Acta Biochimica Polonica

Vol. 43 No. 4/1996

743–752

QUARTERLY

# Some useful hosting centers and databases for molecular biology on the World Wide Web

Magdalena Bryś

*Department of Cytobiochemistry, University of Łódź, St. Banacha 12/16, 90-237 Łódź, Poland*

Received: 20 June, 1996; accepted: 6 August, 1996

Key words: hosting centers, databases, molecular biology, www

**This article describes the information contained within World Wide Web of potential uses for molecular biologists. The aim is to provide a basic description of the server provided services of bioinformatics, major program coordinator activities, and current contents of genome nucleic acid and protein databases.**

The World Wide Web (WWW) is a powerful global information system merging networked information retrieval and hypertext. It gives access, using hypertext links, to the documents and information contained in all the existing WWW servers around the world. To access a WWW server, a software program or "browser", which displays hypertext documents, has to be run on a local computer.

WWW documents written in hypertext markup language (HTML) contain links to other documents in the form of uniform resource locators (URLs). These links are displayed as highlighted (or colored) text or graphics and a mouse click will either display the linked file or establish connection to the WWW server that hosts the file or can make it available.

The first character (xxxx://) of each URL address specifies the type of protocol used (e.g. http — hypertext transfer protocol). The next part of the address specifies the host computer and corresponds to the host name included in Internet e-mail addresses [1].

As a service to molecular biology there is a list of http URLs of interest to genome, nucleic acids and protein studies. The list of presented WWW servers consists of two parts: Information services and major program coordinators, and Specific database hosting centers, with partition into Genome, Nucleic acids, and Proteins. This combination is, of course, incomplete but may be helpful to many researchers in their current work.

WWW beginners should consult the http://www.eit.com/web/www.guide/ and read some helpful documents presented in Entering the World-Wide Web: A Guide to Cyberspace [2].

## INFORMATION SERVICES AND MAJOR PROGRAM COORDINATORS

### ANGIS (Australian National Genomics Information, Australia)

http://morgan.angis.su.oz.au/

This server offers: biocomputing services to university, hospital, government and industry, and operates from the Australian Genomic Information Centre (AGIC), and others [2].

### ANU (Australian National University, Australia)

http://life.anu.edu.au/

This Bioinformatics Hypermedia Server offers in its Molecular Biology part access to: Databases (French WWW genomic service, BLAST Database Searches — sequence similarity searching, CompoundKB database —

981 metabolic intermediate compounds, Codon usage tables & EMBL (European Bioinformatics Institute mirror), DNA Data Bank of Japan, EC enzyme database, EMBL data, ESTDB — Expressed Sequence Tag Database, GenBank — the NIH database of all publicly available nucleotide and protein sequences including supporting bibliographic and biological information, Généthon Human Genome Centre, Microbial germplasm, PIR — Protein Information Resource, Prosite — dictionary of protein sites and patterns, PDB — Protein Data Bank, REF52 2D Gel Protein Database, REBASE — Restriction Enzyme Database, SWISS-PROT — annotated protein sequence database, SWISS-2DPAGE — two-dimensional polyacrylamide gel electrophoresis database), Bibliographies/Tutorials, Software, Newsgroups, Special topics, and others [2].

**BIGROC (The Bioinformatics Unit and Genome Resource Core, Weizmann Institute of Science, Israel)**

http://bioinformatics.weizmann.ac.il/

BIGROC at the Weizmann Institute of Science is a nationally supported center, which deals with the design and development of tools for Bioinformatics, especially in the areas of Molecular Biology and Genome Analysis. This server offers: Human Chromosome 17 Database, HotMolecBase — Molecules for Medicine (a database of proteins and other biologically active molecules, their involvement in diseases and their potential medical applications), Mirror sites of Molecular Biology databases, BioMOO (the biologists' virtual meeting place), An E-Mail server to search for structural patterns in DNA and protein uncharacterized sequences, and others [2].

**BIMAS (BioInformatics & Molecular Analysis Section, Bethesda, U.S.A.)**

http://bimas.dcrt.nih.gov/

The BIMAS provides guidance, support and resources to scientists throughout the NIH in the genomic and genetic analysis fields of BioInformatics. This server gives access to: BioMagResBank Sybase Gateway (a database of NMR-derived protein structures), GenoBase Database Gateway (a molecular biology database which incorporated and linked the contents of several large datasets), and others [2].

**BIOSCI**

http://www.bio.net/

BIOSCI is a set of electronic communication forums — the bionet USENET newsgroups and parallel e-mail lists — used by biological scientists worldwide. BIOSCI is currently supported in the U.S.A. by a grant from DOE with equipment from NSF (National Science Foundation). The UK BIOSCI node is supported by the Daresbury Laboratory. BIOSCI promotes communication between professionals in the biological sciences [2].

