JAERI-Data/Code
2004-001

バイオインフォマティクス整備
－データベース及びアプリケーションWeb実行システム－

2004年3月

木村　英雄・酒井　智

日本原子力研究所
Japan Atomic Energy Research Institute

本レポートは，日本原子力研究所が不定期に公刊している研究報告書です。
入手の問合わせは，日本原子力研究所研究情報部研究情報課（〒319-1195　茨城県那珂郡東海村）あて，お申し越しください。なお，このほかに財団法人原子力弘済会資料センター（〒319-1195　茨城県那珂郡東海村日本原子力研究所内）で複写による実費頒布をおこなっております。

This report is issued irregularly.
Inquiries about availability of the reports should be addressed to Research Information Division, Department of Intellectual Resources, Japan Atomic Energy Research Institute, Tokai-mura, Naka-gun, Ibaraki-ken, 319-1195, Japan.

編集兼発行　　日本原子力研究所

JAERI-Data/Code 2004-001

バイオインフォマティクス整備

ーデータベース及びアプリケーションWeb実行システムー

日本原子力研究所計算科学技術推進センター
木村 英雄・酒井 智*

(2004年1月27日受理)

現在、生物学の研究は、ゲノムとよばれる遺伝子のセットを用いて生命全体の働きを明らかにしていく段階に入っている。数千におよぶ遺伝子がもたらす生命の全体像を明らかにするには、遺伝子の相互作用から生み出される莫大な情報を解析しなくてはならない。今までの実験による解析方法では、この量の情報解析を現実的な時間で行うのは不可能である。そこで注目されているのが、コンピュータによる解析である。大量の遺伝子にもとづく解析や、タンパク質立体構造のシミュレーションなど、コンピュータによってのみ行える研究が必要とされている。バイオインフォマティクスと呼ばれる分野は、生物学と情報学とが融合した新しい分野であり、今急速に発展している。
本書では、バイオインフォマティクス分野の支援としてITBL利用推進室が行っている、データベースの整備、及びWeb経由でアプリケーションを実行するためのシステム(アプリケーション実行Webシステム)について報告する。

日本原子力研究所(関西駐在):〒619-0215 京都府相楽郡木津町梅美台8-1
*財団法人高度情報科学技術研究機構

JAERI-Data/Code 2004-001

Bioinformatics
- Date base and Application Execution Web System -

Hideo KIMURA and Tomo SAKAI*

Center for Promotion of Computational Science and Engineering
(Kansai Site)
Japan Atomic Energy Research Institute
Kizu-cho,Souraku-gun,Kyoto-fu

(Received January 27, 2004)

Research for the biological systems has reached to the stage that clarifies an organism as a whole from genome, a set of genes. To accomplish the researches, one needs to face a lot of data retrieved from genome sequences. Conventional methods, namely experiments in wet-labs, are, however not designed for dealing with genome scale data. For those data analyses, computer based researches, such as data mining and simulation, are suitable. As a result, bioinformatics, a new discipline combining expertise of biology and information science, is emerging. Here, we report a development of a data base and a web based system for application execution (Execution of Application on web system). Those are one of the efforts to support bioinformatics research by Office of ITBL Promotion.

Keywords: Bioinformatics, Genome

---

* Research Organization for Information Science & Technology

目次

Contents

1．はじめに

バイオインフォマティクスとよばれる分野は、生物学と情報が融合して成立した新分野である。本分野は成立してまだ間もないため、研究を行うためのコンピュータ環境整備を今後急速に進めていくことが求められている。このような背景の中、ITBL 利用推進室では、バイオインフォマティクス研究を推進していくための支援として、現在以下に挙げる２つのコンピュータ環境整備を行っている。

第１は、生物学汎用データベース(DB)の整備である。生物関連のデータは、Web で一般公開されているものが多い。これらのデータを集めた生物汎用 DB を構築し、利用しやすい環境を整備することが、バイオインフォマティクスの研究を行う上で必要となる。

第２は、バイオインフォマティクス研究で利用される代表的なアプリケーションの利用支援環境の構築である。今まで実験を基盤に研究を行っていた生物学研究者にとって、コンピュータ環境での研究を受け入れることは用意ではない。それは、実験を基盤にした研究とは異なった方法を新たに身につけていかなくてはならないため、高い障害を感じるためである。そのため、コンピュータ環境を利用した研究を推進していくためには、それに対応した支援が必要となる。

本報告書では、これら ITBL 利用推進室で行っているバイオインフォマティクス関連の整備について報告する。

2．データベース整備

2.1.背景

現在、数々の生物種ゲノム配列を読む作業が急速に進められおり、ゲノム配列のデータ量は膨大な数にのぼる。これらのデータすべてを個々の研究機関で管理することは、実質上不可能であるため、これらデータは世界の代表的な機関で一括して管理し、Web にて公開されている。研究を行うために必要なデータは、これら Web サイトからダウンロードすることで入手可能であるが、膨大なデータから必要なデータを探し出す作業は簡単ではない。できるだけ有効に必要なデータを取得し管理していくことが今求められている。

ITBL では、スーパーコンピュータを複数使用した仮想研究所を構築することにより、数多くの計算を実行するバイオインフォマティクス分野の研究を支えている。ITBL ユーザが更なる生物学研究を進めていくためには、ゲノム配列データを ITBL 環境にて有効に管理していく必要がある。ITBL 利用推進室では、公開されている数多くのデータベースのうち、比較的使用頻度の高いデータベースを取得し、できるだけ多くの研究者が利用できる環境を整備する。

2.2.概要

一般公開されているデータベースのうち、比較的頻繁に使用するデータベースを ITBL 環境へダウンロードする。一般公開されているデータベースは定期的に更新されていることより、定期的なダウンロードを実施し、常に最新状態に保つようにする。取得対象となるデータベースは以下の通りである。

- GenBank
- Swiss-Prot
- PDB(Protein Data Bank)
- NDB(Nucleic Acid Database)

これらのデータベースは、全世界の研究者が実験によって決定したデータが集約されている。世界各地から集められた膨大なデータは Web にて一般公開され、自由に閲覧することができるため、研究者が独自に決定した塩基配列データをこれらのデータベースと照合させることで、データの特徴を把握することができる。現在、次々と新しいデータが決定されており、データベースのデータ量は刻々と増え続けている。

2.3.対象データベース

2.3.1.GenBank

GenBank はアメリカの NCBI(National Center for Biotechnology Information)が管理している。GenBank には、全世界の研究者が実験によって決定した DNA、及び cDNA の塩基配列データが集約されいる。2002 年 8 月までに登録されたデータ数は約 18,197,000 entries であり、FTP 経由でダウンロードすることが可能である。

