

「高度好熱菌丸ごと一匹プロジェクト 基本的生命現象の機能発見からシステム生物学へ」 に向けて必要な技術開発

倉光 成紀*

技術解説

Technologies Necessary for the Future

“ Whole-Cell Project - Systematic Analyses of Basic Biological Phenomena - ”

Key Words : structural genomics, functional genomics, discovery of protein (gene) function, imaging, system biology

1. はじめに

ヒトを含めた生物のゲノム解析が可能な時代が到来した。それによって各生物の全体像を考えながら研究することが可能な時代になった。この時代を、「『平坦だと思っていた地球が、実は丸かった』と気付いた歴史的大変革の時代」に例える考え方があ

る(図1)。ヒトは約60兆個の細胞からできており、各細胞には合計約1mの長さ(ほぼ、身長に相当する長さ)のDNAが核内に存在する(20歳までで、全身のDNA複製速度を平均すると、秒速10kmにも達する)。各細胞のDNAには、23,000種類以上の遺伝子(タンパク質)が存在することがわかっているが、

そのうちの約半分は機能未知である。すなわち、人類がこれまでに研究して来たのは、全遺伝子の約半分であった。このような状況下では、病気の治療を論理的に行うことが難しい局面が多発するのも納得できる。

各細胞が必要とするDNAの遺伝情報が、巧妙に調節されつつ、RNAポリメラーゼの中心とするタンパク質群によってmRNAになり、リボソームでタンパク質群が作られることになる。そのようにして作られたタンパク質群が、細胞内の様々な代謝物質その他を作り、細胞内反応を調節する。細胞内全体のmRNAは、合成ヌクレオチドをチップ上で合成する技術(トランスクリプトーム解析)などにより、ある程度の精度で、定量できるようになった。

作られたタンパク質の多くは、リン酸基、糖鎖、その他様々な修飾を受けて、多様な機能を発揮している。そのため、それらの修飾タンパク質群を含めるとタンパク質の種類は膨大になり、現在の解析技術(プロテオーム解析)では、細胞内タンパク質の全体像を把握することは難しい。

さらに、これらタンパク質群で作られる低分子代謝物質についても、少なくとも2万種類存在することはわかっているが、現在の解析技術(メタボローム解析)で、それらの全貌は把握できておらず、定量できているのはその1割にも満たない。

さらに、ヒトの生命現象を理解するには、組織レベル、個体レベルでの理解が必要となる。ヒトの病気が、60兆個の全細胞を構成する分子の立体構造と機能に基づいて原子レベルで理解できる時代が到来すれば、病気の治療方法や予防方法も飛躍的に進歩するはずであるが、現在の技術レベルでは難しい。

ならば、一つの細胞だけでも理解したいと考え、ちょうど十年前の本誌に、「『高度好熱菌丸ごと一匹プロジェクト 基本的生命現象の系統的解析』

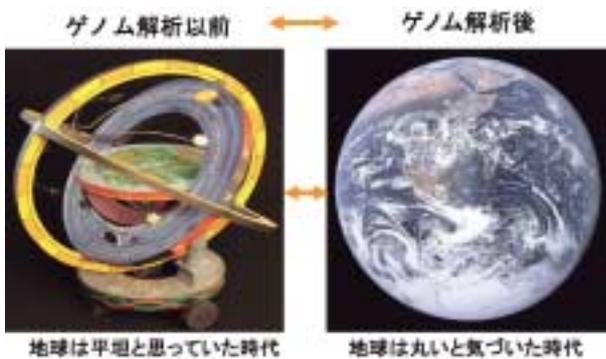


図1. 生物のゲノム解析以前と、ゲノム解析後の違いの比喻



*Seiki KURAMITSU

1949年9月生

大阪大学理学研究科・生物科学専攻・博士課程修了(1977年)

現在、大阪大学理学研究科生物科学専攻生体分子機能学研究室 教授 理化学研究所・放射光システム生物学研究グループ・グループディレクター 理学博士 生物化学・生物物理学・分子生物学

TEL : 06-6850-5435

FAX : 06-6850-5442

E-mail : kuramitsu@bio.sci.osaka-u.ac.jp

へ向けてのポランティア」を報告させていただいた¹⁾。多くの生物に共通であるにもかかわらず、機能未発見の状態に残されている約500種類のタンパク質の基本的生命現象が発見可能であるとともに、次世代のシステム生物学へ向けた学問的基盤が、広い専門領域の研究者の協力によって確立できると期待される。そのコンセプトは実にシンプルだが、プロジェクトは極めて壮大である。

2. 「高度好熱菌丸ごと一匹プロジェクト」のコンセプト

細胞一つからなるモデル生物を利用し、細胞を構成する生体分子すべての立体構造や分子機能を基にして、一つの細胞システム全体の生命現象を、原子レベル(物理化学的レベル)で理解することを目指すのが「高度好熱菌丸ごと一匹プロジェクト」である^{2,3)}。細胞が生きていくために必要な基本的遺伝子(タンパク質)は約1,500種類であり、極限環境下で現在生育している微生物は、そのような基本的に必要な遺伝子(タンパク質)だけは保持するように自然に進化して来たことが、これまでになされた約1,000種類の生物種のゲノム解析の結果からわかる。その1,500種類のタンパク質には、DNA複製・修復・転写・翻訳のサブシステムに關与するタンパク質群、アミノ酸・糖質・ヌクレオチド・脂質の合成や分解のサブシステムに關与するタンパク質群など、基本的なタンパク質群が含まれている。

