



北海道公立大学法人  
**札幌医科大学**  
Sapporo Medical University

**札幌医科大学学術機関リポジトリ *ikor***

SAPPORO MEDICAL UNIVERSITY INFORMATION AND KNOWLEDGE REPOSITORY

Title	遺伝情報解析システムの構築とタンパク質リピート配列の研究
Author(s)	大柳, 俊夫;松嶋, 範男
Citation	札幌医科大学保健医療学部紀要,第 1 号: 1-8
Issue Date	1997 年
DOI	10.15114/bshs.1.1
Doc URL	<a href="http://ir.cc.sapmed.ac.jp/dspace/handle/123456789/6593">http://ir.cc.sapmed.ac.jp/dspace/handle/123456789/6593</a>
Type	Journal Article
Additional Information	
File Information	n1344919211.pdf

- コンテンツの著作権は、執筆者、出版社等が有します。
- 利用については、著作権法に規定されている私的使用や引用等の範囲内で行ってください。
- 著作権法に規定されている私的使用や引用等の範囲を越える利用を行う場合には、著作権者の許諾を得てください。

## 遺伝情報解析システムの構築とタンパク質リピート配列の研究

大柳俊夫, 松嶋範男

札幌医科大学保健医療学部一般教育科

### 要 旨

分子生物学の目覚ましい進歩により、核酸の塩基配列ならびにタンパク質のアミノ酸配列のデータが蓄積されつつある。これらのデータから、情報という立場で生物学の知識の体系化や生命の原理の法則化に取り組む新しい学問領域として「生命情報科学」もしくは「遺伝情報科学」が誕生した。その中の重要な研究領域として、データ解析に中心をおいた「遺伝情報解析」が急速に発展しつつある。本稿では、まず遺伝情報解析の概略と著者らが構築を進めている遺伝情報解析システムについて説明した。そして、構築システムを利用したこれまでの著者らの研究成果として、1) 配列解析（ロイシン・リッチーリピートとホモポリマーの配列解析と分子進化）、2) 立体構造予測（S100Bとアメロジェニンタンパク質）、3) タンデムリピートと核酸の相互作用モデル（AlgR3とUS11タンパク質）に関する研究の概要を報告する。

<索引用語> 生命情報科学、遺伝情報解析、配列解析、立体構造予測、タンデムリピート

### 緒 言

分子生物学の目覚ましい進歩とヒトを含む多くの生物種のゲノムプロジェクトにより、生命システムの基本的な要素である核酸の塩基配列やタンパク質のアミノ酸配列のデータが急速に蓄積されつつある。これらのデータは、人類共通の財産として無料で公開されており、コンピュータネットワークやCD-ROMなどを媒体として誰もが自由に利用できるようになってきている。今後、これらの膨大なデータ（情報）から生体システムを解明する研究が益々盛んになることは疑う余地がなく、21世紀初頭の中心的研究課題になると考えられる。すでに、情報という立場から、生物学の知識の体系化や生命の原理の法則化に取り組む新しい学問領域として「生命情報科学」もしくは「遺伝情報科学」が誕生している。その中の重要な研究領域として、データ解析に中心を置いた「遺伝情報解析」が発展しつつある<sup>1, 2)</sup>。

本稿では、まず遺伝情報解析の概略と著者らが保健医療学部開設以来構築を進めている遺伝情報解析システムについて説明する。さらに構築システムを利用したこれまでの研究成果について報告する。

### 遺伝情報解析とは

遺伝情報科学は、個体の発生・分化の機構と生物の進化などの生命の原理を解明することを目的に、これまでの分子・細胞生物学の知識と遺伝情報の解析による新発見の体系化を行う新しい学問領域である。その中心的な研究テーマである遺伝情報解析は、核酸の塩基配列やタンパク質のアミノ酸配列の一次構造情報から、生物学的に意味のある機能、立体構造、分子進化を明らかにすること、を目指しているものである。機能、立体構造、分子進化は互いに密接に関連しており<sup>3)</sup>、例えばタンパク質は、アミノ酸のポリペプチドが個々のタンパク質に特有な3次元構造に折り畳まれてはじめて機能を果たすことができる。

遺伝情報解析により生物学的に意味のある情報を抽出するためには、一次構造に関する膨大なデータばかりでなく、機能、立体構造、進化に関するこれまでの莫大な研究成果も同時に扱わなければならない。したがって手作業による解析はもはや困難であり、膨大なデータの管理ならびに関連情報の整理のための高性能なコンピュータを道具とする解析システムの構築が必要不可欠となっている。

