

実験医学

EXPERIMENTAL MEDICINE

別刷

株式会社 羊土社

〒101 東京都千代田区神田司町2-15
田中ビル TEL 03(295)4550(代)

ゲノム解析とデータベース

宮澤三造

ゲノム解析計画に必須なデータベースについて考える。各種データベースの現状および DNA データバンクのゲノム解析などへの対処、すなわち①データ収集・処理システムの改善、②関係データベースへの移行を述べる。またゲノム解析計画を遂行する上で解決せねばならないデータベースに関する諸問題が議論される。

はじめに

ゲノム解析は総体としてのゲノム全体を、DNA 配列決定によりその全遺伝情報を解析しようとする試みである。当然のことながら遺伝子地図 (genetic map or linkage map) や物理的地図 (physical map) などの染色体上の位置情報が必要となる。物理的地図はその解像度により Restriction map, DNA 配列データなどに分かれる。これらは当然データベースとして管理されねばならない。また DNA 配列の遺伝的多型を記述するものとして Restriction Fragment Length polymorphism (RFLP) や、遺伝変異、特にヒトの遺伝病に関するデータベースも有用である。また解析にあたって各種の細胞系、染色体、クローンなどを維持しなくてはならない。その管理のためのデータベースも必要となる。また実験を効率よく進めるためのツールも必要となろう。以上のようなゲノム解析の特質は、ゲノムデータという質的側面とゲノムデータゆえの量的側面に起因する。現在バイオテクノロジーの分野でどのようなデータベースが存在するか、DNA データバンクのゲノム解析への対処、ゲノム解析を遂行するために必要なデータベース構築における問題点を述べる。

〔キーワード〕

ゲノム解析, データベース, DNA データベース, DDBJ, GenBank, EMBL, 関係データベース, 分散データベース, ネットワーク, Biocomputing, Bioinformatics

Sanzo Miyazawa, 国立遺伝学研究所遺伝情報研究センター遺伝情報分析研究室

1. データベースの現状

近年データベースの必要性が認識され、各種のデータベースが構築されている。そこで、データベースのデータベースが作成された。LiMB¹⁾もそのひとつで、分子生物に関連したデータベースのリストである。表 1 は LiMB をもとにして作成された主なデータベースのリストである。ゲノム解析計画の発足とも関係して、細胞、株やクローン、プローブの維持、提供を目的としたバンクも多い。このようなバンクにおいてもそれに付随した情報の管理、提供が要求される。特に管理しているクローンなどの数が膨大である時、系統だった管理が必要である。例えばヒトゲノムの場合、少なくとも 90,000 (コスミドクローンで重複なし) から 600,000 のクローンが必要となろう。American Type Culture Collection ではそのための準備段階として、酵母菌のゲノムからの約 1,300 の重複する lambda clone および数千のヒト染色体クローンを用い、クローンの維持配布だけでなくデータ管理に関する研究をスタートさせた²⁾。一方、遺伝子地図データベースに関しては、Human Gene Mapping Library (HGML) が有名である。HGML は LIT (文献情報), MAP (遺伝子地図), PROBE (プローブのデータ), RFLP (RFLP のデータ), CONTACT (研究者のリスト) の 5 つのデータベースからなり、エール大学の計算機を用いて世界の研究者に公開されている³⁾。Bachmann が集積した大腸菌の遺伝子地図⁴⁾⁵⁾、小原、磯野などの物理的地図⁶⁾は近々計算機可読な形でリリースされると聞く。配列データベースは、DDBJ⁷⁾/EMBL⁸⁾/GenBank⁹⁾が汎用の DNA データベースとして有名で

