

国際蛋白情報データベースの設立: 蛋白合成, 特に DNA 転写機構 および蛋白工学への情報論的アプローチ

Establishment of International Protein Information Database: Approaches
by informational sciences for protein synthesis, transcription,
and protein engineering

- | | | |
|-------|---|-------|
| 代表研究者 | 東京理科大学総合研究所教授
Prof., Faculty of Science, Science Univ. of Tokyo
Akira TSUGITA | 次田 皓 |
| 協同研究者 | 東京理科大学総合研究所顧問
Advisor, General Laboratory, Science Univ. of Tokyo
Masao KOTANI | 小谷 正雄 |
| | 東京理科大学理工学部教授
Prof., Faculty of Science and Technology, Science Univ. of Tokyo
Jinya OTSUKA | 大塚 仁也 |
| | 東京理科大学理学部教授
Prof., Faculty of Science, Science Univ. of Tokyo
Kazuo SATAKE | 佐竹 一夫 |
| | 東京理科大学総合研究所客員教授
Prof., Faculty of Pharmaceutical Sciences, Science Univ. of Tokyo
Morio IKEHARA | 池原 森男 |
| | 自治医科大学医学部講師
Lecture, Faculty of Medicine, Jiji Medical Univ.
Hiroshi KIHARA | 木原 裕 |
| | 京都女子大学家政学部教授
Prof., Faculty of Home Economics, Kyoto Women's College
Tasuo OOI | 大井 龍夫 |
| | 名古屋大学理学部教授
Prof., Faculty of Science, Nagoya Univ.
Shozo OSAWA | 大沢 省三 |
| | 国立遺伝学研究所教授
Prof., National Genetics Laboratory
Akira ISHIHAMA | 石浜 明 |

A public domain protein sequence database, the Japan International Protein Information Database (JIPID) has been established. The database is equipped with two Vax II 750 minicomputers. The computer system consists of a work station, which is able to make Lan system of 15 terminals. The system is designed to be accessed by two campus telephones and three outgoing telephones as well as by an IBM Lan system. An international data file exchange system has also been established through BITNET (JNET) and DIALCOM. In addition, an Evans-and-Sutherland

computer graphics system, and several IBM and NEC personal computers are equipped. To date, JIPID has produced about 2,600 sequence data files and merged them into an international database, PIR-International. This international database includes data files from NBRF (USA), MIPS (W. Germany) as well as JIPID. JIPID is one of three official nodes and covers Asian and Oceanian subnodes (China, Taiwan, South-Korea, India, Pakistan and Australia) for assimilation and distribution of sequence data. JIPID also has a connection with DDBJ, a Japanese nucleic acid data base. PIR-International established a standard format for sequence data exchange. Together with three nucleic acid databases PIR-International shares journal scanning and has established an standardized direct entry form to databases.

Beside of the sequence database, JIPID initiated the following several databases; Bio-macromolecule-NMR, Variant database, Enzyme database, Electro carrier database, Oxygen carrier database and Thermostability database. These databases are built in a way of a summarized data file-web in which data items are inter-exchangeable by C language pointers, in order to economize computer memory and to avoid overlapping input. JIPID and NBRF are collaborating to establish an *E. coli* genome database followed by T4 phage, yeast, and rice genome databases.

研究目的

1) 国際的に統一された蛋白質情報データベースを完成し、従来より活動をしていた米国ワシントンにある NBRF のデータベースを包括した国際的蛋白質データベースの設立をする。

2) 上記国際データベースの中でアジア・オセアニア地区のデータの収集・配布をするセンターを発足し、またヨーロッパにおいてもこのようなセンターを発足させ、前記米国のノードとともに三極の on line システムを完成する。