なお、GenBank は「GenBank/DDBJ/EMBL/ 国際塩基配列データベース*」を構築している三大国際 DNA データバンクのひとつである。

【参考 URL1】NCBI HOME：http://www.ncbi.nlm.nih.gov/

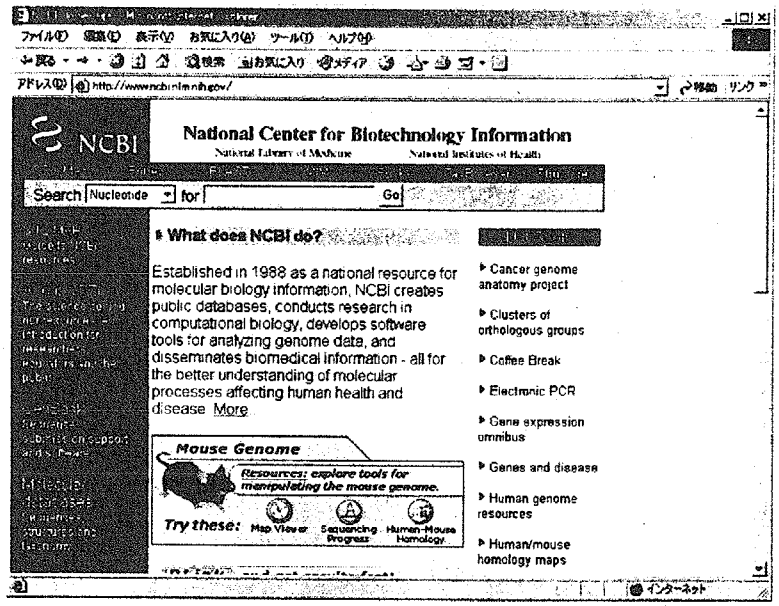

図 1 NCBI HOME

【参考 URL2】GenBank FTP SITE：ftp://ftp.ncbi.nih.gov/genbank/

なお、GenBank のサンプルデータを次頁図 2 に示す。

*世界の研究者が実験によって決定した DNA、あるいは cDNA の塩基配列データを GenBank、DDBJ、EMBL の三大データバンクが、三者間で定めたデータ構築規範に沿って収集・編集し、コンピュータファイルのかたちで提供するもの

```
LOCUS       CA997680              755 bp    mRNA    linear   EST 07-JAN-2003
DEFINITION  rf99h12.y1 Meloidogyne hapla J2 pAMP1 v1 Meloidogyne hapla cDNA 5'
            similar to SW:UN87_CAEEL P37806 UNC-87 PROTEIN. [3] SW:UN87_CAEEL
            TR:Q9TYS6 ;, mRNA sequence.
ACCESSION   CA997680
VERSION     CA997680.1  GI:27542551
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Claire Murphy and Dr. James McCarter
            at Washington University, St. Louis. J2 were provided by Dr.
            Valerie Williamson of the University of California at Davis
            (vmwilliamson@ucdavis.edu).
            Seq primer: -40RP from Gibco
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1..755
                     /organism="Meloidogyne hapla"
                     /db_xref="taxon:6305"
                     /clone_lib="Meloidogyne hapla J2 pAMP1 v1"
                     /dev_stage="J2"
                     /lab_host="DH10B"
                     /note="Vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
                     The library was constructed by Claire Murphy and Dr. James
                     McCarter at Washington University, St. Louis. The cDNA was
                     made by using Dynabead oligo-dT priming (Dynal). PCR based
                     library using a modified protocol from the SMART PCR cDNA
                     Synthesis Kit from Clontech. Directionally cloned into the
                     UDG sites of pAMP1. J2 were provided by Dr. Valerie
                     Williamson of the University of California at Davis
                     (vmwilliamson@ucdavis.edu)."
BASE COUNT     258 a    139 c    164 g    194 t
ORIGIN
        1 cttataagga atgaccagct tcggtaccaa ccgtcgtgaa acaacaaaaa tgttggatac
       61 agctcacccg gaatattctc acgaaagcag catcgatcaa acgagcattc cataccaaat
      121 gggatcaaac cgttatgctt cacaaaaggg aatgacttgt tttggacagc cacgttggga
      181 ggtgcttgac ccaagtatta gctaccagaa ccgtaaatca caaggaatgg tccgtctcca
      241 atctggtaca aaccggttcg cctctcaagc aggcatgaca ggtttcggaa ctccaaggaa
      301 cactacatac gaggcagagt ctggcgaact tccatatgaa gatatgaaga gtcagaaacg
      361 attatcccat gacaggctgg ttggaataag ggagactctc aaaagttgat gactggattt
      421 ggtactcctc gtgacgtgaa aggcaaacat ttgaagcgta tttgggagtt ggaatatcca
      481 gaggaggctg aaatctcttt ggatcgactt taaagaattt tggagagata agaaggagga
      541 aagaaataaa tagtggaaag gaaggcaacg acattttgac aaaattttgg cacacatttt
      601 cttctcattt aaatacattt ttggcaaaaa aataattttt ttggcttttt tggctccatt
      661 attctctccc tattctctca tttggacaaa catttggaca aaatttaact tggatgctta
      721 aataaaagaa aaggatgaga gaaaaataaa ataaa
//
```

図 2 GenBank サンプルデータ

2.3.2.Swiss-Prot

Swiss-Prot はスイスの SIB(Swiss Institute for Bioinformatics)とイギリスの EBI(European Bioinformatics Institute)が共同で管理している。Swiss-Prot には、全世界の研究者が実験によって決定したタンパク質のデータが集約される。2002/12 までに登録されたデータ数は 120,606 entries である。

【参考 URL1】Swiss-Prot : http://www.ebi.ac.uk/swissprot/

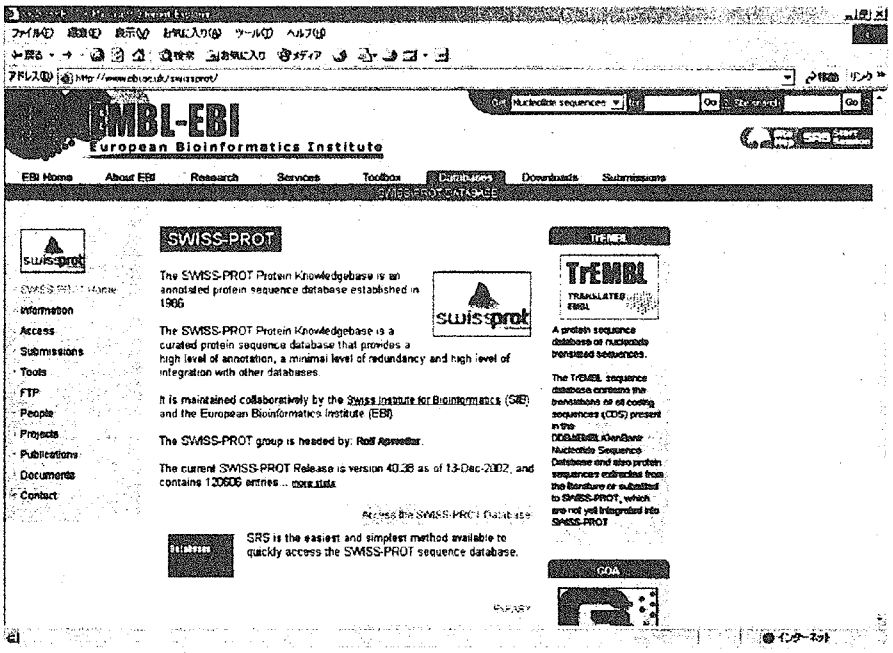

図 3 Swiss-Prot

【参考 URL2】Swiss-Prot FTP SITE : ftp://ftp.ebi.ac.uk/pub/databases/swissprot/

なお、Swiss-Prot のサンプルデータを次頁図 4 に示す。

```
ID   12S1_ARATH     STANDARD;      PRT;   472 AA.
AC   P15455; Q9FFH7;
DT   01-APR-1990 (Rel. 14, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   12S seed storage protein precursor.
GN   CRA1 OR AT5G44120 OR MLN1.4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA.
RA   Pang P.P., Pruitt R.E., Meyerowitz E.M.;
RT   "Molecular cloning, genome organization, expression and evolution of
RT   12S seed storage protein genes of Arabidopsis thaliana.";
RL   Plant Mol. Biol. 11:805-820(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=97471969; PubMed=9330910;
RA   Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA   Miyajima N., Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT   features of the 1.6 Mb regions covered by twenty physically assigned
RT   P1 clones.";
RL   DNA Res. 4:215-230(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   SSP consortium (Salk/Stanford/PGEC).";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 420-472 FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS IS A SEED STORAGE PROTEIN.
CC   -!- SUBUNIT: HEXAMER. EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC       BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC       DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEIN (GLOBULINS)
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37247; AAA32777.1; -.
DR   EMBL; X14312; CAA32493.1; -.
DR   EMBL; AB005239; BAB10979.1; -.
DR   EMBL; AY070730; AAL50071.1; -.
DR   EMBL; Z17590; CAA79005.1; -.
DR   PIR; S08509; S08509.
DR   InterPro; IPR000459; Seedstore_11s.
DR   Pfam; PF00190; Seedstore_11s; 1.
DR   PRINTS; PR00439; 11SGLOBULIN.
DR   PROSITE; PS00305; 11S_SEED_STORAGE; 1.
KW   Seed storage protein; Multigene family; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    282       ACIDIC CHAIN (BY SIMILARITY).
FT   CHAIN      283    472       BASIC CHAIN (BY SIMILARITY).
FT   DISULFID   112    289       INTERCHAIN (ALPHA-BETA) (POTENTIAL).
FT   CONFLICT   167    167       E -> Q (IN REF. 1).
FT   CONFLICT   356    356       V -> E (IN REF. 1).
SQ   SEQUENCE   472 AA;  52595 MW;  700B468E4D251994 CRC64;
     MARVSSLLSF CLTLLILFHG YAAQGGQGGQ QFPNECQLDQ LNALEPSHVL KSEAGRIEVW
     DHHAPQLRCS GVSFARYIIE SKGLYLPSFF NTAKLSFVAK GRGLMGKVIP GCAETFQDSS
     EFQPRFEGQG QSQRFRDMHQ KVEHIRSGDT IATTPGVAQW FYNDGQEPLV IVSVFDLASH
     QNQLDRNPRP FYLAGNNPQG QVWLQGREQQ PQKNIFNGFG PEVIAQALKI DLQTAQQLQN
     QDDNRGNIVR VQGPFGVIRP PLRGQRPQEE EEEEGRHGRH GNGLEETICS ARCTDNLDDP
     SRADVYKPQL GYISTLNSYD LPILRFIRLS ALRGSIRQNA MVLPQWNANA NAILYVTDGE
     AQIQIVNDNG NRVFDGQVSQ GQLIAVPQGF SVVKRATSNR FQWVEFKTNA NAQINTLAGR
     TSVLRGLPLE VITNGFQISP EEARRVKFNT LETTLTHSSG PASYGRPRVA AA
//
```