**CGAT (Canadian Genome Analysis and Technology Program, Montreal, Canada)**

http://cgat.bch.umontreal.ca/

This program funds and supports multidisciplinary genetic researchers in Canada, having as its objective the analysis of the human and other selected genomes including the development of related technologies and informatics and the study of corresponding medical, ethical and legal issues. This website is an information resource and communications nexus for CGAT grantees and other genomics researchers in Canada [2].

**DOE (US Department of Energy, Office of Energy Research, U.S.A.)**

http://www.er.doe.gov/

The programs of the Office of Energy Research fund basic research in order to advance fundamental scientific knowledge, as well as train future scientists. The Biological and Environmental Research program (BER) of the Office of Health and Environmental Research utilizes the unique multidisciplinary capabilities and facilities of the DOE National Laboratories and supports biological and environmental research projects and research infrastructure at National Laboratories, universities and other research institutes. This server offers: Activities and Research Areas (Basic Life Sciences, Health Effects, Environmental Processes and Effects Research, and others) [2].

**EBI (European Bioinformatics Institute, Hinxton, U.K.)**

http://www.ebi.ac.uk/

EBI is an EMBL Outstation, and all activities previously based at the EMBL Data Library are located at the EBI. Its main activity is the distribution of the EMBL Nucleotide Sequence database, SWISS-PROT database, The Nucleic Acid Database (NDB), Radiation Hybrid mapping database, IMGT database and a list of software of general interest in molecular biology and genetics. EBI is a major distributor of molecular biological databases produced by other groups in Europe and world-wide (*via* — ftp://ftp.ebi.ac.uk/pub/databases/ !!!): 3D ali (structure-based sequence alignments), Alu (ALU sequences and alignments), Berlin RNA (5S rRNA sequences), Blocks (Protein Blocks Database), CpGisle (CpG islands database), Cutg (codon usage tabulated from GenBank), dbEST (expressed sequence tags), dbSTS (sequence tagged sites), DSSP (secondary structure assignments of PDB files), ECDC (*Escherichia coli* database collection), Enzyme (database of EC nomenclature), EPD (eukaryotic promoter database), FlyBase (*Drosophila* genetic map database), FSSP (families of structurally similar proteins), HaemA (Haemophilia A database), HaemB (Haemophilia B database), HLA (HLA class I and II sequence database), HSSP (protein structure-sequence alignments), Kabat (proteins of immunological interest), LiMB (list of molecular biology databases), Lista (Yeast protein coding sequences), Methyl (site-specific methylation), Misfolded (deliberately misfolded protein models), NRL3D (sequence-structure database), NRSub (non-redundant *Bacillus subtilis* genome database), Nucleosomal DNA (nucleosomal DNA sequences), p53 (p53 mutations), PDB (Brookhaven protein structures database), PDB Select (representative list of PDB chain identifiers), PIR (protein sequence database), PKCDD (protein kinase catalytic domain sequence database), Primers (fully tested and optimized primers for PCR reactions), Prints (protein motif fingerprint database), Prodom (protein sequence modules — recurring domains), PROSITE (prosite pattern database), PUU (database of structural domains), RDP (ribosomal database project), REBASE, RELibrary (comprehensive restriction enzyme lists), RepBase (prototypic human repetitive DNA sequences), RLDB (reference library database), rRNA (small subunit rRNA sequences), SBASE (protein domain database), SmallRNA (compilation of small RNA sequences), SRP (signal recognition particle database), TFD (transcription factor database), TransFac (eukaryotic cis-acting regulatory DNA elements and transacting factors), TransTerm (translational termination signal database), tRNA (database of tRNA sequences), Yeast (Yeast chromosome database), and others [2–8].

**EMBL (European Molecular Biology Laboratory, Heidelberg, Germany)**

http://www.embl-heidelberg.de/

The aim of this server is to provide an on-line information service of the activities at EMBL [2].

**EMBnet (Molecular Biology Network, Basel University, Switzerland)**

http://www.ch.embnet.org/

EMBnet is a science-based group of 26 collaborating nodes throughout Europe. The combined expertise of the nodes allows EMBnet to provide a service to the European molecular biology community which encompasses more than can be provided by a single node [2].

**ExPASy (Geneva University Hospital and University of Geneva, Switzerland)**

http://expasy.hcuge.ch/

This molecular biology server is dedicated to the analysis of protein and nucleic acid sequences and enables access to following databases: SWISS-PROT, PROSITE, SWISS-2DPAGE, SWISS-3DIMAGE (3D images of proteins and other biological macromolecules), ENZYME and SeqAnalRef (sequence analysis bibliographic reference database). ExPASy offers also TOOLS (access to many protein analysis tools), Swiss-Model (automated knowledge-based protein modeling server), Melanie (software packages for 2-DPAGE analysis), informations about 2-DPAGE services and courses, and gateways to other molecular biology servers [2, 9–11].