遺伝情報解析システム

著者らが現在構築を進めている遺伝情報解析システムの現状を説明する。

(1) ハードウェア環境

本システムで準備しているコンピュータのハードウェア環境は、表1 (a) に示すようにサン社 (SUN) のワークステーション (infoss1 と infoss2)、シリコングラフィックス社 (SGI) の高性能グラフィックスワークステーション (physsg1 と physsg2)、それと2台のDOS/Vパソコン (infopc1 と infopc2) である。これら6台のコンピュータはすべて学内LANに接続しており、ネットワークを通してさまざまなデータを共有している。

表1 遺伝情報解析システムのハードウェア環境

(a) コンピュータ		
機種(ホスト名)	OS(CPU)	主な利用
SUN SS 10 (infoss1)	SunOS 4.1.3 (SuperSPARC)	ネットワークアクセスサーバー、プリントサーバー
SUN SS 5 (infoss2)	Solaris 2.3 (microSPARC)	データベース管理
SGI Indigo2 (physsg1)	IRIX 5.3 (R4400)	分子モデリング、カラープリントサーバー
SGI Indigo2 (physsg2)	IRIX 5.3 (R4000)	計算サーバー
Century P5-100 (infopc1)	Windows 95 (Pentium)	文献データベース管理、カラープリントサーバー
SHARP PC-A335 (infopc2)	Windows 95 (Pentium)	文献データベース管理、イメージ取り込み

(b) 周辺機器		
品目	機種	備考
プリンタ	富士フイルム PictroGRAPHY3000	高画質カラープリンタ
	SUN NeWSprinter 20	高速PostSCRIPTモノクロレーザープリンタ
イメージスキャナ	EPSON MJC3000C	A2版、長尺紙対応カラープリンタ
	EPSON GT9500	600×600dpi
ハードディスク	各メーカー	総容量16GB

表2 遺伝情報解析システムのソフトウェア環境

(a) データベース	
データベース	内 容
GenBank	核酸塩基配列
SWISS-PROT	タンパク質アミノ酸配列
PIR	タンパク質アミノ酸配列
PDB	タンパク質および核酸等の立体構造
CyanoBase	ラン藻 <i>Synechocystis</i> sp. strain PCC6803ゲノムの全塩基配列

(b) プログラム	
ソフトウェア	内 容
[立体構造予測]	
SYBYL	タンパク質・核酸のモデリングを含む、汎用3次元分子設計支援システム
MMS	タンパク質・核酸の立体構造表示およびモデリング
AMBER	エネルギー計算および分子動力学計算
PROCHECK	タンパク質の立体構造評価
MolScript	タンパク質の立体構造のポストスクリプトファイルへの変換
[配列解析]	
CLUSTALW	配列アライメントの計算が中心のパッケージ
PHYLIP	分子進化の解析が中心のパッケージ
TREETOOL	分子進化の系統樹表示
GCG	配列アライメント、分子進化などのための汎用配列解析パッケージ
FASTA	配列のホモロジー検索
REPRO	繰り返し配列の検索
[文献データベース構築]	
EndNote	文献データベース構築

infoss1では、電子メールやWWWブラウザによる情報収集・交換のサービスを提供している。infoss2では表2 (a) に示す各種データベースの管理を行っている。

また、physsg1はグラフィックス専用、physsg2は計

算サーバーの役割を担っている。infopc1とinfopc2は、文献データの整理と論文作成を主な目的としている。

また、表1 (b) に示すように、周辺機器として情報を保存するための16GBのハードディスク、画像入力装置として高解像度のイメージスキャナ、出力装置として高速なポストスクリプト対応モノクロプリンタ、高画質カラープリンタ、ポスター作成のためのA2版・長尺サイズ対応のカラープリンタなどを準備している。

(2) ソフトウェア環境=データベース+プログラム

利用可能なデータベースを表2 (a) に示す。これらのデータベースはinfoss2で管理しており、GenBank、SWISS-PROT、PIRの最新データは、定期的に米国NIH (National Institutes of Health) からインターネットを通して取り込むようにしている。PDBは現在その一部のデータを保存しており、将来的にはすべてを取り込む予定である。CyanoBaseは、日本のかずさDNA研究所がラン藻ゲノムの全塩基配列を決定して公開しているデータベースで、今後の配列解析で新知見の発見が期待されるものである。