3) 種々の蛋白質情報、活性、物理的性質、変異などのデータベースの作成をする。

研究経過

1) 計算機

NBRF の蛋白質配列データベース (PIR) および他の核酸配列データベース, EMBL, GenBank などで使われている Vax-vms の中型計算機, Vax II-750 を 1 台購入した。昭和 61 年度当時としては Vax 型計算機を購入することに十分の理解が我が国では得られなかった。このことは学内および共同研究者間においても、学会においても非常に困難なことであった。この段階で本助成金において援助されたことは、本研究活動の成否（時期を含めて）を決める基本となったことを強調したい。

我々が Vax を購入した後昭和 61 年度以降は我々の宣伝、対米経済関係などの面から急速に

Vax 型計算機が一般的になっていった。この Vax 第 1 号機は 5 個の端末をもっていたが、昭和 62 年度文部省私学助成金により Vax 第 2 号機が、研究成果の上に導入され、Lan ステーションでこの 2 台の Vax が連結されさらに 5 台の端末が増加された。また昭和 62 年エヴァンス・サザランドのグラフィックスの装置も購入された。平成元年新しい建物への移転と同時にさらに 5 台の端末計 15 台の端末機を備えるに至った。この他 IBM PC 2 台、NEC PC 3 台も購入された。IBM PC を通じて理科大中央の計算機センターと直結し、また国際通信網 BITNET に IBM を通して、また J-NET-Vax 計算機の二様の連続をした。この他 VENUS-P-KDD を通じて DIALCOM 国際ネットとも連結した。学内との連絡は 2 台の内線モデムを通し、また学外とは 3 台の電話・モデムを通じての連結が可能となっている。

2) 施設、人員

昭和 61 年度、東京理科大学野田の 6 号館 4 階に Vax 2 台と 10 台の端末、それに付属して、Lan ステーション、グラフィックスすべてを 100 m² の一室に整備し、空調設備を設置した。

平成元年 3 月中旬、野田地区に生命科学研究所が新設され、研究所の 4 研究領域の一つの領域として 250 m² の面積が与えられ CPU 室、国際通信室、国内通信室、端末室、付属図書室、書庫が得られた。現在約 30 名の登録入力員（アルバイ

ター)と、専属助手1名、兼任講師1名、事務員1名、大学院生2名、研究生2名の計8名で作業が進められるに至っている。また空調設備の完備のため、4月以降は24時間のコンピューター運転を行っている。この他入力ロボット1台を試作し、補助的なPC NEC 3台、IPB 2台、複写機2台など設置も完備してきている。

3) データベース

最重要で緊急の研究課題として蛋白質の配列のデータベースを心掛けた。次田、小谷は後述のCODATAの蛋白質データバンクコーディネーション作業委員会を通して相互に互換性を重視したスタンダードフォーマットの基準化を試み、昭和62年に基準化に成功した。さらに古くから歴史があり、世界で唯一の公的なデータバンクである米国のNBRF (National Biomedical Foundation)と、我々と期を一つにして出来つつある西独のMIPS (Martinsried International Protein Sequence Database)に呼びかけてそれを既存の蛋白質配列データベースに一本化する努力を重ね三極のデータベース間で最終的取り決めを見た(昭和63年)。このフォーマットについては核酸データベースおよびソフトを取り扱う機関などとの協議を重ね合意の得られたものを発表し¹⁾、かつこのフォーマットを使って3極のデータベースの作成を行う合意に達した。このため従来NBRFの持つソフトの改変を行い、昭和63年初頭からこれを用いて配列データの入力を開始した。JIPIDにおいて現在までの入力数は約1,700ファイルにのぼり、蛋白データファイルの全数約12,000件の内の1/7に当たることを考えるとかなり良い成績であると考えられる。MIPSも約1,500件のファイルを主としてEMBLの核酸データの転換をして入力している。したがって三極のデータベース間の共同研究は一応の成功を見たものと考えられる。現在JIPIDでは、後で述べるごとくアジア・オセアニア地域からの入力も行っている。

