	平成 14 年度	平成 15 年度	平成 16 年度	平成 17 年度	計
	7,000 (7,195)	3,500	5,100	5,268	27,958

(別紙) 論文のリスト (グループとしての成果を表す代表的な論文10~20編程度)

1	<u>Terada, T., Nureki, O., Ishitani, R., Ambrogelly, A., Ibba, M., Söhl, D., and Yokoyama, S., "Functional convergence of two lysyl-tRNA synthetases with unrelated topologies", Nat. Struct. Biol., 9(4), 257-262 (2002).</u>	1IRX
2	<u>Vassilyev, D. G., Sekine, SI., Laptenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S., and Yokoyama, S., "Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution", Nature, 417(6890), 712-719 (2002).</u>	1IW7
3	<u>Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T., and Yokoyama, S., "Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form", Mol. Cell, 10(2), 359-371 (2002).</u>	1KN0
4	<u>Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, JH., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S., "Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains", Cell,</u>	1IVO
5	<u>Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassilyev, D. G., and Yokoyama, S., "Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution", Nature, 420(6911), 43-50 (2002).</u>	1H38
6	<u>Sekine, SI., Nureki, O., Dubois, D.Y., Bernier, S., Chênevert, R., Lapointe, J., Vassilyev, D. G., and Yokoyama, S., "ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding", EMBO J., 22(3), 676-688 (2003).</u>	1J09, 1N75, 1N77, 1N78
7	<u>Yamauchi, E., Nakatsu, T., Matsubara, M., Kato, H., and Taniguchi, H., "Crystal structure of a MARCKS peptide containing the calmodulin-binding domain in complex with Ca<sup>2+</sup>-calmodulin", Nat. Struct. Biol., 10(3), 226-231 (2003).</u>	1IWQ
8	<u>Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S., "Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme", Cell, 113(3), 383-394 (2003).</u>	1J2B
9	<u>Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., and Yokoyama, S., "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion", Nat. Struct. Biol., 10(6),</u>	1J1U
10	<u>Okabe, M., Tomita, K., Ishitani, R., Ishii, R., Takeuchi, N., Arisaka, F., Nureki, O., and Yokoyama, S., "Divergent evolutions of trinucleotide polymerization revealed by an archaeal CCA-adding enzyme structure", EMBO J., 22(21), 5918-5927 (2003).</u>	1UET, 1UEU, 1UEV
11	<u>Shimbo, I., Nakajima, R., Yokoyama, S., and Sumikura, K., "Patent protection for protein structure analysis", Nat. Biotechnol., 22(1), 109-112 (2004).</u>	
12	<u>Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S., and Vassilyev, D. G., "Structural basis for substrate selection by T7 RNA polymerase", Cell, 116(3), 381-</u>	1S0V
13	<u>Matsubara, M., Nakatsu, T., Kato, H., and Taniguchi, H., "Crystal structure of a myristoylated CAP-23/NAP-22 N-terminal domain complexed with Ca<sup>2+</sup>/calmodulin", EMBO J., 23(4), 712-718</u>	1L7Z
14	<u>Kise, Y., Lee, SW., Park, SG., Fukai, S., Sengoku, T., Ishii, R., Yokoyama, S., Kim, S., and Nureki, O., "A short peptide insertion crucial for angiostatic activity of human tryptophanyl-tRNA synthetase", Nat. Struct. Mol. Biol., 11(2), 149-156 (2004).</u>	1ULH
15	<u>Nakamura, H., Kumita, H., Imai, K., Iizuka, T., and Shiro, Y., "ADP reduces the oxygen-binding affinity of a sensory histidine kinase, FixL: The possibility of an enhanced reciprocating kinase reaction", Proc. Natl. Acad. Sci. USA, 101(9), 2742-2746 (2004).</u>	
16	<u>Artsimovitch, I., Patlan, V., Sekine, S., Vassilyeva, M. N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassilyev, D. G., "Structural basis for transcription regulation by alarmone ppGpp", Cell, 117(3), 299-310 (2004).</u>	1SMY, 1T0X
17	<u>Kinebuchi, T., Kagawa, W., Enomoto, R., Tanaka, K., Miyagawa, K., Shibata, T., Kurumizaka, H., and Yokoyama, S., "Structural basis for octameric ring formation and DNA Interaction of the human homologous-pairing protein Dmc1", Mol. Cell, 14(3),</u>	1V5W

