

第4部および第7部報告

タンパク質の構造・機能研究の  
総合戦略の提言

平成13年3月26日

日本学術会議第4部および第7部

この報告は、第18期日本学術会議第4部および第7部附置の小委員会「タンパク質の構造・機能研究の総合戦略」で審議した結果を取りまとめ、第4部および第7部報告として発表するものである。

日本学術会議第18期第4部会員  
(平成13年3月26日現在)

部長	大瀧 仁志	立命館大学工学部教授 東京工業大学名誉教授
副部長	土居 範久	慶應義塾大学工学部教授
幹事	岩槻 邦男	放送大学教授 東京大学名誉教授
幹事	郷 信広	京都大学大学院理学研究科教授
会員	青木 謙一郎	東北大学名誉教授
	赤岩 英夫	群馬大学学長
	池内 了	名古屋大学大学院理学研究科教授
	入倉 孝次郎	京都大学防災研究所教授
	岩槻 邦男	放送大学教授 東京大学名誉教授
	岩村 秀	放送大学教授 東京大学名誉教授
	上野 健爾	京都大学大学院理学研究科教授
	江澤 洋	学習院大学理学部教授
	大瀧 仁志	立命館大学工学部教授 東京工業大学名誉教授
	岡田 守彦	筑波大学先端学際領域研究センター教授
	岡本 和夫	東京大学大学院数理科学研究科長
	郷 信広	京都大学大学院理学研究科教授
	小林 俊一	理化学研究所理事長 東京大学名誉教授
	斎藤 常正	放送大学宮城学習センター客員教授 東北大学名誉教授
	坂元 昴	文部省メディア教育開発センター所長 東京工業大学名誉教授
	佐藤 文隆	京都大学大学院理学研究科教授
	柴田 徳思	高エネルギー加速器研究機構・放射線科学センター長
	土居 範久	慶應義塾大学工学部教授
	西田 篤弘	日本学術振興会理事
	野上 道男	日本大学文理学部教授 東京都立大学名誉教授
	廣田 榮治	総合研究大学院大学学長 分子科学研究所名誉教授
	星 元紀	慶應義塾大学工学部教授 東京工業大学名誉教授
	松原 謙一	財団法人国際高等研究所副所長
	松本 忠夫	東京大学大学院総合文化研究科教授
	村橋 俊一	大阪大学大学院基礎工学研究科教授
	森脇 和郎	総合研究大学院大学副学長 国立遺伝学研究所名誉教授
	矢原 一郎	株式会社医学生物学研究所伊那研究所長
	吉原 経太郎	北陸先端科学技術大学院大学副学長
	吉村 功	東京理科大学工学部教授
	米倉 伸之	東京大学名誉教授
	米沢 富美子	慶應義塾大学工学部教授

日本学術会議第18期第7部会員

(平成13年3月26日現在)

部長	遠藤 實	埼玉医科大学副学長 東京大学名誉教授
副部長	鴨下 重彦	賛育会病院長 東京大学名誉教授
幹事	内田 安信	明倫短期大学長 東京医科大学名誉教授
幹事	橋本 嘉幸	財団法人佐々木研究所長 東北大学名誉教授
会員	青野 敏博	徳島大学副学長
	赤沼 安夫	朝日生命糖尿病研究所所長
	浅野 茂隆	東京大学医科学研究所先端医療研究センター長病院長
	渥美 和彦	東京大学名誉教授
	安樂 泰宏	帝京科学大学理工学部教授 東京大学名誉教授
	内田 安信	明倫短期大学長 東京医科大学名誉教授
	遠藤 實	埼玉医科大学副学長 東京大学名誉教授
	折茂 肇	東京都老人医療センター院長
	金子 敏郎	千葉大学名誉教授
	鴨下 重彦	賛育会病院長 東京大学名誉教授
	黒川 清	東海大学医学部長 東京大学名誉教授
	黒川 高秀	昭和大学医学部教授 横浜市北部病院長
	古賀 憲司	奈良先端科学技術大学院大学物質科学教育研究センター教授
	小林 宏行	杏林大学医学部長
	小林 義典	日本歯科大学歯学部教授
	瀬崎 仁	摂南大学薬学部長 京都大学名誉教授
	高倉 公朋	東京女子医科大学長 東京大学名誉教授
	高橋 清久	国立精神・神経センター総長
	多田 啓也	NTT 東日本株式会社東北病院顧問 東北大学名誉教授
	田中 平三	東京医科歯科大学難治疾患研究所教授
	角田 文男	労働福祉事業団岩手産業保健推進センター所長岩手医科大学名誉教授
	中村 紀夫	日本交通科学協議会長
	橋本 嘉幸	財団法人佐々木研究所長 東北大学名誉教授
	秦 順一	慶応義塾大学医学部教授
	久道 茂	東北大学医学部長
	平野 寛	杏林大学医学部教授
	藤村 重文	社団法人全国社会保険協会連合会東北厚生年金病院長
	堀内 博	東北大学名誉教授
	本郷 利憲	東京都医学研究機構 東京都神経科学総合研究所理事
	松木 明知	弘前大学医学部教授
	武藤 輝一	新潟国際情報大学長 新潟大学名誉教授
	矢崎 義雄	国立国際医療センター総長
	渡辺 洋宇	金沢医科大学呼吸器外科科長客員教授 金沢大学名誉教授

日本学術会議 第4部および第7部附置小委員会  
(平成13年3月26日現在)

世話人、座長

安楽泰宏(帝京科学大教授、第7部会員、生化学研究連絡委員会委員長) anraku@ntu.ac.jp  
郷 信広(京都大学大学院理学研究科教授、第4部会員、生物物理学研究連絡委員会委員長)  
go@qchem.kuchem.kyoto-u.ac.jp

幹事

新井賢一(東京大学医科学研究所所長・教授、生化学研究連絡委員会委員) ken-ichi@ims.u-tokyo.ac.jp  
磯辺俊明(東京都立大学大学院理学研究科教授、生化学研究連絡委員会委員)  
isobe-toshiaki@c.metro-u.ac.jp  
中野明彦(理化学研究所生体膜研究室主任研究員、生化学研究連絡委員会委員)  
nakano@postman.riken.go.jp  
難波啓一(ERATOプロジェクト総括責任者) keiichi@crl.mei.co.jp  
松尾 洋(理化学研究所、チームリーダー) matsuo@jota.riken.go.jp  
若槻壮市(高エネルギー加速器研究機構・物質構造科学研究所、教授) soichi.wakatsuki@kek.jp

委員

板井昭子(医薬分子設計研究所、所長) itai@immd.co.jp  
上田国寛(京都大学化学研究所教授、生化学研究連絡委員会委員) ueda@scl.kyoto-u.ac.jp  
甲斐荘正恒(東京都立大学大学院理学研究科教授) kainosho@nmr.chem.metro-u.ac.jp  
京極好正(福井工業大学教授) yckyo@attglobal.net  
倉光成紀(大阪大学大学院理学研究科教授、生化学研究連絡委員会委員) kuramitu@bio.sci.osaka-u.ac.jp  
谷口直之(大阪大学大学院医学系研究科教授、生化学研究連絡委員会委員)  
proftani@biochem.med.osaka-u.ac.jp  
月原富武(大阪大学蛋白質研究所教授、生物物理学研究連絡委員会委員) tsuki@protein.osaka-u.ac.jp  
永井克也(大阪大学蛋白質研究所所長・教授、生化学研究連絡委員会委員) k\_nagai@protein.osaka-u.ac.jp  
中村春木(大阪大学蛋白質研究所教授、生物物理学研究連絡委員会委員)  
harukin@postman.protein.osaka-u.ac.jp  
新村信雄(日本原子力研究所先端基礎研究センター研究主幹) niimura@kotai3.tokai.jaeri.go.jp  
二井将光(大阪大学産業科学研究所教授) m-futai@sanken.osaka-u.ac.jp  
本庶 佑(京都大学医学部長・教授、生化学研究連絡委員会委員) honjo@mfour.med.kyoto-u.ac.jp  
柳田敏雄(大阪大学大学院医学系研究科教授) yanagida@phys1.med.osaka-u.ac.jp  
横山茂之(東京大学大学院理学系研究科教授、生化学研究連絡委員会委員)  
yokoyama@y-sun.biochem.s.u-tokyo.ac.jp

## 要 旨

### 1 報告書の名称

タンパク質の構造・機能研究の総合戦略の提言

### 2 報告書の内容

#### 1) 作成の背景

科学技術会議政策委員会は「平成13年度科学技術振興に関する重点指針」(平成12年6月29日発表)において、ゲノムサイエンスのさらなる振興とその拡大推進の一環として、バイオインフォマティクスおよびタンパク質の構造・機能研究をとりあげてを提言した。この指針は、ゲノムサイエンスの国際的な研究動向およびその成果の社会還元を目指すコンセンサスと軌を一にするものであり、時宜を得た政策決定である。

また、中央省庁再編の一環として発足した総合科学技術会議においても、本年3月22日に提出された「科学技術に対する総合戦略について」に対する答申の中で、プロテオミクス、タンパク質の立体構造解明、バイオインフォマティクス等が重要政策としてとりあげられている。日本学術会議第4部、第7部およびタンパク質科学の推進に深い関心をもつ研究連絡委員会は、この国家的科学政策の策定を評価し、その実効ある施行を強く願うものである。ここに、第4部および第7部は、附置小委員会の調査ならびに検討所見を踏まえ、21世紀におけるタンパク質の構造・機能研究の在り方、特に新しい「目的提示、使命達成型」プロジェクト研究の展開にかかわる諸問題を整理し、その効率的実行に関する提言をとりまとめた。

#### 2) 提言の内容

##### 新しいタンパク質科学の振興に関わる公的議論の必要性

プロテオミクス、構造ゲノミクス、構造生物学、バイオインフォマティクス等の新しいタンパク質科学の振興に関わる学術行政を、専門家の参加を得て論議する公的委員会の設置を提言する。

##### 目的提示、使命達成型の新しいタンパク質科学のための研究拠点の整備

プロテオミクス、構造ゲノミクス、バイオインフォマティクスの3分野に目的提示、使命達成型の新規国家プロジェクトを立ち上げ、その使命達成のための複数の適正規模の研究拠点の新設を提言する。

##### 新しいタンパク質科学の振興、推進、支援に関わる研究推進機構の必要性

プロテオミクス、構造ゲノミクス、構造生物学、バイオインフォマティクス等の新しいタンパク質科学の振興、推進、支援に関わる次の3つの機能を持つ研究推進機構を設けることを提言する。(1)上記の国家プロジェクトの使命達成のための研究拠点を運営し、(2)上記のプロジェクト等で用いる技術を競争的環境下で開発するための「課題公募、新技術創成移転型」研究を支援し、また、(3)使命達成型になじまない多様なタンパク質の個性に即した多数の個別研究を支援する。