これら基本的生命現象に關与する各酵素の活性部位は、いずれの生物種のものでもほとんど同じであることがわかっている。そのため、一つの生物種の酵素反応について立体構造や分子機能が深く理解できれば、他生物種の理解に役立つ。しかし、その際の注意点として、複数のタンパク質が相互作用する過程が含まれているサブシステムの研究は、一つの生物種で研究する必要があることを、かつて経験した。そこで約15年前に、(1)高温環境で生息し、(2)簡便な遺伝子操作系が確立した生物、という2つの条件を満たすモデル生物として高度好熱菌 *Thermus thermophilus* HB8 (図2) を選ぶことにした。約80 °Cの高温で生息する生物のタンパク質は、安定性が高く、立体構造解析や分子機能解析に適している。立体構造解析については、本生物種のタンパク質が結晶化の成功率が高く、リボソームや、

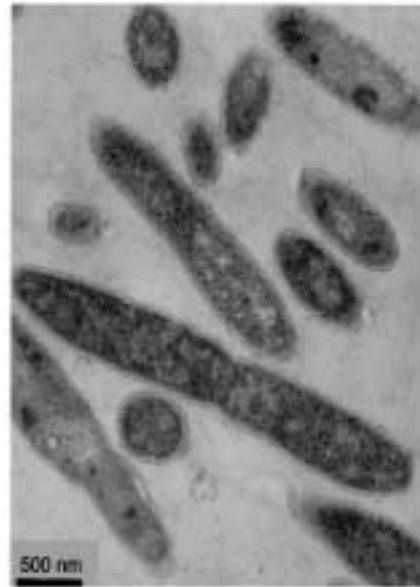


図2 . 高度好熱菌 *T. thermophilus* HB8 の電子顕微鏡像

RNAポリメラーゼ、種々の膜タンパク質などの成果からも、巨大分子や膜タンパク質のX線結晶解析に適していることが経験的に知られている。また、分子機能解析についても、分子機能を改変してタンパク質を利用する際に、遺伝子操作でアミノ酸残基を置換するが、それによってタンパク質が変性することは起こったことが無い。また、反応機構を解析する際に、pHを変えたり(触媒基がプロトンの出し入れをする酵素分子の場合、触媒基の $pK_a - 2 < \text{pH} < pK_a + 2$ の範囲で測定する必要がある)、温度や圧力を変化させる方法も、タンパク質が安定なので利用できる。さらに、タンパク質の精製や保存も、室温で行うことが可能なので、取り扱いが簡単である。

3. 「高度好熱菌丸ごと一匹プロジェクト」の進捗状況、および、必要とされる技術

研究の最終目標を達成するまでに、研究は以下の4段階で並行して進行すると考えられる(図3)

第1段階: タンパク質その他生体分子のゲノムワイドな立体構造解析。

第2段階: タンパク質その他生体分子のゲノムワイドな機能解析。

(機能未知遺伝子(タンパク質)の機能推定を含む。)