図 4 Swiss-Prot サンプルデータ

2.3.3. PDB(Protein Data Bank)

PDB はアメリカの RCSB(Research Collaboratory for Structural Bioinformatics) consortium の以下の 3 機関で管理している。PDB には、全世界の研究者が実験によって決定した生体高分子の立体構造座標に関するデータが集約される。

- Rutgers,the State University of New Jersey
- SDSC(San Diego Supercomputer Center)
- CARB(Center for Advanced Research in Biotechnology)

【参考 URL1】PDB：http://www.rcsb.org/pdb/index.html

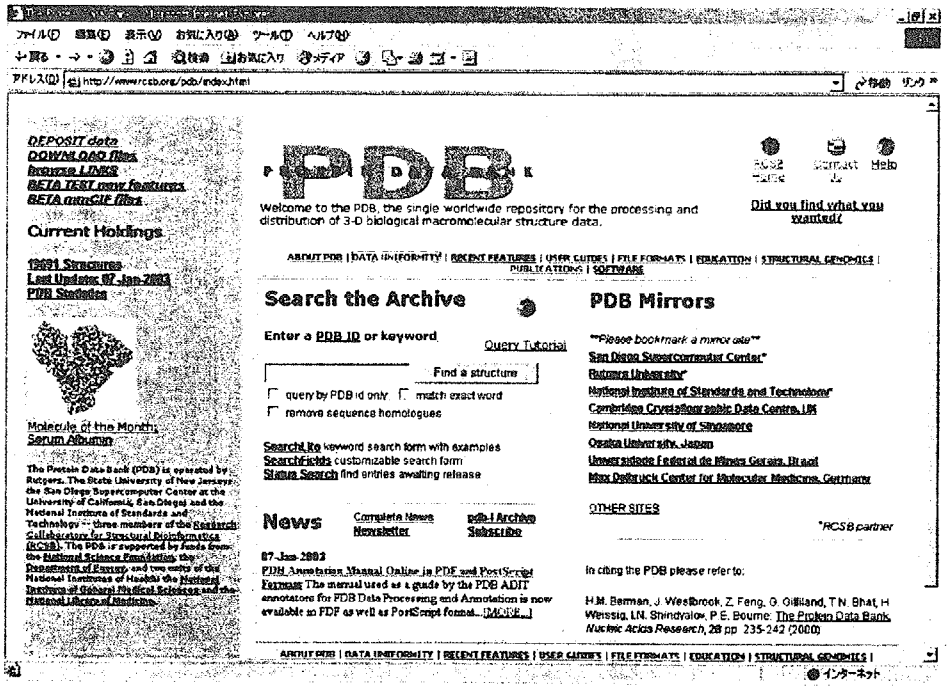

図 5 PDB

【参考 URL2】PDB FTP SITE：ftp://ftp.rcsb.org/pub/pdb/

なお、PDB のサンプルデータを次頁図 6 に示す。

```
HEADER    BLOOD CLOTTING                          28-AUG-01   1JV8
TITLE     NMR STRUCTURE OF BPTI MUTANT G37A
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: TRYPSIN INHIBITOR;
COMPND    3 CHAIN: A;
COMPND    4 ENGINEERED: YES;
COMPND    5 MUTATION: YES
SOURCE    MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE    3 ORGANISM_COMMON: BOVINE;
SOURCE    4 ORGAN: PANCREAS;
SOURCE    5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE    6 EXPRESSION_SYSTEM_COMMON: BACTERIA;
SOURCE    7 EXPRESSION_SYSTEM_STRAIN: HB101;
SOURCE    8 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE    9 EXPRESSION_SYSTEM_PLASMID: PTI103
KEYWDS    BPTI, G37A MUTANT, CONFORMATIONAL STRAIN
EXPDTA    NMR, 23 STRUCTURES
AUTHOR    J.L.BATTISTE,R.LI,C.WOODWARD
REVDAT    2   27-FEB-02 1JV8    1       JRNL   REMARK
REVDAT    1   12-SEP-01 1JV8    0
JRNL      AUTH   J.L.BATTISTE,R.LI,C.WOODWARD
JRNL      TITL   A HIGHLY DESTABILIZING MUTATION, G37A, OF THE
JRNL      TITL 2 BOVINE PANCREATIC TRYPSIN INHIBITOR RETAINS THE
JRNL      TITL 3 AVERAGE NATIVE CONFORMATION BUT GREATLY INCREASES
JRNL      TITL 4 LOCAL FLEXIBILITY.
JRNL      REF    BIOCHEMISTRY                V. 41  2237 2002
JRNL      REFN   ASTM BICHAW US ISSN 0006-2960
REMARK    1
REMARK    1 REFERENCE 1
REMARK    1  AUTH   R.LI,J.L.BATTISTE,C.WOODWARD
REMARK    1  TITL   NATIVE-LIKE INTERACTIONS FAVORED IN THE UNFOLDED
REMARK    1  TITL 2 BOVINE PANCREATIC TRYPSIN INHIBITOR HAVE DIFFERENT
REMARK    1  TITL 3 ROLES IN FOLDING.
REMARK    1  REF    BIOCHEMISTRY                V. 41  2246 2002
REMARK    1  REFN   ASTM BICHAW US ISSN 0006-2960
                                        .
                                        .
                                        .
                                        .
                                        .
                                        .
ATOM    880  C   GLY A  57      14.764   1.851  -0.267  1.00  0.00           C
ATOM    881  O   GLY A  57      15.448   2.414   0.566  1.00  0.00           O
ATOM    882  H   GLY A  57      13.268   1.427   2.244  1.00  0.00           H
ATOM    883 1HA  GLY A  57      13.563   0.142   0.215  1.00  0.00           H
ATOM    884 2HA  GLY A  57      12.704   1.400  -0.673  1.00  0.00           H
ATOM    885  N   ALA A  58      15.141   1.774  -1.513  1.00  0.00           N
ATOM    886  CA  ALA A  58      16.434   2.381  -1.941  1.00  0.00           C
ATOM    887  C   ALA A  58      17.594   1.766  -1.155  1.00  0.00           C
ATOM    888  O   ALA A  58      17.739   2.100   0.009  1.00  0.00           O
ATOM    889  CB  ALA A  58      16.294   3.869  -1.618  1.00  0.00           C
ATOM    890  OXT ALA A  58      18.319   0.971  -1.731  1.00  0.00           O
ATOM    891  H   ALA A  58      14.574   1.315  -2.169  1.00  0.00           H
ATOM    892  HA  ALA A  58      16.582   2.247  -3.001  1.00  0.00           H
ATOM    893 1HB  ALA A  58      16.725   4.453  -2.418  1.00  0.00           H
ATOM    894 2HB  ALA A  58      16.809   4.087  -0.694  1.00  0.00           H
ATOM    895 3HB  ALA A  58      15.248   4.117  -1.515  1.00  0.00           H
TER     896      ALA A  58
ENDMDL
CONECT   78  866
CONECT  206  593
CONECT  478  801
CONECT  593  206
CONECT  801  478
CONECT  866   78
MASTER      79    0    0    1    2    0    0  620585  23    6    5
END
```

図 6 PDB サンプルデータ

2.3.4. NDB(Nucleic Acid Database)

NDB は PDB のデータのうち、核酸データ、及び核酸と相互作用するタンパク質のデータを集めたデータベースである。したがって、PDB の部分集合データベースとなる。