##### ゲノム統合科学の研究基盤の拡大と教育機構の整備

ゲノム統合科学が領域融合型の新しい自然科学の基礎となることの認識に立ち、大学などを中心とする教育、研究機構を整備することを提言する。

## 目次

1. 序	新しいタンパク質科学の創出に向けて	1
2. ゲノム情報に依拠するタンパク質科学の特質		2
2.1. ゲノム科学のタンパク質科学への展開		2
2.2. タンパク質科学の特質とそのための研究体制		2
3. プロテオミクス研究の目標と課題		3
3.1. プロテオミクス研究の目的		3
3.2. プロテオミクス研究の意義と波及効果		4
3.3. 国内外での研究の現状		5
4. 構造ゲノミクスおよび構造生物学の目的と課題		5
4.1. 構造ゲノミクスおよび構造生物学の目的		5
4.2. 構造ゲノミクスおよび構造生物学の意義と波及効果		6
4.3. 国内外での研究の現状		6
5. バイオインフォマティクスの目的と課題		7
5.1. バイオインフォマティクスの目的		7
5.2. バイオインフォマティクスデータ(BID) 整備の意義と波及効果		8
5.3. 国内外での研究の現状		9
6. 当面の検討課題と提言		9
6.1. 新しいタンパク質科学の振興に関わる公的議論の必要性		9
6.2. 目的提示、使命達成型の新しいタンパク質科学のための研究拠点の整備		9
6.3. 新しいタンパク質科学の振興、推進、支援に関わる研究推進機構の必要性		9
6.4. ゲノム統合科学の研究基盤の拡大と教育機構の整備		9
付録		
1. プロテオミクスセンターがモデル研究として取り組む研究課題案		11
2. 構造ゲノミクス研究および構造生物学のための技術開発および課題遂行		12
2.1. 技術開発投資		13
A. ハイスループット構造解析のための要素技術開発		13
A1. 無細胞系タンパク質発現システム		13
A2. タンパク質の自動結晶化技術の開発		13
A3. ハイスループットタンパク質結晶構造解析ビームラインの開発と設置		14
A4. 次世代2次元X線検出器の開発		14
A5. NMRによる高効率、高精度構造解析技術の開発		14
B. 構造解析が困難なタンパク質の解析技術開発		16
B1. 糖鎖タンパク質大量生産技術の開発		16
B2. 膜タンパク質の大量調製、結晶化技術の開発		16
B3. 生体超分子複合体の大量調製技術の開発		16
C. 超分子複合体の単粒子構造解析あるいは超微結晶構造解析技術		16
C1. 第四世代放射光XFELによる構造生物学計画		16

C2. 極低温電子顕微鏡によるタンパク質複合体の高分解能単粒子構造解析技術の開発	17
C3. 次世代2次元電子線検出器の開発	17
D. タンパク質の構造機能連関を水素の配置を中心に詳細に探る技術開発	18
D1. 大強度パルス中性子線を用いた構造生物学	18
2.2. 実行研究予算	18
A. 網羅的構造ゲノミクス(拠点型)	18
A1. 複数の特定生物系、細胞系あるいは機能単位系についての網羅的なタンパク質構造解析	18
B. 生物学的重要なタンパク質の構造機能研究(個別研究)	18
B1. 生物学的に重要な個別の系についての機能研究を絡めた構造解析	18
3. バイオインフォマティクスおよび計算生物学における技術開発	18
4. ハイスループット細胞4次元プロテオミクス	20
用語解説	21

参考図1 ゲノム科学総合分野の研究連関

参考図2 タンパク質科学戦略構想

参考図3 プロテオミクスの研究分野

参考図4 プロテオミクス領域研究機関(新設)(案)  
「プロテオミクス研究センター」(仮称)の構想

## 1. 序  新しいタンパク質科学の創出に向けて

生命科学は、生命現象の素過程とその制御機構を、物質、情報、エネルギーなどの視点から研究し、

これに基づいて生命の普遍性と多様性、決定性と可塑性の分子基盤、さらには生命の起源と進化などの諸問題を解明することを目標としている。今世紀の生命科学研究は、生命が織りなすさまざまな現象が、ゲノムに刻まれた情報によって生産されるタンパク質によって演出されていることを明らかにしたが、その生体内での動態や個々のタンパク質の構造と機能の関連、ならびにその相互作用が作り出す「情報ネットワーク」の実態は未解明であり、生命の物語を分子の言葉で十分に理解する段階には至っていない。

ヒトや多様な生物の一生を規定するゲノム情報の全貌が明らかになりつつある現在、その膨大な情報を背景として、新しいタンパク質科学が生まれ出ようとしている。

その第1は、細胞を舞台として、生命活動を演出しているタンパク質の総体、すなわち「プロテオーム」を包括的に解析対象とする分野である。タンパク質社会学として誕生したこの次世代の生命科学領域は「プロテオミクス」と命名された。プロテオミクスは、タンパク質の発現情報を迅速かつ大量に読み取る手段を駆使し、生命活動のある瞬間や特定のステージで機能しているタンパク質群の種類と動態、構造、局在と分子集合状態、相互作用を大規模かつ体系的に解析することで生命活動の分子基盤を明らかにすることを目指している。

第2は、生命活動を担う基幹的タンパク質の構造と機能を、ゲノム情報に基づきつつ、網羅的に、あるいは個々の機能システムを個別にその動的側面を重視する立場から解明し、得られた情報を統合的に解釈する分野、すなわち「構造生物学」と呼ばれる研究領域である。多くの細胞における普遍的生物機構には遺伝子発現・複製、タンパク質合成、信号伝達処理、物質代謝、エネルギー代謝、物質輸送などに加えて、これらから派生して進化した運動、感覚から脳機能に至る高次機能などがある。関与する分子機械の種類と数が増えてシステムが高次になればなるほど、このような生物機構を統御するネットワークと扱われる情報量は指数関数的に膨大になる。一般的にはこれが、生命特有の興味深い特徴、すなわち非常に緻密で精巧である一方、柔軟で曖昧な部分を持った仕組みの基盤を造りあげていく要因となると考えられている。ここで、ゲノムが生産しうる全ての分子機械の構造に焦点をあてて網羅的に解析する組織的研究プロジェクトは、「構造ゲノミクス」プロジェクトと呼ばれるようになってきている。ただし、ゲノム情報が利用できるようになった新しい時代の構造生物学においても、個々の機能システムについて動的な構造機能連関に深い洞察をもたらす個別の構造生物学的研究は、従来にも増して重要である。このような個別の構造生物学研究と、構造ゲノミクスプロジェクトとを、バランスよく同時に推進することが、生命現象を包括的に理解する基盤づくりに直結する。

第3は、種々のタンパク質データベースを構築しそれを活用するバイオインフォマティクスと呼ばれる新しい研究分野である。プロテオミクスおよび構造ゲノミクスの研究進展が、ハイスループットのデータ生産とその多様化を促すことにより、大量のタンパク質関連の情報が蓄積される。バイオインフォマティクスと計算生物学は、これらの生データを整理し研究者が利用しやすい公共データベースを構築、運営するとともに、それらデータベースの高度な利用技術開発を緊急に展開することによって、タンパク質の構造機能相関に基づいた、高次構造、生化学的機能、相互作用機序、細胞内局在、動態などの知見を体系化し、理論的予測手法を確立していかなばならない。こうしたデータベースと計算生物学による細胞内、細胞間情報ネットワークの体系的な解析、予測によって、初めて生命現象の包括的理解が可能となろう。

上記の3分野はゲノム情報を一次情報とするゲノム統合科学の基幹となる新領域であり（参考図1参照）、わが国の科学政策を国の内外に明確に示す形において、省庁の壁を超えてその研究体制が整

備され、拠点施設の充実とともに個別分散研究支援の拡充、さらには人材育成のための教育制度の拡充と柔軟化が図られる必要がある。

## 2. ゲノム情報に依拠するタンパク質科学の特質

### 2 - 1. ゲノム科学のタンパク質科学への展開

ゲノム情報に依拠し、新しいタンパク質科学を包括するゲノム統合科学は、21世紀の人類社会が最も期待を寄せる生命科学技術である。これまでに得られた小規模のゲノム情報からも、生命の実像とその多様性にかかわる認識がおおきく変革されてきた。今後10年間のゲノム統合科学の研究成果はこの趨勢に拍車をかけ、われわれの生命観、物質観、宇宙観を抜本的に啓明することとなる。また、ゲノム統合科学は、生命科学の領域に留まることなく、物理学、化学、情報科学、宇宙科学の諸分野と融合し、新しい自然科学の建設の基盤となるであろう。その汎学術性と巨大性のゆえに、ゲノム統合科学は、個人や国、地域を超えた国際協力の精神と地歩に立ち、共同研究推進体制と成果公開、共有の理念を掲げなければならない。

わが国の政府、行政機関はこの基本的命題に深い認識と理解を示している。2000年5月2日に開催された日米科学技術協力協定に基づく第8回合同高級委員会（共同議長、中曽根弘文科学技術庁長官およびニール・レーン米大統領補佐官（科学技術担当））は、その共同コミュニケにおいて、ヒトゲノム解読計画の終了にともなう第2ステージのゲノム統合科学が、ヒトの全タンパク質の構造・機能の解明を目指す国際プロジェクトの発足に向けて両国が協力し、国際的なイニシャチブをとることを宣言している。

国際生化学・分子生物学連合は、第18回総会（2000年7月、英国バーミンガム市）の特別講演者として、J. Craig Venter 博士（Celera Genomics、USA）を招待し、“Decoding the Human Genome”の成果を公式に評価した。同氏はヒトゲノムの科学がようやくその端緒についたことを述べ、プロテオミクスが次なる緊急のターゲットであることと、その研究のグランドデザイン、波及効果につき熱弁をふるった。多くの識者が指摘しているとおり、わが国は、和田昭允博士（東大名誉教授）による先駆的提唱にもかかわらず、国際ヒトゲノム計画においてイニシャチブを取り損ね、その実践実行においても先進するところが少なかった。科学者の総意の結集が未熟であったことや、政府、行政機関と学会、産業界との間に事実認識、波及効果の評価に温度差があったためと思われる。ゲノム統合科学が21世紀の人類社会が最も期待を寄せる科学技術であることに鑑み、わが国は、そのもてる総力を結集し、明快な理念と実行責任を開示した研究構想を立て、国際協力体制を構築していくことが望まれる。

### 2 - 2. タンパク質科学の特質とそのための研究体制

このようにゲノム情報解読に始まったゲノム科学は、遺伝子産物であるタンパク質の科学へと拡大しつつあるが、タンパク質にはその本質に由来する特質があり、その特質を踏まえた科学技術振興策を必要としている。DNA上の遺伝情報は、そこに書き込まれた情報の内容に無関係な均一な記録様式に特徴があり、そのために塩基配列読み取りはルーチン化、工場化が可能であった。それに対してタンパク質は情報内容の表現であり、多様性に第一の特徴がある。そのためタンパク質の研究は、従来から個々の特徴に即して、対応する生命現象の素過程を解明する意識の下、個別の研究が丁寧に進

められて来た。今後とも、そのスタイルの研究、すなわち、個々の研究者、研究室の独自性、独創性に依存した個別研究は、その重要性を減じることはない。しかし、従来型のこの研究と並列して、ゲノム情報に基づき、細胞内に存在するタンパク質を総体として研究する新しいスタイルの研究の可能性と必要性が付け加わった。このスタイルの研究は、タンパク質の持つ本質的な多様性にも関わらず、多種類のタンパク質を研究対象にしなければならない使命を持つ。これは今までのタンパク質研究になかった側面である。そのための技術も概して未成熟であり、新しい研究戦略を必要としている。

網羅性を追及するこの新しいスタイルの研究は、技術が概して未成熟であるにもかかわらず、一般に目的を提示し、その使命の達成を図る。従って、このスタイルの研究には、(1)そのための技術開発と、(2)各時点での最適技術を組み合わせ、網羅的研究を実際に遂行することの二つの面がある。技術開発の現段階は、概して多様な選択肢の検討を必要としている。そこでは一極集中的な研究投資方針を採ると、技術の選択肢が減り、結果的に国際的競争力を失う危険性が大きい。明確な目標を設定し、しかし多くの研究者、研究室が多様な可能性を競争的環境下でそれぞれに追求する目的設定並列競争型研究が適している。(2)の網羅的研究には、それぞれ固有の明確な使命を持った拠点を設立してその遂行にあたるのが適切である。