高度好熱菌 丸ごと一匹 プロジェクト
— 基本的生命現象の系統的解明 —

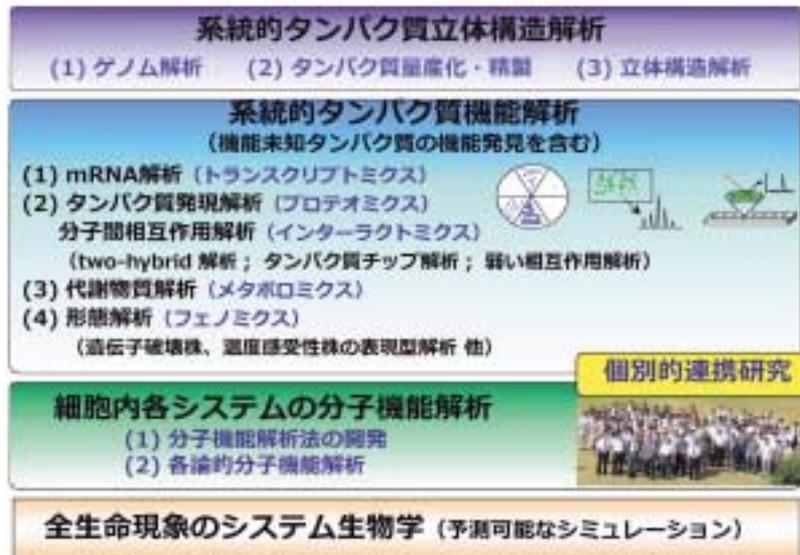


図3. 高度好熱菌丸ごと一匹プロジェクトの研究ステップ

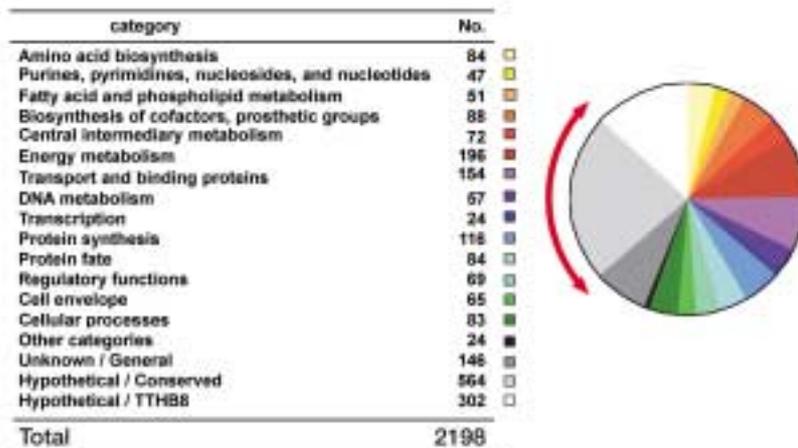


図4. 高度好熱菌のゲノムにふくまれる遺伝子(タンパク質)の分類
各ジャンルの遺伝子を円グラフに示す。その の部分(約564種類)は、多くの生物に共通で基本的生命現象に
関与する遺伝子が、機能未発見の状態でも残されていることを表す。

第3段階：細胞内各システムの各論的個別解析。
(機能未知遺伝子(タンパク質)の
機能発見を含む。)

第4段階：予測可能なレベルのシミュレーション
(システム生物学)

3-1. 第1段階：タンパク質その他生体分子のゲ
ノムワイドな立体構造解析
まずゲノム解析を行った。その結果、本高度好熱

菌は約2,200種類の遺伝子(タンパク質)からなる
ことがわかった(図4)。遺伝子数が約1,500以下
の生物は、マイコプラズマのように、寄生しなけれ
ばならなくなるようである。そのため、本高度好熱
菌は、周囲から栄養分を摂取しつつ自力で生きて行
くことが可能な、もっとも遺伝子数の少ない単純な
生物と言えるであろう。遺伝子数が少ないことを反
映して、例えば、活性酸素種を分解して無毒化する
サブシステムの場合(図5)、多くの生物種で、各

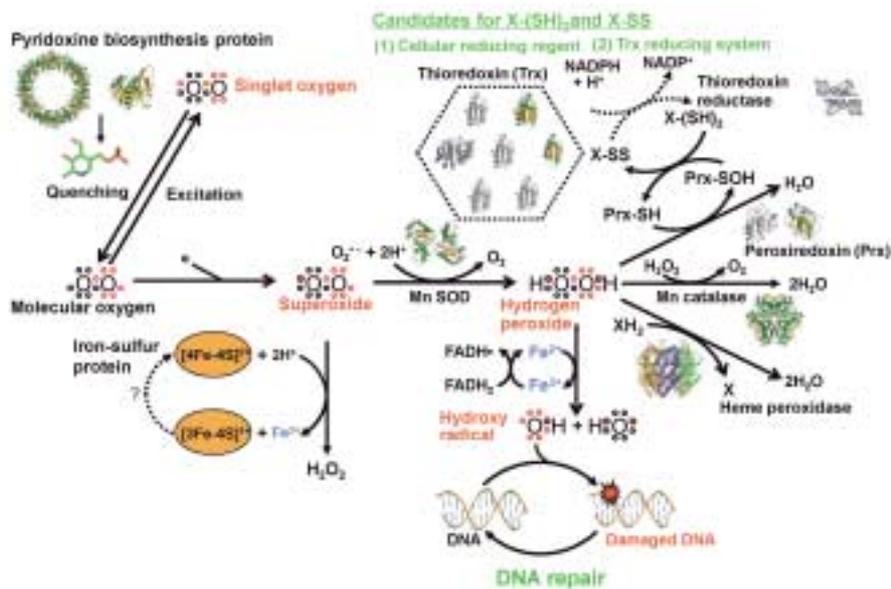


図5 . 活性酸素種を無毒化する代謝経路(サブシステム)

表1 . 活性酸素種を無毒化する系のタンパク質数の比較

生物種	遺伝子数	スーパーオキシドジスムターゼ	カタラーゼ	ペルオキシシドクシレン	ペルオキシダーゼ
ヒト	36,469	2	1	6	8
シロイヌナズナ	32,026	8	3	10	17
酵母	5,855	2	2	4	4
大腸菌	4,537	3	1	3	3
高度好熱菌	2,238	1	1	1	1

表2 . タンパク質の立体構造解析の進捗状況

	タンパク質数 (遺伝子数)
(1) タンパク質発現用プラスミド	2059
(2) タンパク質発現(量産化)	1450
(3) タンパク質精製	944
(4) タンパク質結晶化	682
(5) X線による回折データ収集	460
(6) 立体構造解析	360
他グループとの集計	360 + (106) = 466

反応過程に複数の酵素が存在するが、本高度好熱菌では一つずつしか存在しない(表1)。このようなモデル生物は、そのシステムの解析が容易である。そのため、モデル生物として研究に適している。

