

第4章

タンパク質の全基本構造の解明

《「タンパク3000」プロジェクト》

ゲノム科学総合研究センター（GSC）でタンパク質の基本立体構造の解析に関わったのが、横山茂之（現横山構造生物学研究室上席研究員）のグループである。NMRとX線により、タンパク質の構造解析に取り組んだ。このグループの特徴は、国家プロジェクトである「タンパク3000プロジェクト」（2002-2006年度）の中心となったことである。さらに後継の「ターゲットタンパク研究プログラム」（2007-2011年度）と「創薬等支援技術基盤プラットフォーム事業」（2012-2016年度）も担当した。

「タンパク3000」は2007年3月に終了したが、その後の10年の研究によって、実は、ライフサイエンスの歴史において極めて重大な成果を上げていたことが明らかになった。2007年以降2015年まで、アメリカで同様のPSIプロジェクトが継続されたが、タンパク質の新しい立体構造タイプ（フォールド・ファミリー）は発見されなかった。つまり、「タンパク3000」の時代に、国際的な目的であった「タンパク質の全基本構造の解明」は、すでに達成できていたことが判明したのである。

横山グループは2008年のGSC終了を受けて、生命分子システム基盤研究領域（SSBC、横山領域長）として独立した。その後、2013年の第3期中期計画によって、白水美香子をリーダーとする構造・合成生物学部門として、ライフサイエンス技術基盤研究センター（CLST）へと統合され、新たに横山構造生物学研究室も設置された。

第1節 構造ゲノム科学・構造プロテオミクス

タンパク質の全基本構造の解明

2003年、ヒトゲノムの全塩基配列が解読された。ほぼ同時期に、mRNAすなわち完全長cDNAも解読された。そして2015年、タンパク質の基本構造を解明する国際的なプロジェクトも成功裡に終了した。

次世代シーケンサーの登場により、現在の生物学・医学・生命科学では、ゲノムやRNAを前提とした分析的な研究手法が標準化している。病気や成長など生命現象の何らかの変化を、DNAやRNAの変化と関連づけるアプローチである。一方、ゲノム編集技術の登場によって、ゲノムを自由に書き換えてそれが作り出す変化や影響をみる、という合成的なアプローチも登場した。タンパク質に関しても、同じようなアプローチが登場すると予想される。

DNA、mRNA、タンパク質と並べたとき、DNAとmRNAは情報を担う物質である。一方、その遺伝情報が実際の機能を担う物質へと翻訳されたものがタン

パク質である。つまり、タンパク質は生命の基本物質であり、その多様性が今回全て解明されたということである。

もう少しはっきりさせよう。遺伝情報をもとにタンパク質は作られる。しかしその種類は無限にあるわけではない。生命体には限られた種類のタンパク質しか存在しないのである。そのことが今回はっきりした。タンパク質の種類は多様であるが、有限の種類しか存在しないことがはっきりしたのである。

生命体の基本物質であるタンパク質。この物質群は、特殊な閉じた世界を作り上げていると言える。その閉じた物質群が、生命体を構成し、多彩な生命活動を展開する。このことが明確になり、物質という面を表に出したアプローチも始まり情報科学を駆使した合理的な創薬法も登場している。

構造生物学と「タンパク3000」

2002年度から始まった「タンパク3000」は、国をあげた国家プロジェクトであり、理化学研究所は、研究者・研究グループ・研究センターのレベルを超え、組織全体として取り組んだ。推進本部を設置して、理事が本部長を務めた。横浜研究所と播磨研究所に加えて、和光本所も参加した。ゲノム科学総合研究センターでも、タンパク質構造・機能研究グループを中心に、多くの研究者が参加した。数千種類のタンパク質の立体構造を解明するため、横浜研究所には大規模NMR施設が、また播磨研究所にはハイスループットファクトリー（HTPF）が設置されたのである。

このような理研の全所的な取り組みにより、国際的な潮流を主導するとともに、結果においても日本は最も多くの成果を上げた。なかでも理研は、日本が担当した全タンパク質の半分以上（約2500構造）の構造解析を達成し、世界一の貢献を果たした。一つの研究機関だけでここまで実績を上げた例は珍しい。

構造生物学の発展史を振り返ると、NMR（核磁気共鳴）、無細胞タンパク質合成、放射光などの重要技術の進歩が、「タンパク3000」の5年間と軌を一にしていることがわかる。つまり、このプロジェクトは技術の進歩を著しく加速し、構造生物学におけるインフラ整備、人材育成にも大きく貢献したのである。

構造ゲノム科学（ゲノミクス）とは

和田昭允が提唱した「オミックススペース」の概念によれば、ゲノム（DNA）を最下層とし、セントラルドグマに従って、その上にトランスクリプトーム（RNA）、さらにその上にタンパク質（プロテオーム）が位置する。これらの分子的基础の上に、低分子の代謝化合物（メタボローム）、細胞・組織・個体の表現形（フェノーム）が位置する。このような階層を貫くライフサイエンスを展開するために重要な役割を担ったのが、理研のゲノム科学総合研究センター（GSC、1998-2013、和田が初代センター所長）であった。

このGSCにおいてタンパク質の階層を担当したのが、横山茂之をリーダーとするタンパク質構造・機能研究グループ（PRG）であった。彼らが目指したのは、タンパク質の立体構造に基づいて、その機能を解明することであった。それが構

造ゲノム科学（構造ゲノミクス、構造プロテオミクス）である。

ここで、呼び名に「ゲノム」がある理由は、ゲノムやトランスクリプトームからの遺伝暗号に基づくアミノ酸配列情報の取得や、遺伝子発現（転写・翻訳）の情報を基礎として研究を展開するからである。理研では、ゲノム解析、トランスクリプトーム解析とほぼ同時期に、この構造ゲノミクス・構造プロテオミクスの研究も始まった。