以上をまとめ、タンパク質研究の特質に鑑み、上述の個別研究および目的設定並列競争型研究の支援をし、また使命達成型研究拠点の運営に関わる省庁の壁を超えた研究推進機構の設立を提言する。

### 3. プロテオミクス研究の目標と課題

#### 3-1. プロテオミクス研究の目的

プロテオミクスは、細胞内に発現しているタンパク質全体の動態、構造、細胞内局在、相互作用等を体系的に研究する領域であり、次節で述べる構造ゲノム学とともに、次世代の生命科学研究の中核をなす領域である(参考図1・2参照)。その柱の1つはゲノムの発現、活動状況を把握するための「タンパク質発現プロファイルの体系的解析」である。ある瞬間の生命活動の分子的基盤を理解するためには、その瞬間を演出しているプロテオームの実像を知ることが不可欠である。プロテオミクスでは、細胞や組織に存在する機能タンパク質群の動態を時間軸に沿って大規模に解析することで、生命活動の実態を計測し、解析する。また、血液などの体液や各種の組織、器官に存在するタンパク質全体を俯瞰し、その動態を追跡するタンパク質のプロファイリング研究は、総合的な健康診断や臨床診断、新たな医薬品開発のための創薬ターゲットの同定、さらには医薬品の作用、副作用をモニターするための高度医療技術としても注目される。

プロテオーム研究が目指すもう1つの目標は、タンパク質機能の解明である。この目的に向けたプロテオミクスは、タンパク質の細胞内局在、分子集合状態、あるいはリガンドに依存した動的なタンパク質相互作用を包括的に解析する。ゲノム情報によって生産されたタンパク質は、適切な修飾を受けた後に細胞各所に輸送され、他のタンパク質や遺伝子、その他の生体成分と相互作用を繰り返すことによって生命現象を演出する。また、翻訳されたタンパク質への糖鎖やリン酸基の付加や脱離、プロセッシングなどの修飾は、細胞内でのタンパク質の高次構造形成や機能発現、代謝、さらには相互作用に基づくシグナルネットの構築に不可欠である。このような修飾と相互作用のネットワークは、タンパク質の発現プロファイルをデータベース化して大規模に解析するプロテオミクスデータバンクによってはじめて研究することができる。例えば、生体内に存在するスプライズゾーム複合体やヌクレアソームなどの細胞内小器官やシグナルに依存して形成される複雑な転写装置の構造解析、タンパク

質の高次構造形成を触媒するシャペロンタンパク質による生合成後のフォールディング機構など、従来の方法では困難であった生命現象の分子的基盤が次々に明らかにされている。脳研究の分野でも、脳の高次機能を支えるタンパク質分子の1つである N-methyl-D-aspartate (NMDA) 受容体複合体や、シナプス後膜を構成する高密度タンパク質複合体の解析に適用され、脳機能に関する重要な知見を与えている。生命は単なる分子集団の総和ではなく、高度に組織化された有機体であることを考慮すると、プロテオミクスは分子の組織化の過程を明らかにし、生命を構築するための原理を研究するための強力なツールである。

### 3 - 2 . プロテオミクス研究の意義と波及効果

プロテオミクスの技術革新と研究開発、とりわけタンパク質機能と細胞情報ネットワークの解析は、ゲノムと生命に関する膨大な情報を生み出すと同時に、将来の産業技術創成のシーズとなることが期待される。医薬領域では、糖尿病や高血圧などのいわゆる生活習慣病や老化にともなう疾病、癌、環境ストレスによるアレルギー疾患などの多遺伝子疾患の原因の解明や創薬、総合的な臨床診断、遺伝子組換え動植物の生体への影響評価などへの応用のほか、タンパク質新素材の開発やバイオインフォマティクスを中心とした新規情報産業の創成が期待できる。プロテオーム研究は、さらに高度な医療を目指すゲノムの多型解析、インフォマティクスを駆使した構造ゲノム解析とともに、次世代の生命科学と関連産業の中核を担う重要な基盤と位置づけられる。国際的な研究動向に照らし、プロテオミクス研究の振興推進のため、適正規模の複数の中核研究開発拠点の設置が必要である。この研究拠点は、多次元の包括的タンパク質分離システムと質量分析計各種を専用ラインに敷設し、自動制御による連続運転により、特定の生物種、細胞集団のタンパク質のプロファイリングを経常的に行うほか、一分子光学計測技術等による各タンパク質の細胞内局在、動態、相互作用を時間軸に沿って4次元的に観察記録し、それらの統合データベース化を行う。そのためには、プロテインチップの開発、従来の2次元電気泳動ゲル(シート)の改善、タンパク質スポットの高速自動分離分析装置システムのロボット化や、多次元液体クロマトグラフィーと質量分析計を複合化した次世代の高性能自動タンパク質分離解析システムなどのプロテオーム迅速解析のための基本技術、機器の革新、さらには一分子光学計測のための光学顕微鏡高感度化や超高感度 HDTV を応用した自動記録解析技術などの要素技術開発、改善、汎用化を緊急に行うことが必要である。この中核研究開発拠点は、同時に、産官学にまたがる新技術創成、移転の場として運営し、機能システムごとに分散して進められる個別型研究の財政的支援をも含めて、内外の緊密な連携を保ちつつ研究推進をはかることが必要である。そのために、前述の拠点支援体制のもとに省庁の壁を越えた「研究課題公募・新技術創成移転型」の研究振興、支援機構を設置することを提言する。

### 3 - 3 . 国内外での研究の現状

すでに欧米ではプロテオーム研究が21世紀の生命科学研究の先端として位置づけられ、欧州共同体ならびに豪州においては国家プロジェクトとしてプロテオーム解析のための中核研究開発拠点の設立などの基盤整備が行われている。また、米国では産官学共同研究体制の中で、政府の支援のもとに

多くのベンチャー企業が起業されている。民間では、国際的製薬企業であるグラクソウェルカム社が英国本社ならびに豪州に同様の研究所を設立し、ロッシュなどの大手企業でもプロテオーム研究に特化した研究開発ユニットを整備している。わが国においてもここ数年、理化学研究所や大阪大学蛋白質研究所、東京大学医科学研究所などの研究機関あるいは日本生化学会などを母体として、プロテオーム研究の現状や課題についてのシンポジウムや国際会議が数多く開催されている。一方、政府が出資するプロテオーム研究関連予算としては、プロテオミクス基礎技術整備のためのプロジェクトが経済産業省(NEDO)ならびに文部科学省において個別に実施されているのみである。

わが国においては、ヒトやマウスを対象とした完全長 cDNA ライブラリーの構築が世界に先駆けて進行し、その成果を国内外の研究に広く活用している。また、構造ゲノム科学の分野では、理化学研究所を中心に大規模な基盤整備が進められている。さらに我が国は、プロテオーム研究を支援するタンパク質科学の領域において伝統的に強固な基盤をもち、来春に向けて関連学会の統合による「日本蛋白質科学会」設立の準備も進められている。本提言は、こうした学術的背景や人的資産を最大限に活用し、生命科学研究領域での我が国の優位性を確保するとともに、この分野における国際貢献と国際競争力を保持し続けるために、21世紀初頭の科学技術政策でのプロテオーム研究の位置付けならびに必要性を明確に策定し、その基盤整備のための緊急な対応を提言するものである。

#### 4. 構造ゲノミクスおよび構造生物学の目的と課題

##### 4 - 1. 構造ゲノミクスおよび構造生物学の目的

多様な生命活動を支える物質的実体は細胞内、細胞間に形成されているタンパク質からなる分子機械ネットワークである。それはゲノムに記録された設計情報に基づいて構築されるが、ゲノムの持つ情報量に比べて桁違いに大きな階層的情報量を創り出すため、生命の非常に複雑で生命らしい機能が産み出される。特定の細胞機能や生物個体の多様な生命活動の場においては、数十、数百、あるいは数千種の分子機械が相互作用してネットワークを形成している。したがって、その機能を解明し、医療、創薬、その他様々な産業利用に役立てるためには、これらのネットワークに關与する全てのタンパク質の立体構造と機能を明らかにすることが必要である。この目的に沿った動きの1つが、最近のこの分野の大きな研究動向となった目標提示型の網羅的研究、すなわち構造ゲノミクス研究である。それと並行して、網羅的研究は個々のタンパク質の機能に深く立ち入って洞察を与える余裕がないため、従来から主要な生物機能システムごとに個別に遂行されてきた動的構造機能連関を重視した構造生物学的研究は、この分野の両輪のもう片側として、必須のものである。

構造ゲノミクス研究では目的を提示し、その使命の達成を図る。しかし、従来個別的研究を前提に開発されてきた技術は網羅的研究に必ずしも適合していないため、その推進にはタンパク質の大量生産、精製、結晶化、迅速構造解析法等、構造決定に至る多くの工程のそれぞれにおいて新しい技術革新が必要であり、現状では多様な可能性の追求が必要である。従って、構造ゲノミクス研究には、(1) そのための技術開発と、(2) 現段階での最適技術を組み合わせ、網羅的構造決定を実際に遂行することの二つの面がある。技術開発の現段階では一極集中的な研究投資方針を採ると、技術の選択肢が減り、結果的に国際的競争力を失う危険性が大きい。米国ではNIHの方針の下、複数の研究グループが競争的環境のなかで分散して関連技術の開発に携わりながら網羅的構造決定も推進しつつある。

このため、リスクも少なく結果として領域全体の成果も期待できる。技術開発課題としては、網羅的高速構造決定のための技術要素だけではなく、放射光X線、電子線、NMR、中性子線など、それぞれの実験領域における先端技術開発をも含めなければならない。

#### 4 - 2 . 構造ゲノミクスおよび構造生物学の意義と波及効果

構造ゲノミクス研究を効率よく進めるには、複数の先端技術開発プロジェクトを有機的に統合し、その成果を活用して、少数の拠点で行われる組織研究プロジェクトを効率的に運営しなければならない。もちろんその成果は構造生物学の個別研究プロジェクトにも活用される。短期的技術開発の課題としては、無細胞系タンパク質大量生産、精製システムの確立、糖鎖タンパク質、膜タンパク質、生体超分子複合体の大量生産技術の開発、安定同位体標識技術の開発、タンパク質自動結晶化技術および膜タンパク質の結晶化技術の開発、ハイスループットを可能にするタンパク質結晶構造解析ビームラインの設置、NMRスペクトル自動解析技術等の解析技術の高速化、自動化、等がある。中・長期的課題としては、1 GHz 級高磁場NMR装置の開発、次世代2次元X線検出器の開発、極低温電子顕微鏡によるタンパク質複合体の高分解能単粒子構造解析技術の開発が必要であろう。また、第四世代の放射光、X線自由電子レーザー(XFEL)による、1ミクロン以下の超微小結晶やウイルス等の単粒子の構造解析などの実験計画とそのフィージビリティの検討なども要請される。