大阪大学、理化学研究所、その他約20の大学や研究機関との共同研究により、約2,000種類のほぼ全てタンパク質について立体構造解析を試みた。我々の結果と他のグループの結果とを合計すると、全タンパク質の約21%について立体構造解析が完了し(表

2) 本高度好熱菌はタンパク質の立体構造解析が最も進んだ生物種となった。現在の技術レベルでも、25~30%までは立体構造解析が可能のように思われる。残されたタンパク質の多くは、「膜タンパク質」や「複合体タンパク質」である。図6は、タンパク質精製までステップの成功率を、タンパク質の機能別にジャンル分けしたものである。棒グラフの長さはそのジャンルに属するタンパク質数で、棒の中の横棒は、タンパク質精製に成功した数を示す。これを見ると、膜タンパク質(ジャンル15)の成功率が低いことがわかる。その主な原因は、タンパク質の発現が低いことであった。しかし最近、Mistic(membrane-integrating sequence for translation of integral membrane protein constructs from *Bacillus subtilis* (図7))を利用すれば、4~8回膜貫通型を含めて膜タンパク質の約70%を発現させることが可能となり、タンパク質発現技術は飛躍的に進歩した。一方、細胞内で複合体を形成するタンパク質については、高度好熱菌に遺伝子操作を行って、タンパク質のC末端側に特異的なタグ配列(His-tagなど)を持たせる。そして、細胞内で発現させたそのタンパク質をtag配列で捕捉すると同時に、それに結合しているタンパク質を質量分析計で同定する。それによって強く結合したタンパク質がわかれば、それら大腸菌体内などで共発現させるようにすれば、複合体タンパク質の立体構造解析の成功率は上

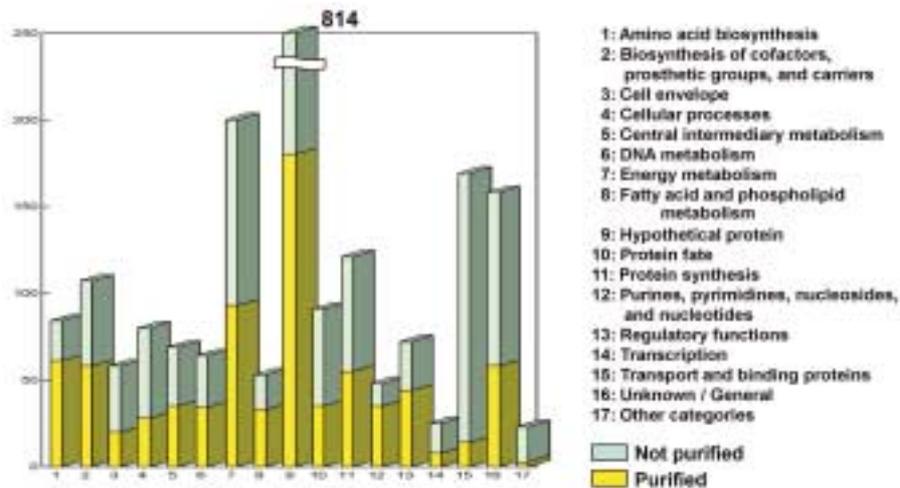


図6. タンパク質の精製ステップまでの完了率
棒グラフの高さは各ジャンルの全タンパク質数。
途中の横線は、精製ステップまで完了したタンパク質数を表す。

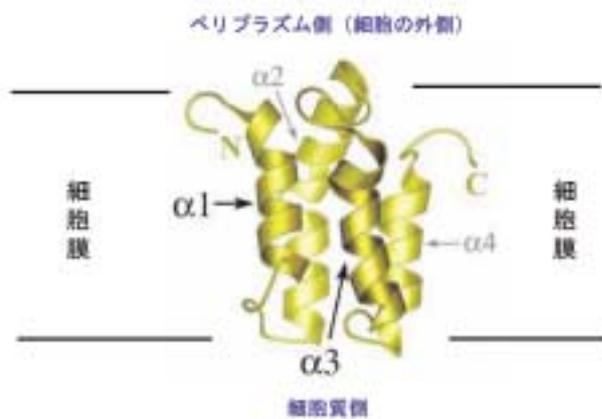


図7. 膜タンパク質の発現に有効な
“Mistic”タンパク質ドメイン



図8. タンパク質のドメイン単位の立体構造 (フォールド) のパターンは、わずか1,000 ~ 4,000種類しか存在しない。その立体構造予測は70%以上成功する時代になった。

昇すると思われる。リボソームのように、非常に多くのサブユニットからなるタンパク質などの場合には、細胞を大量に培養して、直接精製する方が良いでしょう。

100万種類以上のタンパク質のアミノ酸配列が知られており、タンパク質の立体構造予測に必要な難形となるタンパク質 (ファミリー数) は約20,000種類存在するが、ドメイン単位 (100から200アミノ酸残基からなる) のペプチド主鎖の折れたたまりのパターン (フォールド) はわずか1,000から最大でも4,000種類しか存在しないことが知られている (図8)。そのようなフォールド単位での、タンパク質主鎖の立体構造予測技術は成功率約70%と、飛躍的に上がった。予測された立体構造が正しいか

否かの判断は、(1) 生物種間で保存されたアミノ酸残基が空間的に集まっていることと、(2) 重要と思われるアミノ酸残基を置換して機能を解析すれば、確認できるであろう。