表1 ゲノム解析に関係したデータベース一覧

データベース名	作成者または組織名
1. データベースのデータベース	
* LiMB (Listing of Molecular Biology Databases)	C. Burks et al., LANL
Directory of Biotechnology Information/Resources	American Type Culture Collection
2. Cell, strain, clone, probe バンク	
Database for the Repository of Human Probes and Libraries	American Type Culture Collection
Hybridoma Data Bank on Immunoclonal	Hybridoma Data Bank
NIGMS Human Genetic Mutant Cell Repository	Coriell Institute for Medical Research
Catalog of Cell Lines	
Caenorhabditis Genetics Center	M. Edgley, Univ. of Missouri
Microbial Culture Information Service	Lab. of the Government Chemist, Cornwall House, U.K.
Microbial Strain Data Network	Biotechnology Center, U.K.
E. coli ordered clones	Kohara, Akiyama & Isono, 遺伝研
3. 遺伝子地図, 物理的地図	
* The Howard Hughes Medical Institute	Human Gene Mapping Library
Human Gene Mapping Library	
Gene Communications	J. Schmidtke, Gene Communications
RFLPD	Centre d'Etude du Polymorphisme Humain, France
Genetic Variations of Drosophila melanogaster	D. L. Lindsley, Univ. of California, San Diego
E. coli Genetic Map	B. J. Bachmann, Yale Univ.
4. DNA 配列データベース	
* Restriction Enzyme Database	R. Roberts, Cold Spring Harbor Laboratory
* Cloning Vector Sequence Database	W. Gilbert, Whitehead Institute
* Eukaryotic Promoter Database	P. Bucher, Stanford Univ.
* DDBJ/GenBank/EMBL	The DNA Data Bank of Japan/GenBank /The EMBL Data Library
5. 蛋白質配列, 炭水化物構造データベース	
* SWISS-PROT Protein Sequence Data Bank	A. Bairoch of C.M.U and EMBL Data Library
* Protein sequence database of PIR	Protein Identification resource, NBRF in collaboration with MIPS and JIPID
* PRFSEQDB, PRFLITDB	Protein Reserach Foundation in Osaka
* Amino Acid and Nucleotide Sequences of Proteins of Immunological Interest	E. Kabat, NIH
Carbohydrate Structure Database	S. Doubet, Complex Carbohydrate Research Center
6. Gel plot データベース	
Quest 2D Gel Protein Database	J. I. Garrels, Cold Spring Harbor Laboratory
Gene-Protein Index of E. coli	F. C. Neidhardt, Univ. of Michigan Medical School
7. 原子座標データベース	
* Protein Data Bank	Protein Data Bank, Brookhaven National Laboratory
8. 文献, ファクトデータベース	
Plasmid Prefix Registry	E. Lederberg, Stanford Univ.
MEDLINE and Backfiles	National Library of Medicine
BIOSIS Previews (R) and The Taxonomic Reference File	BIOSIS
CAS and CAS ONLINE Registry File	Chemical Abstracts Service

* DDBJ 計算機でアクセス可能なデータベース

ある。蛋白質配列データベースとしては、Dayhoff の Atlas of Protein Sequence and Structure¹⁰⁾に端を發

する Protein Identification Resource(PIR) のデータベースがある。PIR¹¹⁾のデータベースは配列の類似性

Summary

Databases are considered in relation with the project of genome analysis. The present status of databases in biotechnology, and what the DNA data banks, DDBJ/EMBL/GenBank, prepare for genome data of large volume are reported. Problems that must be solved to promote genome projects are discussed.

に基づき蛋白質を Superfamily, family, subfamily に分類し付加的情報を与えて利用価値が高い。また、各種の生物の蛋白質の二次元のゲルプロットのデータベースは蛋白質の同定に有用であろう。大腸菌に関しては Neidhardt¹²⁾によってこの種のデータベースが計画されている。