またPIR-Internationalに用いられている入力および解析のソフトはFORTRANで書かれている。FORTRANは計算機によって互換性がない

ため、C言語で書き直しを開始した。現在Feature Tableの項については完成を見た。

研究成果

1) 蛋白配列データベース

蛋白配列データベースにおいてはその一次配列が主な入力項目である。(3)のデータベースの項に書いたように配列データ入力は1,700件、またそれに付随して作製した核酸の入力は1,200件である。

ただし、蛋白質を理解するにはその立体構造、生物活性、物理的性質などを知る必要があり、かつそれらをデータベース化し、しかも蛋白配列データベースと相関性をもつ構造にし、かつ他の計算機への互換性を考えてPointerを含んだC言語でソフトを作ることとした。またこのため配列データ、生物活性データ物理特性データの共通項を各データベースで一度入力するとPointerを使って他データファイルに移行できるようにした。このため各々のデータファイルには一度入力した場合他のファイルに同じ入力を繰り返さなくてよく、メモリーの節約と重複入力に伴う誤りを避けることが出来る。現在これらのソフトを備えた配列データベースの作製している。

2) 生物活性データベース

一般に生物活性といっても多種多様な生物活性があるためそのすべてを一つのフォーマットに統一することはできない。このためまず酵素活性データベースから着手した。このデータベースはさらに一般酵素データベースと個別酵素データベースに分けられ、前者は一般的な反応機構をはじめ、数項目の共通な一般的性質を後者は個々の配列データベースに対応する個々の性質を入れることとした。ここで強調しておきたいことは前述の配列データベースでは1本の蛋白鎖がデータの単位となっているが、活性データベースでは活性を持っている単位、すなわち1本鎖から複数本鎖のデータから作られていることである。

酵素活性データベースに続いて電子伝達蛋白質、酸素運搬蛋白質のデータベースを作った。これらのフォーマットは前述の配列データベースのスタンダードフォーマットを原形として修飾した

Table 1. Biological activity database

ENZYME		ELECTRON CARRIER PROTEIN		OXYGEN CARRIER PROTEIN	
GENERAL	SPECIFIC	GENERAL	SPECIFIC	GENERAL	SPECIFIC
ACCESSION NUMBER TITLE EC-NUMBER SYSTEMATIC-NAME ALTERNATE-NAME DATE SOURCE DISTRIBUTION	ACCESSION NUMBER TITLE EC-NUMBER SYSTEMATIC-NAME ALTERNATE-NAME DATE SOURCE HOST EXPRESSION	ACCESSION NUMBER TITLE ALTERNATE-NAME DATE DISTRIBUTION	ACCESSION NUMBER TITLE ALTERNATE-NAME DATE SOURCE	ACCESSION NUMBER TITLE ALTERNATE-NAME DATE DISTRIBUTION	ACCESSION NUMBER TITLE ALTERNATE-NAME DATE SOURCE
ORGANIZATION-G COFACTORS-G REACTION	ORGANIZATION COFACTORS REACTION SUBSTRATE	ORGANIZATION-G COFACTORS-G REACTION	ORGANIZATION COFACTOR REACTION	ORGANIZATION-G COFACTOR-G FUNCTION	ORGANIZATION COFACTOR FUNCTION
EQUILIBRIUM-CONSTANT-G PH-RANGE ACTIVATOR-G INHIBITOR-G ACTIVITY-G	EQUILIBRIUM-CONSTANT OPTIMAL-PH ACTIVATOR INHIBITOR ACTIVITY THERMOSTABILITY PHYSICAL	REDOX-POTENTIAL	REDOX-POTENTIAL		
PHYSICAL-G		PHYSICAL-G SPECTRUM	PHYSICAL SPECTRUM	PHYSICAL-G SPECTRUM	PHYSICAL SPECTRUM
COMMENT-APPLICATION-G	PATHOLOGY COMMENT-APPLICATION GENETIC SUMMARY-P FEATURE-P SEQUENCE-P SUMMARY-N FEATURE-N SEQUENCE-N	COMMENT-APPLICATION-G	COMMENT-APPLICATION GENETIC SUMMARY-P FEATURE-P SEQUENCE-P SUMMARY-N FEATURE-N SEQUENCE-N	COMMENT-APPLICATION-G	COMMENT-APPLICATION GENETIC SUMMARY-P FEATURE-P SEQUENCE-P SUMMARY-N FEATURE-N SEQUENCE-N
CROSS-REFERENCE(S)	CROSS-REFERENCE	CROSS-REFERENCE(S)	CROSS-REFERENCE	CROSS-REFERENCE(S)	CROSS-REFERENCE