生体分子の違い

生体分子の機能は、基本的には、全てその立体構造から生み出されている。しかし、タンパク質と核酸には大きな違いがある。核酸の場合、DNAやRNAの塩基配列に意味があり、そこに記された情報は、情報学的に分析できる。しかしタンパク質の場合、意味があるのはアミノ酸配列ではなく、タンパク質分子がとる全体の立体構造そのものである。

要するに、核酸は情報であり、タンパク質は物質（もの）である。他の生体分子、例えば脂質（生体膜）やいろいろな低分子化合物なども、タンパク質との相互作用を通じて機能を発揮している。このようなことを見れば、タンパク質が生命の最も基本的な物質であることは明らかである。

構造ゲノム科学・構造プロテオミクスが目指したのは、生命現象を分子的に解明するための基盤を確立することであった。さらに、自然科学全体の視点からいえば、物質科学と生命科学を真に融合させて、例えば、観察と直感に基づく実験設計から、理論的予測に基づく実験設計へとパラダイム転換させることであった。

分子生物学における構造学派と情報学派

理研における構造ゲノム科学・構造プロテオミクスは、柴田武彦および横山茂之の提案から始まった。柴田によれば、分子生物学は、本来は、構造学派（二重らせん構造、タンパク質構造など）と情報学派（遺伝学など）が車の両輪であった。ところが、塩基配列のような情報解析と比べ、構造解析のスピードははるかに遅く、構造学派はずっと後ろから追いかける形になっていた。ゲノムの塩基配列が解読されればその差は開くばかりで、構造学派と情報学派のアンバランスは深刻化する恐れがあった。そのような事態を回避するため、先鋭的な戦略が立てられ、それに基づくプロジェクトが具体的に提案された。

はじめは、ゲノム塩基配列情報の真の意味を理解するには立体構造情報が必須であるという趣旨から、「ゲノム・デコーディングプロジェクト」と命名された。しかし、ゲノム解析（塩基配列の読み取り）も目処が立たないうちに、このような命名は誤解を与えるとの指摘を受け、「タンパク質基本構造解明計画」と変更された。

タンパク質の基本的な立体構造

1992年、英国のチョシア（Cyrus Chothia）は、「タンパク質の立体構造の多様性は無限ではなく、比較的少数の系統（フォールド・ファミリー）に分類され、

その総数は約1000と推定される」という仮説を提唱した。この仮説は、タンパク質の立体構造の全体像を明らかにしようとする機運を高めた。しかし、どれだけの数のタンパク質構造を実験的に決定すれば全体像がつかめるのか、それが不明では、プロジェクトとして企画するのは難しい。

理研では、横山が次の項で述べるようなタンパク質立体構造のモジュラリティーに注目して、ヒト・マウスの「ドメイン」にフォーカスした戦略を立てつつあった。一方、大阪大学教授の倉光成紀は、まったく異なる視点から、ゲノムサイズの小さい高度好熱菌のタンパク質の網羅的構造解析を提案していた。横山と倉光は、これらの二つの戦略を組み合わせることで、立体構造の多様性に関する全体像が得られると確信した。

柴田、井上頼直、飯塚哲太郎らの主任研究員は、理研における構造生物学の振興を願っており、このアイデアに関心を寄せ、強力に支援した。方法論としては、高度好熱菌のタンパク質はSPRING-8による結晶構造解析で決定し、ヒト・マウスのドメインは無細胞タンパク質合成で調製してNMRで構造決定する。技術的インフラも登場したところであり、絶好のタイミングであった。

ドメインからの基本構造解明計画

ヒトなどの多細胞真核生物は、数万種のタンパク質を持つと考えられていた（実際は、ゲノム解析により、約2万数千の遺伝子を持つことが後に判明する）。それらのタンパク質のアミノ酸配列には、高い類似性（ホモロジー）を示す領域（数十～数百アミノ酸残基）が見いだされ、シーケンス・ドメインとして同定することができる。

ドメインは、一つの生物種の中において複数種のタンパク質の間で保存されたり、生物種を超えて保存されたりすることが普通に見られ、一つの系統（ファミリー）を形成する。ドメインには立体構造も保存されており、何らかの分子機能に対応すると予想される。ドメインはまた、立体構造と分子機能を担うモジュール（単位）として組み合わせられ、さまざまなタンパク質を構成する（すなわち、いろいろな目的に使い回される）。この性質をタンパク質立体構造の「モジュラリティー」とよぶ。

あるドメインが異なるドメインと組み合わせられていたり、あるいは、異なる順序で組み合わせられていたりすることも多く、各ドメインには、他のドメインと立体構造を形成する際にも独立性が保たれていると期待される。そこで、ドメインを基本として切り出すことで、少数の構造解析で全体像に迫る、というのが第一の戦略である。

cDNAの比較

ドメインのタンパク質試料を発見するには、cDNAが必要である。ヒトゲノム解析の終了前という時期ではあったが、わが国の特色として、完全長cDNAの収集が推進されていた。そこで、東京大学医科学研究所の菅野純夫博士、かずさDNA研究所の大石道夫博士、小原収博士からヒトの完全長cDNAライブラ

リーの提供を受けた。他方、理研GSCでは、林崎良英主任研究員からマウス、篠崎一雄主任研究員からシロイヌナズナ（植物）について、完全長cDNAの提供を受けた。この完全長cDNAライブラリーの構築という国際的にも大きなアドバンテージを活かし、ドメインのアミノ酸配列を網羅的に検討したのである。

例えばヒト・マウスとシロイヌナズナの間を比較すると分かるが、ヒトとマウスのように、進化的な距離が近づくほど（つまり分化してからの歴史が浅いほど）、アミノ酸配列のホモロジーが高い。また、ドメインは、スペーサーとなるアミノ酸配列を介して連結され、ヒトとマウスの間でも、スペーサーのアミノ酸配列の保存性は、ドメイン内部よりも有意に低いことが多い。このため、配列保存性の高い領域（すなわちドメイン）を同定しやすい。