このような立体構造予測技術の向上は、それは、タンパク3000プロジェクトを初めとして、世界中の多くの研究者が貢献したおかげである。タンパク質の立体構造情報を収納したPDB (http://www.pdbj.org/index_j.html) の登録数も5万件を超えた。それによって、例えばD. Bakerらのグループは、タンパク質立体構造の予測⁵⁾だけでなく、新規な立体構造のタンパク質をデザインした上で実際に作製したり⁶⁾、自然界には存在しない反応を触媒する新規酵素を作製することにも成功し

ている⁷⁾。

3-2 . 第2段階：タンパク質その他生体分子のゲノムワイドな機能解析（機能未知タンパク質の機能推定を含む）

機能未知タンパク質の機能推定には、立体構造解析がもっと役立つことが経験的にわかった。立体構造解析がなされた約50種類の機能未知タンパク質のうち、約60%の分子機能推定が可能であり、各論的機能解析によって分子機能が確認された。しかし、立体構造解析の成功率は約20%なので、約90%の機能未知タンパク質の機能推定はできないことになる。そこで、mRNAの発現（トランスクリプトミックス）、タンパク質の発現（プロテオミックス）、

代謝物質の増減（メタボロミックス）、分子間相互作用（インターラクトミックス）、遺伝子破壊株や表現型（フェノミックス）のようなゲノムワイドな機能解析法が、機能未知タンパク質の機能推定に威力を発揮することになる（図9）。

そのために、高度好熱菌の遺伝子操作に必要な薬剤耐性遺伝子を、世界で初めて耐熱化することに成功した（図10^{8),9)}。

プロテオーム解析やメタボローム解析の際に、超高精度のESI-FT-ICR型質量分析装置を利用すると、有効数字6桁のデータが得られ、解析精度の高いデータが得られる。通常の質量分析装置の精度と比較したのが図11である。このESI-FT-ICR型の質量分析は、大阪大学学内ではリノベーションセンターを