もので、相互のフォーマットがかなり類似しているように作られている。Table 1 にこれら 3 種のフォーマットを示してある。引き続いて他の活性データベース例えば毒性データベースなどを作りつつある (文献 2)。

3) 物性データベース

蛋白質の物性を表すデータベースも多様になる。手始めに熱耐性のデータベースを作ってみた。熱耐性は最近爆発的に活発になってきた蛋白質工学にとって構造の変化に伴う活性の変化と共に欠くことのできない特性である。このため熱耐性データベースを試作した。(Table 2 参照, 文献 2, 3, 6)

4) 変異データベース

最近の site specific 変異手法の確立を発端とした人工的に蛋白質を変化させ、有用な蛋白質を作り出す蛋白質工学の発展、さらにはこの手法を用いて構造と生物活性および蛋白質の安定性の分子機構を解明することに変異データベースは重大な貢献をしつつある。これらのデータは現在のところデータベース化されず膨大な文献が放置されつつある現状である。これらのデータを上記の蛋白配列、核酸配列、生物活性、蛋白質物性などのデータベースと統合性をもたしたフォーマットに入れることを試みた。ここでは蛋白質および核酸の配列

Table 2. Physical database Variant database

ACCESSION NUMBER	<u>ACCESSION NUMBER</u>
TITLE	TITLE*
EC-NUMBER	EC-NUMBER
ALTERNATE-NAME	<u>ALTERNATE-NAME</u>
DATE	DATE
SOURCE	SOURCE
HOST	HOST
ORGANIZATION	<u>EXPRESSION</u>
COFACTORS	<u>METHOD</u>
REACTION	ORGANIZATION
OPTIMAL-PH	<u>ORGANIZATION-V</u>
TRANSITION-NUMBER	ACTIVITY
TRANSITION-STATE-N	<u>ACTIVITY-N</u>
#Ttrs-STEP-N	PHYSICAL
#DELTA-H-STEP-N	<u>PHYSICAL-V</u>
#DELTA-CP-STEP-N	PATHOLOGY*
#DELTA-G-25-STEP-N	COMMENT-APPLICATION*
CONDITIONS	SUMMARY-P
#Methods	<u>SUMMARY-V</u>
#HDS(C(a))	FEATURE-P*
#HDS(C(na))	SEQUENCE-P
#vH	<u>VARIATION-P</u>
#FC	SUMMARY-N
#HITI	<u>SUMMARY-V</u>
#Reversibility	FEATURE-N*
#pH	SEQUENCE-N
#Ionic strength	<u>VARIATION-N</u>
#Buffer	CROSS-REFERENCE
#Salt	
#Sample concentration	
SUMMARY	
SEQUENCE-P	
CROSS-REFERENCE	

の変換、付加、欠損、組合せなどを、これらの配列のデータベースのコードと変換の指示でソフトを用いて自由に入力すること、さらにはその変化に伴った野性型のものから変化した結果の生物活性、蛋白質物性をその各々のフォーマットに準じて入力してあり、ソフトによってこれらの活性又は

物性を比較できるように仕組まれている。ここでは前項の 2), 3) で述べたデータベースは、変異データベースの活性および物性のデータの野性株型データベースであるといえる。ここでもメモリーの節約と重複入力の省力化が考えられている。