また、ネットワーク経由で学内の文献データベースサーバーにアクセスし、医学・生物学関連文献データベースであるMEDLINEを利用できる環境にある。

利用可能な代表的なプログラムを表2 (b) に示す。立体構造予測に関連するプログラムがSYBYL、MMS、AMBER、PROCHECK、MolScriptである。SYBYLは低分子から高分子までを同時に扱うことのできる多機能の汎用3次元分子設計支援システムである。MMSは、タンパク質や核酸の立体構造を線画としてグラフィックス表示し、マウス操作で動的に立体構造を変化させることができ、構造モデリングを支援するプログラムである。MMSはモデリング以外の機能はほとんど持ち合わせていないが、SYBYLと比較するとその作業は軽快である。AMBERはSYBYLやMMSで予測した立体構造からエネルギー極小化や分子動力学により、より合理的な構造を与えるプログラム、PROCHECKは立体構造の総合的な質と局所的な信頼性を評価するプログラムである。MolScriptは、分子の立体構造をワイヤフレーム、リボン、ボールアンドスティックなどの形式で高精度に表示するためのポストスクリプトファイルを生成するプログラムである。これら5つのプログラムはphyssg1とphyssg2で利用できる。MMS、PROCHECK、MolScriptはフリーソフトである。

配列解析に関連するプログラムとしてCLUSTALW、PHYLIP、TREETOOLを用意している。CLUSTALWは複数配列のアライメントのためのプログラムパッケージで、近隣結合法による分子進化の解析も行える。一方PHYLIPは、最大節約法、近隣結合法などさまざまな系統分析法のソフトを含んだ分子進化解析用パッケージで

ある。TREETOOLはCLUSTALWおよびPHYLIPで計算した分子進化の系統樹を描くためのプログラムである。これら3つのプログラムは、すべてinfoss1とinfoss2で利用できるフリーソフトである。

EndNoteは文献データベース構築用のプログラムで、MEDLINEで検索した結果を容易にデータベース化して整理し、論文等で引用することができる。infopc1とinfopc2で利用可能である。

この他にも、独自プログラム開発のための環境やネットワークへのアクセスのためのプログラムを種々用意している。また、本システムで利用可能なコンピュータとOSの組み合わせは、世界的に普及している一般的なものであり、世界中の研究者が開発したプログラムのほとんどすべてをこの環境で利用できる状況にある。

なお配列解析では、前述の3つのプログラム以外にGCG、FASTA、REPROなども利用している。GCGは、配列のアライメント、分子進化の解析、RNA 2次構造予測、タンパク質2次構造予測などさまざまな配列解析のソフトが1つのパッケージになった世界的に著名なソフトウェアである。東京大学医科学研究所ヒトゲノム解析センターの計算機利用者登録をすることでGCGが利用可能であり、著者らも登録し、ネットワーク経由でヒトゲノム解析センターのコンピュータにログインして利用している。FASTAは、新しく決定した未知の配列のホモロジー配列を既知の配列データベースから検索するプログラム、REPROは配列中の繰り返しセグメントを検索するプログラムである。FASTAとREPROの利用では電子メールが必須で、電子メールで配列を送信すると、検索結果が電子メールで送られてくる。

(3) システムの利用

本システムは、教育・研究の利用のための「オープンなシステム」であり、学内の学生・教職員への開放を基本原則として構築を進めてきた。現在、保健医療学部のほぼ全教員・研究補助員のユーザー登録を終えており、教育・研究のために自由に利用し、各々の専門分野で成果をあげている<sup>4~14)</sup>。また、一部の医学部教員もユーザー登録しており、タンパク質の立体構造の解析や高画質カラープリンタへの出力等に活用している。学生は、カリキュラム上の「物理学実験」で、MMSを利用してタンパク質の立体構造科学の学習に役立てている。

解析システムを利用した研究

本解析システムを利用して得られた、アミノ酸配列解析、タンパク質の立体構造予測、タンデムリピートと核酸との相互作用モデル、に関する研究成果について述べる。なお、以下で説明する図2、4、5のタンパク質の立体構造はいずれもMolScriptで描いた<sup>15)</sup>。