#### 4 - 3 . 国内外での研究の現状

構造ゲノミクスという、ゲノム情報を補完し、大規模に拡張、展開するこの巨大科学は、個々の機能システムについての分散個別型構造生物学研究とともに、その成果として莫大な規模の社会貢献と経済効果が見込まれるため、すでに世界各国で多くのプロジェクトがスタートし、各国政府が本格的な支援を開始している。国内では、理化学研究所が、播磨研究所と横浜研究所ゲノム科学総合研究センターとの統合化されたプロジェクトとして、様々な生命現象に注目して、高等動植物の完全長 cDNA や高度好熱菌ゲノムからターゲットとするタンパク質を調製して、X線結晶構造解析とNMRによって構造ゲノミクスを行う「構造ゲノム科学イニシアティブ」を開始している。他方、新設の生物情報解析研究センターでは、膜タンパク質に特化した構造ゲノミクスのプロジェクトが、新規技術開発を中心として、企業からの広範な参加も得て、スタートしている。さらに、農林水産省の研究所で、イネの構造ゲノミクスが計画され、スタートしようとしており、また、国立大学などを中心にショウジョウバエの発生に関する構造ゲノミクスなどの企画も開始されている。これらの計画研究とは独立に、リボソーム、転写装置、複製装置、輸送装置などの超タンパク質複合体の構造が、高級な生物機械装置として作動する原理を、その機能状態との関連で解明するあらたなプロジェクトを発足する必要がある。今後さらに様々な細胞機能や生物種についての構造ゲノム科学研究を推進することが強く望まれるが、それぞれが比較的巨額の研究予算を必要とする組織研究であるため、政府や業界による的確な研究支援が重要である。また、プロジェクトの効率的な推進をはかるためにも、組織研究の運営、支援形態、分散研究との連携体制に注意深い考慮を必要とする。

ドイツでは、ヒト cDNA の基礎、応用研究を目指す巨大プロジェクトの中で、構造ゲノミクス・プロジェクト(Protein Structure Factory) が本格的に開始している。カナダでも、超好熱菌の構造ゲノミク

ス・プロジェクトがスタートしていたが、今年から本格的にスケールアップした。同じく今年、米国 NIH の NIGMS (National Institute of General Medicine and Sciences) が、全米のエリアをカバーするように、7つの構造ゲノミクス・プロジェクトを選んでおり、研究助成を開始するところである。そこでは、結核菌の網羅的なタンパク質構造研究やシグナル伝達系の系統的な構造解析等、医学上重要な細菌もしくは特定の細胞機能が主としてターゲットとして選ばれた。また、個々のプロジェクトは、大規模構造解析をスタートするとともに、それぞれ技術開発面で特定の課題を明示して取り組んでいる。引き続き来年も、2-3のプロジェクトを追加する予定である。さらに、フランスは、3つの構造ゲノミクス・プロジェクトを開始することを決定した。それぞれ、病原菌、酵母、および高等動物を対象とする。イギリスの Wellcome Trust は、SNPs コンソーシアムとよく似た仕組みで企業を集めて、構造ゲノミクス・コンソーシアムを開始することを発表した。これらの多くに共通することは、機能ゲノミクス、機能プロテオミクスなどの網羅的な機能研究との連携が重視されていることである。また、立体構造の特許化、創薬など、産業との連携も重要なファクターとなっている。

理化学研究所の NMR による構造決定プロジェクトは当初約千種類の基本構造を決めることを目標として先行的に出発したが、その後の議論の中で NIGMS は、国際的な構造ゲノミクスプロジェクトの協力により、それぞれのタンパク質ファミリーから1個以上、全体で約10,000の構造を10年で決めることを目指すとしている。課題の設定面だけではなく、実施計画にも特徴がある。NIH も英国 Wellcome Trust も、学術的に意義のある競争的なプロジェクトを選抜し、数年間にわたって研究資金援助をし、各グループの成果の厳密な評価に基づいて、次のより大規模な集積型の資金援助を行うという二段階構想で進んでいる。我が国でもこれにならば、一極集中をさげ、まず第一段階のパイロットプロジェクトを競争的環境で複数スタートすることが、この分野の健全な発展を促す重要な施策となるであろう。それとともに、機能システムごとの動的構造機能連関を重視した分散個別型研究に対するバランスのとれた支援と連携が、この分野の振興の重要な鍵となるであろう。適切な施策が実施されれば、我が国がこの分野で世界をリードするポテンシャルは非常に高い。

## 5. バイオインフォマティクスの目的と課題

### 5-1. バイオインフォマティクスの目的

プロテオミクスと構造ゲノミクス研究は、タンパク質の構造・機能に関する様々なデータのハイスループット生成をもたらす。したがって、

- (1) それらの多種多様大容量のデータを整理統合し、公共のインフラストラクチャとしてのデータベースを構築する。
- (2) それぞれの種類データの記述形式、項目、及び内容記述のためのオントロジー（または統制された語彙）に関する業界標準を速やかに提案、改善する。
- (3) 必要なデータを、一定の質を確保しつつ効率的かつ網羅的に収集し、安定して提供するための組織と体制を整える。

などの諸課題を解決しなければならない。

今後公共データベース化への需要が高まるデータの種類として、既に国際的な協力体制のもとにデータベース化されている核酸塩基配列、タンパク質立体構造座標に加えて、二次元電気泳動法等によるタンパク質発現プロファイル、two-hybrid システムによるタンパク質間相互作用データ、質量分析、NMR 等による生体分子間相互作用データ、タンパク質変異、多型データ及び表現型との相関データな

どがある。また、タンパク質立体構造決定のハイスループット化に伴ない、最終データとしての座標のみならず、例えば NMR による距離制約情報などの中間段階のデータ（いわば、ゲノムのドラフト配列に対応するドラフト構造）の組織的収集と公開に対する需要が高まっている。既に欧米を中心に、そのためのデータ記述形式の検討などが始まっている。

一方で、これらのデータベースを有効に活用するためのバイオインフォマティクス、計算生物学研究の展開が求められる。プロテオミクス研究からもたらされる生体分子間相互作用データや、構造ゲノミクスおよび構造生物学によって解明されるタンパク質およびタンパク質複合体の立体構造データ等、タンパク質に関する多様なデータを利用、解析して生物学的に有用な知識を獲得するための諸方法を開発することが必要となる。その際、生物学者と数学者、情報科学者等他分野の研究者との間の有効な協力関係が進展することが期待される。そして、それらのデータ利用技術開発と連携しながら、分子シミュレーション、分子間ネットワークシミュレーションなどの、生物学上の個別的または原理的な問題に関する理論的研究を進めることが重要である。

## 5 - 2 . バイオインフォマティクスデータ (BID) 整備の意義と波及効果

タンパク質に関する多様なデータを有効に組織化するには、どのような形式と項目でデータを記述するかということが非常に重要になる。今後、タンパク質の構造・機能に関連する多様な記述のための標準的オントロジーが求められるであろう。日本は構造データ生産の加速化と同時にこの点に関しても貢献を果たすべきであろう。以上のようなミッションをもつ公共データベースを構築、運営するためには、適正規模の BID センターが国内に一つ乃至複数必要であろう。そこにおいて、データ生産者であると同時に利用者である生物学および医学の研究者が主導して、必要な情報技術を有効に利用しながらデータベースを構築する必要がある。その際、データベースセンター間の密接な連携、データ生産者に過度の負担をかけず網羅的なデータ収集を可能とする体制と研究者コミュニティ内の合意が重要である（参考図 1 及び 3 参照）

プロテオミクスおよび構造ゲノミクス関連の諸研究は、タンパク質の立体構造データとともに、タンパク質 生体分子間相互作用に関するデータを大量にもたらす。それらのデータ解析を通じて、相互作用を行なう分子内原子群の空間配置パターンに関する経験則が蓄積することが期待され、医薬品の合理的、効率的設計やナノデバイスの開発を推進するために役立つであろう。さらに、今後は、立体構造比較によるフォールドの全体的類似性の検討、機能未知タンパク質の立体構造比較に基づく機能推定が広く行なわれるようになるであろう。

## 5 - 3 . 国内外での研究の現状

上述のデータベース構築、利用技術の開発は、プロテオミクス、構造ゲノミクスの基盤となるだけでなく、新しい理論生物学や機能プロテオミクスの延長線上に現れてくるシステム生物学の建設に寄与する。例えば、薬理学との関係で新たな脚光を浴びる可能性のあるプロテオミクスの課題として、SNPsに伴なうアミノ酸残基置換が及ぼす分子間相互作用（分子機能）への影響の予測がある。これは、SNPs研究により蓄積する遺伝子型の多型と個体レベルの表現型（疾患感受性、薬剤応答性等）の違いの相関のメカニズムを、分子レベルで理解するための指標をあたえることになるであろう。また、近年の分子擬態の予言から興味をもたれはじめたトリペプチドアンチコドンの解明もタンパク質化学の構造・機能・反応性の相関機構を研究する新しい分野を拓くものと思われる。

個別タンパク質の構造・機能情報と生体分子間相互作用情報が蓄積される中で、それら分子レベルの現象と細胞、組織レベルの現象とを橋渡しする理論の構築が求められる。システム生物学は、生体分子群を要素として、それらの相互作用のネットワークから構成される生体システムの全体（例えば、

シグナル伝達系等の分子系あるいは細胞全体)を計算機上でシミュレートし、生物学上の仮説の検証と現象の予測を行なうことを目的とする。米欧日で先駆的な試みが既に始められている。その実用上の価値は現時点でまだ明らかでない。しかし、分子レベルでの知見を医療、創薬に役立てる上で実験的アプローチを補完するものとして、特にヒトを対象とした仮説の検証における可能性が期待される。

## 6．当面の検討課題と提言

### 6 - 1．新しいタンパク質科学の振興に関わる公的議論の必要性

プロテオミクス、構造ゲノミクス、構造生物学、バイオインフォマティクス等の新しいタンパク質科学の振興に関わる学術行政を、専門家の参加を得て論議する公的委員会の設置を提言する。

### 6 - 2．目的提示、使命達成型の新しいタンパク質科学のための研究拠点の整備

プロテオミクス、構造ゲノミクス、バイオインフォマティクスの3分野に目的提示、使命達成型の新規国家プロジェクトを立ち上げ、その使命達成のための複数の適正規模の研究拠点の新設を提言する。

### 6 - 3．新しいタンパク質科学の振興、推進、支援に関わる研究推進機構の必要性

プロテオミクス、構造ゲノミクス、構造生物学、バイオインフォマティクス等の新しいタンパク質科学の振興、推進、支援に関わる次の3つの機能を持つ研究推進機構を設けることを提言する。(1)上記の国家プロジェクトの使命達成のための研究拠点を運営し、(2)上記のプロジェクト等で用いる技術を競争的環境下で開発するための「課題公募、新技術創成移転型」研究を支援し、また(3)使命達成型になじまない多様なタンパク質の個性に即した多数の個別研究を支援する。

### 6 - 4．ゲノム統合科学の研究基盤の拡大と教育機構の整備

ゲノム統合科学が領域融合型の新しい自然科学の基礎となることの認識に立ち、大学、研究所などを中心とする教育、研究機構を整備することを提言する。

## 付録

本付録では本文で記述した各種の研究課題、技術開発要素を具体的に例示し、そのために必要な予算額のイメージを得ることを目的にその概略を与える。

### 1. プロテオミクスセンター（仮称：参考図4）がモデル研究として取り組む研究課題（案）

「ヒト幹細胞の分化、組織形成とその異常に起因するプロテオームの動態変化と機能ネットワーク解析ならびにプロテオミクス新技術の開発」

プロテオミクスは、生命現象の変化を演出するタンパク質の発現変動を大規模に俯瞰（プロファイリング）して機能タンパク質群をクラスターとして把握し、その中に含まれるタンパク質相互の繋がり（リンケージ）を解析することで生命現象を担う分子機構や代謝経路を明らかにする。生命科学の従来の方法論とは異なるこうしたプロテオミクスの演繹的な研究戦略を適用する実験モデルとして、多彩な分化過程が知られているヒトの幹細胞を選び、細胞分化と組織形成過程、ならびにその異常に起因するがんなどの疾病と薬物投与によるプロテオームの動態と相互作用変化を網羅的に調査することで、細胞機能ネットワークの解明とデータベースの構築を目指す。また、この目標の達成に向けたプロテオミクス新技術の研究開発を実施し、その適用性について評価、研究する。