2. DNA 塩基配列データベース

1982年欧州に EMBL Data Library⁸⁾, 米国に

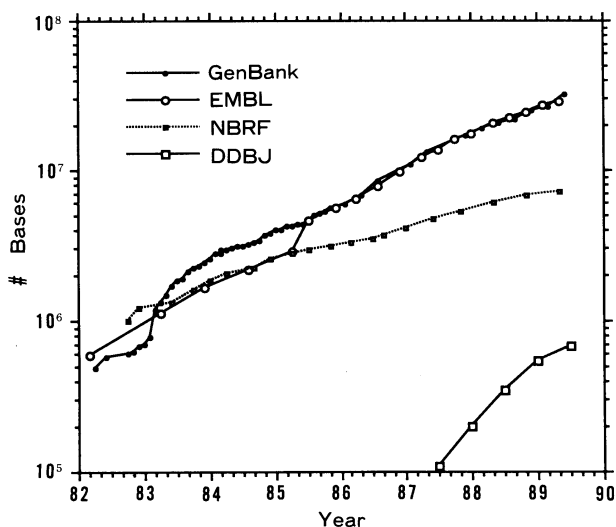


図1 GenBank, EMBL, NBRF (DNA), DDBJ 各データベースに収集されている塩基数の変遷
DDBJ データベースは GenBank, EMBL に含まれている

GenBank⁹⁾が DNA データバンクとして国の援助の下で設立された。日本では、DNA Data Bank of Japan (DDBJ)^{7)13)~16)}が1986年国立遺伝学研究所遺伝情報分析研究室に設立された。現在 DDBJ/EMBL/GenBank は密接に協力し、共通の accession number システムによるデータ入力の分担およびデータベースの再構築を行っている。

図1は GenBank と EMBL データベースに収集されている塩基数の変遷を示したものである。1985年の中頃以降、GenBank と EMBL データベースの塩基数がほぼ同じであることに気付かれると思う。この時期

表2 GenBank (Release 60, 6/89) の内容

	エントリー数	塩基数	レポート数
霊長類	4424	5328439	5520
齧歯類	4499	4418314	5377
他の哺乳類	898	1091508	1023
他の脊椎動物	1135	1175963	1353
無脊椎動物	1875	2117269	2208
植物	1660	2409054	1950
細胞小器官	829	1305885	969
細菌	2482	3726367	3172
構造RNA	1150	178351	1362
ウイルス	2508	4114238	3356
ファージ	420	514107	659
合成DNA	773	281432	838
無注釈データ	3664	5147857	4609
合計	26317	31808784	32396

表3 GenBank (Release 60, 6/89) に収集された生物種の例

生物名	エントリー数	塩基数	%ゲノム
ヒト (Homo sapiens)	4309	5149477	0.17
マウス (Mus musculus)	2942	2509480	0.08
ラット (Rattus norvegicus)	1283	1568821	0.05
ウシ (Bos taurus)	391	454121	0.015
ウサギ (Oryctolagus cuniculus)	281	328005	0.01
ニワトリ (Gallus gallus)	608	678932	0.1
アフリカツメガエル (Xenopus laevis)	253	252999	0.07
ショウジョウバエ (Drosophila melanogaster)	676	932236	0.6
線虫 (Caenorhabditis elegans)	66	96283	0.1
酵母菌 (Saccharomyces cerevisiae)	629	971530	5
枯草菌 (Bacillus subtilis)	147	211350	6
大腸菌 (Escherichia coli)	883	1212858	26
Liverwort Marchantia polymorpha chloroplast	1	121024	100
タバコ クロロプラスト	1	155844	100
Epstein-Barr ウイルス	1	172282	100
アデノウイルス2型	1	35937	100
バクテリオファージ lambda	1	48502	100
バクテリオファージ T7	1	39936	100

から GenBank と EMBL は重複がないようにデータ入力を分担し、お互いにデータを交換している。現在はデータ収集の方法が学術雑誌単位であることから、学術雑誌の分割により入力の分担を行っている。DDBJ はおもに日本で出版される雑誌を中心に入力している。しかし、直接提出されたデータについては他のデータバンクに転送することなく受理したデータバンクで入力処理をしている。