(1) 配列解析

a. ロイシン・リッチーリピートの配列解析と分子進化

ロイシン・リッチーリピート（以下、LRR:Leucine-Rich Repeat）は、タンパク質の相互作用やシグナル伝達に参与するモチーフの1つとして注目を集めており、LRRsを含むタンパク質が、細菌から植物、動物にいたる多くの生物種の核内、細胞質内、細胞外、細胞外マトリックスで発見されている<sup>17,18)</sup>。その数は生物種の違いも考慮すると既に100を越えており、現在もなお増え続けている。LRRsがさまざまな生物種で数多く発見されていることは、LRRsが機能的に非常に重要な構造モチーフであることを示唆している。

LRRsは、タンパク質のアミノ酸配列中でタンデムな繰り返しをしており、その繰り返し単位の中にロイシンもしくは疎水性の高いアミノ酸残基があるパターンで出現している。図1にタンパク質全体が16個のLRRsで構成されているブタのRNase inhibitor (RI)のアミノ酸配列とRI中のLRRsのコンセンサス配列を示す。RIでは、2つのタイプの繰り返し単位が交互に出現していることが明らかのため、コンセンサス配列が2つ定義されている<sup>17)</sup>。

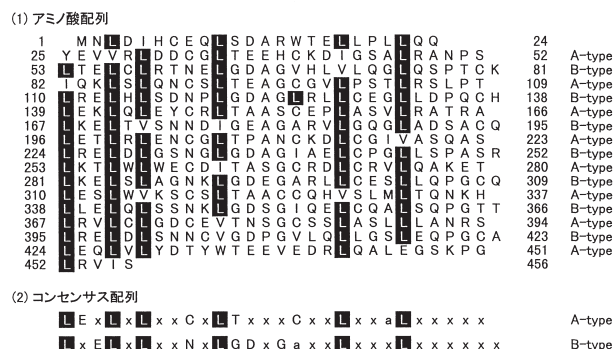


図1 ブタのRNase inhibitorのアミノ酸配列と2つのタイプのLRRコンセンサス配列

LRRsの繰り返し単位の長さは、短いものでアミノ酸20個、長いものでは29個（例えばRIのB-type）と変化に富んでいる。このため、すべてのLRRsのコンセンサス配列を定義することは容易ではなく、KobeとDeisenhoferによるコンセンサス配列の提案はあるが、これはLRRsを同定するにはまったく不十分なものである<sup>17,18)</sup>。

以上のようなLRRsに関するこれまでの研究から、LRRsの一次構造の多様性を明らかにすることが、LRRsの分子進化ならびに機能や立体構造を研究する上で必要不可欠なものであると考え、LRRsの配列解析に着手してきた<sup>19)</sup>。

その第一歩として、これまで唯一X線結晶構造解析が行われたRIのアミノ酸配列と立体構造（図2）の対応関係をコンピュータグラフィックスの画面上で詳

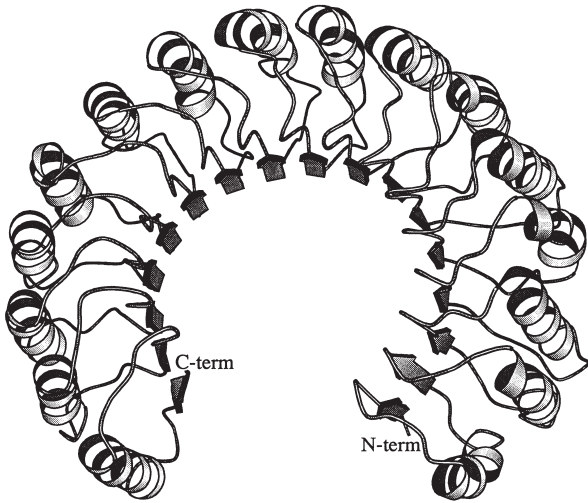


図2 R Iの立体構造<sup>16)</sup>

細に調べた。その結果、馬蹄形というR Iの特殊な立体構造の安定化に、βシート（図2の矢印）部分とαヘリックス（図2のリボン）部分の疎水性のアミノ酸残基が寄与していることが明らかとなった。つぎに、LRRsを含む147個のタンパク質のアミノ酸配列を収集した。この147個の中には同じタンパク質で生物種の異なるものも含まれており、タンパク質の種類としては85種である。これら147個のタンパク質に対し、疎水性のアミノ酸残基の出現パターンに着目した配列解析を行った結果、以下のことが明らかとなった<sup>20)</sup>。