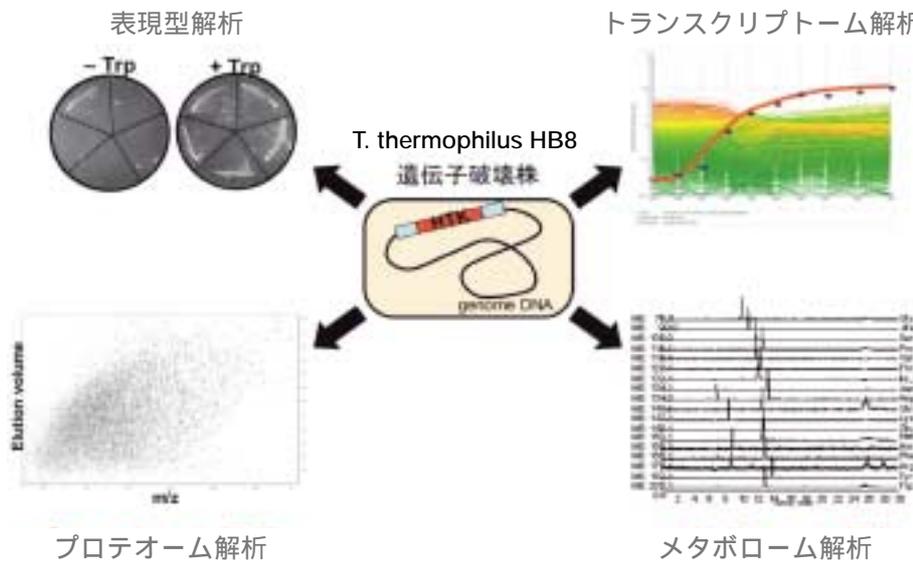


図9 . 遺伝子（タンパク質）のゲノムワイドな機能解析
これらの方法を駆使して、機能未知遺伝子（タンパク質）の機能推定が可能になる。

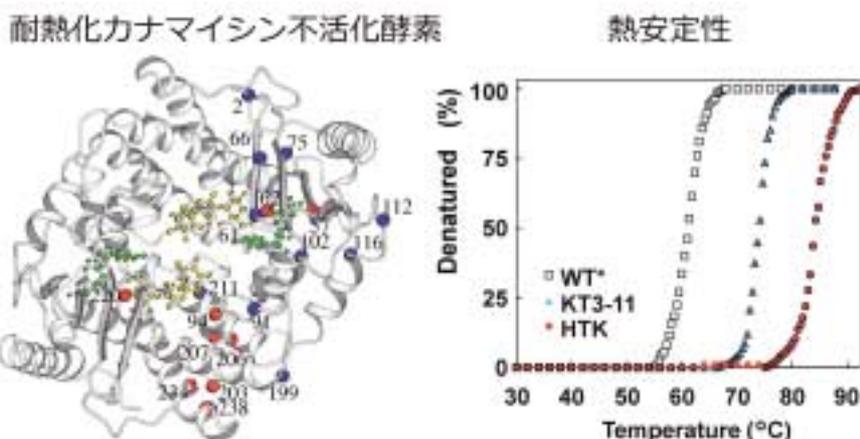


図10 . カナマイシン不活化酵素の耐熱化

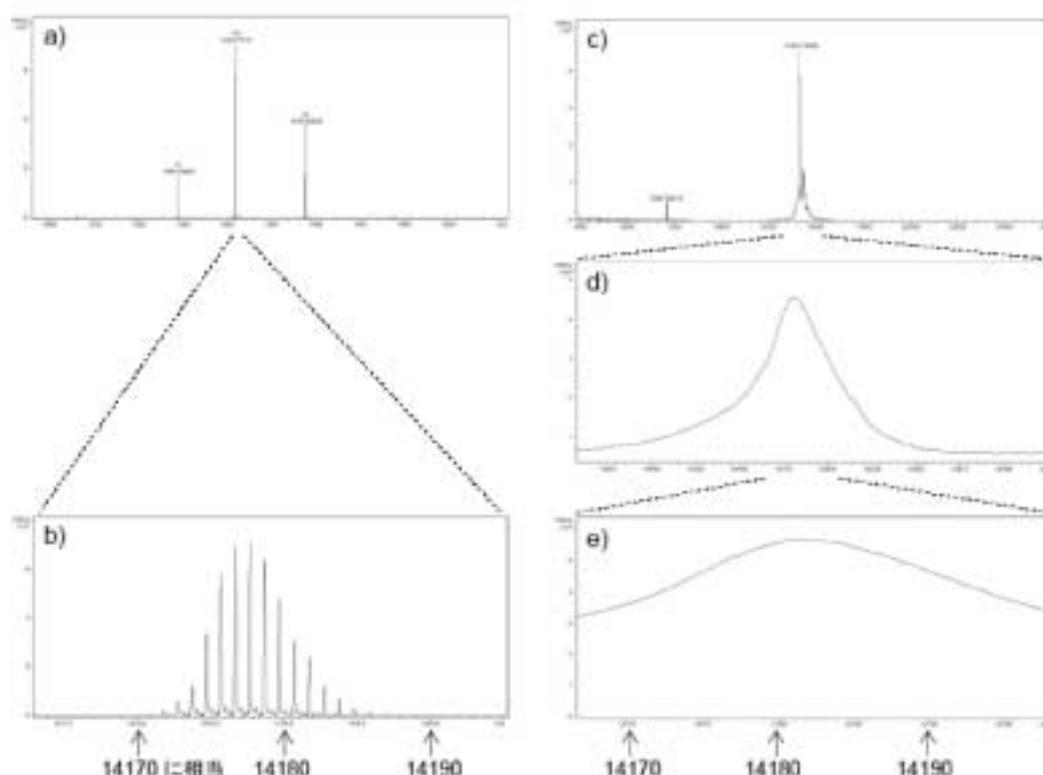


図11．質量分析計のESI-FT-ICR-MS とMALDI-TOF-MS との比較

質量分析計による全長タンパク質の分子量測定

a) ~ b) エレクトロスプレー・フーリエ変換イオンサイクロトロン共鳴質量分析(ESI-FT-ICR MS)

c) ~ e) マトリックス支援レーザー脱離イオン化飛行時間質量分析(MALDI-TOF MS)

分子量は、計算値が 14176.7659 (最も存在比の高い、モノアイソトピックの質量) であるが、

測定値*は、ESI-FT-ICR MS が 14176.700 ± 0.023 、

MALDI-TOF MS が 14182.260 ± 2.636

* 測定値は20回測定した平均値、 \pm は標準偏差

通じて行われ、学外にはNPO・阪大理工学テクニサーチ (<http://www.handairigaku-techno.or.jp>) を通じて、公開されている。

3-3 . 第3段階：細胞内各システムの各論的個別解析 (機能未知タンパク質の機能発見を含む)

細胞内に存在する各サブシステムの機能解析である。細胞内には図5の例でしめすようなサブシステムが約200種類存在するが、これまでに精製したタンパク質を利用すれば *in vitro* でサブシステム全体の解析を行うことが可能となり、質量分析計やイメージング技術を利用した *in vivo* での解析と比較することによって、さらなる調節機構の存在の有無を確認することも可能となる。この各サブシステムの機能解析には、ゲノムワイドな研究が始まる遥か以

前から行われてきた“各論的”分子機能解析(図12)が含まれている^{10,11)}。その各論的研究には、アミノ酸置換に耐えるだけの十分なタンパク質の安定性や、反応機構を解析するために、少なくともpH5~9(常温)のpH領域での十分な安定性なども必要となるが、安定性が高いタンパク質はそれらの解析に十分耐える。

このようにして解析されたタンパク質の中には、実に奇妙なものが多数存在する。その一例は、DNA合成に必要なはずのdNTPを分解する6量体酵素(図13)である。この酵素は、1種類のdNTPに対しては活性を示さないが、複数の基質が存在すると強い活性を示す¹²⁾(図14)。このような酵素が存在すると、今後のシステム生物学的解析を *in vitro* で行う際に、対応する *in vivo* での解析条件を見い出すのが難しくなる。