例えば、造血幹細胞に由来する形態学的免疫学的分類が可能な25形態の細胞について、その分化過程を時系列に沿って10ステップに区切り、それぞれのステップにおけるプロテオーム変動を網羅的に解析すると仮定すると約60万個のタンパク質シグナル（スポットまたはピーク）の同定作業が必要である。これと平行して、変動が観察されたタンパク質クラスターの相互作用リンケージを包括的に明らかにして細胞分化や組織形成を規定する分子機構や代謝経路（細胞機能ネットワーク）を明らかにすることをプロジェクトの目標とする。さらに、各種の組織や器官形成の異常、薬物投与の影響をプロテオームの動態と相互作用変化を指標として解析することで、その過程の分子機構の解明と創薬のための基本戦略の創成を目指す。この研究の過程では、細胞内に存在するさまざまな機能性タンパク質の動態や複合体組成をはじめとして、シグナル伝達系などを構成する従来未知のタンパク質の修飾や相互作用が次々に明らかになることが想定される。この成果は、構造ゲノミクスや構造生物学が対象とすべき多くの新たなターゲットを提供するだけでなく、「メカニズムに基づいた創薬」へ向けた合理的なアプローチを加速すると考えられる。上記の課題を達成するために最新の質量分析装置を導入し、現在の技術でこの目標を5年間で達成するためには、施設建設費を切り離れた初期設備費で10～15億円、年間経費として約10億円の経費が必要と試算される。

一方、このプロジェクトを効率良く推進するためには、実験科学の研究現場に密着し、共同で研究を推進するバイオインフォマティクス部門が不可欠である（本提案書5項および付録）。したがって、センター全体の組織体制は、試料調製からデータ解析までを一貫して行うバイオインフォマティクスと連携したプラットフォームとして整備し、実験科学と情報科学を駆使して目標を達成する目的志向型の研究組織とする必要がある。また、センター周辺には、プロテオミクスに関連した技術開発を専門に行う研究チームから構成するコアテクノロジー部門を設置する。コアテクノロジー開発部門では、上記の発現プロテオミクスならびに機能プロテオミクス研究に必要なプロファイリング技術の

高性能化と自動化、ならびにタンパク質の細胞内局在や機能性複合体の組成、修飾などに依存したタンパク質相互作用を効率よく迅速に解析する新規リンケージ解析法の開発研究を実施する。これらの研究には、タンパク質の細胞内での動態を蛍光プローブと超高感度 HDTV 技術などを応用して「on-site」で直接観察する挑戦的な方法論の開発や、観察された記録をデータベース化して共有するシステムの整備も含まれる。こうした技術開発は、本プロジェクトの推進を加速するとともに、我が国のプロテオミクス研究の裾野を拡げ、新しい人材の育成と医薬領域などへの新たな応用を生み出すことで、新規産業育成のためのシーズとなることも期待される。この部門には、ロボティクスや光学検出系などを専門とするハード、ソフトウェアの技術者や研究者を配置し、生命科学者との密接な連携のもとに研究開発を行う体制を整備するとともに、現在まだ挑戦的なレベルにある質量分析計や磁気、光学検出計などの科学計測機器や、チップ技術などを中心とするナノテクノロジーを積極的に導入して現在困難とされる技術課題のブレークスルーを計る。こうした研究開発を積極的に支援するためには、年間20～30億円の研究資金が期待される。

以上のプロテオミクス研究センター（仮称）をコアとして、センターでのプロジェクト研究を支援する付属施設として「研究支援施設」の設置が望まれる。この施設の業務としては、（1）生物材料として使用可能な各種培養細胞やモデル動物の提供、（2）無細胞タンパク質合成システムによるタンパク質標品の提供と抗体作成、（3）ナノスケールレベルの超微量試料の操作と分析を可能とするマイクロフルイディクスや微量装置の設計ならびに加工、（4）定量的プロテオミクスやタンパク質の細胞内動態解析のための同位体作成と標識化、などが想定される。この施設の経常的な経費としては、概算で年間5～10億円程度が見込まれる。

さらに、センター機能を最大限に活用する目的で、国公私立の大学、研究機関あるいは民間の研究機関が利用する「応用、共同研究部門」、ならびにプロジェクトの研究成果の社会への還元を目指す「企業化促進研究開発部門」を周辺に配置して、全体を事務部門で統括する（参考図4）。センター周辺に付設する施設群はいずれも共同研究機関の活力によって時限で運用を計ることとし、事務部門では、センターで得られた成果について積極的に知的所有権の獲得を目指す。

最後に、本提案のプロテオミクス研究センター（仮称）は、我が国の次世代ゲノム科学の中核的研究拠点として、欧米ならびに環太平洋地域のゲノム関連研究機関との密接な連携のもとに運営されることが強く望まれる。とりわけ、アジア諸国で設立またはその準備が進められている多くの研究機関との情報交換および人材交流を積極的に推進することで、21世紀のゲノム科学の中心的役割を担うとともに、将来の人的資源育成を指向するアジアの拠点となることが期待される。

## 2. 構造ゲノミクス研究および構造生物学のための技術開発および課題遂行

タンパク質の構造・機能研究の総合戦略構想にも述べたように、均一な記録様式を特徴とするゲノムDNAの配列解析のルーチン化高速化に比べて、情報内容の4次元的表现型であるがために多様に第一の特徴があるタンパク質は、その機能および構造解析をルーチン化高速化することは遙かに難しい。それゆえ、従来の重要な生物学的ターゲットに的を絞った個別研究の効率化のためはもちろんのこと、網羅的研究であるために高速化ルーチン化がその成功度を本質的に左右する構造ゲノミクスには、構造解析技術のハード、ソフト両面にわたる先端技術開発による支援が必須である。

そのための技術開発投資として、短期、中期、長期に期待される開発目標と必要投資額の概算を、

いくつかの項目別に以下にまとめた。また、今後5年間にわたる、網羅的研究と個別研究の研究実行予算案を概算として最後に添付した。

## 2 - 1 . 技術開発投資 ( 短期、 3 年程度 ; 中期、 5 年程度 ; 長期、 7 ~ 1 0 年 )

### A . ハイスループット構造解析のための要素技術開発

A 1 . 無細胞系タンパク質発現システム ( 短期 : 1 5 億円程度 ) : ゲノム解読に DNA を効率よく増幅する PCR 技術が不可欠であったように、タンパク質立体構造解析にはタンパク質を高純度で大量に調製する技術が不可欠である。現在は大腸菌、酵母、昆虫細胞などに目的タンパク質の DNA を含んだ大量発現ベクターを組み込んで発現させるのが一般的である。しかし、生きた細胞をそのまま利用すると、発現をするべき外来タンパク質と細胞生理とのコンフリクトのため、多くの困難を生じる。また、細胞という閉じた系であるため、タンパク質の発現や正しいフォールディングを助けるための方策が取りにくいことも、大きな問題である。このような問題を解決する次世代のタンパク質発現法として、細胞のタンパク質合成系を取り出し、その効率的な稼働に必要な全ての条件を試験管内に整え、無細胞系でタンパク質を大量発現する技術開発が、理化学研究所と愛媛大学において進められている。ここで強調すべきことは、無細胞系によるタンパク質大量発現法の技術開発は、個別の基礎研究として、競争的な資金を獲得して、十数年をかけて行われたもので、大規模なプロジェクトを開始してからその中で行うには適さない、リスクをかけるタイプの開発の例である。当然、そのような基礎研究の成果を大規模化するには、スケールアップという別の観点からの様々な技術開発が要求される。ここで、多種類のタンパク質を発現して、さらに先に進むべきものをスクリーニングする場合、ベクターにクローニングすることなく、PCR で増幅した DNA を用いることができるので、ハイスループット化に非常に有効である。X 線結晶構造解析のためのセレノメチオン導入や、NMR のための安定同位体標識 ( 特に重水素化や位置特異的標識 ) などにも、威力を発揮することが示されている。実際は、タンパク質の生合成過程は、それらが本来局在する場所を反映して異なり、複雑なプロセッシングを受ける場合も多いので、今後は、これらの問題への対処が必要である。膜タンパク質などへの応用にも期待がかかる。なお、無細胞系のタンパク質合成キットの市販は開始されているが、合成量などに問題が残されており、今後も引き続き、研究者による基礎研究が必須である。単純なキット化にとどまらず、多種類のタンパク質の発現データのデータベース化や、公的あるいは民間の受託合成システムの確立など、一般の研究室でも容易に使用できるシステムを早急に整えることが必要である。

A 2 . タンパク質の自動結晶化技術の開発 [ 結晶化スクリーニング自動化技術開発、新しい結晶化法、結晶化セル、結晶化装置の開発 ] ( 中期 : 2 0 億円程度 ) : X 線タンパク質結晶解析技術のなかで、結晶化は各タンパク質の強い個性のためにもっとも多く試行錯誤を必要とし、時間と労力と豊かな経験を要求するステップである。良質な結晶さえできればその後のデータ収集の高速化と構造解析の自動化は進んでおり、結晶化が迅速構造解析のボトルネックとなっている。最近、結晶化ロボットや結晶化状況の自動判別記録システムの開発が主に外国企業で進められているが、我が国でもこの分野の技術開発に投資し、多くのアイデアを投入して本格的に取り込むべき時期である。また、試行錯誤的なスクリーニングに頼らず、タンパク質溶液の接する固体表面微細形状の工夫による不均一核形成

法など、系統的かつ論理的な結晶化技術の実現にも力を入れるべきである。

A 3 . ハイスループットタンパク質結晶構造解析ビームラインの開発と設置 [ ビームラインの完全自動化、実験制御ネットワーク経由リモート化 ] ( 短期 : 開発費 10 億円、設置費 1 億円 / ビームライン × 20 本、新規ビームラインの建設 8 億円 / ビームライン × 10 本 ) : 数々の構造ゲノミクス計画において膨大な数のタンパク質の構造解析を遅延なく進めるには、高度に自動化され、高精度で高速データ収集の可能なタンパク質結晶構造解析ビームラインが多数必要である。現状では、実験者が液体窒素デューワーから予め凍らせたタンパク質結晶を取り出し、回折装置にマウントし、位置合わせをして、実験ハッチからの退室操作を行うため、X線露光以外にかかる時間がビームタイムの大半を占める。ロボティクスを活用した自動回折装置や次世代X線検出器が実用化されれば、ビームタイム効率を 5~10 倍改善できる。数ミクロンの微小結晶の位置合わせには、タンパク質結晶の蛍光像を高倍率同軸顕微鏡と画像解析により可視化し、ミクロン精度のゴニオシステムを開発して使用する。既存ビームラインのハイスループット化だけでなく、多数の新規ビームライン建設が望まれる。

A 4 . 次世代 2 次元 X 線検出器の開発 [ 2 種類の 2 次元 X 線検出器の開発、性能評価 ] ( 中期 : 20 億円程度 ) : 現在はイメージプレートや CCD (Charge Coupled Device) が、X線検出器として広く放射光実験施設で使用されている。CCD は読み取り時間が 10 秒以内のため、データ収集サイクルタイムを大幅に短縮し、クライオ技術の進歩と相まって、構造解析の速度と精度を飛躍的に高めた。ただ CCD はダイナミックレンジが狭い。欧米諸国ではピクセルアレイ検出器の開発が進められているが、タンパク質結晶解析に必要な大面積を実現することには困難が予想され、高強度 X 線の入射に対する非線形性応答も心配されている。そこで、全く新しい方式の 2 次元検出器、いずれも国内技術である HDTV-FEXS (Field Emitter X-ray Sensor)、あるいは HDTV-X線 HARP 管 (High-gain Avalanche Rushing amorphous Photoconductor) を用いた次世代の X 線検出器の開発が望まれる。これらの検出器により 2 桁以上の性能向上が期待でき、構造ゲノミクスの飛躍的な加速に寄与すると予想される。