- ① LRRsの各繰り返し単位は、LRRs全体にわたって保存の良い部分（Highly Conserved Segment:HCS）とそれ以外の部分（Less well Conserved Segment:LCS）の2つの部分から構成されている。R Iでは、HCS部分にβシートが、LCS部分にαヘリックスがそれぞれ含まれている。
- ② HCSとして、11残基のものを1,395個、12残基のものを196個、それら以外に欠失などのために不完全なHCSを34個同定した。
- ③ 1,599個のLCSを同定し、それらを疎水性残基の出現パターンで分類した結果、1,576個（全体の約98%）のLCSが7つのパターンに分けられた。

解析結果に基づき、LRRsを含むタンパク質を検索するプログラムを作成した。そのプログラムを使って、SWISS-PROT（Release 32）とPIR（Release 46）のデータベースを検索した結果、それらのデータベースに登録されている既知のLRRsを含むタンパク質すべてと、LRRsを含む新しいタンパク質17種を見いだした。LRRsを同定するアルゴリズムに関する論文を作成中である。

また、LRRsの分子進化に関連する研究として、以下の結果を得ており、これに関しても論文作成の準備中である。

- ① R IのLRRsは、2つのLRRs（図1のA-typeとB-type）でさらに高次の繰り返し単位を構成しており、この高次の繰り返し単位で遺伝子内分子進化を調べた結果、R I中のLRRsの遺伝子内重複の妥当なモデルを構築した。
- ② R I以外の高次の繰り返しを持つLRRsを含むタンパク質を同定し、遺伝子内重複の妥当なモデルを構築した。

b. ホモポリマー

転写因子の一つであるヒトTATA結合タンパク質TFIIDは、グルタミンが連続して38回繰り返すアミノ酸配列を持っている（図3）。このように、一次構造において1種類のアミノ酸だけが連続して繰り返すホモポリマーが多く存在することが知られている。これらのホモポリマーの立体構造および機能、進化的側面についていくつか議論はされているが、いまだあまり明らかではない。ここでは、ホモポリマーを”単一アミノ酸の10回以上の繰り返し”として定義した。この条件に合致するホモポリマーをもつタンパク質を、20種類のアミノ酸すべてについてデータベースから検索し、配列解析を行った<sup>21)</sup>。これらの解析から、次のことが明らかとなった。

1	MDQNNLSLPPYAQGLASPPQGAMTPGPIFSP	30
31	MMPYGTGLTPQPIQNTNSLSILEEQRRQQQ	60
61	QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ	90
91	QQQQQQAVAAAAVQQSTSQQAATQGTSGQAPQ	120
121	LFHSGTLTTAPLPGLTPLYSPMTPMTPT	150
151	PATPASLSSGLVPLQNLVSTVNLGOKLDD	180
181	KTIALRARNAEYNPKRFAAVIMRIREPRT	210
211	ALIFSSGKMVCTGAKSEEQSRLAARKYARV	240
241	VQKLGFPKFLDFKIQELMFGSDVVKFPIRL	270
271	EGLVLT HQQFSSYEPELQNMVGLIYRMIKPRI	300
301	VLLIFVSGKVVLGTGAKVRAEIIYEAFENIYP	330
331	ILKGFRTT	339

図3 ヒトTATA結合タンパク質TFIIDのアミノ酸配列、反転部分は38回グルタミンが繰り返したホモポリマーを示す。

- ① ホモポリマーが、約400のタンパク質アミノ酸配列中に存在することが見いだされた。
- ② アミノ酸の種類によって、ホモポリマーをもつタンパク質の数に大きな違いがある。例えば、トリプトファン、システイン、メチオニンのホモポリマーをもつタンパク質は今のところ存在しない。
- ③ ホモポリマーは、そのアミノ酸配列を表すために複数種類のコドンがそれぞれ平均して使われているのではなく、その使用頻度には偏りがある場合が多い。ホモポリマーの長さが遺伝病の発症と密接に関連していることが最近の研究から明らかにされてきており<sup>22)</sup>、今回の解析結果が医療の分野にとって大変意義のあるものと考え、論文としてまとめている最中である。

(2) 立体構造予測

a. S100Bの立体構造予測

S100は、リン酸化の調節等のさまざまな機能に関する調節タンパク質であり、カルモデュリンと同様にカル

シウム結合タンパク質である<sup>23)</sup>。S100Bは、ホモダイマー ( $\beta\beta$ ) からなり、サブユニットあたりEFハンドが2個存在する。著者らはX線溶液散乱の結果を満足するS100Bの立体構造を予測した。