図12. 分子機能解析法のチャート

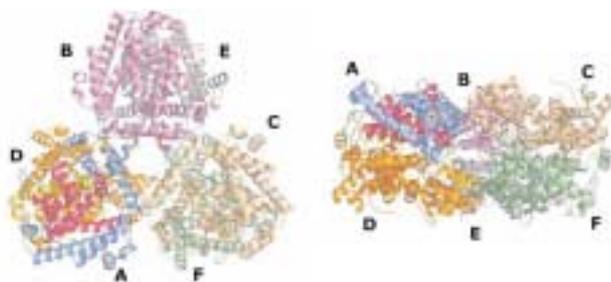


図13. 複数の類似基質が共存すると活性発現する奇妙な酵素

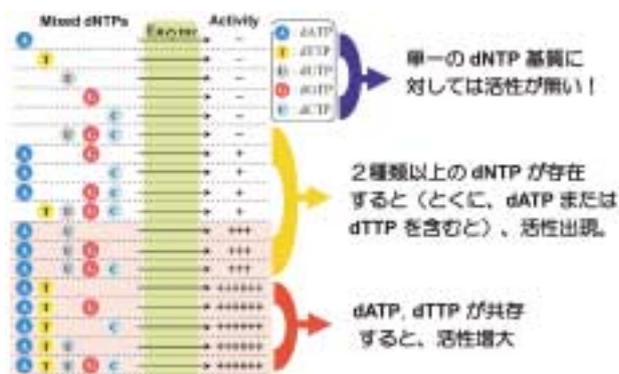


図14. 複数の類似基質が共存すると活性発現する奇妙な酵素

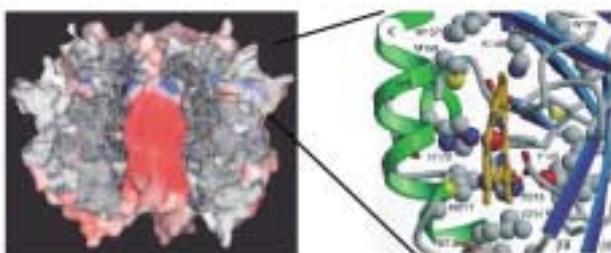


図15. 亜塩素酸分解酵素を発見

また立体構造解析に成功したことによって、ヘムの結合が推定され、さらに亜塩素酸分解酵素であることがわかった5量体酵素の場合には(図15)、活性部位は図の右上に存在するが、それとは別に、負に帯電した大きな溝が中央部に存在する¹³⁾。現時点での測定条件では、酵素活性が低いことから、この溝に結合して活性上昇を引き起こす生体分子が発見できるのではないかと期待している。

しかし、タンパク質の立体構造や分子機能には、一般性があるにもかかわらず、説明のできない現象がまだ数多く残されている。それらの例は文献¹⁰⁾に記した。また、最近では、同じ活性をもつ酵素分子でも、分子によって、さらに時刻によって、活性が変化することが知られている。また、高密度の細胞内では、生体分子が希薄溶液中と異なる挙動を示すことも知られている。各生体分子の濃度だけでなく、活動度(活量)の情報を知る方法の技術開発も望まれる。さらに、細胞内では平衡状態が成り立っていないので、非平衡の熱力学取り扱いも必要となる。このような現象を理解するためには、新たな学問基盤の確立や、分子・細胞機能解析法の技術開発が必要となる。

さらにこの第3段階では、タンパク質のみならず、すべての生体成分の物性を研究者間で共通した一定の条件下で解析し、それら「どのような構造をした、どのような機能をもった」生体分子が「細胞内のどの場所に」「どれだけの分子数」存在するかなどを