18	<u>Hanawa-Suetsugu, K., Sekine, SI., Sakai, H., Hori-Takemoto, C., Terada, T., Unzai, S., Tame, J. R. H., Kuramitsu, S., Shirouzu, M., and Yokoyama, S.</u> , "Crystal structure of elongation factor P from <i>Thermus thermophilus</i> HB8", Proc. Natl. Acad. Sci. U S A, 101(26), 9595-9600 (2004).	1UEB
19	<u>Perederina, A., Svetlov, V., Vassilyeva, M. N., Tahirov, T. H., Yokoyama, S., Artsimovitch, I., and Vassilyev, D. G.</u> , "Regulation through the secondary channel-structural framework for ppGpp-DksA synergism during transcription", Cell, 118(3),	1TJL
20	<u>Saito, K., Kigawa, T., Koshiba, S., Sato, K., Matsuo, Y., Sakamoto, A., Takagi, T., Shirouzu, M., Yabuki, T., Nunokawa, E., Seki, E., Matsuda, T., Aoki, M., Miyata, Y., Hirakawa, N., Inoue, M., Terada, T., Nagase, T., Kikuno, R., Nakayama, M., Ohara, O., Tanaka, A., and Yokoyama, S.</u> , "The CAP-Gly Domain of CYLD Associates with the Proline-Rich Sequence in	1IXD
21	<u>Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tomo, Y., Hayami, N., Terada, T., Shirouzu, M., Osanai, T., Tanaka, A., Seki, M., Shinozaki, K., and Yokoyama, S.</u> , "Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1", Plant Cell, 16(12), 3448-3459	1WID
22	<u>Kobayashi, T., Sakamoto, K., Takimura, T., Sekine, R., Vincent, K., Kamata, K., Nishimura, S., and Yokoyama, S.</u> , "Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion", Proc. Natl. Acad. Sci. USA, 102(5), 1366-1371 (2005).	1WQ3, 1WQ4, 1VBN
23	<u>Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K., and Yokoyama, S.</u> , "Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid", Nature Methods, 2(3), 201-206 (2005).	
24	<u>Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tomo, Y., Hayami, N., Terada, T., Shirouzu, M., Tanaka, A., Seki, M., Shinozaki, K., and Yokoyama, S.</u> , "Solution structure of an Arabidopsis WRKY DNA binding domain", Plant Cell, 17(3), 944-956 (2005).	1WJ2
25	<u>Sharma, M. R., Barat, C., Wilson, D. N., Booth, T. M., Kawazoe, M., Hori-Takemoto, C., Shirouzu, M., Yokoyama, S., Fucini, P., and Agrawal, R. K.</u> , "Interaction of Era with the 30S ribosomal subunit: Implications for 30S subunit assembly", Mol. Cell, 18(9), 319-329 (2005).	
26	<u>Baba, D., Maita, N., Jee, JG., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M.</u> , "Crystal structure of thymine DNA glycosylase conjugated to SUMO-1", Nature, 435(7044), 979-982 (2005).	1WYW
27	<u>Fukunaga, R. and Yokoyama, S.</u> , "Aminoacylation complex structures of leucyl-tRNA synthetase and tRNA <sup>Leu</sup> reveal two modes of discriminator-base recognition", Nat. Struct. Mol. Biol., 12(10), 915-922 (2005).	1WZ2
28	<u>Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S., and Cusack, S.</u> , "The crystal structure of leucyl-tRNA synthetase complexed with tRNA <sup>Leu</sup> in the post-transfer-editing conformation", Nat. Struct. Mol. Biol., 12(10), 923-930 (2005).	2BYT, 2BTE