ただし、同じシーケンス・ファミリーに属すると分類されても、タンパク質の違いや生物種の違いによって、アミノ酸配列は少しずつ異なる。メンバーの非常に多い（すなわち大きな）シーケンス・ファミリーでは、アミノ酸配列の多様性も高いことが多く、全てが同一の機能に対応するとは限らない。そこで、立体構造解析を体系的に進め、ドメインの機能上や構造上の差異、あるいは進化の経緯までを捉え、全体像を明らかにしようと考えた。

高度好熱菌丸ごと一匹プロジェクト

進化的に距離が離れた生物間でも、同様の機能を担うタンパク質であれば、活性部位の立体構造が互いによく似ていることがある。つまり、立体構造はアミノ酸配列よりも保存性が高いのである。実際、基本的な酵素類は、一つの触媒ドメインだけで比較的大きなファミリーを形成しており、それらのタンパク質は、細菌、古細菌から真核生物まで保存されている。

基本的タンパク質群では、より複雑な複合体や、翻訳後修飾によって機能的には高度になっても、基本となる触媒ドメインの立体構造は保存されると考えられる。そこで、まず、主に触媒ドメインから成るような基本的なタンパク質群を対象として、立体構造を網羅的に解析する。これが第2の戦略であり、そのために、基本的なタンパク質群からなる生物（高度好熱菌など）を選択したのである。

1995年、倉光教授は、「高度好熱菌丸ごと一匹プロジェクト」を提案していた。これは、85℃までの高温で生育する高度好熱菌 *Thermus thermophilus* HB8（大島泰郎博士が日本の温泉で単離した細菌）をモデル生物として、ゲノム解析による全遺伝子の塩基配列の決定、網羅的結晶構造解析による全タンパク質の立体構造の決定、立体構造を活用した機能解析を目指すという、画期的な提案であった。

倉光提案は、独立栄養で増殖することを重視して細菌を選び、また、ヒトを含む全生物への普遍化を見据えたものであった。高度好熱菌など極限環境に生育する微生物のゲノムは、通常の細菌と比較すると小さく、遺伝子数は約2000で、大腸菌の半分である。おそらく極限環境への適応の過程で、必須な遺伝子が濃縮された結果であり、その約2000の遺伝子は生物に広く共通する基本セットであると推定できる。

したがって、これらを明らかにすることで、生命の基本（ヒトまで共通する）

を解明できるのではないかと期待したのである。さらに、高度好熱菌由来のタンパク質は、常温菌のタンパク質よりも熱的に安定であり、単離精製、生化学解析、結晶化に有利であると想定された。

超好熱性古細菌

以上の二つのプロジェクトに加えて、「タンパク3000」の開始に際して、宮野雅司主任研究員の提案で、超好熱性古細菌*Pyrococcus horikoshii* OT3の構造ゲノム科学に取り組むプランが追加された。これは産総研・製品評価技術センターの河原林裕博士らがゲノム塩基配列の決定を報告したものであり、日本の強みを活かすものであった。古細菌は、極限環境に生育するものが多く、ゲノムは小さく（約2000の遺伝子）、高度好熱菌と同様に構造ゲノム科学に適した性質を持つ。しかも、真核生物（ミトコンドリア、葉緑体を除く）の祖先と見なされており、遺伝情報系の類似度が高い。

第2節 「構造ゲノム科学」の開始

「ターゲット選択」に関する国際的な綱引き（1995-1999年）

この時期、アメリカでもよく似た議論が進められていた。ラトガース大学のモンテリオーネ（Gaetano Montelione）は、アミノ酸配列より立体構造の方が保存性が高いので、「機能未知タンパク質」の立体構造を解析することで、機能のヒントをつかもうと提案した。実際、カリフォルニア大学バークレー校のキム（Sung-Hou Kim）は、小規模なパイロットプロジェクトながらそれを実行して見せ、世界を驚かせた。

これを受けて、NIH米国立一般医科学研究所（NIGMS）のカスマン（Marvin Cassman）やノーヴェル（John Norvell）は会議を重ね、構造解析の「ターゲット・セレクション」会議を開く段階になった。そこでキムは、肺炎の原因菌として知られるマイコプラズマが、最小のゲノム（約550kb）であることに注目し、その機能未知タンパク質の結晶構造を網羅的に決定するプロジェクトを提案しようとしているとの話が伝わってきた。モデル生物が設定されてしまうと、その後の機能研究も同じ生物で進む可能性が高い。そこで倉光らは、あくまでも高度好熱菌という日本由来のモデル生物を提案することに決めたのである。会議には倉光、横山、河原林が出席した。

会議では、生物種を特定しない提案もあった。それはインフォマティクスの立場からで、メリーランド大学のモルト（John Moulton）や立体構造の評価アルゴリズムDALIの開発者サンダー（Chris Sander）らによるものであった。それらも含めて議論され、10年程度のプロジェクトで1万構造を適切に選択して決定しようという方向に収束し、生物種を限定するアプローチもそこに含めることになった。また、このプロジェクトを、構造ゲノム科学（structural genomics）とよぶことも決まった。

ストラクチュロームとGSC (1997-1999年)

「構造ゲノム科学」のプロジェクト化が国際的に議論されている段階ではあったが、理研はいち早くプロジェクトを開始させた。これによって、その後の「タンパク3000」より予算規模は小さいものの、基礎となる体制が確立されたのである。

1997年、播磨研究所において、高度好熱菌 *T. thermophilus* HB8 の構造解析である「ストラクチュローム・プロジェクト」を7年計画でスタートさせた。すでに述べた倉光の「高度好熱菌丸ごと一匹プロジェクト」の全体は、①ゲノムの塩基配列の決定、②タンパク質結晶構造の網羅的解析、③立体構造に基づく未知機能の探索、と続くものであったが、①のゲノム解析が終盤を迎え、②の構造解析の段階へと向かいつつあった。そこで、SPring-8の振興も目指して、「構造」を強調して、ストラクチュローム (structurome) という名前となった。