表 2 は GenBank (Release 60, 6/89) に収集されている内容を示す。また表 3 は各生物種について収集されている塩基数を示す。表を比較すると、表 3 にあげたような数少ない特定の生物種に集中して解析されてきたことがわかる。とは言ってもヒトの場合で全ゲノムの約 0.17% でしかない。大腸菌の場合でも約 26% である。収集されている全塩基数は大腸菌ゲノムの約 7 倍に相当する約 32MB である。

ゲノム解析計画への対処—DNA データバンクの場合

DNA データバンクが対処せねばならない事柄には、ゲノムデータという質的側面と量的側面がある。量的側面はもちろん多量のデータを処理せねばならないということである。ヒトゲノムの場合、現在管理しているデータの 100 倍のデータを処理せねばならない。処理の高速化が不可欠である。現在のデータ入力処理は専門家の人手にたよる部分が多い。データバンクは入力処理の自動化を目指して以下のようなプロジェクトを共同で実行しつつある。①学術雑誌にデータの提出を論文掲載の条件とするよう協力要請。Nucl. Acids Res., をはじめ Proc. Natl. Acad. Sci., Cell, Biochem., BBA などがこのシステムを採用しつつある。DNA 配列だけでなく、他のデータベースに関しても類似のシステムを学術雑誌が採用することが必要であろう。②著者自身によるデータエントリーを支援するソフトウェアの開発。このソフトのテスト版 (IBM-PC 版) はすでに GenBank により開発されている。1990 年から実際に使用されるであろう。DDBJ はこのプログラムを NEC-PC9801 に移植する計画である。③データバンク間でのデータ交換、処理の自動化およびデータ受理の自動化のため、データ交換の手続き (protocol) を標準化¹⁷⁾。データエントリーソフトウェアは、このプロトコルに従ってデータを作成する。

従来、遺伝子解析は興味ある部分に関して、DNA 配

列解析などを実行することにより研究されてきた。個々の解析は通常たかだか数キロベース程度である。このような段階ではこれまでのデータ管理、報告された配列を 1 つのエントリーとして管理する方法 (図 2a 参照) でも、なんとかなった。しかしゲノムデータにおいては、配列に関する遺伝子地図および物理的地図情報、配列相互の重複情報、配列のマージに必要な情報などを取り扱わねばならない。現在データの検索は通常エントリー単位でなされている。遺伝子名、蛋白質名、absolute and relative map positions による配列の検索が必要とされよう。また遺伝子名、蛋白質名などの用語は変更されることが多い。データベースはこのような変更が容易でなければならない。このような要求に答えるべく、データバンクは以下のようなデータベースの再構築を推進している。①現在のフラットファイルデータベースから関係データベース [Relational Database (RDB)] へ移行。DDBJ/EMBL/GenBank は実質的に同一のデータベースを構築する。GenBank はトランザクション処理のためのソフトウェアおよびデータ入力、更新のためのフロントエンドソフトウェア¹⁷⁾の開発をほぼ終え、1989 年 9 月よりテスト期間、1990 年 1 月より RDB に移行する計画である。EMBL は 1990 年中に、DDBJ も GenBank の協力を得て 1990 年中に移行する計画である。データベース管理ソフトウェアとしては Sybase を使用する。関係データベースは表の集まりから構成される。遺伝情報にはもちろん階層的構造もあるので関係データベースで表現するには難点もあるが、現時点で利用できる管理システムとしては最適であろう。② feature table の更新¹⁸⁾。データの注釈部分 (feature table) を、より計算機処理が可能なるよう明確に定義されたキーワードを用い、また書式も複雑な DNA 遺伝情報を的確に表現できる feature table に 1990 年中に移行する計画である (図 2b 参照)。