**Harvard Biological Laboratories (Boston, U.S.A.)**

http://golgi.harvard.edu/

This server gives access to Selected Model Organism Databases (*Drosophila*, human, mice and rats, prokaryotes, yeasts, zebrafish), GCG Software Documentation, Selected Biomolecular & Biochemical Databases (GenBank, PDB, SWISS-PROT+PIR+GenBank search, submit DNA sequence to GenBank with BankIt, MSDS — Material Safety Data Sheets, Vectordb —

a sequence database of recombinant DNA vectors, Codon Usage Tables for > 450 species), Selected Other Biological Databases (Biological Journals, Culture Collections, Taxonomic Databases, REBASE, BIOSCI), and others [2].

**HGMIS (Human Genome Management Information System, U.S.A.)**

http://www.ornl.gov/TechResources/Human_Genome/

This page is maintained by the HGMIS for the US Department of Energy Human Genome Program and contains material about the history, progress, research, and resources of the Human Genome Project [2].

**IBT (Albert B. Alkek Institute of Biosciences and Technology, Houston, U.S.A.)**

http://keck.tamu.edu/

IBT undertakes creative research on molecular aspects of agriculture and medicine, provides a forum for the exchange of ideas between the agricultural and medical communities, and encourages transfer of technology and medical discoveries from the laboratory to the marketplace [2].

**INFOBIOGEN (Informatics for Biomolecules and Genomes, Villejuif, France)**

http://www.infobiogen.fr/

INFOBIOGEN is a French national center for Bioinformatics. Its server offers: shortcuts to some databases, catalogues, and access to other WEB services (biological and medical servers, AGORA Informations — Health and Medical Research, and others) [2].

**Johns Hopkins University (Baltimore, U.S.A.)**

http://www.gdb.org/

This server offers Prot-Web which is a Collection of Protein Databases: OWL — a non-redundant protein sequence database which ties a number of databases together, NRL 3D — a sequence-structure database, PIR, EC Enzyme, GenQuest — The Q server — an integrated interface to the sequence comparison server at the Oak Ridge National Lab designed for rapid and sensitive comparison of DNA and Protein sequence to existing DNA and Protein sequence databases and rapid retrieval of the full database entries of any sequence found in course of a search, REBASE. The other biological databases at this server are: TBASE — The Transgenic/Targeted Mutation Database, DSHB — The Developmental Studies Hybridoma Bank, The DOE Human Subjects Database, GDB — The Genome Data Base, and ECACC — The European Collection of Animal Cell Cultures [2, 12].

**MedWeb — Genetics and Molecular Biology**

http://www.gen.emory.edu/medweb/

This server offers many tools, databases, electronic newsletters and journals, etc. [2].

**NCBI (National Center for Biotechnology Information, U.S.A.)**

http://www.ncbi.nlm.nih.gov/

Its main activity is access to: GenBank, dbEST, dbSTS, MMDB — Molecular Modelling Data Base, NCBI Taxonomy database, OMIM — Online Mendelian Inheritance in Man (catalogue of human genes and genetic disorders which contains textual information, pictures, and reference information). This server contains also BLAST, BankIt Sequence Submission Tool, and Entrez Search System which allows access to three databases: the National Library of Medicine's MEDLINE database, the NCBI protein database, and the NCBI nucleotide database. The Protein and Nucleotide entries in Entrez have been compiled from a variety of sources [2, 13, 14].

**NCHGR (National Center for Human Genome Research, U.S.A.)**

http://www.nchgr.nih.gov/

NCHGR is one of 24 institutes, centers or divisions that make up the National Institutes of Health, the federal government's primary agency for the support of biomedical research. NCHGR presides the Human Genome Project for the NIH. This server gives access to Genomics and Genetic Data, Grant Information, Offsite Resources, and The Human Genome Project [2].

**NIH (National Institutes of Health, Bethesda, U.S.A.)**

http://www.nih.gov/

This server offers: News & Events (the NIH Calendar of Events, special reports, and employment information), Health Information (a selection of some NIH health resources such as CancerNet, AIDS information, Clinical Alerts, and others), Grants and Contracts, Scientific Resources (intramural research news and information including special interest groups, online library catalogues and journals, research training information, NIH research labs on the web and computer and network support for NIH) [2].