1998年に、和光地区にゲノム科学総合研究センター (GSC) が設置された。横山は大型NMR施設を整備する提案を続けてきたが、GSCにおいて、ヒト・マウスの機能ドメインをターゲットとする構造ゲノム科学と合わせることで、その実現を目指すことになった。横浜キャンパスに、第1期のNMR棟 (後に西NMR棟) の建設が進められ、2000年4月より、NMR装置も徐々に設置されていった。

国際的な「構造ゲノム科学」の開始 (2000年)

NIHのノーヴェルらは、ヒトゲノムプロジェクトのような国際的なプロジェクトにすることを目指し、英国のウェルカムトラストと組んで、2000年4月、最初の国際会議を英国ヒンクストンのゲノムキャンパスで開催した。理研からは、横浜から横山が、播磨から神谷信夫が出席した。

この会議では、塩基配列解析と立体構造解析の違いを考慮せず、現実にもそぐわないポリシーが提案された。構造解析の終了を自動判別し、解析者の意向に関わらず公開する、ゲノムのドラフト配列のアナロジーでフォールドのみ決まれば公表する、といった極端な意見も出た。こうした主張に対し、各国から参加した結晶構造解析を専門とする実験科学者 (横山と神谷を含む) は、強い疑義を呈したのである。

NIHは「即時公開」を「構造解析終了後3週間以内に公表する」と改めてプロジェクト化の準備を進めたが、これには知的財産獲得競争という側面もあった。NIHは、新規の立体構造に基づいてインシリコ・スクリーニングを行い、その結合化合物を機能解析して薬効を確認すれば特許化、知的財産化できると踏んでいた。ともあれNIHは、国際会議を通じて「国際プロジェクト」としての位置づけを強化し、アメリカにおける構造ゲノム科学プロジェクトをPSI (タンパク質構造戦略) の名でスタートさせた。

このような流れを受け、理研の研究者は、サイエンスとしての議論をもっと深める必要性を痛感した。そこで2000年11月、研究発表の場として日本で第1回の構造ゲノム科学国際会議 (ICSG2000) を企画・実行することになった。この国際会議を最初に日本で開催したことは、世界における理研の存在感を高める上

で極めて重要な意味を持っていた。これに合わせて、先に英国で開催された国際会議の第2回も開き、政策面でも理研の主体性の確保に努めたのである。

国際ゲノム科学機構 (ISGO) の創設 (2001年度前半)

2001年に文部科学省が誕生し、理研GSCにおける構造ゲノム科学プロジェクトを、大学の研究者も含めたナショナルプロジェクト化する動きが始まった。時を同じくして、アメリカではNIHのPSIが本格的に開始し、英国やEUでも検討が進み、大きな動きとなってきた。2001年4月、アメリカ・バージニア州エアリーハウスにおいて、日米英の研究費助成機関による政策会議が開かれた。日本から文部科学省と理研、アメリカは国立衛生研究所 (NIH)、英国はウェルカムトラストである。目的は、すでに述べた国際会議での議論をとりまとめ、合意文書を作成することであった。

文部科学省からは、坂田東一審議官、田中敏ライフサイエンス課長が出席し、会議の中核となった。研究者サイドでは、理研から横山、宮野、倉光らが、大学からは、大島、安楽泰宏、月原富武、三木邦夫、田之倉優、若槻壮市、田中勲、産業技術総合研究所からも京極好正らが参加した。

文部科学省の最大の関心事は、データの即時公開の原則にあった。前項で述べたように、NIHでは、PSIで決定する新規のタンパク質構造について、インシリコ・スクリーニングと機能解析による特許出願を想定していた。アメリカではそのような分野が進み、短期間でも出願が可能（実施例のデータは後で追加するとしても）と思っており、国際ゲノム科学プロジェクト全体（すなわち文部科学省が助成するプロジェクトも含まれる）に対して、3週間以内でのデータ公開を求めている。ところが、日本で同様の特許出願をするには、もっと長い期間が必要であった。将来のゲノム創薬の芽が、アメリカによって一網打尽にされてしまうという危機感があった。そこで、坂田審議官は、NIHやウェルカムトラストからの強い説得にも負けず、6カ月の猶予期間を勝ち取ったのである。

ただし、後に（「タンパク3000」の実施中）、日米欧の特許庁の会議が開催され、タンパク質の構造座標はコンテンツであり、特許としては認められない、という方針が確定し、特許・知財に関する状況が大きく変化した。このため、「基礎科学」と「産業応用」の「一石二鳥」という作戦を変更し、タンパク質構造の情報を活用して、時間をかけて創薬を行っていくこととなった。

エアリーハウス会議では、この6カ月の公表猶予期間を含めて、さまざまなルールが決められ、文書にまとめられた。さらに、タンパク質の構造・機能解析を国際協調と協力の下に進めるため、ロスアラモス国立研究所のターウィリガー (Thomas Terwilliger) の提案によって、国際構造ゲノム科学機構 (ISGO) の設立が合意された。ターウィリガー、横山、ハイネマン (Udo Heinemann) が執行委員会を構成した。

エアリーハウス文書に定められた諸ルールの肉付けも進められ、例えば、進行中の研究について、ターゲットのリスト、それぞれのタンパク質の発現・精製、結晶化、X線回折データ収集、構造決定、PDB登録等、各ステップの進捗等を、

PDBのデータベースで公表することになった。

「タンパク3000」へ（2001年度後半）

構造ゲノム科学の国家プロジェクト化に向けて、理研横浜研究所では、第2期のNMR棟（中央NMR棟）の建設が進められた。並行して、NMRによるタンパク質立体構造解析に必要な安定同位体標識タンパク質試料を調製するための技術開発（無細胞タンパク質合成法など）が進められた。一方、播磨研究所では、SPRING-8に構造解析用の偏向電磁石ビームラインの建設が進められた。並行して播磨では、前述の「ストラクチュローム」プロジェクトにおいて、ゲノムDNAを用いて各タンパク質を、大腸菌を宿主とする遺伝子組換えで発現させ、熱安定性を活かして単離・精製するという作業手順を構築した。2001年、X線結晶構造解析用の微生物タンパク質試料を調製・結晶化するため、ハイスループット棟が完成した。