3. ゲノム解析とデータベース

ここでは、ゲノム解析計画を遂行する上で解決せねばならないデータベースに共通した問題について議論したい。

1) 学術用語 (terminology) の整備—検索およびデータベース間のクロスレファレンスのためのキーワードとして

(2a) GenBank/DDBJ format

```
LOCUS      HUMPALB      614 bp ss-mRNA      PRI      15-JUN-1989
DEFINITION Human prealbumin mRNA, complete cds.
ACCESSION  M10605
KEYWORDS   prealbumin.
SOURCE     Human liver, cDNA to mRNA, clone PA7.
ORGANISM   Homo sapiens
            Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
            Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE  1 (bases 1 to 614)
AUTHORS    Wallace, M.R., Naylor, S.L., Kluge-Beckerman, B., Long, G.L.,
            McDonald, L., Shows, T.B. and Benson, M.D.
TITLE      Localization of the human prealbumin gene to chromosome 18
JOURNAL    Biochem. Biophys. Res. Commun. 129, 753-758 (1985)
STANDARD   simple staff review
COMMENT     Draft entry and sequence in computer readable form for [1] kindly
            provided by M.R. Wallace, 26-DEC-1985.
FEATURES   from to/span description
            pept      26      469      prealbumin precursor /nomgen="PALB"
            /map="18q11.2-12.1" /hgml_locus_uid="LC0122E"
            sigp      26      85      prealbumin signal peptide
            matp      86      466      prealbumin
            mRNA      <      1      614      prealb mRNA
            refnumbr   1      1      numbered 1 in [1]; zero not used
BASE COUNT 148 a 162 c 155 g 149 t
ORIGIN     Chromosome 18q11.2-q12.1; 247 bp upstream of AluI site.
            1 cagaagtcca ctcattcttg gcaggatggc ttctcatcgt ctgctctcc tctgccttgc
            ...
            ...
            601 ttctgtgaa aggc
//
```

(2b) GenBank/DDBJ New Feature Table

```
FEATURES      LOCATION/QUALIFIERS
sig_peptide   26..85
mat_peptide   86..466
              /product="prealbumin"
CDS           26..469
              /product="prealbumin" /gene="PALB"
              /map="18q11.2-12.1" /hgml_locus_uid="LC0122E"
mRNA          <1..614
BASE COUNT    148 a 162 c 155 g 149 t
```

図2 GenBank/DDBJ format の例

(b)は1990年より使用する新 feature table を示す。EMBL も同様のフラットファイルを用い管理している

データベース化するにあたって必要とされるものの1つに学術用語の整備があげられる。生物名、DNA プローブ名、制限酵素名、遺伝子名、蛋白質名、Ezyme code number など数多い。矛盾がないようデータを記述するためには、これら学術用語の統一は欠くことができない。用語はデータベースの検索のためのキーとして使用されるだけでなく、データベース間でのクロスレファレンスにおいても使用される。例えば遺伝子名は、配列と遺伝子地図のクロスレファレンスのために使用できる。よって用語は個々のデータベース内だけで矛盾なく使用されるだけでなく、関連したデータ

ベースがすべて統一した命名を採用することが望ましい。またデータベースは旧名などを別名として管理する必要がある。標準化の必要な分野の同定、標準設定委員会との情報交換を確立せねばならない。論文雑誌の簡略名の標準化に関しては、米国国立医学図書館(NLM)が標準化委員会として機能している。DDBJ/GenBank/EMBL/PIRは、1989年からNLMの設定した簡略名を採用している。

2) データベース間のクロスレファレンス

ゲノム解析には複数のデータベースが必要である。これらのデータベースを一体化し1カ所で集中管理す

ることは不可能であろう。それゆえデータベース間のクロスレファレンスが要求される。DNA データバンクはデータ入力の支援データベースとして生物分類、学術雑誌名データベースを共同で整備してきた。またクロスレファレンスのため遺伝子名を使用する際、それが適切な用語か否かを調べるため遺伝子地図データベースをアクセスする。新遺伝子名の場合は登録せねばならない。これはデータベース間でクロスレファレンスを試みる際、何が必要となるかの例である。すなわちクロスレファレンスのために、関連データベースへのアクセス (retrieval と update) が必要とされる。現在 DNA データバンクは、電子郵便を用い情報を交換し支援データベースを集中管理している。ゲノム解析におけるデータ量は、はるかに高速な分散アクセスを必要とするであろう。