S100Bの $\beta$ サブユニットは、すでにX線結晶構造解析により立体構造が解明されているウシ小腸のビタミンD依存性カルシウム結合タンパク質 (Calbindin D<sub>9k</sub>) と比較的高いホモロジー (互いに同一なアミノ酸は28%) を示す。そこで、 $\beta$ サブユニットの立体構造を、Calbindin D<sub>9k</sub>のホモロジーモデリングによって求め、次にこの予測構造を用いて、ホモダイマー ( $\beta\beta$ ) の合理的な立体構造を予測した (図4)。この予測立体構造の原子座標を用いて計算した慣性半径 ( $R_o$ ) は、X線溶液散乱の実験値とほぼ等しい値を示した。さらに、核磁気共鳴法により決定された立体構造と非常に似たものであった<sup>24)</sup>。

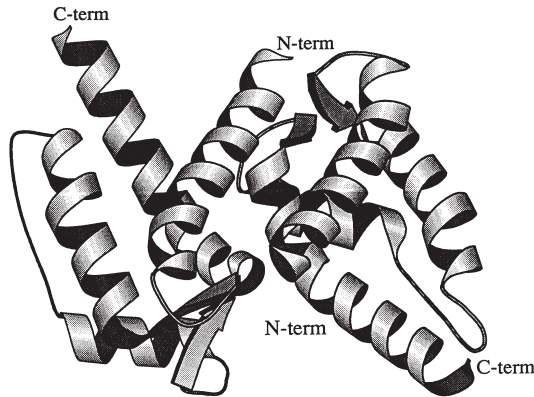


図4 S100Bの予測立体構造

#### b. アメロジェニンの立体構造予測

歯のエナメル質形成を制御する細胞外マトリックスタンパク質であるアメロジェニンは、Pro-Xaa-ProあるいはPro-Xaa-Xaa-Proの配列が繰り返している (Xaaは任意のアミノ酸1残基を表す)<sup>25)</sup>。最初にX線小角散乱により慣性半径を測定した。つぎにその値を満足する合理的な立体構造をコンピュータグラフィックスによる分子モデリング法により予測することを試みた。これらの結果から、ブタ・アメロジェニンの酵素分解物である20KDaフラグメントは、2回折れ曲がってポリプロリンIIあるいは $\beta$ -シート構造の3本が互いによりあった非常に伸びた束構造をとることを提案した (図5)<sup>26)</sup>。

### (3) タンデムリピートと核酸との相互作用モデル

#### a. AlgR3

緑膿菌に存在するアルギン酸合成に関与する遺伝子のプロモーターを活性化する遺伝子調節タンパク質 AlgR3/AlgP は、Ala-Ala-Lys-Pro (アミノ酸1文字記号でAAKP) のテトラペプチドが連続して40回あるい



図5 アメロジェニンの予測立体構造

は45回繰り返す特徴的なアミノ酸配列 (分子全体の3/4を占める) を持っている<sup>27)</sup>。これらの繰り返し単位には塩基性アミノ酸である Lys, Argが存在するため、DNAと直接相互作用することが考えられる。これらの合成ペプチドを用いたDNAとの結合実験は、確かにDNAと相互作用することを示した。実際のアミノ酸配列により近い繰り返し配列 (AAKP)<sub>5</sub>AAKTAの立体構造を予測した結果、 $\alpha$ ヘリックスがプロリンによってねじれたコイルドコイル構造として特徴づけられ、DNAの大溝 (major groove) に合理的にフィットした<sup>28)</sup>。

#### b. US11

単純ヘルペスウイルスの遺伝子US11のコードするタンパク質は、ウイルス感染細胞から2本鎖DNAセルロースカラムを用いて分離された核酸結合タンパク質であり、C末端側にPro-Arg-Xaa (アミノ酸1文字記号でPRX) が連続して23回繰り返す特徴的な配列をもっている<sup>29)</sup>。この配列の合成ペプチドは2本鎖DNAに直接結合することをDNAセルロースを用いて示した。また、コンピュータグラフィックスによる立体構造予測研究から、PRX反復配列は、ポリプロリンと $\beta$ -ターンが繰り返したコイルドコイル構造をとって2本鎖DNAあるいはRNAのらせん状の溝に結合できることを見いだした<sup>30)</sup>。

AlgR3およびUS11において提案したコイルドコイル構造は、ヘリックス-ターン-ヘリックス、ロイシンジッパー、亜鉛フィンガー等とは異なる新しいDNA結合モチーフであると考えられる。