国家プロジェクト化にあたっては、月原富武、三木邦夫、田之倉優、若槻壮市、田中勲ら理研以外の構造生物学研究者も大きな寄与を果たした。和田昭允GSCセンター所長との会談を通じて、理研におけるプロジェクト（網羅的解析）との差別化を明確化し、大学等の研究者の参加が必須であることを確認した。そして、生物学的な課題に取り組みながら、大学等の全体で、500構造の決定を目指すことになった。

こうして、新世紀重点研究創生プラン（RR2002）の一つとして、「タンパク3000プロジェクト」が始まった。アメリカなどとの国際競争を意識し、「我が国初のゲノム創薬の実現等を目指し、我が国の研究機関の能力を結集して、平成14年度から5年間でタンパク質の全基本構造の3分の1（約3000種）以上のタンパク質の構造およびその機能を解析し、特許化まで視野に入れた研究開発を推進することを目的とする」とされた。要するに、「ヒトゲノム解析では7%に留まった日本の寄与を、タンパク質解析では対等なレベルまで高め、ゲノム創薬・知財獲得では遅れを取らない」という政策的な意図であった。

「タンパク3000プロジェクト」では、理研は網羅的なタンパク質立体構造解析を担当し、2500構造を決定することになった（「タンパク質基本構造の網羅的解析プログラム」）。一方、大学等の研究者は、生物学的なタンパク質機能を重視した、次のような八つの「タンパク質の個別的解析プログラム」を推進することになった。

- ① 発生・分化とDNAの複製・修復（中核機関：東京大学大学院農学生命科学研究科）
- ② 転写・翻訳（中核機関：北海道大学大学院先端生命科学研究院）
- ③ 転写・翻訳（中核機関：横浜市立大学大学院国際総合科学研究科）
- ④ 翻訳後修飾と輸送（中核機関：高エネルギー加速器研究機構物質構造科学研究所）
- ⑤ タンパク質高次構造形成と機能発現（中核機関：京都大学大学院理学研究科）
- ⑥ 細胞内シグナル伝達（中核機関：北海道大学大学院薬学研究院）
- ⑦ 脳・神経系（中核機関：大阪大学蛋白質研究所）
- ⑧ 代謝系（中核機関：大阪大学大学院理学系研究科）

第3節 「タンパク3000」の実施と成果 (2002年度からの5年間)

構造プロテオミクス研究推進本部 (RSGI)

2001年、理研は「タンパク3000」を実施するため、吉良爽理事を本部長、横山を副本部長とする体制をとり、構造プロテオミクス研究推進本部 (RSGI) を設置した。後に本部長は、大熊健司理事に引き継がれた。

ターゲットとするタンパク質の選択、試料の発現・精製、X線結晶構造解析では、結晶化、回折データの測定、位相決定、構造モデル構築、NMR解析では、安定同位体標識、スペクトルデータの測定、シグナル帰属、構造計算など、全ての進捗状況をデータベースにより管理した。また、決定した構造座標は蛋白質構造データバンク (PDB: Protein Data Bank) に登録した。PDBは国際的な協力で運営される唯一のタンパク質立体構造データベースであり、日本からは大阪大学蛋白質研究所が運営に参加している。これらのプロジェクト進捗状況は、企業との共同研究を除き、毎週1回、RSGIの名前で、PDBが開設した共通データベースに登録して公開した。全世界から進捗の全てを監視された状態で、緊張感を維持してプロジェクトを推進したのである。

ヒト・マウスのドメイン構造

プロジェクトでは、日本の持つアドバンテージであったヒト・マウス等の完全長cDNAライブラリー (合計約21万クローン) をフルに活用してターゲット選択を行った。Pfamといったバイオインフォマティクスのツールを用いて、複数のタンパク質で配列が保存されている領域、すなわち「ドメイン」をリストアップし、発現実験によるスクリーニングの対象とした。さらに、横山・木川隆則が開発してきた無細胞タンパク質合成法を活用して、cDNA全長から、ターゲットとする「ドメイン」をコードする領域を迅速に同定する技術を開発・採用した。それがまさに「網羅的解析プログラム」を成功に導いた理由であった。

無細胞タンパク質合成法というのは、遺伝子組換えによる合成法とは異なり、大腸菌から調製した抽出液 (「無細胞タンパク質合成系」とよばれる) を反応容器 (数十マイクロリットル~数十ミリリットル) に入れ、ATPやアミノ酸等の基質を供給し、鋳型DNAから転写・翻訳を行い、タンパク質を合成する生化学反応である。この方法では、鋳型DNAとして、プラスミドだけでなく、PCRで増幅した (直線状の) リニアDNA断片を用いることができるので、手間のかかるクローニングなしにタンパク質を合成できる。

「ドメイン」を同定するというのは、タンパク質全体のアミノ酸配列の中から、ドメインの開始と終了のアミノ酸残基を特定することである。これについてはまず、ターゲットとする「ドメイン」について、複数 (通常は四つずつ) の開始点と終了点の候補を設定し、それらをコードするリニアDNA断片を調製した。そして、このPCRからの無細胞タンパク質合成過程と、合成産物の評価を全て自

動で行う装置（ロボット）を開発した。これを使って、cDNAライブラリー全体におけるアミノ酸配列レベルの「ドメイン」から、実際に、分子量2万数千以下で、「独立に立体構造（フォールド）を形成するドメイン」を選び出して、それらをほぼ全て同定することに成功した。この段階までの成果だけでも、タンパク質の立体構造の全体像を解明する上で大きな貢献となった。