【参考 URL1】NDB：http://ndbserver.rutgers.edu/NDB/index.html

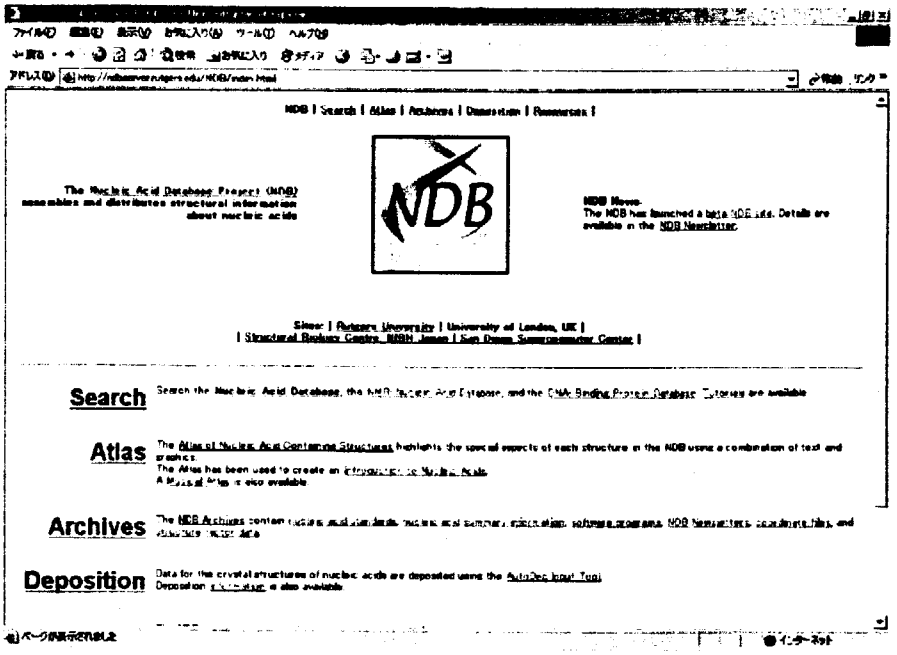

図 7 NDB

【参考 URL2】NDB FTP SITE：ftp://beta-ndb.rutgers.edu/NDB/

なお、NDB のサンプルデータを次頁図 8 に示す。

```
HEADER    DEOXYRIBONUCLEIC ACID                   04-FEB-97   311D
TITLE     THE ROLE OF HYDROGEN BONDING IN MINOR-GROOVE DRUG-DNA
TITLE    2 RECOGNITION. STRUCTURE OF A BIS-AMIDINIUM DERIVATIVE OF
TITLE    3 HOECHST 33258 COMPLEXED TO THE DODECANUCLEOTIDE
TITLE    4 D(CGCGAATTCGCG)2
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: DNA (5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-
COMPND   3 3');
COMPND   4 CHAIN: A, B;
COMPND   5 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 SYNTHETIC: YES
KEYWDS    B-DNA, DOUBLE HELIX, COMPLEXED WITH DRUG
EXPDTA    X-RAY DIFFRACTION
AUTHOR    G.R.CLARK,D.W.BOYKIN,A.CZARNY,S.NEIDLE
REVDAT   2  27-NOV-00 311D   5
REVDAT   1  12-MAR-97 311D   0
JRNL      AUTH   G.R.CLARK,D.W.BOYKIN,A.CZARNY,S.NEIDLE
JRNL      TITL   STRUCTURE OF A BIS-AMIDINIUM DERIVATIVE OF HOECHST
JRNL      TITL 2 33258 COMPLEXED TO DODECANUCLEOTIDE
JRNL      TITL 3 D(CGCGAATTCGCG)2: THE ROLE OF HYDROGEN BONDING IN
JRNL      TITL 4 MINOR GROOVE DRUG-DNA RECOGNITION
JRNL      REF    NUCLEIC ACIDS RES.          V. 25  1510 1997
JRNL      REFN   ASTM NARHAD  UK ISSN 0305-1048
REMARK   1
REMARK   2
REMARK   2 RESOLUTION. 2.2 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : X-PLOR

                                        .
                                        .
                                        .
                                        .
                                        .
                                        .

CONECT  492  491  493  497
CONECT  493  492  494
CONECT  494  493  495  498
CONECT  495  494  496  500
CONECT  496  495  497
CONECT  497  492  496
CONECT  498  494  499
CONECT  499  498  500  501
CONECT  500  495  499
CONECT  501  499  502  506
CONECT  502  501  503
CONECT  503  502  504  507
CONECT  504  503  505  509
CONECT  505  504  506
CONECT  506  501  505
CONECT  507  503  508
CONECT  508  507  509  510
CONECT  509  504  508
CONECT  510  508  511  515
CONECT  511  510  512
CONECT  512  511  513
CONECT  513  512  514  516
CONECT  514  513  515
CONECT  515  510  514
CONECT  516  513  517  518
CONECT  517  516
CONECT  518  516
MASTER     210    0    1    0    0    0    0    6  593    2   30    2
END
```

図 8 NDB サンプルデータ

2.4.環境

来年度よりITBL計算機のアプリケーションサーバにデータベース環境が用意されるが、現在は教育用PCクラスタでデータベース環境を試験的に構築している。従って、以降に記述するデータはすべて、教育用PCクラスタでデータベースを構築した場合のものである。

なお、教育用PCクラスタのマシン構成を表 1に示す。

表 1 教育用PCクラスタ構成

制御ノード	本体	Compaq ProliantDL380
	CPU	PentiumIII 1.13GHz × 1
	メモリ	1GB
	HDD	72GB × 6 (RAID5後実容量330GB)
	LAN	100Base-TX × 2 1000Base-SX x1ポート Myrinetボード × 3
	DAT	COMPAC 20/40GB DAT
	CDRW	CDRW-SX24B
計算ノード	本体	Compaq ProliantDL360
	CPU	PentiumIII 1GHz × 2
	メモリ	1GB
	ディスク容量	36GB
	LAN	100Base-TX × 2 Myrinetボード × 1

図 9 教育用PCクラスタ概観

2.5.取得対象データ

各データベースのデータは、テキスト形式のファイル(以下フラットファイル)で管理されている。そのため、データを取得する場合は、各データベースの FTP サイトへアクセスし、フラットファイルをすべて FTP 経由でダウンロードする。

各データベースのデータ量は多く、全データをダウンロードするためには相当な容量のディスクが必要となるため、比較的利用率の高いデータにしぼってダウンロードを行う。ダウンロード対象となるデータについて表 2 に示す。なお、取得先(FTP)とは、オリジナル FTP サイト内における取得対象データの保存場所を指す。

表 2 取得対象データ保存場所

データベース名	取得先(FTP)
GenBank	ftp:// ftp.ncbi.nih.gov/genbank/*.seq.gz
Swiss-Prot	ftp:// ftp.ebi.ac.uk/pub/databases/swissprot/release/*.dat
PDB	ftp:// ftp.rcsb.org/pub/pdb/data/structures/divided/pdb/*
NDB	ftp:// beta-ndb.rutgers.edu//NDB/coordinates/*

また、ダウンロードするデータの容量、及びダウンロードの所要時間について表 3 に示す。

表 3 データ容量、及びダウンロード所要時間

データベース名	容量(GB)	所要時間(h)
GenBank	12.6	13.5
Swiss-Prot	0.305	0.5
PDB	2.8	8.8
NDB	1.1	3.7

(2002/12/21 時点)

GenBank、PDB、NDB についてはダウンロード対象データは圧縮されている。従ってこれら 3 つのデータベースについては、ダウンロード後に全データを解凍しなければならない。解凍後の容量を表 4 に示す。

表 4 データ容量(解凍後)

データベース名	容量(GB)
GenBank	79.6
PDB	10.9
NDB	2.4

(2002/12/21 時点)

2.6.データ更新

各データベースでは、新しいデータが次々と登録されているため、管理サイトでは定期的にデータを更新している。したがって、常に最新のデータを整備するために、管理サイトでデータベースが更新されるタイミングで、逐次新しいデータをダウンロードする。

2.7.ダウンロード処理

2.7.1.処理の流れ

データをダウンロードする処理の流れは以下の通り。

① FTP サイトへ接続し、対象データをダウンロードする。

② 実際に使用するデータを格納するディレクトリへ、ダウンロードしたデータをコピーする。

(ダウンロードしたデータと、実際に使用するデータは区別して管理するため)

③ ダウンロードしたデータが圧縮されている場合は、データを解凍する。

データを FTP 経由でダウンロードする場合は、シェルスクリプト(以下、ダウンロードシェルスクリプト)を使用する。ダウンロードシェルスクリプトで行う処理は、データベースに特化した情報(FTP サイト情報等)以外は、どのデータベースにおいてもほぼ同一である。従って、以降 PDB を代表としてシェルスクリプトについて解説する。