A 5 . NMR による構造解析技術の開発 ( 中期 : 50 億円程度 ) : NMR はタンパク質の立体構造情報を溶液中、固体 ( 結晶 )、或いは、生体膜中において入手できる極めて有効な手法である。溶液内におけるタンパク質の立体構造決定が報告されて以来 10 数年の歴史しかない新しい手法であるため、未だ、今後の大きな技術的発展が期待できる。

ハイスループット構造解析については、当分、溶液 NMR が中心になると考えられる。現在までに、装置の高磁場化、同位体標識試料 ( 特に均一な  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  標識 ) の調製技術の発展、多次元多核種パルス技術の開発、NOE を中心とする NMR パラメータを用いる立体構造決定ソフトウェアの開発等に支えられて、分子量 3 万程度迄のタンパク質の構造決定が可能となっている。ことに、分子量が 2 万程度以下の低分子量タンパク質を対象とする構造決定は、急速にルーティン化しつつあり、立体構造決定手法として着実な発展を遂げている。立体構造決定に必要な一連の NMR スペクトル測定を自動的に実施するシステムは既に確立している。今後、NMR 測定による良好な試料のスクリーニング、データ測定時間の短縮 ( 多次元測定の次元縮小等 )、シグナル帰属の自動化、構造決定の自動化 ( データベースの活用も含む ) などを図ることにより、低分子量タンパク質のハイスループット構造決定手段として十分に競合的な手法なると期待される。

従来の NOE 法では困難な解析（ドメイン配向決定、比較的高分子量な試料の解析など）でも、試料をバイセル、ファージなどを用いて配向させ、residual dipole coupling などを用いることで、構造解析が可能になりつつあるが、さらに多様な配向手法の開発が必要である。さらに、化学シフトや水素結合の直接情報など、様々な NMR パラメータも利用する総合的なシステムの構築と自動化によって、ハイスループット解析の対象を広げる開発が強く望まれる。ここで、NOE を主な NMR パラメータとして利用する場合は、特に分子量 3 万を越える試料で、プロトン・シグナルの重なりなどに起因する問題を解決することが高効率、高精度のために必須である。ひとつの方法として、位置、立体選択的重水素標識技術を拡張して、パラメータ測定用に選択した観測用プロトン以外の水素を総て重水素する次世代の安定同位体標識技術（測定感度を低下させることなく、分子質全体のプロトン密度の減少を達成し、分子量限界の向上を可能とする）の開発が進められている。この位置、立体的同位体標識アミノ酸の合成技術を 20 種類すべてのアミノ酸について確立する開発が急務である。

溶液 NMR は、分子間相互作用、動的特性など、タンパク質の機能的な側面の解析にも極めて有用で、そのための開発も重要である。ハイスループットなりガンド・スクリーニングや、結合インターフェースの同定など、有用性も確立してきた。安定同位体標識も、残基特異的標識法、インテンによるセグメント標識法などが開発され、タンパク質の機能部位などに注目する解析を可能にしている。さらに、多数のサブユニットを含むタンパク質、膜タンパク質など、従来は対象外であったような超分子システムについても、TROSY 法による分子量限界の緩和、分子間相互作用インターフェースの特定手法の開発など、今後の発展を期待させる開発が続いている。今後も、従来の枠に捕らわれない斬新な方法論開発が強く望まれる。

固体 NMR による立体構造決定については、溶液 NMR と比べると遅れていたが、低分子量タンパク質のシグナル帰属が可能になってきた。さらに、大口径の高磁場超伝導磁石の開発が急速に進みつつある。溶液 NMR や X 線結晶構造解析など他の手法でも困難が大きい膜タンパク質などの試料を対象とする手法として大きな期待がかかる。さらにはタンパク質の特定領域に関する高精度な局所情報の取得手段としての開発も重要である。

ハードウェアについては、ひとつは、超伝導マグネットの高磁場化による感度、分解能の改良が重要である。特に、900 MHz を越える高磁場の超伝導マグネットの開発には、従来の金属超伝導線材に加えて、酸化物の高温超伝導線材の利用が必須といわれるだけに、極めて大きな技術的チャレンジである。当面の目標は、1 GHz への到達であるが、これには、高感度、高分解能に加えて、TROSY の発明により、分子量の制約を大きく緩和するという期待がもたれており、高分子量系への NMR の適応を推進するためにも、極めて重要な開発目標である。他方、クライオプローブによる感度の著しい改善が大きい。3 倍の感度上昇は、測定時間を 10 分の 1 にするため、従来の不可能を可能に変えることがある。現在、600 MHz までのクライオプローブが開発されているが、今後は、800 MHz 以上のクライオプローブの開発が強く望まれる。

## B．構造解析が困難なタンパク質の解析技術開発

B 1．糖鎖タンパク質大量生産技術の開発 [ ヒトのタンパク質輸送系、糖鎖修飾系タンパク質群の構造、機能、相互作用解析とその応用 ] （中期：30 億円程度）：ヒトゲノム情報を生物機能発現の理解に結び付け医薬産業面に応用するためには、生理活性を持ったタンパク質を大量に発現精製する技

術開発が求められる。中でも糖タンパク質の生理活性の保持には正しい糖鎖修飾が不可欠である。その大量発現系の確立にはまず、それらを司る細胞内タンパク質輸送系と糖鎖修飾系の包括的理解、すなわち、輸送過程におけるタンパク質複合体形成、タンパク質間相互作用、シグナル伝達系を介した制御機構の解明が不可欠である。そこで、小胞体、ゴルジ体間のタンパク質輸送および糖鎖修飾系をターゲットとして、放射光X線構造生物学による構造ゲノム科学プロジェクトを展開し、その構造知見に基づいて糖タンパク質の大量発現系を開発することが必要である。

B 2 . 膜タンパク質の大量調製、結晶化技術の開発 [ 大量調製法、結晶化法探索と確立、100 個の膜タンパク質結晶化 ] ( 中期 : 4 0 億円程度 ) : ゲノムのコードするタンパク質の約三分の一は膜タンパク質であるが、その大量調製と結晶化は可溶性タンパク質に比べて極めて困難で、結晶化成功率は非常に低い。中でも大量調製は個々のタンパク質毎に方法を模索している段階であり、成功事例を増やすのが当面の課題である。自然に大量に発現しているタンパク質の大量調製法のみならず、無細胞系、大腸菌、昆虫細胞等による大量発現系による大量調製法の開発を個々の膜タンパク質について進めることが重要である。標準的な膜タンパク質の結晶化法では、様々なアルキル鎖長を持つ界面活性剤を試行錯誤的に試すのみであるが、最近、Lipidic Cubic Phase 法や、抗体を用いる結晶化法などがいくつかの成功を収めている。しかしまだ、全ての膜タンパク質に有効な方法として確立したわけではない。膜タンパク質の構造解析をルーチン化するためには、既存の大量調製、結晶化法を組織的に改良するだけでなく、新たな方向性を模索し、さまざまなアイデアに基づく技術開発を進めるため、複数の拠点で技術開発研究を行えるよう支援することが重要である。

B 3 . 生体超分子複合体の大量調製技術の開発 [ ハイブリッドタンパク質等タンパク質複合体の大量調製、結晶化 ] ( 中期 : 4 0 億円程度 ) : 生体超分子複合体も大量調製が最大の課題である。膜タンパク質の場合と同様に、自然に大量に発現しているタンパク質の大量調製法のみならず、無細胞系、大腸菌、昆虫細胞等による大量発現系による大量調製法の開発が必要である。効率の良い共発現系の開発や、チトクロム酸化酵素のようなハイブリッドタンパク質複合体の大量発現系の構築などは今から推進すべき課題である。

### C . 超分子複合体の単粒子構造解析あるいは超微結晶構造解析技術

C 1 . 第四世代放射光 X F E L による構造生物学計画 [ 加速器、光源の開発、予備実験装置の建設、ナノスケールタンパク質試料の操作技術、高分解能単分子 X 線解析法の開発 ] ( 長期 : 8 0 億円程度 ) : 第四世代放射光として XFEL の準備研究が各国で進められている。XFEL は、フェムト秒オーダーのパルス性高輝度 X 線を産み出し、高時間分解能で動的特性を追う実験に威力を発揮する。また、その超高輝度により、1 ミクロン以下の超微小結晶やウイルス等の単粒子の構造解析などを実現させる可能性も秘めている。ドイツ DESY 内 HASYLAB では、2010 年の実験開始を目標に XFEL の技術設計計画書が作成されつつあり、構造生物学研究課題の検討が始まっている。米国スタンフォード大学でも、XFEL の応用研究テーマを全世界から公募し 6 件を採択、その内の 2 件が生物関連である。研究開発費申請の審査が通過すれば 2006 年の実験開始を目指す。日本では高エネルギー加速器研究機構内で XFEL の開発研究の検討が始められた。それと並行して、生体オルガネラ単体や超微小結晶の XFEL に

よる構造解析法等の、先端的構造生物学実験の提唱とその可能性の検討、さらにそれに必要な実験装置、試料のハンドリング、解析法の検討が行われている。

C 2 . 極低温電子顕微鏡によるタンパク質複合体の高分解能単粒子構造解析技術の開発 [ タンパク構造解析用に特化した極低温電子顕微鏡の高度化、および解析ソフト開発 ] ( 中期 : 7 0 億円程度 ) : 電子顕微鏡は、タンパク質分子像を直接観察記録するため結晶化の必要がなく、原理的には個々のタンパク質の動作中の構造を見ることができる。装置としての分解能も X 線回折法と同様に高い。ただ、水素結合など弱い力で構造を保持しているタンパク質では電子線照射損傷が無視できず、照射量を低く抑えるために像の S/N と分解能に制限が生じる。それでも、試料温度 10K 以下での像観察により照射損傷を抑えて S/N を向上させ、数万個の分子像をそろえて平均する事により、最近では 0.5nm を越える分解能が得られるようになった。この解析法により、例えばリボソームの構造とそのタンパク質合成機械としての動きが明らかにされつつある。電界放射型電子銃やエネルギーフィルターなどの比較的新しい技術も像分解能の向上に役立っている。結晶化の必要が無く直接分子像が記録できるため、いかに巨大なタンパク質複合体でも動作中の構造が解析可能であり、これは X 線結晶解析法にも NMR にも原理的に期待できないユニークな特徴である。しかも装置や解析ソフトの技術革新により原子分解能を達成する高いポテンシャルを持つ。また X 線に比べて散乱断面積が大きいことにより、1 ミクロン以下のタンパク質結晶から分解能 0.2nm の回折データが得られるなど、現在の電子顕微鏡でさえ第三世代放射光 X 線ビームラインを超える能力を発揮しており、XFEL に期待されるのと同等の実験が短期に実現できる可能性を持つ。このようにさまざまな技術革新の萌芽が予想される中、人材と資源を集中して、タンパク質構造解析用に特化した電子顕微鏡、2 次元検出器、解析ソフト開発を推進することが望まれる。