3) 分散データベース

ゲノム計画には多数の研究者が参加する。データバンクがデータ入力を代行するシステムは、データ量と専門知識が必要とされることの2点において、もはや機能しえない。研究者自身によるデータエントリーが要求される。参加する研究者は日米欧にまたがっている。それゆえデータベースの retrieval だけでなく、update も可能な分散アクセス(データは1カ所に集中して存在し、アクセスはネットワークに接続された複数の計算機上で可能である形態)が要求される。しかし真の分散データベース管理システム(データそれぞれも複数の計算機に分散されて存在し、同じトランザクションで複数の計算機のデータにアクセス可能な形態)を日米欧で実現するには困難な点が多い。数カ所のセンターがおのおのの地域において分散アクセス可能なデータベースを管理し、センター間では完全ではないにしても、実質的に同一のデータベースが維持できるように互いにデータの更新情報(transaction)を交換するシステムが現在望みうるベストであろう。DDBJ/EMBL/GenBank が構築しつつある関係データベースは、まさにこのようなシステムである。各種データベース間でこのようなシステムを完成させるにはまずデータ交換のための protocol を設定しなくてはならない。1988年 NLM に設立された Center for Biotechnology Information はデータベース間でのデータ交換の protocol 設定に意欲をみせている。

4) 計算機ネットワーク

分散データベースの構築には、信頼性の高い高速のデータ交換が欠かせない。米国では全米の大学、政府機関、企業にある多数の計算機(1986年次で約2,000ノード)を高速デジタル回線(56 Kilobits per second)で接続した Internet, NSFNet(1.54Mbps), その他特定分野のネットワークが急速に拡大されつつある。これらのネットワークはすべてゲートウェーを通して相互に接続され、巨大な広域ネットワークを構成している。ネットワークに属する計算機間では、ファイル転送はもちろん、どの計算機でも直接アクセスすることが可能であり、ネットワークはあたかも1つの計算機であるがごとく機能する。このネットワークに接続された計算機間では、分散アクセスを容易に実現できる。もちろん現在、研究者へのデータベースの提供は、このネットワークを通じてなされている。一方、日本ではネットワークは無に等しい。大学間ネットワークの N1 protocol は Internet の TCP/IP protocol にはるかに機能が劣り、N1 から Internet へのゲートウェア接続はほとんど意味がない。DDBJ はこのような現状をふまえ、DNA 分散データベースの共同構築のため1989年、遅くとも1990年には直接 Internet に DDBJ 計算機を接続することを計画している。日米欧のデータベース間で分散アクセスを可能にするためにも、このネットワーク接続は意味をもつ。国内の関係機関もこのネットワークに接続される必要がある。

おわりに

現在米国は、ゲノム解析のためデータベースおよび Biocomputing/Bioinformatics の分野の育成に努めている。最近 NLM はヒトゲノムデータベース構築の予備段階として、大腸菌のデータベース構築計画を発足させた。LANL(Los Alamos National Laboratory)では実験家のツールとして画像解析、画像データを含む遺伝子マッピングデータのデータベースを完成させつつある¹⁹⁾。データベースに関しては明らかにエンジニアリングの問題である。一方、ここでは触れるスペースがなかったが、ゲノム解析では DNA 配列の情報解析の発展が要求される。L. Hood は計算機科学者と共同でホモロジーサーチのためのハードウェア(TRW chip)を開発した²⁰⁾。今後 DNA、蛋白質配列のパターン解析がますます盛んになろう。