2.7.2.ダウンロードシェルスクリプト

ダウンロードシェルスクリプトは、以下のスクリプトから構成される。

- **get_pdb.sh**
  オリジナル FTP サイトのデータ取得先より、データをダウンロードする。
- **cp_pdb.sh**
  実際に使用するデータを格納するディレクトリへ、ダウンロードしたデータをコピーする。
- **unzip_pdb.sh**
  ダウンロードしたデータは圧縮されているため、全データを解凍する。
- **pdb.sh**
  get_pdb.sh、cp_pdb.sh、unzip_pdb.sh を使用して、ダウンロード全体の流れを統制する。
- **new2pdb.sh**
  1 世代古いデータ、及び実際に使用するデータを更新する。

ダウンロードシェルスクリプトを使用した、ダウンロード処理の流れを図 10 に示す。

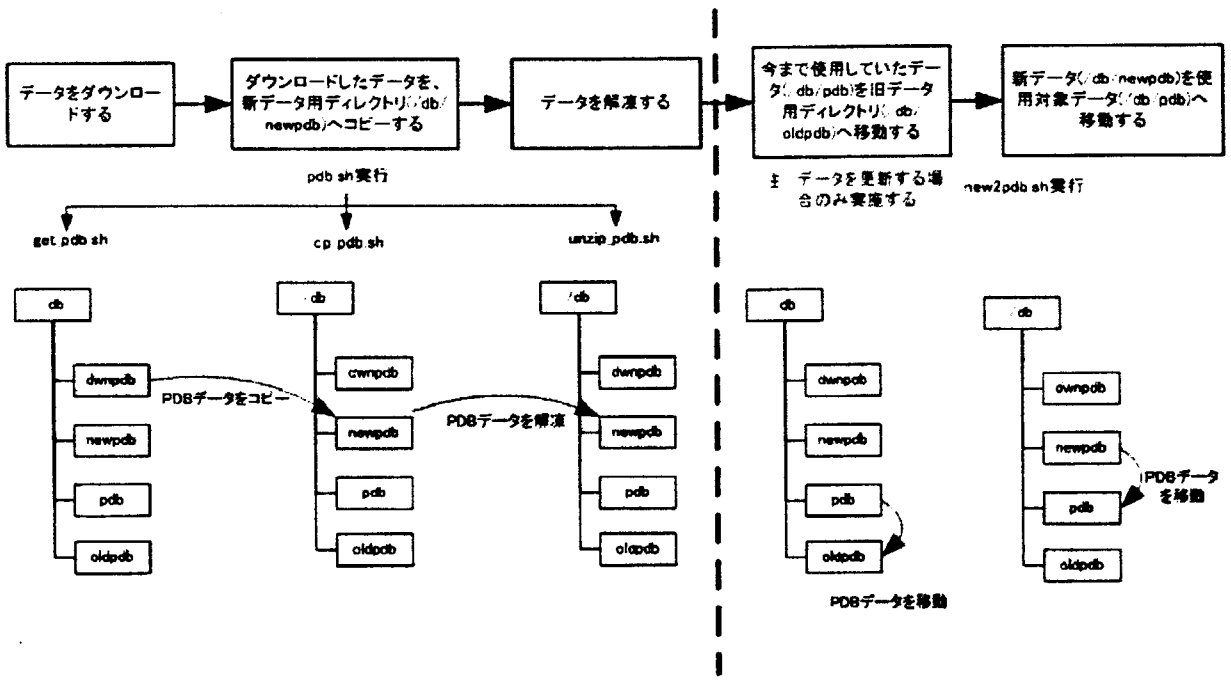

図 10 ダウンロード処理の流れ

2.7.3.プログラム

2.7.3.1. get_pdb.sh

```
#!/bin/sh
#
# get_pdb.sh:
#       Download files (called from pdb.sh)
#
# Author:Sakai Tomo
#
# Version History
#    1.0.0          : 2002/06/17 : first release by Sakai Tomo
#

echo "Start get_pdb.sh:"`date`

cd $DLBASE
find . -type d -print -exec chmod u+w {} ¥;
cat <<"EOF" | lftp
open ftp.rcsb.org
user anonymous sakait@apr.jaeri.go.jp
cd /pub/pdb/data/structures/divided/pdb
ls -ltr
mirror -c -s -e --parallel=1 -v
EOF

echo "End get_pdb.sh:"`date`
```

2.7.3.2. cp_pdb.sh

```
#!/bin/bash
# Copy PDB data to PDB root directory
# This shell has two arguments
# First argument is, for example "/db/dwnpdb/data/structures/divided/pdb"
# Second argument is, for example "/db/newpdb"

echo "Start cp_pdb.sh:"`date`

# Define parameters
dir=$1
cpdir=$2
#dir="/db/dwnpdb/data/structures/divided/pdb"
#cpdir="/db/newpdb"
dirlist="/tmp/dir_list"
filelist="/tmp/file_list"

# Coping PDB data
cd $dir
echo Curent Directory = $dir
echo
find . -type d -print > $dirlist

cur=.

while read pdb_dir
do
        echo pdb_dir = [ $pdb_dir ]
        if [ ! $pdb_dir = $cur ] && [ ! $pdb_dir = "nonCIF" ]; then
                echo Now copy files in [ $pdb_dir ] directory to $cpdir
                cd $pdb_dir
                find . -type f -print > $filelist
                while read file_name
                do
                        cp $file_name $cpdir 2>> $LOGBASE/$CPERROR$DATE.log
```

```
                done < $filelist
                echo Finish copying files from [ $pdb_dir ] directory to $cpdir
                cd ..
        fi
done < $dirlist

echo All files are copied
echo

# Remove last dir_list and file_list
rm -f $dirlist
rm -f $filelist

echo "End cp_pdb.sh:"`date`
```

2.7.3.3. unzip_pdb.sh

```
#!/bin/bash
# gzip files in /db/newpdb directory

echo "Start unzip_pdb.sh:"`date`

# Define parameters
dir="/db/newpdb"
dir_str="/db/newpdb/structure_factors"
#log="/var/log/dataget/pdb/unzip/pdb_unzip_"$DATE.log

cd $dir

for i in 0 1 2 3 4 5 6 7 8 9
do
        if [ $i -eq 1 ]
        then
                for j in 0 1 2 3 4 5 6 7 8 9 a b c d e f g h i j k l m n o p q r s t u v w x y z
                do
                        pdbfile="pdb"$i$j"*"
                        echo "gzip of pdb"$i$j" files"
                        gzip -d $pdbfile 2>> $LOGBASE/$UNZIPERROR$DATE.log
                done
        else
                pdbfile="pdb"$i"*"
                echo "gzip of pdb"$i" files"
                gzip -d $pdbfile 2>> $LOGBASE/$UNZIPERROR$DATE.log
        fi
done

echo "End unzip_pdb.sh:"`date`
```