C 3 . 次世代 2 次元電子線検出器の開発 [ 画素サイズ 5 ミクロン、画素数 4000x4000、読みとり速度 10 秒 ] ( 中期 : 2 0 億円 ) : 電子顕微鏡によるタンパク質複合体構造解析のための像記録媒体は、その分解能において現在もまだ、現像、水洗、乾燥の必要な写真フィルムに優るものがなく、データ収集効率が悪い主要原因となっている。電子線用の CCD 検出器として 2000x2000 画素のものが実用化されているが、電子線 光変換面での電子線二次散乱によるスポットの広がりのため、現状では画素サイズが 25 ミクロンより小さいものはない。これでも回折データ収集には大変有効であるが、像拡大率 5 万倍とした時に 0.5nm に相当する像分解能は、高分解能構造解析 ( 分解能 0.3nm 以上 ) のための像記録には不十分であり、必要な分解能を得るためには最低でも 0.1nm に相当する 5 ミクロン画素が必要である。画素サイズ 5 ミクロン、画素数 4000x4000 程度の CCD 検出器を早急に開発し、電子顕微鏡の標準画像記録媒体として実用化する事が、今後ますます重要性を増す巨大なタンパク質複合体の構造解析のために強く望まれる。

D . タンパク質の構造機能連関を水素の配置を中心に詳細に探る技術開発

D 1 . 大強度パルス中性子線を用いた構造生物学 [ ビームライン設計と時分割可能な中性子用イメージプレートの開発、および準備調査研究 ] ( 長期 : 2 0 億円程度 ) : 中性子回折によるタンパク質の構造研究は、タンパク質の構造形成、タンパク質間相互作用、エネルギー変換、信号伝達処理など、

生命活動のすべてにわたって重要な役割を果たす水素原子や水素イオンを他の原子と同様に可視化できるという点で、大変優れた特徴をもつ。また、準弾性散乱や非弾性散乱解析により、タンパク質の機能に深く関わる構造の動力学的特性を定量化することが可能である点でもユニークな測定法である。最近の中性子用イメージプレートの開発によりデータ収集効率と精度は飛躍的に向上したが、現在利用可能な中性子源の強度が低いためデータ収集に週単位の時間が必要で、効率的な構造解析からはほど遠いのが現状である。高強度の中性子源として核破砕中性子源計画が欧米にあり、国内でも日本原子力研究所と高エネルギー加速器研究機構が実施する「大強度陽子加速器計画」による大強度パルス中性子源計画の詳細検討が始まっているが、その実現によってビーム強度が2桁向上すると期待されており、そうなるとう中性子回折データ収集が大幅に改善され、また準弾性散乱や非弾性散乱解析が普通に実施されるようになり、タンパク質の構造・機能研究が水素原子、水素イオン、水和構造および動力学的特性を含んだ新しい切り口で展開されることになる。大強度パルス中性子源による中性子実験では時分割可能な中性子用イメージプレートや分光器の開発等、新しい技術開発が必須であり、これを進めるべきである。

## 2 - 2 . 実行研究予算 ( 5 年間 )

### A . 網羅的構造ゲノミクス ( 拠点型 )

A 1 . 複数の特定生物系、細胞系あるいは機能単位系についての網羅的なタンパク質構造解析 ( 約 3 0 0 億円 ) : 5 ないし 6 の拠点を整備し、それぞれが特定の生物系、細胞系あるいは機能単位系を選び、網羅的タンパク質構造解析を 5 年間にわたって推進する。それぞれのプロジェクトの継続性、拡張可能性は、5 年間の成果によって判断する。

### B . 生物学的重要タンパク質の構造機能研究 ( 個別研究 )

B 1 . 生物学的に重要な個別の系についての機能研究を絡めた構造解析 ( 約 5 0 0 億円 ) : 生物学的重要課題に取り組む 1 0 0 グループ程度に対し、それぞれ 5 年間で計 5 億円程度の研究支援を行う。

## 3 . バイオインフォマティクスおよび計算生物学における技術開発

バイオインフォマティクスは、ゲノミクス及びプロテオミクスの研究プロジェクトを構成する不可欠の要素である。あらゆるゲノミクス、プロテオミクス関連研究プロジェクトは、その一部として、研究現場に密着してプロジェクトを推進するためのバイオインフォマティクス活動を組み込んだ形で計画されるべきである。

一方で、それら個別プロジェクトのバイオインフォマティクス活動を集約し、構築された様々なデータベースの管理、運営を支援するとともに、新たな方法論やアルゴリズム開発を支える学術的な役割を果たすためのセンターが必要とされる。日本およびアジアのバイオインフォマティクスセンターとして、米国 NCBI (National Center for Biotechnology Information) または欧州 EBI (European Bioinformatics Institute) の規模、内容に匹敵するものが期待される。例えば、NCBI は約 150 人の人員で運営されており、現在その規模を倍増中である。このようなセンターを運営するために、施設建設費を別として年間 30~40 億円程度の予算が必要となろう。

同時に、これら使命達成型のプロジェクトにおけるバイオインフォマティクス活動とは独立に、基礎的、萌芽的な課題を追求するためのバイオインフォマティクス、計算生物学研究を振興する必要がある。全国の生物科学関連の大学学科、公的研究機関に 50 前後のバイオインフォマティクス、計算生物学研究グループが配置され、多様な課題での研究を展開する中で分野全体としての日本のレベルを高めることが期待される。それが同時に、この分野の人材の不足の解消につながるであろう。米国の大学には約 50 のバイオインフォマティクスのコースがあるとの報告がある。

そして、これら多数の研究グループにより、多様な観点からのタンパク質関連データベースが開発され、サービスが提供されることが期待される。内容として例えば、シグナル伝達系、代謝系、転写制御ネットワーク等の、具体的な観点からの様々な生物学的課題に関するデータベース、タンパク質構造・機能の分類のためのデータベース、タンパク質分子の運動に関するデータベース、など多彩な内容のものが想定される。これらの活動を推進するために、年間総額 10 億円程度の研究費の支援が期待される。

一方、特に計算生物学においては、大規模なシミュレーション計算によって行なわれる、個々の分子の動的な構造解析、構造予測、エネルギー解析を始めとして、細胞内、細胞間タンパク質ネットワークの解析、予測等は、プロテオミクス研究や構造ゲノミクス研究において得られた知見の体系化を行うために必須な研究である。これらの研究推進に対しても、年間総額 10 億円程度の研究費の支援が期待される。

これらの活動を進める上で、計算機ハードウェアの利用環境を、個々のデータベース構築、運営およびシミュレーション技術開発研究との 1 対 1 の関係から開放して、数箇所集中管理する新しい利用環境を提案する。このことにより (1) 計算機利用が効率化される、(2) 個々のデータベース事業が大型計算機の維持に振り回されることを防げる、(3) ハード、ソフト両面の激しい変化に対応しやすくなる、等の利点が生まれる。

このような共同利用計算機設備は、データベースサービスの提供と同時に、計算生物学および関連する計算機科学の研究上必要となる大規模計算のための CPU、メモリ資源の提供をも行なうことが期待される。これらの点を考慮し、共同利用計算機は大型でかつ変化に柔軟に対応できるものでなければならない。具体的には、超並列クラスタマシンが想定される。例えば、米国エネルギー省(DOE)は、30TFLOPS - 1, 250MHz - 12, 000CPU - 12TBytes memory のシステムを構築しようとしている (2004 年までに 100TFLOPS 達成を計画)。米国科学財団(NSF)は、様々な分野での科学技術計算のために 6TFLOPS - 1GHz - 2, 728CPU - 2.7TBytes memory の共同利用システムを有している。

日本国内では、従来、ベクトル計算型スーパーコンピュータが主流であったが、バイオインフォマティクスおよび計算生物学における大規模な演算の実施は、ベクトル型計算機よりもむしろ並列型計算機がより有効な場合が多い。このため、日本国内においても、超並列計算機の導入を積極的に行い、当面の共同利用計算機システムとして、5~10TFLOPS - 1GHz - 2,000~5,000CPU - 2~5TBytes memory 以上の性能を有するものが期待される。また、その程度の規模の共同利用計算機サービスを提供するためのセンター(群)の運営のための費用(施設建設費を除く)は、計算機の導入のために 40~60 億円程度、計算機運用保守を含む運営費として年間 30~40 億円程度が必要であろう。

#### 4. ハイスループット細胞 4 次元プロテオミクス (中・長期の課題)

ゲノム統合科学を基盤とした細胞生物学がめざすところは、細胞内、細胞間の分子機械ネットワークの動態を 4 次元的に詳細に把握することである。すなわち、現在の技術で実行可能なタンパク質発現プロファイリング (一般的な意味でのプロテオミクス) によって得られる膨大なデータを、さらに有機的に活用してその解釈を深めるべく、発現している個々のタンパク質の局在、動態、そして相互作用を空間的・時間的に追跡、解析し、複雑で精緻な生命活動を支える分子機械ネットワーク制御の仕

組みを明らかにしなければならない。そのためには、細胞の各場所に局在するタンパク質個々についての単分子観察技術の高度化、および単分子同定技術の開発が不可欠の要素となるであろう。そこで、前述の「1. プロテオミクスセンターの研究課題」で提案したものに加えて、さらに中長期にわたる研究課題として、例えば、細胞内局在タンパク質の単分子N末端アミノ酸配列解析技術であるとか、細胞内局在タンパク質同定のためのタンパク質コピー数増幅技術の開発に真剣に取り組むことが望まれる。これらは、現在の技術の延長上にはない革新的かつ挑戦的な技術開発目標であり、生物のもつ超分子機械の働きをうまく改変して組み合わせるか、あるいはナノテクノロジーの技術革新に期待する以外にないと思われる程度に、現在のところ実現への道ははるかに遠い夢物語ではある。しかし、我が国の生命科学研究を推進し、その応用としての医療、産業分野における圧倒的な優位をもたらすためには、この分野での我が国の研究開発レベルの高さと先進性を有効に活用し、また十分な財政支援により、是非真剣に取り組まなければならない課題であろう。おそらく年間10億円レベルの研究開発投資が必要とされるが、将来的には桁違いに大きな経済効果を産むと期待される。

また、前述の網羅的タンパク質発現プロファイリングとともに、網羅的な細胞内タンパク質動態、局在、相互作用観察記録と解析およびデータベース化については、現状の解析速度は遅いものの現在の技術レベルでも実行は可能である。複数の特定生物系、細胞系あるいは機能系について、拠点体制を整えて推進することがのぞましい。それとともに、従来型の生物学的重要ターゲットに的を絞った分散個別型研究にも財政的支援を拡充して、両者をバランス良く推進すべきであろう。それぞれ、年間数十億円の財政支援が適当と考えられる。

## 用語解説

【CCD】電荷結合素子。2次元光学像検出器であるが、蛍光物質などを用いることによりX線や電子線の2次元検出器の素子としても用いられる。

【CPU】中央処理装置(Central Processing Unit)。計算機の構成部品であって、計算機への指令の解釈と実行を行なう。

【FLOPS】Floating Point Operations Per Second の略。CPU の計算速度を表わす単位の一つ。1秒当りにCPUにより実行される浮動小数点演算の数を示す。分子シミュレーションでは浮動小数点演算が多用されるため、一般にFLOPS数の大きなCPUの使用が望まれる。

【HDTV】High Definition Television (高品位テレビ)の略称。

【Lipidic Cubic Phase 法】脂質、水混合相のひとつで立方格子状の網目構造を形成し、膜タンパク質の結晶化に有効な方法として注目されている。

【NMR】核磁気共鳴法。タンパク質や核酸などの生体高分子の立体構造、および特定領域の原子間相互作用の解析を、結晶化することなく溶液試料を対象として行うことができるのが特徴。解析対象分子の分子量には限界がある。