変異データベースは上記の人工変異に限らず、天然の変異データベースも包含している。蛋白配列データベース PIR では、95%以上の相同性をもつ配列については、同一データベースに入力する方針をとっているため、一つのアミノ酸の変化で重要な性質の変化を起こすようなものについては十分な配慮がなされていない。このため、特に医学的に問題となるような天然変異のデータに注目して天然変異データベースも作製し始めた。こ

のデータベースではほぼ人工変異データベースのフォーマットを踏襲している。Table 2 の右に変異データベースの入力フォーマットを示してある。ここでも前述のごとく pointer によるデータの取り込みを頻用してある。(文献 4~7, 21, 22)

5) NMR データベース

ここ二、三年前までは蛋白の立体構造は主として X 線解析によっていて、このためのデータ Reposit は PDB (米国) によって整えられている。しかし近年に NMR の測定から立体構造の解析が可能となった。我々はこれらの NMR のデータを、文献を主としたデータベースと Reposit データベースに分けて作製した。NMR 情報データベースのフォーマットは前同様スタンダードフォーマットを準用し、後者は PDB とほぼ同様

Table 3. Database web

PIR(PSQ)	NAQ	ENZYME SPECIFIC	VARIANT	THERMOSTABILITY
ENTRY-CODE ACCESSION-NUMBER TITLE EC-NUMBER ALTERNATE-NAME INCLUDES SOURCE HOST COMMENT	ENTRY-CODE ACCESSION-NUMBER TITLE SOURCE HOST COMMENT	ACCESSION-NUMBER TITLE* EC-NUMBER SYSTEMATIC-NAME ALTERNATE-NAME* DATE SOURCE HOST EXPRESSION ORGANIZATION COFACTORS REACTION SUBSTRATE EQUILIBRIUM-CONSTANT OPTIMAL-PH ACTIVATOR INHIBITOR ACTIVITY THERMOSTABILITY	ACCESSION-NUMBER TITLE* EC-NUMBER ALTERNATE-NAME DATE SOURCE HOST EXPRESSION METHOD ORGANIZATION ORGANIZATION-V ACTIVITY ACTIVITY-V PHYSICAL PHYSICAL-V PATHOLOGY* COMMENT-APPLICATION*	ACCESSION-NUMBER TITLE* EC-NUMBER ALTERNATE-NAME DATE SOURCE HOST ORGANIZATION COFACTORS REACTION OPTIMAL-PH TRANSITION-NUMBER TRANSITION-STATE-N CONDITIONS
GENETIC SUMMARY-P FEATURE-P SEQUENCE-P	GENETIC SUMMARY-N FEATURE-N SEQUENCE-N	GENETIC SUMMARY-P FEATURE-P SEQUENCE-P SUMMARY-N FEATURE-N SEQUENCE-N	SUMMARY-P SUMMARY-V FEATURE-P* SEQUENCE-P VARIATION-P SUMMARY-N SUMMARY-V FEATURE-N* SEQUENCE-N VARIATION-N CROSS-REFERENCE	SUMMARY-P FEATURE-P SEQUENCE-P CROSS-REFERENCE
CROSS-REFERENCE	CROSS-REFERENCE	CROSS-REFERENCE	CROSS-REFERENCE	CROSS-REFERENCE

なフォーマットを用いて立体構造の軸，原子間隔を含めたものとなっている⁸⁾。

6) 文献データベース

蛋白配列データベースと核酸配列データベースに入力してある学術論文はかなり重複している。さらに上記のデータベース群を合わせて考えると，重複する文献が数多くあることが予想される。このため上記すべてのデータベース間で統一した文献データベースを独立させることが考えられた。(このため Table 1, 2, 3 に示されたデータベースから文献の項は除かれている。) 勿論符号とコード名で，このデータファイルから各々のデータベースに引用できるように作られている。

7) データベースの統合

上述のすべてのデータベースを統合し，C言語の Pointer を用いて相互のデータベース間で併用しあって各々のデータベースのメモリーを極力縮小し合うように努力した。ただしどのようにしてもその固有のデータベースで独自に入力しなくてはならない項目がある。このような場合の優先順位は蛋白配列→核酸配列→生物活性，蛋白物性→NMR→文献のようにした。Table 3 には独自に入力するものを黒地ぬきで，借用できるものを普通字で示し，その際でも借用するに際して修飾を要するものを*で示してある^{6), 23)}。