2.7.3.4. pdb.sh

```
#!/bin/sh
#
# pdb.sh:
#       Get public pdb db (from ftp.rcsb.org/pub/pdb/data/structures/divided/pdb)
#
# Author:Sakai Tomo
#
# Version History
#   1.0.0       : 2002/06/17 : first release by Sakai Tomo
#   1.0.1       : 2002/10/4  : cope with structure_factors data
#


#
# set common variable
#   modify when you download from other sites
#
#   BINDIR    : Program installed directory
#   LOGBASE   : log file directory
#
#   DLSITE    : FTP site name
#   SITEDIR   : FTP site download directory       (files FROM)
#   DLBASE    : local machine download directory (files TO)
#
BINDIR=/usr/local/dataget/pdb/bin
LOGBASE=/var/log/dataget/pdb

DLSITE=ftp.rcsb.org
#DLSITE=198.202.75.77
SITEDIR=/pub/pdb/data/structures/divided/pdb
DLBASE=/db/dwnpdb/data/structures/divided/pdb
NEWDATA=/db/newpdb

LOCKFILE=/tmp/pdb.lck

export BINDIR LOGBASE
```

```
export DLSITE SITEDIR DLBASE NEWDATA
export LOCKFILE

#
# set log file name from current date
#
YEAR=`date '+%Y'`
MON=`date '+%m'`
DAY=`date '+%d'`
HOUR=`date '+%H'`
MIN=`date '+%M'`
DATE=$YEAR$MON$DAY$HOUR$MIN ; export DATE

#
# Set log file name
#
OUTLINELOG=outline/pdb_outline_
DLLOG=dl/pdb_dl_
FMTLOG=fmt/pdb_fmt_
CPLOG=cp/pdb_cp_
CPERROR=cp/error/pdb_cp_error_
UNZIPLOG=unzip/pdb_unzip_
UNZIPERROR=unzip/error/pdb_unzip_error_

export OUTLINELOG DLLOG FMTLOG CPLOG CPERROR UNZIPLOG
UNZIPERROR

########################################
# Start Process
########################################

#
# Trap interupt
#    cannot catch SIGKILL(9)
#
trapclean() {
```

```
    rm -f $LOCKFILE
    exit 1
}
trap trapclean 1 2 13 15

#
# Check if another download shells are running
#
if [ -f $LOCKFILE ]; then
    echo "Another download shell is running. exit." >
$LOGBASE/$OUTLINELOG$DATE.log
    exit
fi
echo "Start pdb.sh:"`date` > $LOGBASE/$OUTLINELOG$DATE.log

# create lock file
touch $LOCKFILE

#
# Check if specific directory exists
#
if [ ! -d $DLBASE ]; then
    echo "[ERROR]:$DLBASE not exist. End script." >>
$LOGBASE/$OUTLINELOG$DATE.log
    exit
fi

#
# Download PDB Data
#
$BINDIR/get_pdb.sh > $LOGBASE/$DLLOG$DATE.log
echo "get_pdb.sh is done. logfile is $LOGBASE/$DLLOG$DATE.log" >>
$LOGBASE/$OUTLINELOG$DATE.log

#
# Copy downlod files(pdb) to /db/newpdb
```

```
#
$BINDIR/cp_pdb.sh $DLBASE $NEWDATA > $LOGBASE/$CPLOG$DATE.log
echo "cp_pdb.sh is done. logfile is $LOGBASE/$CPLOG$DATE.log" >>
$LOGBASE/$OUTLINELOG$DATE.log

#
# Unzip
#
$BINDIR/unzip_pdb.sh > $LOGBASE/$UNZIPLOG$DATE.log
echo "unzip_pdb.sh is done. logfile is $LOGBASE/$UNZIPLOG$DATE.log" >>
$LOGBASE/$OUTLINELOG$DATE.log

#
# Send mail to report termination
#
$BINDIR/mail.sh >> $LOGBASE/$OUTLINELOG$DATE.log

#
# Remove locked files
#
rm -f $LOCKFILE
echo "End pdb.sh:"`date` >> $LOGBASE/$OUTLINELOG$DATE.log
```

2.7.3.5. new2pdb.sh

```
#!/bin/bash
# Move /db/pdb datas to /db/oldpdb
# Move /db/newpdb datas to /db/pdb
# Make /db/newpdb directory

#
# set log file name from current date
#
YEAR=`date '+%Y'`
MON=`date '+%m'`
DAY=`date '+%d'`
HOUR=`date '+%H'`
MIN=`date '+%M'`
DATE=$YEAR$MON$DAY$HOUR$MIN ; export DATE

LOGDIR=/var/log/dataget/pdb/new2pdb
LOG=$LOGDIR"/new2pdb_$DATE.log"
ERROR=$LOGDIR"/error/error_new2pdb_$DATE.log"

# set common variable
OLDPDB="/db/oldpdb"
PDB="/db/pdb"
NEWPDB="/db/newpdb"

# Address to reporting the status
TO="sakait@apr.jaeri.go.jp to-mo@muc.biglobe.ne.jp"

echo "Start new2pdb.sh:"`date` > $LOG
rm -rf $OLDPDB 2> $ERROR
echo "Remove "$OLDPDB >> $LOG
mv $PDB $OLDPDB 2>> $ERROR
echo "Move to "$OLDPDB" from "$PDB >> $LOG
mv $NEWPDB $PDB 2>> $ERROR
echo "Move to "$PDB" from "$NEWPDB >> $LOG
mkdir $NEWPDB 2>> $ERROR
```

```
echo "Make directory "$NEWPDB >> $LOG

cat <<"EOF" | sendmail $TO
Finish Update PDB data and old PDB data (^o^)v
.
EOF

echo "Sned mail of termination to "$TO >> $LOG
echo "End new2pdb.sh:"`date` >> $LOG
```

2.7.4.今後の展開

来年度(平成15年度)より、ITBL計算機のアプリケーションサーバにデータベース環境が整うため、今後は整備環境を教育用PCクラスからITBL計算機のアプリケーションサーバへ移行する。現状GenBank、Swiss-Prot、PDB、NDBを整備対象データベースとしているが、必要に応じ対象データベースを増やしていく必要がある。また、データを検索し取得するITBL独自の環境を整備し、更なる利便性の向上を目指す。

3．アプリケーション実行 Web システム

3.1.背景

バイオインフォマティクスの分野で使用するアプリケーションは数多く存在する。これらのアプリケーションの多くは Web にて公開され、比較的簡単に入手できるものが多い。しかし、これらのアプリケーションの使用方法は、独自に習得しなければならないのが現状である。生物学の研究を行う研究者にとって、アプリケーションの使用方法を習得するために多くの時間を費やすことができないため、アプリケーションの使用方法の習得を比較的簡単に行うための支援が求められている。

3.2.目的

現在、アプリケーションを実行するための環境を用意している Web サイトはいくつか存在する。コマンドによる実行を行うアプリケーションに対し、入力支援を行う Web 画面を用意し、Web 経由でアプリケーションを実行するサイトである。これらの Web サイトを利用することで、ユーザは手軽にアプリケーションを利用することができる。また、使用方法がわからない場合は、付属のヘルプ画面を参照するなどして、利用することもできる。ここで問題になるのが、実際の研究とアプリケーションの関連付けである。研究者が行っている研究に対し、アプリケーションをどのように使用すれば、新しい成果を得ることができるのかについては、研究内容が多種多様であるため、一概に決めることができない。しかし、汎用的な使用方法を抽出し、Web にて公開することで、自身の研究への応用が比較的簡単になることが考えられる。

本システムは、Web サイトで比較的簡単に研究への応用方法を解説することで、アプリケーションの使用方法を習得するための支援を行うことを目的としている。

【参考】バイオインフォマティクス関連の情報を提供している既存 Web サイト(一例)

- ゲノムネット
  URL：http://www.genome.ad.jp/Japanese/
- DNA Data Bank of Japan(DDBJ)
  URL: http://www.ddbj.nig.ac.jp/Welcome-j.html
- JBiC
  URL: http://www.jbic.or.jp/katudo/jigyo.html
- NCBI
  URL：http://www.ncbi.nlm.nih.gov/
- RCSB
  URL：http://www.rcsb.org/index.html
- EBI
  URL：http://www.ebi.ac.uk/

3.3.前提

本システムは、本報告書を執筆する段階において開発中であるため、現地点で明確になっている事項について報告するものとする。

3.4.利用形態

ユーザの端末よりブラウザを使用し、本システムの Web サイトへアクセスする。ユーザは使用方法を習得したいアプリケーションを選択し、Web ページの解説を閲覧する。必要に応じ、実際にパラメータを入力し、Web 経由でアプリケーションを実行する。実行結果は Web で表示されるため、逐次ブラウザで確認することが可能である。
利用形態概念図を図 11 に示す。

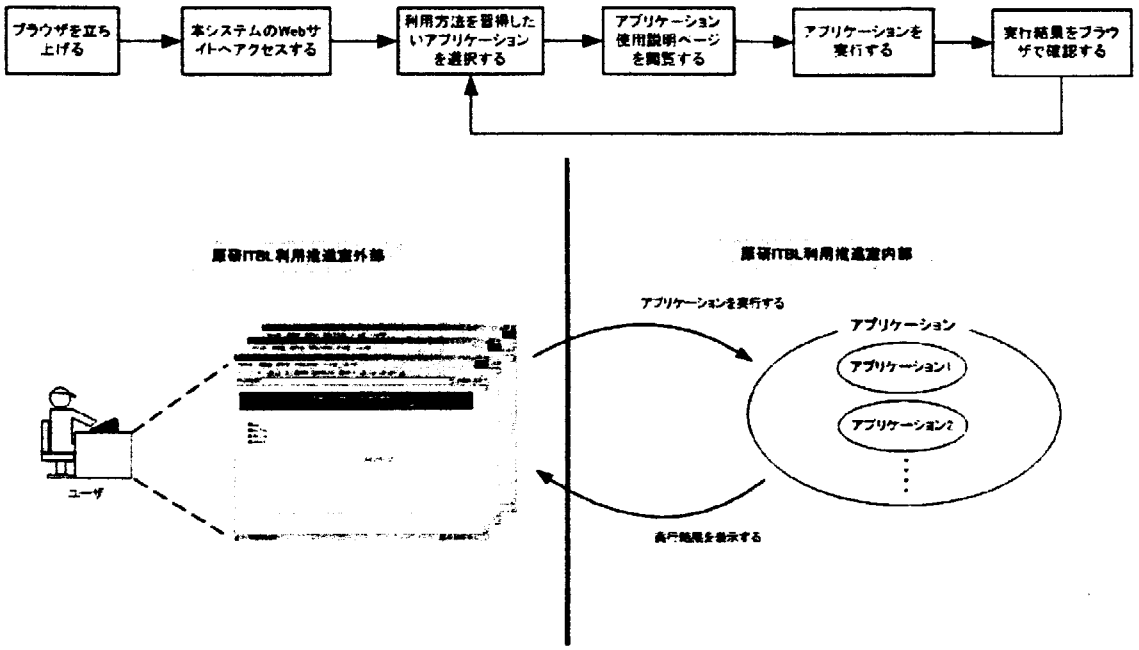

図 11 利用形態概念図

3.5.環境

3.5.1.Web サーバ

■役割

ユーザに対し、アプリケーションを使用するためのインターフェースを提供する。
ユーザはブラウザで Web サーバへアクセスし、本システムを使用する。
具体的には、以下のことを実現する。