【Proteomic Signature】特定の生命現象や疾患で特異的に変動する蛋白質群を俯瞰できる地図またはリスト。発現プロテオミクス研究(2を参照)のまとめとして作成され、目的の生命現象や疾患のマーカーとなるだけでなく、リストされたタンパク質の相互の繋がりを知ることで、その生命現象や疾患の原因を探る手掛りとなる。

【X線結晶解析法】対象分子を結晶化して回折データを収集し解析することによって、分子の立体構造(構成原子の立体配置)を解く手法。結晶化さえ可能なら対象分子の分子量に限界はない。

【X線自由電子レーザー(XFEL)】直線型電子加速器により加速されたパルス電子ビームの軌道を周期的に波打たせることで、放射光を高度に干渉させ、超高輝度のパルスX線を発生する装置。現在まだ稼働中のものはなく、各国とも設計構想段階にある。SPring-8など現在稼働中の第三代放射光施設で得られる高輝度X線に比べて時間平均輝度で一千倍から一万倍、ピーク値で10億倍から100億倍の高輝度が得られる予定。

【2次元X線検出器】主にX線結晶解析のための回折データ収集に使われる検出器。数十cm四方の大面積と数秒以内の読みとり速度が望まれる。イメージプレート(IP)やCCDを使ったものがあるが、それぞれ、IPはデータ読みとり速度が遅い、CCDは大面積のものがつくりにくいといった短所がある。

【2次元電気泳動法】1つ1つの蛋白質分子がもつ電荷と大きさの差を利用して、複雑な蛋白質混合物を二次元的に分離する電気泳動法。現在の技術では、20cm x 20cm程度の大きさのゲル(担体)を使用して約4,000種類の蛋白質を丸いスポットとして分離、検出できる。

【2次元電子線検出器】電子顕微鏡にとりつけて回折データや画像を記録する装置。100kVから300kVに加速された電子線の高いエネルギーのため、検出面での2次散乱の広がりが大きく、現状では画素サイズが25ミクロン程度にとどまる。この分解能の制限により高分解能の画像解析に使える検出器が存在せず、その開発が強く望まれる。

【アルキル鎖】脂質分子の足に当たるCH<sub>2</sub>の繰り返した疎水性鎖状残基。

【イメージプレート】X線や電子線の入射エネルギーを蓄え、のちにレーザーでスキャンすることにより、輝尽発光という原理で蓄えられたエネルギーに比例した発光強度を示す。X線フィルムの約百

倍の感度を持ち、ダイナミックレンジも広いため、非常によく定量性を持った2次元像検出器。

【エネルギーフィルター】電子顕微鏡で、電子線照射により試料から散乱される電子線のうち、特定のエネルギーを持った電子線のみを選んで結像させる装置。非弾性散乱電子を除くことによって像の背景ノイズを低減して像質および像分解能を高めたり、特定のエネルギーロス電子線のみを使うことで、試料の元素マッピングなどを行うことができる。

【エピトープタグ】遺伝子工学技術などを利用して天然の蛋白質に人為的に導入する標識。特定のアミノ酸配列をもつペプチド鎖を目的蛋白質分子の一部に標識として導入することで、その蛋白質の分離や検出を容易にする目的で使用される。

【オントロジー】概念の定義の集積。バイオインフォマティクスの文脈では、ある研究領域において個々の事象、概念を指すために標準的に使用されるべき専門用語のセット（語彙）の公式的な定義、記述のこと。例えば、Gene Ontology Consortium (<http://www.geneontology.org/>)では、遺伝子、蛋白質の機能などに関する注釈（アノテーション）を行なう際に使用するべき標準的語彙を定義している。

【機能プロテオミクス】蛋白質の機能を相互作用を基礎にして解析するプロテオミクスの分野。主として蛋白質間あるいは蛋白質と他の生体成分の親和性を利用した方法が用いられ、相互作用プロテオミクスとも呼ばれる。

【共鳴プラズモン共鳴分析】「表面プラズモン共鳴」と呼ばれる分光学的な検出法を利用して、分子間相互作用をフェムトモルレベルの超微量試料で定量的に分析する方法。

【極低温電子顕微鏡】電子線照射損傷をできるだけ抑えるため、試料を数K〜100Kの低温に保って像観察、像記録ができる電子顕微鏡。

【計算生物学】計算機によるデータ解析とシミュレーションを主要な研究手段とする、生物学研究のアプローチ。

【構造ゲノミクス】特定生物種のゲノムから発現される数千から数万種におよぶタンパク質を網羅的かつ迅速に構造解析し、ゲノム情報を基礎生物学のみならず、医薬開発、産業利用などに役立つ、より有益なものにしようとする学問分野。現状ではX線結晶解析法とNMR構造解析法が使われる。

【構造生物学】タンパク質や核酸などの生体高分子の立体構造、構成原子の立体配置、原子間の相互作用、構造変化などを詳細に解析することにより、それらの分子に関わる生体機能のしくみを深く理解しようとする学問分野。X線結晶解析法、NMR構造解析法、電子顕微鏡法、中性子回折法、光学顕微鏡法、一分子解析法、表面探針顕微鏡法、その他多くの分光法や熱解析法なども幅広く組み合わせて使われ、対象となる生体高分子や複合体の動作機構について総合的な理解を得ることを目的とする。

【シグナル伝達系】細胞内外の様々な信号伝達を司るタンパク質群がなすネットワークシステム。

【質量分析法】イオン化した物質を真空中で分離し、質量を測定する方法。近年、蛋白質やペプチドなどの生体高分子をイオン化する方法と高精度で質量を測定する方法が開発され、プロテオームの研究に不可欠な分析法となっている。

【小胞体、ゴルジ体】細胞内小器官で、蛋白質の折りたたみ、S-S結合、糖修飾を司る。

【情報内容の4次元的表现型】ゲノム情報がDNAの塩基配列という一次元情報であるのに対し、ゲノム情報の翻訳により発現するタンパク質は立体構造を形成しその時々刻々の動作が生体機能を支えるという意味での4次元。

【生体超分子複合体】複数のタンパク質分子や核酸分子の複合体で、単分子の機能を越えた複雑な機能を有することから超分子と呼ばれる。生命機能のシステムとしての性質を持つ最小単位でもある。

【タンパク質機能リンケージ】生体内での蛋白質相互の機能的な繋がり。例えば生物の代謝地図では、さまざまな蛋白質や酵素が機能的に繋がって特定の代謝経路や回路を作り、生物が生きるためのエネルギーや物質などを作り出していることが知られている。

【タンパク質ファミリー】アミノ酸配列および機能が似通ったタンパク質の一群。

【単粒子構造解析】タンパク質や核酸等の生体高分子やその複合体の電子顕微鏡像を個々の粒子像として多数記録し、それぞれの向きを解析し、計算機トモグラフィと同様の方法で立体構造を求める画像解析法。

【データベース】関連するデータを一定の形式で集積したもの。通常、データベース管理システム(DBMS)と呼ばれるソフトウェアを通じて、利用者によるデータの追加、削除、更新、検索を可能にする。多くのDBMSは、大量データからの高速な検索を可能にするためのしくみ、データベース内のデータ間の一貫性を保証するためのしくみ、計算機の障害発生後のデータベースの復旧を容易にするためのしくみ、等を内蔵している。現在最も一般的なデータベースの形式は、関係(relational)データベースである。関係データベースは、互いに関連付けられた多数の表の集まりから成る。表は行(レコード)と列(フィールド)から成り、各行は個々の事例に対応し、各列は事例に関する様々な属性に対応している。データベースの形式には他に、オブジェクト指向データベース、階層型データベース、などがある。

【電界放射型電子銃】電子顕微鏡の電子線源で、フィラメントを高温に加熱することにより放射される熱電子を加速する従来の電子線発生原理を使わず、電界をかけることによってのみフィラメントから電子を引き出して加速する電子線源。干渉性と輝度が高く、得られる電子顕微鏡像の像質が高いため、高分解能の構造解析に適している。

【バイオインフォマティクス】計算機と情報技術、情報科学、計算機科学的手法を活用することによって、生物学上の多種多様で大量なデータを効果的に組織化し、それらデータの解析を通じて生物学的に有用な情報を抽出しようとする、生物学研究上の一群のアプローチの総称。特にゲノミクス、プロテオミクスを支える重要な構成技術として進展してきており、DNAの塩基配列データと関連情報、蛋白質のアミノ酸配列、立体構造、機能データと関連情報は主要な処理、解析対象である。

【ハイスループット(high-throughput)化】多試料を短時間に高速で分析処理するための技術革新。ゲノムやプロテオームに関する膨大な情報を大規模に解析するために重要である。

【発現プロテオミクス】ゲノム情報をその最終産物である蛋白質の発現情報として解析するプロテオミクスの分野。ある状態の細胞などに存在している蛋白質を大規模に分離して解析することで、生命現象の変化や疾病などの異常をもたらす蛋白質群を同定する目的などで行われる。

【発現プロファイル】ある状態の細胞や組織に存在している蛋白質の種類や量などを俯瞰できる地図。発現プロテオミクス研究によって、2次元電気泳動や質量分析法などを利用して作成される。

【ピクセルアレイ検出器】半導体2次元X線検出器の一種で、CCD型X線検出器にかわる次世代検出器として注目されている。

【不均一核形成法】タンパク質などを結晶化する際、容器の表面に結晶の核形成を促進する構造をつくり、結晶成長を制御する方法。

【プロテオーム解析】ある状態の細胞、組織、器官、個体に発現している蛋白質の総体(プロテオーム)についての情報を把握するための研究。

【分子間ネットワークシミュレーション】相互作用のネットワークを構成する分子の系の挙動を、計

算機上でシミュレートすること。

【分子機械】タンパク質分子や核酸分子が、原子を部品として立体構造が組み上げられ特定の機能を果たすナノスケールの機械であるとして呼ばれる名前。ナノテクノロジーでは、様々な分子を設計して人工的に分子機械を製作することを目指している。

【分子シミュレーション】物理化学的な法則に従って、分子の時間、空間的挙動（運動、状態変化）を計算機上でシミュレートすること。計測、実験が困難、非現実的あるいは不可能であるような分子の挙動を理解、予測することを可能にする。

【マイクロフルイディクス】微小容量の溶液や超微量の試料を操作する技術。半導体チップに刻んだ溝などを利用して、核酸や蛋白質などの生体試料を電気泳動法などで分離、検出する試みが行われている。

【膜タンパク質】細胞膜や細胞内小器官の脂質膜を貫通あるいは膜に結合したタンパク質。イオンポンプ、物質輸送装置、イオンチャネル、信号分子受容体などとして、細胞内、細胞間の信号伝達処理、物質輸送を司る。

【無細胞系タンパク質大量発現システム】大腸菌、酵母、昆虫細胞などに目的タンパク質のDNAを含んだ大量発現ベクターを組み込んで発現させる従来のやり方ではなく、細胞のタンパク質合成系を取り出し、その効率的な稼働に必要な全ての条件を試験管内に整え、無細胞系でタンパク質を大量発現する技術。

【メモリ（計算機の）】メインメモリ（主記憶）。CPUの中に含まれ、CPUによる演算処理に必要なデータとプログラムが一時的に格納される記憶装置。

通常ハードディスクなどの二次記憶装置に保存されているデータとプログラムは、実際の演算処理の際にはメインメモリに移される。一般に、メモリの容量が大きいほど、一度に計算に使用できるデータの量が大きくなる。

【目標提示型の網羅的研究】構造ゲノミクスやプロテオミクスのように、特定細胞が発現するタンパク質種の網羅的プロファイリングや網羅的構造解析を目的とした研究。ある期間内にどれだけの領域を網羅するかをあらかじめ設定し研究を進める。