主として分子量が2万以下のドメインについては、安定同位体標識試料を用いて、多次元（2次元-4次元）のNMRスペクトルを測定した。

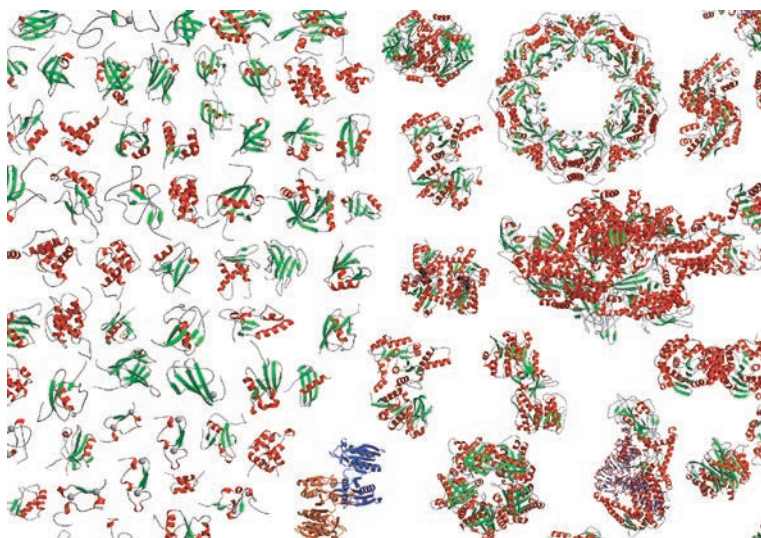
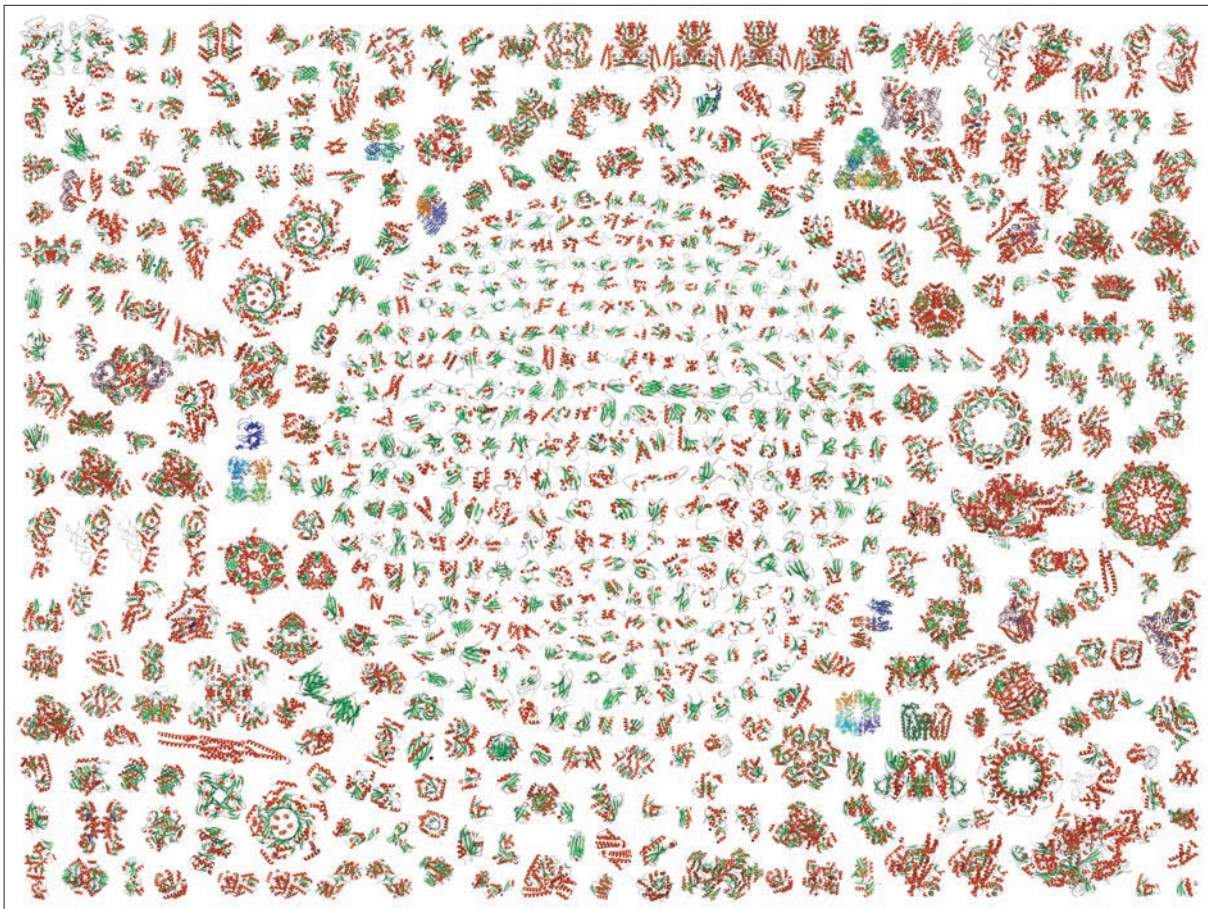
測定したNMRスペクトルの解析（「シグナル帰属」）をほぼ自動的に行うためのソフトウェアKUJIRAも開発した（小林直宏が担当し、現在も大阪大学で開発を続けている）。次の段階で、溶液中の立体構造を解析するためのソフトウェア開発のために、スイスのヴェートリッヒ研究室出身のギュンタート（Peter Güntert）博士が加わり、開発グループを率いた。ギュンタートは、ソフトウェアCYANAを開発し、立体構造解析をほとんど自動化することに成功した。これらの解析システムは、国際的に見てユニークかつ高水準で、網羅的立体構造解析に大きく貢献した（ギュンタートは現在、ドイツ・ゲーテ大学）。