文献

- 1) Burks, C., Lawton, J.R. & Bell, G.I.: Science, 241 : 888, 1988
- 2) Office of Energy Research, Office of Health and Environmental Research : The Human Genome Initiative of the U.S. Department of Energy, DOE/ER-0382
- 3) Howard Hughes Medical Institute: New Haven Human Gene Mapping Library Chromosome Plots, Number 4, HGM9.5, 1988
- 4) Bachmann, B.J.: *Escherichia coli* and *Salmonella Typhimurium* cellular and Molecular Biology, Ed. in chief : F.C. Neidhardt (Washington, D.C. : American Society For Microbiology), pp. 807-876, 1987
- 5) Genetic maps 1987, S.J. O'Brien, Ed., Cold Spring Harbor Laboratory, 1987
- 6) Kohara, Y., Akiyama, K. & Isono, K.: Cell, 50 : 495-508, 1987
- 7) Miyazawa, S.: The Interface between Computational Science and Nucleic Acid Sequencing, Santa Fe Institute Studies in the Sciences of Complexity, Eds. G. Bell & T. Marr (Reading, MA : Addison-Wesley), vol. VIII, pp.47-61, 1989
- 8) Cameron, G.N.: Nucl. Acids Res., 16, 1865-1867, 1988
- 9) Burks, C. et al.: The Interface between Computational Science and Nucleic Acid Sequencing, Santa Fe Institute Studies in the Sciences of Complexity, Eds. G. Bell and T. Marr (Reading, MA : Addison-Wesley), vol. VIII, 1989
- 10) Dayhoff, M.O.: Atlas of Protein Sequence and Structure 1978, Vol. 5, Supplement 3, Ed. Dayhoff, M.O., (Washington, D.C.: Nat. Biomed. Res. Found.), 1978
- 11) George, D.: The Interface between Computational Science and Nucleic Acid Sequencing, Santa Fe Institute Studies in the Sciences of Complexity, Eds. G. Bell and T. Marr (Reading, MA : Addison-Wesley), vol. VIII, 1989
- 12) Phillips, T.A., Vaughn, V., Bloch, P.L. & Neidhardt, F.C.: *Escherichia coli* and *Salmonella Typhimurium* cellular and Molecular Biology, Ed. in chief : F. C. Neidhardt (Washington, D.C. : American Society for Microbiology), pp. 919-966, 1987
- 13) DDBJ : DDBJ ニュースレター No.8(遺伝研, DDBJ), 1989
- 14) 宮澤三造 : DDBJ 計算機利用の手引, (遺伝研, DDBJ), 1988
- 15) Miyazawa, S.: The Manual of the FLAT Database and Sequence Analysis System for DNA and Proteins, Version 1.1, (Mishima : DDBJ), 1989
- 16) Hayashida, H.: The Manual of the QANALYS Sequence Analysis System for Molecular Evolution, Version 1.0 (beta), (Mishima : DDBJ), 1989
- 17) Cinkosky, M.J., Nelson, D. & Marr, T.G.: A technical Overview of the GenBank/HGIR Database (Los Alamos, NM : GenBank/HGIR), 1989
- 18) DDBJ, EMBL Data Library & GenBank : The DDBJ/EMBL/GenBank Feature Table : Definition, version 1 (Mishima, Japan : DDBJ), 1988
- 19) Marr, T.G.: personal communication
- 20) Research News : Science, 244 : 655-656, 1989

〈著者プロフィール〉

1975年名古屋大学大学院博士課程物理学専攻満了, 1975~1979年九州大学理学研究科生物学教室研究生, 1979~1985年 NIH, 1986~現在まで遺伝研, 大学院では分子進化, NIH では蛋白質のコンフォメーションを勉強し, 現在は DNA データベースと DNA, 蛋白質配列解析に興味をもっている. 非実験家で計算機を使用して研究している.