8) 大腸菌全遺伝子データベースなどの作成

生物の全遺伝子のデータベース作成については，ヒト全遺伝子のデータベースの作成のような国際的大プロジェクトがあるが，その方向に向けての研究開発には多くの課題が残されている。JIPID ではこのラインにそった最初の試みとして大腸菌全遺伝子のデータベースの作成を試み，個々に発表されている遺伝子の重複を，その配列，遺伝子の物理的地図，生物的地図，制限酵素地図などを統合した全遺伝子マップを発表した。入力された全ヌクレオチド配列は全体の約半分を占め，25%がリンクされている。これと連動して各遺伝子に対応する蛋白質のデータベースも NBRF と協力の下に上述の Pointer で引用出来るよう作られている⁹⁾。

これに続いて T4 ファージ全遺伝子のデータ

ベース化も進められ，ほぼヌクレオチド 85%，蛋白 95% をカバーするに至っている。続いて MIPS を中心として酵母菌の全遺伝子を中心として酵母菌の全遺伝子を，また JIPID を中心にしてイネ全遺伝子のデータベース化に向けて努力をしている。

9) 国内の協力態勢の確立

蛋白配列データベースの内での学術雑誌から情報の抽出には，蛋白配列の JIPID での国際的分担雑誌の配分を，さらに国内協力者とこの分担雑誌の分担を協力する態勢を確立した。北海道大学薬学部の有坂文雄，農水省生物資源研・野津祐三，相模中研・近藤淑，静岡大学教養・八木達彦，阪大蛋白研・戸田弘子各氏らの積極的協力を受けている。

また核酸のデータベース DDBJ (国立遺伝研) とはテープの交換，直接入力データの交換などについて on line で行っている。

また従来 DDBJ が行ってきた PIR のテープの国内配布を JIPID が PIR の公的 node となった 1989 年を期して直接行うようになった。

10) アジア・オセアニア地区の協力態勢の進展

中国北京の CODATA 中国国内委員会の協力を得て，生体高分子委員会が発足し，主として北京生物物理科学院，北京大学，上海生物化学科学院から組織され，データ入力，データ配布の JIPID のサブノードとして位置づけられた。

中国台湾は 1990 年からの CODATA 加盟が 1988 年秋に決定され，国内委員会が 1989 年より発足し，JIPID のサブノードが台北科学院を中心にスタートした。台湾内部のテープの配布は既に始まっている。

韓国は 1990 年からの CODATA 加盟が 1989 年春決定され，国内委員会の発足をまって JIPID のサブノードが出来る準備が進められている。

インドは CODATA の中で Poona 大学を中心に Bio-informatics の活動が 1988 年から発足し，Poona 大学にサブノードが設けられた。