「網羅的解析プログラム」の大きな成果

これらの成果を総合して、理研構造プロテオミクス研究推進本部RSGIは、2007年3月末までに、NMRにより1342構造（高等生物：1285構造、微生物：13構造）、X線結晶解析により1333構造（高等生物：142構造、微生物：1144構造）、合計2597構造を決定し、目標値2500を達成した。PDB登録数2597は、アメリカNIHのPSIによって同時期に行われた九つのセンターの合計登録数を上回って、ISGO全体の約55%という大きな割合を占めた。さらに、ヒト・マウスのタンパク質については、PSIでもほとんど構造解析されておらず、RSGIの独壇場であった。例えば2005年に全世界で（構造ゲノム科学に限らず）PDBに登録されたヒト由来あるいはマウス由来のタンパク質のNMR構造のうち、約70%がRSGIによって決定されたものであった。これらの寄与は、単一の研究機関として、群を抜いていた。

当時、配列相同性（アイデンティティー）が30%以上のタンパク質の構造はホモロジーモデリングが可能と考えられていた。この考え方にしたがって、タンパク3000プロジェクト推進委員会において「基本構造」の定義が整理された。すなわち、プロジェクトで構造決定したタンパク質に対して、配列の相同性が30%以上であるタンパク質は、その構造をテンプレート（型板）としてホモロジーモデリングが可能であること。また、PDBに構造登録されたタンパク質の全てに対して配列相同性が30%未満であり、ホモロジーモデリングが可能ではなかったタンパク質は、この構造決定によって「新たにホモロジーモデリングが可能になる」タンパク質であること。そして、そのような新たなホモロジーモデリングのテンプレートとなるタンパク質構造を「基本構造」とよぶことにした。

RSGIで決定した構造には、マルチドメインタンパク質の構造も含まれていた。その場合は構造を分割して、ドメインごとにモデリング可能性を調べた。



決定した構造（リボン図）の一部を国旗のように並べたイラスト。中の円に相当する箇所に、NMRにより決定されたタンパク質の構造を、円を囲む外側に相当する箇所に、X線結晶解析により決定されたタンパク質の構造を配置。下の図はその一部を拡大したもの。

他方、機能研究（当初は知的財産取得に必須であるという側面もあった）のために解析したものとして、基質との複合体、活性の変化した変異体の構造や、技術解析のために解析した構造などもあったが、それらはモデリングのテンプレートとしては数えないことになった。以上の差し引きをしても、RSGIとしては、構造解析数（2675）とほぼ同数の「基本構造」を決定したことになる（2500を十分に超えている）。

2007年3月末の段階で、決定した「基本構造」の一つ当たり、新たに、平均で250以上のタンパク質のホモロジーモデリングが可能になった。このような他を圧倒する成果は、ホモログの多い（影響力の強い）タンパク質の構造解析を多く行ったことに起因する。まさにターゲット戦略の勝利といえる。この後、「京」コンピュータやその後継のスーパーコンピュータ、機械学習等を駆使したタンパク質構造モデリングを推進していく基盤が構築できたといえる。このように、RSGIは、そのプロジェクト終了（2007年3月）までに、タンパク質の構造・機能研究に多大な寄与を果たした。

2017年現在、「タンパク3000」終了から10年が経過し、その意義がより明確になってきた。アメリカではPSIが2015年まで継続し、構造既知のタンパク質と配列のアイデンティティーが30%未満のタンパク質を、腸内フローラのメタゲノム解析も駆使して徹底的に探索してきた。ところが、2007年以降、新たなフォールドファミリー（Pfamファミリー等）は発見されなかった。この事実から、すでにタンパク3000の時期に、人類は、タンパク質の立体構造の基本を解明していたことが検証された。

現在、全タンパク質の約70%について、実際に決定された構造、または、信頼できる構造モデルが得られるようになった。残りの30%は、細胞膜の脂質二重膜に埋まった「膜タンパク質」と、単独では立体構造を形成しない「天然変性タンパク質」である。

タンパク質の研究において、現在では立体構造を基礎とするパラダイムが確立している。構造生物学の大きな流れの中で、「タンパク3000」の時期に理研の組織を挙げ、大規模に、かつ集中的に取り組み、短期間でタンパク質の基本構造を解明するという戦略がまさに実を結んだのである。

第4節 高難度タンパク質用の技術開発

「ターゲットタンパク」研究プログラム

網羅的解析を象徴とする「タンパク3000」に対して、後継プロジェクトは、重要なターゲットにフォーカスすることを最大の特徴とした。名前は「ターゲットタンパク」研究プログラムとなり、2007年度-2011年度に行われた。

高難度な膜タンパク質等は、微小な結晶しか得られないため、X線結晶構造解析をするには、放射光施設に数 μm 程度の細いビームを出せるビームラインを作らねばならない。理研では、2009年、SPring-8にマイクロビーム・ビームラインBL32XUを建設し、これを実現した。

一方、高難度ターゲットは、試料調製と結晶化に大きな困難があるので、それらを克服する新規技術も必要になる。理研では、GSCのタンパク質構造・機能研究グループの後継組織である「生命分子システム基盤研究領域」において、「タンパク3000」の技術開発の基盤を活用して、ヒトの膜タンパク質と高分子量複合体の構造生物学のための技術開発を進めた。

さらに、この「ターゲットタンパク」研究プログラムのさらなる後継事業として、「創薬等支援技術基盤プラットフォーム事業」(PDIS)が行われた(2012年度-2016年度)。この事業の大きな特徴は、「ターゲットタンパク」で開発した高度技術を、公募・採択したさまざまなライフサイエンス研究に「支援」として提供し、これと並行して、技術の高度化を図ることにある。大きな拠点としては、「ターゲットタンパク」で構築した技術基盤、解析拠点のSPring-8およびKEKの新規ビームライン、制御拠点の化合物ライブラリーと6カ所のスクリーニングセンター等を運用した。なお、無細胞タンパク質合成法等の開発は、横山構造生物学研究室が主に担当し、新規の膜タンパク質調製技術(特許出願)を開発した。

第5節 大規模NMR施設と技術開発

大規模NMR施設と「タンパク3000」

横浜研究所における高磁場NMR施設の建設・整備について、概略をまとめておく。これは、ゲノム科学総合研究センターにタンパク質構造・機能研究グループ(PRG)がスタートするにあたって、目玉の施設として、補正予算による整備が始まった。