### 遺伝情報解析の将来展望

ベルツが「タンパク質構造の医学への応用の将来の発展に関して、特別の予言をすることは性急であろう。しかし、過去10年が未来の10年の指針となるなら、進歩が加速されることはほぼ間違いない。」と述べているように<sup>31)</sup>、タンパク質の立体構造の情報が今後の医療にと

って重要な役割を担うことは、多くの研究者が認識しはじめている。多くのタンパク質の立体構造が未知である現状において、遺伝情報解析により一次情報から立体構造を予測することは大変意義のあることといえる。

今後、遺伝情報解析の研究への期待がさらに大きくなり、分子生物学、生化学、医学を含む様々な専門分野の研究者がこの研究テーマに取り組むことが予測され、益々発展するものと考えている。

著者らは、タンデムな繰り返し配列を持つタンパク質に着目した遺伝情報解析を進めており、本稿で説明した研究成果の多くはタンデムな繰り返しを持つタンパク質に関連するものである。タンデムな繰り返しを持つタンパク質の数は膨大で、現在もお増え続けている。このような状況から著者らは、“タンデムな繰り返し”を切り口としたタンパク質の構造、機能、分子進化、さらには生体システムの原理の解明を目指している。

### 謝 辞

本解析システムを用いた研究を進める上で多大なご協力、有益なご助言を頂いた、青葉孝昭教授（日本歯科大）、和泉義信教授（山形大工）、大内東教授（北大工）、高橋伸幸助教授（北海道教育大学）、団野源一教授（神戸大農学部）、津田栄博士（工業技術研究所）、中山享講師（長崎大医学部）、新田順子教授（札幌国際大）、松田慎行助教授（北海道教育大）、引地邦男教授（北大理）、Robert H. Kretsinger教授（米国バージニア大）、ならびに北大大学院修士課程学生の北田英俊君、田中剛範君、戸野一博君に厚く御礼申し上げます（五十音順）。また、日頃より分子生物学・生化学に関するさまざまな知識および情報を親切にご教授くださっている山田恵子助教授（札幌医大保健医療学部）、相馬仁講師（札幌医大医学部）に感謝いたします。

### 文 献

- 1) 金久實：遺伝子とゲノムの情報処理. 情報処理 35 : 983-990, 1994
- 2) 金久實：ゲノム情報学. 情報処理 37 : 908-913, 1996
- 3) Schulz GE and Schirmer RH : Principles of Protein Structure, Springer-Verlag, 1979
- 4) 大柳俊夫：広域情報検索 Gopherによるインターネット上の情報検索. 医学のあゆみ 170 : 171-173, 1994
- 5) 松嶋範男：基礎医学および分子生物学研究へのコンピュータ利用. 医学のあゆみ 170 : 225-228, 1994
- 6) 大柳俊夫、山本雅人、大内東：命題論理におけるβ節充足可能性問題の解法. 電気学会論文誌 114-C : 796-804, 1994
- 7) Yamamoto M, Ohyanagi T, Ohuchi A :

Computational results for satisfiability problems. Proc APORS'94 538-545, 1994

- 8) 皆川美紀、嘉屋優子、佐伯和子ほか：看護におけるコンピュータネットワーク時代の到来 - パソコン通信・インターネット活用の展望、電子メールによるインターネットでのコミュニケーション、Gopherによる看護情報の公開. 看護教育 35 : 1078-1092, 1994
- 9) 嘉屋優子、皆川美紀、佐伯和子ほか：札幌医科大学保健医療学部看護学科のインターネット普及活動. 看護 48 : 65-70, 1996
- 10) 嘉屋優子、皆川美紀、佐伯和子ほか：海外の看護系大学によるインターネット上の情報サーバーの分析. 看護研究 29 : 321-330, 1996
- 11) 皆川美紀、佐伯和子、嘉屋優子ほか：看護系大学がインターネット上で公開する教育関連情報について - 米国の5大学の発信情報の分析より -. 看護教育 37 : 1160-1163, 1996
- 12) 佐伯和子、嘉屋優子、皆川美紀ほか：インターネット上での看護系メーリングリストのコミュニケーション構造. 日本看護科学会誌 16 : 67-74, 1996
- 13) Sohma H, Matsushima N, Watanabe T et al : Ca<sup>2+</sup>-dependent binding of annexin IV to surfactant protein A and lamellar bodies in alveolar type II cells. Biochem J 312 : 175-181, 1995
- 14) 松嶋範男、団野源一、松田慎行ほか：タンパク質のX線溶液散乱とコンピュータグラフィックス. Photon Factory News 13 : 19-23, 1996
- 15) Kraulis PJ : MOLSCRIPT : a program to produce both detailed and schematic plots of protein structures. J Appl Cryst 24 : 946-950, 1991
- 16) Kobe B, Deisenhofer J : Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature 366 : 751-756, 1993
- 17) Kobe B, Deisenhofer J : The leucine-rich repeat : a versatile binding motif. Trends Biochem Sci 19 : 415-421, 1994
- 18) Kobe B, Deisenhofer J : Proteins with leucine-rich repeats. Curr Opin Struct Bio 5 : 409-416, 1995
- 19) Ohyanagi T, Matsushima N : Toward evolution of leucine-rich repeats : Determining the consensus sequence of the leucine-rich repeat. The 3rd International meeting of the Society for Molecular Biology and Evolution, 71-72, 1995
- 20) Ohyanagi T, Matsushima N : Sequence analysis and classification of tandem leucine-rich repeats within 147 proteins. to be submitted, 1997
- 21) Tanaka T, Ohuchi A, Ohyanagi T et al :