パキスタンには JIPID のサブノードがカラチ大学内に作られて活動を準備している。

オーストラリアは CODATA 国内委員会がサブノードとなりデータの学術雑誌からの on line 入力、テープの配布が行われている。

この他アジア地区のシンガポール、香港、フィリピン、タイ、マレーシアなどと予備的接触を行いつつデータの収集、テープ配布の協力を呼びかけている。

11) 国際協力態勢

1984年から CODATA (ICSU の常設機関で、国際的な科学と工業のデータベース活動を目的としている) 内に蛋白データバンクの coordination 作業委員会を発足させ、標準フォーマットの作製¹⁾、データベース活動の調査など¹¹⁾の活躍をして来たが、この他に新しい学術雑誌の発刊¹⁰⁾も行い、最終的に米国 NBRF、西独 MIPS、日本 JIPID が一つの蛋白配列データベース PIR-International を設立するに至った。4年間2期にわたる作業委員会の最初の目的を達成した後、この活動をさらに核酸のデータベースの活動、糖データベース、蛋白立体構造のデータベースなどに拡大する目的で、1988年生体高分子データバンクの作業委員会が CODATA の総会で認められるに至った。ここでは核酸データベースが3極の世界的ネットワークを作り、それらと蛋白データベースの3極が対応する。これら6極間のデータベースでは、まず研究者からの直接入力統一フォームを作り、研究者からの直接入力を進めている。次に各3極間で学術雑誌の分担を決め、分担入力を開始して各データでの重複入力を避けるようにした。さらに GenBank と NBRF、EMBL と MIPS、DDBJ と JIPID は入力情報の交換を行って入力の省力化をはかることを開始した。

現在ではこれら6極間の Feature Table、雑誌の引用、用語と生物名の統一のための作業を開始している。

PIR-International のデータテープは南北米大陸は NBRF で、欧州では MIPS で、アジア・オセアニア地区では JIPID で配布することが決定し分担配布網が完成している。

現在までこれらの国際交流には日独交流基金(科技厅)、日米協力研究費(学振)が大きな役割

を果たしてきた。前者では1名が西独から、後者では4名が渡米4名が来日し細かい共同研究の打ち合わせを行って作業の活性化を行っている。

今後の課題と発展

1) データベース web の完成: 配列データベースの Fortran で書かれた従来のソフトを、すべて C 言語に変換しデータの相互の乗り入れ借用を可能にする。このシステムはデータのメモリー節約と重複入力を避ける上で重大な役割を持つ。

2) 他の生物活性、物性、変異データベースなどの入力を遂行し、また上述の配列データベースとの借用、pointer で相互に乗り入れの出来るシステムをより総合的なものとする。

3) 生物データベースに共通な用語の統一: 現在まで作られている各学問分野の統一用語を収集し、統一することを目標として国際的常設委員会が発足したが、この機関と積極的に交流して統一されたデータベースの用語作りをする¹⁷⁾。

4) 高次元構造データベースとの交流: NMR データベースは統一フォーマットの上で作られて上述の webs の中に入っているが、PDB、X線データベースとの交流を行い、一次構造、属性データと立体構造のデータが相互に交流できるようにする。PIR-International からこのための基金が出ている。

文 献

- 1) D. G. George, H. W. Mewes, and H. Kihara: A standardized format for sequence data exchange, *Protein Sequences and Data Analysis*, 1, 27-39 (1987).
- 2) C. S. Jone, A. Tsugita, T. Yagi, K. Imai and K. Takahashi: Development of several non-sequence biological databases, '89 International Protein Engineering Congress.
- 3) A. Tsugita: Current status of protein databanks, *J. Prot. Chem.*, 7, 316 (1988).
- 4) A. Tsugita: Is sequence data enough for protein engineering, *CODATA Bull.* 68, 39 (1988).
- 5) A. Tsugita, A. Ubasawa, C. S. Jone, M. Ikehara, D. G. George, L. -S. Yeh and H. R. Chen: An artificial variant database, '89 International Protein Engineering Congress.
- 6) A. Tsugita: in "Methods in Protein Sequence Analysis" ed. by B. Wittmann-Liebold, Cur-