廣田洋副PDによって、ユニークな建物(1期:2000年竣工、2期:2002年竣工、各20台のNMR装置を設置)がデザインされた。

第1期施設(西NMR棟)は、中央の建物に小型の600MHz装置を5台入れ、まわりに五つのドーム(平面は六角形)を配置した。これが2セットで合計20台であった。各ドームには、高磁場(800MHz以上)の装置を設置した。漏洩磁場が大きいことを想定し、ドームは900MHzの装置でも設置できるように設計された。建物は木造で(基礎部分にも鉄筋を使わない)、ドームの形状は、漏洩磁場の形状におおまかに対応した。

第2期施設はドーナツ型で、実際は、木造の建物に対する消防法上の規制により、二つの半ドーナツ型建物に分けられた。ここには600MHzの装置と800MHz以上の装置を交互に配置した。装置の導入は、2000年から開始され、「タンパク3000」開始後も続けられた。

第1期の800MHzの装置の導入を進めるうちに、第2期には、磁場遮蔽型の800MHz電磁石が開発され、また、普及型の700MHzの装置も開発されたため、それらの導入を進めた。一方でさらなる高磁場化も進められ、ブルカー社もオックスフォード社も、900MHzの開発に目処がついてきた。世界で1台目の900MHz装置は、隣接する横浜市立大学に納入された。ブルカーは少し遅れて、900MHz装置を完成し、理研にも2台が導入された。これに対し、物質・材料研究機構と日本電子は920MHz、930MHzの装置を開発した(物材機構に設置)。

このような大規模な高磁場NMR施設は世界で初めてであったが、廣田、好田真由美らによって運用にさまざまな工夫がなされ、クエンチによる装置の停止もほとんどなく、良好に運転された。装置の故障・修理や新規装置の搬入など、不

可避の使用不可時間を除くと、ほぼ100%の利用率を達成したことは特筆に値する。

大規模NMR施設の外部開放

「タンパク3000」の完了後は、理研の大規模NMR施設は、木川をリーダーとして、外部開放事業を開始し、高磁場NMR利用者を開拓してきた。特に、「タンパク3000」に際して開発したシステム、すなわち、無細胞タンパク質合成法による安定同位体標識タンパク質試料の調製から、NMRデータ計測、立体構造



誤解された「タンパク3000」

ヒトゲノム、完全長cDNAが解読され、さらに、地球生命が生み出す全タンパク質の基本構造も解明された。日本の「タンパク3000プロジェクト」が終了して10年、確認作業とも言えるアメリカのPSIプロジェクトが終わり、ついにそう言える時代が来た。その半分以上を成し遂げた日本はまた一つ、素晴らしい科学の金字塔を打ち立てたことになる。胸を張っていい。ところが、非常に残念なことに、その意義を理解していない専門家が特に日本に多いようだ。それはなぜなのか。

次世代シーケンサーの登場で、生命科学は様変わりした。ほとんどの研究がゲノムやcDNA情報を前提として進められるようになった。これは、問題発見型の伝統的な生物学の研究スタイルと、一線を画しているように見える。2007年5月、著名な科学者が新聞紙面で、若い研究者の創造の芽を奪うやり方だと批判した。

確かに「タンパク3000」の研究手法自体は、ある意味で物量作戦の色彩が強い。まるで工場の流れ作業のように、同じ手順でどんどん処理していく。これ自体に創造性が発揮される可能性は低い。しかも、5年間で578億円という大型予算を投入したことも、槍玉に挙げられるに十分だった。高価なNMRを40台も用意するというのも、事情が分からない科学者には、無謀としか見えなかったであろう。

その上に嫉妬に近い感覚も生まれた。研究予算が限られる外国人も、ねたむ側にまわった。*Nature*の日本特派員（当時）は、2006年9月28日号のニュース欄で、「タンパク3000」を紹介しつつも、「日本のプロジェクトで新しい方法論はゼロ」、「解析したと言っているタンパク質の大半はゴミばかり」などとこき下ろした。この記事の影響は非常に大きかった。先の新聞による主張も、この記事を踏まえて書かれた可能性があるし、*Nature*に書かれた内容を今なお信じている人もいる。

「タンパク3000」のリーダーの一人である横山茂之理研上席研究員と、著名な構造生物学研究者であるThomas Terwilliger、Seiki Kuramitsu、Dino Moras、Joel Sussmanは連名で、2007年の正月明けの*Nature*で、無細胞タンパク質合成法などの発明によっていかに世界に貢献したか、きっぱりと反論したが、誤解はなお残っているようにみえる。「タンパク3000」を批判するなら、その後2015年まで続けられたアメリカのPSIプロジェクトも、批判の対象にすべきである。それより何より、タンパク質の基本構造が有限であるという重大な科学的事実について、いかなる評価をしているのだろうか。

解析までの「NMR立体構造解析パイプライン」を提供するという、国際的にも極めてユニークな存在となっている。

その後、理研NMR施設では、最大で40台を所有していたNMR装置の多くを大学等に移設し、高磁場装置を中心とする10台に絞った。それでもなお、世界最大規模のNMR施設としての地位は保っている。現在も、文部科学省「先端研究基盤共用・プラットフォーム形成事業」の「NMR共用プラットフォーム」として、大学等のNMR施設と連携・協力した外部開放事業を推進している。利用は、アカデミアから産業界まで広く、タンパク質関係以外にも多様な課題に取り組み、高磁場NMR装置の威力を示している。

現在のNMR施設長の前田秀明は、物材機構との共同で、当初の「NMRパーク構想」の流れの中で、高温超伝導材を用いるマグネット・NMR装置の開発を続けてきた。最内層コイルに高温超伝導材を用いてタンパク質NMRスペクトルの測定に世界で初めて成功した。さらに、1GHzを超える装置の開発で、物材機構、神戸製鋼、日本電子と連携して、世界をリードする成果を次々と上げ、注目を集めている。