- Homopolymers in proteins. The 3rd Inter-national meeting of the Society for Molecular Biology and Evolution, 63, 1995
- 22) 三木哲郎：筋緊張性ジストロフィー. 医学のあゆみ 179 : 338-342, 1996
- 23) Schafer BW, Heizmann CW : The S100 family of EF-hand calcium-binding proteins: functions and pathology. Trends Biochem Sci 21 : 134-140, 1996
- 24) Drohat AC, Amburgey JC, Abildgaard F et al : Solution structure of rat apo-S100B ( $\beta\beta$ ) as determined by NMR spectroscopy. Biochemistry 35 : 11577 - 11588, 1996
- 25) 青葉孝昭、田辺孝子、深江允：アメロジェニン-歯のエナメル質形成を制御する細胞外マトリックス蛋白質-. 蛋白質・核酸・酵素 38 : 811-820, 1993
- 26) Matsushima N, Izumi Y, Aoba T : Small - angle X-ray scattering and computer - aided molecular modeling studies of 20kDa fragment of porcine amelogenin : does amelogenin adopt an elongated bundle structure? to be submitted, 1996
- 27) Kato J, Misra TK, Chakrabarty AM : AlgR3, a protein resembling eukaryotic histone H1, regulates alginate synthesis in *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA 87 : 2887-2891, 1990
- 28) Medvedkin VN, Permyakov EA, Klimenko LV et al : Interactions of  $(Ala^*Ala^*Lys^*Pro)_n$  and  $(Lys^*Lys^*Ser^*Pro)_n$  with DNA. Proposed coiled-coil structure of AlgR3 and AlgP from *Pseudomonas aeruginosa*. Protein Eng 8 : 63-70, 1995
- 29) McGeoch DJ, Dolan A, Donald-S et al : Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. J Mol Biol 181 : 1-13, 1985
- 30) 中山享、毎田徹夫、Kretsinger RHほか：PRX型DNA結合蛋白質の立体構造予測.第2回「タンパク質立体構造の構築原理」ワークショップ, 1995
- 31) Perutz M : Protein Structure - New Approach to Disease and Therapy. New York, W.H.Freeman and Company,1992  
(黒田玲子訳:タンパク質-立体構造と医療への応用. 東京、東京化学同人、1995、p255-266)



## Computer-based Analysis System for Research on Genetic Informatics

Toshio OHYANAGI, Norio MATSUSHIMA

Department of Liberal Arts and Sciences, School of Health Sciences, Sapporo Medical University

### Abstract

Vast sequence data of amino acids in proteins and nucleic acids have been increasing explosively along with the progress in molecular biology. Life informatics or genetic informatics using the sequence data is developing as a new field of research which is expected to provide a systematic knowledge in molecular biology and to solve some fundamental principles of life. Most of the main research has been based on sequence analysis which gives information on the structure, function and evolution of proteins and genes.

In this article, first we describe a computer-based analysis system that was constructed in our laboratory. Secondly, we note interesting results of our studies obtained with this system. Our studies contain the sequence analyses of leucine-rich repeats and homopolymers within many proteins, and the structure predictions of S100B, amelogenin, and each of tandem repeats within algR3 and US11 which interact with DNA or RNA.

Key words : Life informatics, Genetic informatics, Sequence analysis, Structure prediction, Tandem repeats