- rent Status of Protein Data Banks, pp. 361-364 (1989), Springer Verlag.
- 7) A. Tsugita, C. S. Jone, A. Ubasawa, K. Satake and M. Ikehara: 蛋白質工学分野における Database, 第1回蛋白質工学年会 (1988).
 - 8) E. L. Ulrich, J. L. Markley and Y. Kyogoku: Creation of a nuclear magnetic resonance data repository and literature database, *Protein Sequences and Data Analysis*, 2, 23-37 (1989).
 - 9) T. Kunisawa, J. Otsuka, A. Tsugita, L.-S. Yeh, D. G. George and W. C. Barker: Genome database of *Escherichia coli*, '89 International Protein Engineering Congress.
 - 10) A. Tsugita: Protein sequence and data analysis, *Protein Sequences and Data Analysis*, 1, 1 (1987).
 - 11) A. Tsugita: Directory of protein sequence and nucleic acid data sources, CODATA Bull., Pergamon Press, 1-59 (1987).
 - 12) H. W. Mewes and A. Tsugita: Additional protein sequence databanks. Comput. Mol. Biol. Sources and Methods for Seq. Anal., ed. by A. Lesk, Oxford Press, 38-40 (1988).
 - 13) A. Tsugita: Nomenclature and Protein Database, The 1989 International Chemical Congress of Pacific Basin Society.
 - 14) H. W. Mewes, D. G. George, W. C. Barker and A. Tsugita: International Cooperation among Protein Sequence Data Banks, Methods in Protein Sequence Analysis, ed. by B. Wittmann-Liebold, pp. 357-360 (1989), Springer Verlag.
 - 15) A. Tsugita: Terminology and Nomenclature of Biology; History and Goal, 情報知識学会.
 - 16) A. Tsugita: Introduction to impact of knowledge of amino acid sequence data on biophysical research, Proc. 13th Taniguchi International Symposium on Biophysics, pp. III-XI, ed. by A. Tsugita (1987) (published 1989).
 - 17) A. Tsugita: 世界における蛋白質・核酸データベースの活動の現状, "Present status of Protein and Nucleic acid Database activities in the world", 情報管理, 37, 843-858 (1989).
 - 18) A. Tsugita: Rapporteur report on Biology Activity, CODATA Special Report, 11, 1-8 (1989).
 - 19) A. Tsugita: The database crisis; An Emerging Japanese Database's Problems and Solutions, Biomolecular Data: A resource in transition (ed. by R. Colwell, Oxford Sci. Pub., 91-95 (1989)).
 - 20) A. Tsugita: CODATA activity for Molecular Biology databases in Japan, CODATA Bull., in press.
 - 21) A. Tsugita: Are Sequence Data Enough for Protein Engineering? CODATA Bull., in press.
 - 22) A. Tsugita, A. Ubasawa, C. S. Jone, M. Ikehara, D. George, L.-S. Yeh and H. R. Chen: An artificial variant database, '89 International Protein Engineering Congress.
 - 23) A. Tsugita: Current Status of Protein Data Banks, CODATA Bull., in press.
 - 24) H. Suzuki, T. Kunisawa and J. Otsuka: Theoretical Evaluation of Transcriptional Pausing Effect on the Attenuation in trp Leader Sequence, *Biophysical Journal*, 49, 425-436 (1986).
 - 25) H. Suzuki, T. Kunisawa and J. Otsuka: Theoretical Analysis of Single-round Transcription Experiments on trp Leader Region, *Biophysical Journal*, 51, 407-412 (1987).
 - 26) T. Kunisawa, K. Horimoto and J. Otsuka: Accumulation Pattern of Amino Acid Substitutions in Protein Evolution, *Journal of Molecular Evolution*, 24, 357-365 (1987).
 - 27) T. Kunisawa and J. Otsuka: A Possible Mode of Protein Evolution: Role of the Anti-sense Strand in the Generation of New Proteins, *Protein Sequences and Data Analysis*, 1, 117-121 (1987).
 - 28) T. Kunisawa and J. Otsuka: Periodic Distribution of Homologous Genes or Gene Segments on the *Escherichia coli* K12 Genome, *Protein Sequences and Data Analysis*, 1, 263-267 (1988).
 - 29) J. Otsuka, H. Suzuki and T. Kunisawa: Regulation in the Synthesis of *Escherichia coli* RNA Polymerase, *Proposed from Sequence Analysis*, 1, 355-361 (1988).
 - 30) K. Horimoto, J. Otsuka and T. Kunisawa: Rapid Evolutionary Repair of Base Mismatches in Stem Regions of Eukaryotic 5S rRNA, *Protein Sequences and Data Analysis*, 2, 93-99 (1989).