- 各アプリケーション用のパラメータ入力画面を用意し、ジョブ実行に必要なパラメータを取得する
- ジョブの実行結果を Web 画面で表示する

■マシン構成

表 5 Web サーバ構成

本体	Sunblade1000
CPU	UltraSPARCⅢ (750MHz)
メモリ	512MB
HDD	18.2GB
FDD	フロッピーディスクドライブ
CD-ROM	CD/DVD-ROM ドライブ
DAT	内蔵 DAT 装置
LAN	Gigabit Ethernet 1000BASE-SX
ディスプレイ	18.1 インチ液晶
マウス/KB	Type6
OS	SunOS 5.8

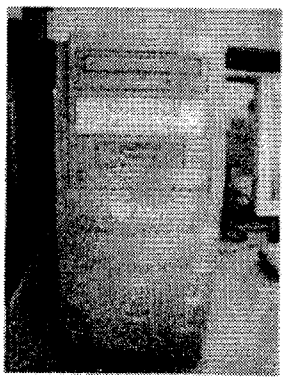

図 12 Web サーバ概観

3.5.2.教育用 PC クラスタ

■役割

ユーザから要求されたジョブを実行する。計算ノードでジョブを実行するための管理は、PBS を使用して行う。

■マシン構成

2.4 環境を参照のこと。

3.6.処理の流れ

ユーザが本システムを使用する場合の、処理の流れは以下の通り。

② 各アプリケーションについて説明した Web 画面を表示する
③ アプリケーション実行画面において、実行パラメータの入力を受け付ける
④ 入力されたパラメータをもとに、アプリケーションを実行するためのジョブを作成する
⑤ 教育用 PC クラスタにおいてジョブを実行する
⑥ 実行結果を Web 画面で表示する
⑦ 実行終了の案内、及び実行結果を表示する Web 画面の URL をユーザへ通知する

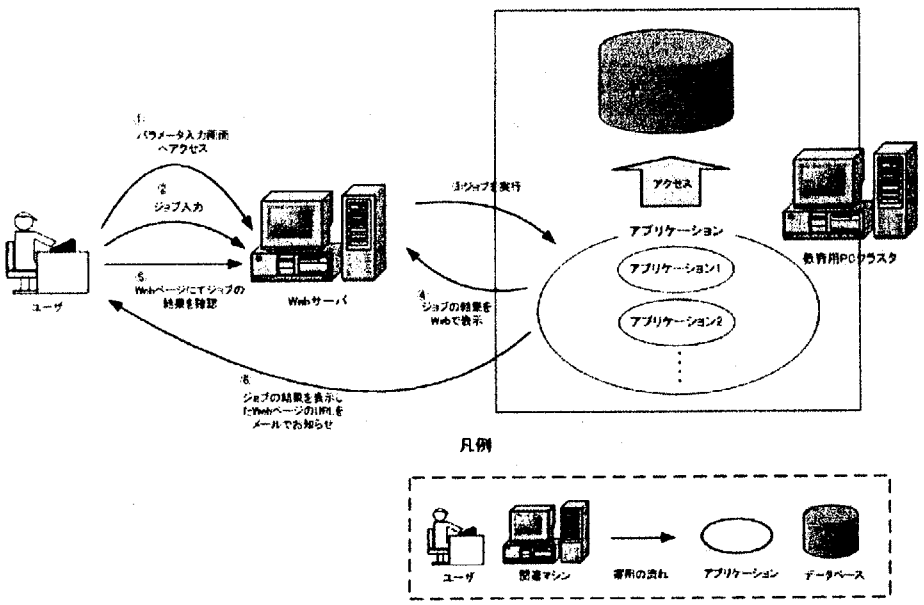

図 13 処理の流れ

3.7.Web 画面表示

ユーザが本システムを利用する場合、導入となる部分がブラウザに表示する Web コンテンツである。Web コンテンツで記述する項目を以下に挙げる。

■ 各アプリケーションの使用説明

現状以下のアプリケーションについての説明を記述する。

- BLAST
- ClustalW
- GenScan
- HMMER

また、各アプリケーションの説明については、さらに以下の項目を記述する。

- 概要
- 使用方法
- 実習

■ サンプルデータ集

アプリケーションの説明の例題で、いくつかサンプルデータを紹介する。これらのサンプルデータを集約し、公開する。公開するサンプルデータは、ファイルとしてダウンロードすることが可能である。

■ データベースの説明

アプリケーションを実行するためには、データベースが必要となる。ユーザはアプリケーションを実行する際に、使用するデータベースを選択することができるが、選択するためには、各データベースの詳細を把握しておく必要がある。そのために、使用対象となっているデータベースの説明を記述する。

■ 用語集

Web コンテンツの中には、バイオインフォマティクス分野の専門用語が数多く含まれている。これらの専門用語の説明を記述する。

■ Q&A

場合に分け、Q&A 形式で使用方法を記述する。

3.8.Web 画面イメージ

現状想定している画面イメージを以下に２つ挙げる。

注)本報告書を執筆段階において、Web 画面を開発中であるため、実際の画面はこれらのイメージをもとにさらに精度を上げたものとなる。

（1）メインページ

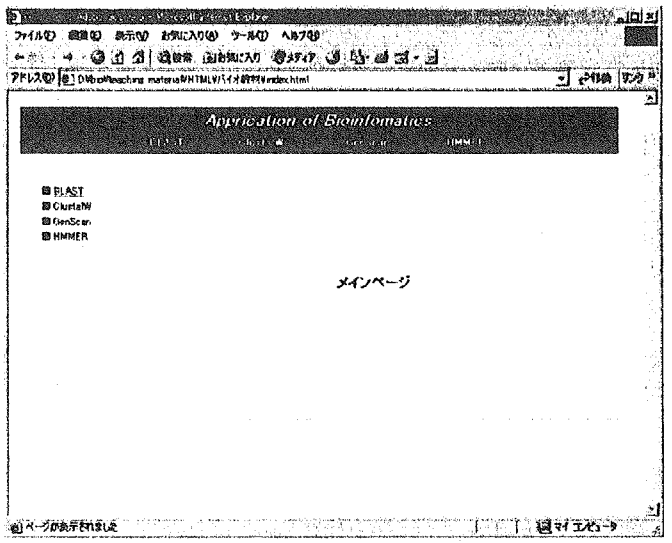

図 14 メインページ

（2）BLAST 使用説明ページ

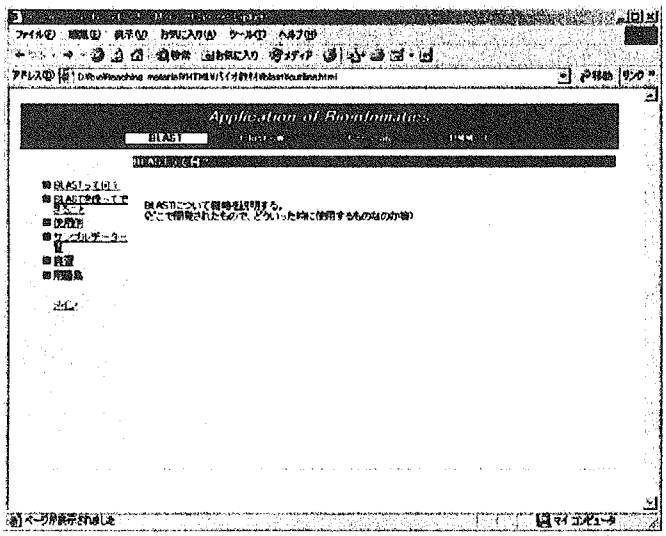

図 15 BLAST 使用説明ページ

3.9.システム機能

本システムでは以下の機能が実現できる。

（1） Web経由でのアプリケーション実行

Web画面で入力したパラメータを基に、教育用PCクラスタのアプリケーションを実行する。教育用PCクラスタでは、アプリケーションの実行を各計算ノードに割り振って行う。これらジョブ管理はPBSを使用して行う。本システムではPBSで実行するためのジョブを作成し、アプリケーションの実行を実現する。