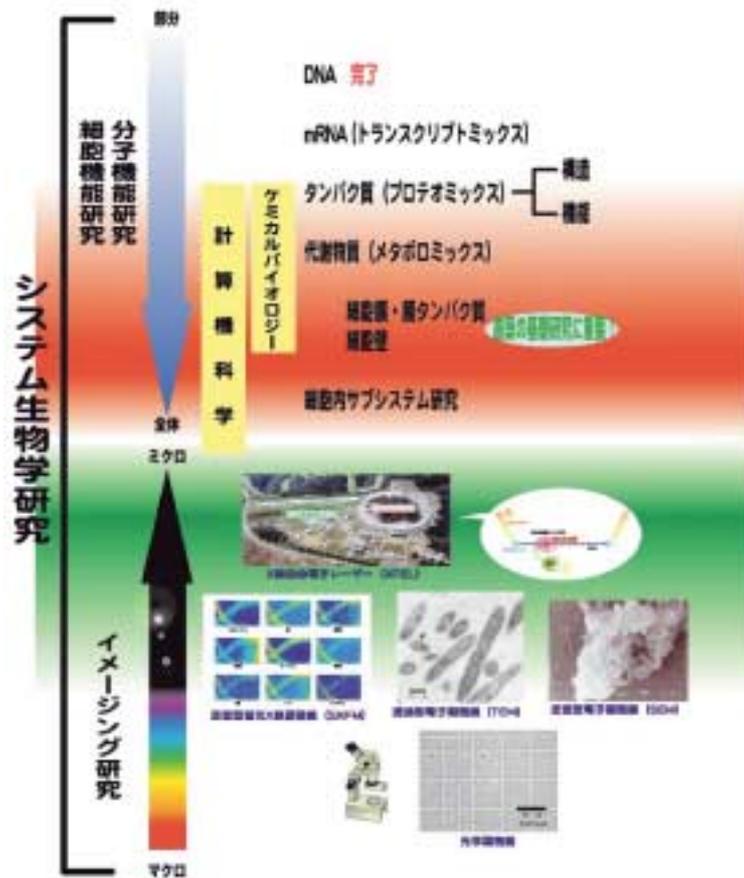


図16 . 究極のイメージングが、予測可能なシステム生物学に繋がる

調べるとともに、それらの「時間依存性」をも調べる必要がある。そして、第4段階のシミュレーションに必要な各論的データを系統的に収集することになる。

3-4 . 第4段階：予測可能なレベルのシミュレーション(システム生物学)へ向けて
第3段階までの膨大な情報を統合し、細胞全体を原子レベルで理解するために、シミュレーション結果と比較しつつ研究を進めることになる(システム生物学)。そして、単なる説明ではなく、予測可能なシステム生物学を目指すことになる。

その際に、細胞を構成する全分子について、究極のイメージングが可能になれば、これまでの「間接的な生命現象の理解」から、「直接的に生命現象を理解する」ことが可能になる。幸い、高度好熱菌 *T. thermophilus* HB8 は、細胞内のタンパク質分子の立体構造解析の成功率がもっとも高い生物種である上に、全タンパク質の種類も約 2,200 と少ない

モデル生物である。このように、*T. thermophilus* HB8 は現時点で、原子分解能の細胞内イメージングへ向けて最適の生物種である(図16)。

4 . おわりに

上述のような研究過程を通して、ヒトを含めたあらゆる生物に共通で、基本的生命現象に関与する機能未知タンパク質(遺伝子)約 500 種類の機能発見が可能となる¹⁴⁾。

このようにして、高度好熱菌でシステム生物学の学問基盤が整備でき、シミュレーションが生命現象の単なる説明ではなく、生命現象を予測できる段階に達すれば、「我々人類は、生命現象を理解できた」と言える時代に一步近づくことになる。ヒトなどの場合には、さらに、組織レベル、個体レベルでの理解が必要となるが、高度好熱菌を利用した学問基盤の整備によって、ヒトの病気の治療や予防なども大きく様変わりすると期待される。

なお、タンパク質発現用プラスミドや遺伝子

破壊株作製用プラスミド(リストは、http://www.brc.riken.jp/lab/dna/ja/TT_plasmid_sum.html)はバイオリソースセンター(<http://www.brc.riken.jp/>)から自由に入手可能なほか、タンパク質発現・精製方法などもホームページ(<http://www.thermus.org/>)で順次公開しつつある。

謝辞

本研究は約15年前から多くの研究者の協力で行われて来ました。とくに、理化学研究所・生命分子システム基盤研究領域・横山茂之領域長(東京大学大学院理学系研究科教授兼務)、理化学研究所・放射光システム生物学研究グループ・新海暁男チームリーダー、海老原章郎チームリーダー(現岐阜大学)、大阪大学理学研究科・増井良治准教授、中川紀子助教、その他多くの方々から感謝します。

参 考 文 献

- 1) 倉光成紀(1998) *生産と技術* **50**, 84-86
(home page: <http://www.thermus.org>)
- 2) Kuramitsu, S. *et al.* (1995) *Protein Eng.* **8**, 964-964
- 3) Yokoyama, S. *et al.* (2000) *Nature Struct. Biol.* **7**, 943-945; <http://www.thermus.org>
- 4) Roosild, T. P. *et al.* (2005) *Science* **307**, 1317-1321
- 5) Bradley, P. *et al.* (2005) *Science* **309**, 1868-1871
- 6) Kuhlman, B. *et al.* (2003) *Science* **302**, 1364-1368
- 7) Roethlisberger, D. *et al.* (2008) *Nature* **453**, 190-195
- 8) Hoseki, J. *et al.* (1999) *J. Biochem.* **126**, 951-956; Hoseki, J. *et al.* (2003) *Biochemistry* **42**, 14469-14475
- 9) Hashimoto, Y. *et al.* (2001) *FEBS Lett.* **506**, 231-234
- 10) 倉光成紀、増井良治、中川紀子(2004)「生物学が変わる! ポストゲノム時代の原子生物学」, 大阪大学出版会; 倉光成紀、杉山政則編(2007)「構造生物学 ポストゲノム時代のタンパク質研究」, 共立出版
- 11) Fersht, A. (1999) "Structure and Mechanism in Protein Science", W. H. Freeman & Company (桑島邦博他訳(2006)「タンパク質の構造と機構」医学出版)
- 12) Kondo, N. *et al.* (2007) *Acta Cryst.* **D63**, 230-239; Kondo *et al.* (2004) *J. Biochem.* **136**, 221-231; Kondo *et al.* (2008) *Extremophiles* **12**, 217-223
- 13) Ebihara, A. *et al.* (2005) *J. Struct. Func. Genom.* **6**, 21-32
- 14) 倉光成紀(2008) *生化学* **80**, 1075