（2） 実行結果表示

教育用PCクラスタで実行した結果をHTMLへ変換し、Webサーバへ転送する。従って、ユーザは実行結果をWeb画面で確認することが可能である。

（3） メール通知

ジョブの実行が終了した場合、実行結果を表示するWeb画面のURLをメールにてユーザへ通知する。従って、ユーザはジョブの実行開始後は、別作業を行いながらジョブの終了を待つことが可能である。

また、ジョブの実行終了通知のほか、異常終了通知など、ユーザにとって必要な情報を、逐次メールにて通知する。

3.10.今後の展開

本システムは、平成14年度末に完成するため、平成15年度よりITBL利用推進室ホームページより一般公開する予定である。現在、BLAST、ClustalW、GenScan、HMMERの4つのアプリケーションを対象としている。ユーザからの要望等により今後対象アプリケーションを増やしていく。また、Webコンテンツの更なる充実も考慮し、必要に応じシステムの拡張を行う。

謝辞

本報告書の執筆の機会を与えて下さったITBL利用推進室の相川室長、ITBL利用推進室の皆様に感謝いたします。また、データベース整備、及びアプリケーション実行Webシステム構築にあたり、数々のご協力やご助言を頂きました量子生命情報解析グループの皆様、ITBL利用推進室の皆様に感謝いたします。

参考文献

1）ITBLプロジェクト　http://www.itbl.riken.go.jp/
2）ITBL利用推進室　http://wwwitblpg.apr.jaeri.go.jp/itblpg/index.html
3）NCBI HOME　http://www.ncbi.nlm.nih.gov/
4）SWISS-PROT　http://www.ebi.ac.uk/swissprot/
5）PDB　http://www.rcsb.org/pdb/index.html
6）NDB　http://ndbserver.rutgers.edu/NDB/index.html
7）DDBJ　http://www.ddbj.nig.ac.jp/Welcome-j.html

国際単位系（SI）と換算表

表1　SI基本単位および補助単位

量	名称	記号
長さ	メートル	m
質量	キログラム	kg
時間	秒	s
電流	アンペア	A
熱力学温度	ケルビン	K
物質量	モル	mol
光度	カンデラ	cd
平面角	ラジアン	rad
立体角	ステラジアン	sr

表2　SIと併用される単位

名称	記号
分，時，日	min，h，d
度，分，秒	°，′，″
リットル	l，L
トン	t
電子ボルト	eV
原子質量単位	u

1 eV=1.60218×10^{-19}J

1 u=1.66054×10^{-27}kg

表3　固有の名称をもつSI組立単位

量	名称	記号	他のSI単位による表現
周波数	ヘルツ	Hz	s^{-1}
力	ニュートン	N	$m\cdot kg/s^2$
圧力，応力	パスカル	Pa	N/m^2
エネルギー，仕事，熱量	ジュール	J	N·m
工率，放射束	ワット	W	J/s
電気量，電荷	クーロン	C	A·s
電位，電圧，起電力	ボルト	V	W/A
静電容量	ファラド	F	C/V
電気抵抗	オーム	Ω	V/A
コンダクタンス	ジーメンス	S	A/V
磁束	ウェーバ	Wb	V·s
磁束密度	テスラ	T	Wb/m^2
インダクタンス	ヘンリー	H	Wb/A
セルシウス温度	セルシウス度	℃	
光束	ルーメン	lm	cd·sr
照度	ルクス	lx	lm/m^2
放射能	ベクレル	Bq	s^{-1}
吸収線量	グレイ	Gy	J/kg
線量等量	シーベルト	Sv	J/kg

表4　SIと共に暫定的に維持される単位

名称	記号
オングストローム	Å
バーン	b
バール	bar
ガル	Gal
キュリー	Ci
レントゲン	R
ラド	rad
レム	rem

1 Å=0.1nm=10^{-10}m

1 b=100fm^2=10^{-28}m^2

1 bar=0.1MPa=10^{5}Pa

1 Gal=1cm/s^2=10^{-2}m/s^2

1 Ci=3.7×10^{10}Bq

1 R=2.58×10^{-4}C/kg

1 rad=1cGy=10^{-2}Gy

1 rem=1cSv=10^{-2}Sv

表5　SI接頭語

倍数	接頭語	記号
10^{18}	エクサ	E
10^{15}	ペタ	P
10^{12}	テラ	T
10^{9}	ギガ	G
10^{6}	メガ	M
10^{3}	キロ	k
10^{2}	ヘクト	h
10^{1}	デカ	da
10^{-1}	デシ	d
10^{-2}	センチ	c
10^{-3}	ミリ	m
10^{-6}	マイクロ	μ
10^{-9}	ナノ	n
10^{-12}	ピコ	p
10^{-15}	フェムト	f
10^{-18}	アト	a

（注）

1．表1－5は「国際単位系」第5版，国際度量衡局1985年刊行による。ただし，1 eVおよび1 uの値はCODATAの1986年推奨値によった。

2．表4には海里，ノット，アール，ヘクタールも含まれているが日常の単位なのでここでは省略した。

3．barは，JISでは流体の圧力を表わす場合に限り表2のカテゴリーに分類されている。

4．EC閣僚理事会指令ではbar，barnおよび「血圧の単位」mmHgを表2のカテゴリーに入れている。

換算表

力	N(=10^{5}dyn)	kgf	lbf
	1	0.101972	0.224809
	9.80665	1	2.20462
	4.44822	0.453592	1

粘度　1 Pa·s(N·s/m^2)=10 P(ポアズ)(g/(cm·s))

動粘度　1m^2/s=10^{4}St(ストークス)(cm^2/s)

圧力	MPa(=10bar)	kgf/cm^2	atm	mmHg(Torr)	lbf/in^2(psi)
	1	10.1972	9.86923	7.50062×10^3	145.038
	0.0980665	1	0.967841	735.559	14.2233
	0.101325	1.03323	1	760	14.6959
	1.33322×10^{-4}	1.35951×10^{-3}	1.31579×10^{-3}	1	1.93368×10^{-2}
	6.89476×10^{-3}	7.03070×10^{-2}	6.80460×10^{-2}	51.7149	1

エネルギー・仕事・熱量	J(=10^7 erg)	kgf·m	kW·h	cal(計量法)	Btu	ft·lbf	eV
	1	0.101972	2.77778×10^{-7}	0.238889	9.47813×10^{-4}	0.737562	6.24150×10^{18}
	9.80665	1	2.72407×10^{-6}	2.34270	9.29487×10^{-3}	7.23301	6.12082×10^{19}
	3.6×10^6	3.67098×10^5	1	8.59999×10^5	3412.13	2.65522×10^6	2.24694×10^{25}
	4.18605	0.426858	1.16279×10^{-6}	1	3.96759×10^{-3}	3.08747	2.61272×10^{19}
	1055.06	107.586	2.93072×10^{-4}	252.042	1	778.172	6.58515×10^{21}
	1.35582	0.138255	3.76616×10^{-7}	0.323890	1.28506×10^{-3}	1	8.46233×10^{18}
	1.60218×10^{-19}	1.63377×10^{-20}	4.45050×10^{-26}	3.82743×10^{-20}	1.51857×10^{-22}	1.18171×10^{-19}	1

1 cal= 4.18605J（計量法）

= 4.184J（熱化学）

= 4.1855J（15℃）

= 4.1868J（国際蒸気表）

仕事率　1 PS(仏馬力)

= 75 kgf·m/s

= 735.499W

放射能	Bq	Ci
	1	2.70270×10^{-11}
	3.7×10^{10}	1

吸収線量	Gy	rad
	1	100
	0.01	1

照射線量	C/kg	R
	1	3876
	2.58×10^{-4}	1

線量当量	Sv	rem
	1	100
	0.01	1

（86年12月26日現在）

バイオインフォマティクス整備 －データベース及びアプリケーションＷｅｂ実行システム－