**OncoLink (University of Pennsylvania, U.S.A.)**

http://cancer.med.upenn.edu/

The University of Pennsylvania Cancer Center is dedicated to provide the most advanced and innovative programs in cancer care, research, and education. This University Cancer Center has established nationally-recognized treatment programs in breast cancer and in melanoma. There are also multidisciplinary programs in gastrointestinal, urologic, lung and gynecologic cancers [2].

**Oxford University (Molecular Biology Data Centre, U.K.)**

http://molbiol.ox.ac.uk/

This is the Molecular Biology Data Information Service. This service provides information on the Data Centre as well as links to other molecular biological services in Oxford and beyond [2].

**SEQNET (Daresbury Laboratory, U.K.)**

http://gserv1.dl.ac.uk/SEQNET/

The SEQNET central node at Daresbury offers access to all major molecular biology sequence and structure databases and programs for U.K. academics. There is also a MasPar for running fast, exhaustive sequence comparisons based on the Smith-Waterman algorithm. All major software packages are provided (e.g. GCG, Serratus, Staden, Pearson, Blast, Phylip) plus a host of other utilities, particularly for protein analysis. The system is connected to the major national and international networks. SEQNET provides a range of training courses, both one and three-day colloquia. Seminars are given regularly and the biannual Genes, Proteins and Computers conference at Chester organized [2].

**SoBS & SoM (Schools of Biological Sciences and Medicine, Southampton, U.K.)**

http://molbiol.soton.ac.uk/

The Schools of Biological Sciences and Medicine are two of the four Schools within the Faculty of Medicine, Health and Biological Sciences. Both Schools have active research programs, and undergraduate Honours Schools [2].

**UK MRC HGMP-RC (UK Medical Research Council Human Genome Mapping Project Resource Centre, U.K.)**

http://www.hgmp.mrc.ac.uk/

The UK MRC HGMP-RC exists to provide specialist resources and services for scientists working on the Human Genome Project. This server gives access to Biology Resources and Services, Computing Resources, Training Courses and Projects, and Lists of Other Genome Resources [2].

**Walter and Eliza Hall Institute of Medical Research (Melbourne, Australia)**

http://www.wehi.edu.au/

This server offers National and International Web Scientific Resources and MHCPEP database (database of peptides known to bind MHC class I or II molecules) [2, 15].

## SPECIFIC DATABASE HOSTING CENTERS

### GENOME

**Baylor College of Medicine (Houston, U.S.A.)**

http://gc.bcm.tmc.edu/

This is the Baylor College of Medicine Human Genome Center server. The most interesting on this server are: Biologists Control Panel (Tumor Gene Database, BCM Sequence Annotation Server, Genome Reconstruction Manager — manage data flow cosmid DNA sequencing, BLAST, CEPH-Généthon map data, DNA Vaccine Web, Entrez, dbEST, GDB, Genome Sequence Database, GenBank, GenBank+ SWISS-PROT+PIR+PDB, Jackson Labs-Mouse Genome Database, PIR, Rat Brain 2D Protein Database, SBASE, SWISS-2DPAGE, Libraries and Literature), BCM Search Launcher — WWW sequence search and analysis software, and YAC (Yeast Artificial Chromosome) Data Search — the most current YAC data from various sources [2].

**CEPH (Centre d'Etudes du Polymorphisme Humain — Fondation Jean Dausset, Paris, France)**

http://www.cephb.fr/

CEPH is a research laboratory which constructs maps of the human genome. Its server offers CEPH Chromosome 21 mapping resource, CEPH — Généthon integrated map, CEPH Genotype database of genotypes for all genetic markers that have been tested in the reference families for linkage mapping of the human chromosomes [2].

**CHLC (Cooperative Human Linkage Center, U.S.A.)**

http://www.chlc.org/

The main options available on this server are: genetic maps showing the positions of genetic markers, and search by name for information on markers [2].

**Chromosome 12 Genome Center (Yale University School of Medicine, Albert Einstein College of Medicine, New York, U.S.A.)**

http://paella.med.yale.edu/chr12/

The Chromosome 12 Genome Center is a collaborative research effort of the Albert Einstein College of Medicine and the Yale University of Medicine. This hosting center offers Chromosome 12 Physical Mapping Project, Description of the Chromosome 12 Informatics Core, and Chromosome 12 Sequencing Project [2].

**Columbia University (Human Genome Project, New York, U.S.A.)**

http://genome1.ccc.columbia.edu/~genome/

This server is run by The Human Genome Project and contains information on maps, databases and software produced by the project. It offers q21 map, BRCA2 map, Human Chromosome 13 Maps and Data, Columbia University Human Chromosome 13 Database [2].

**DKFZ (Deutsches Krebsforschungszentrum, Heidelberg, Germany)**

http://genome.dkfz-heidelberg.de/

The division of Molecular Biophysics examines molecular and genomic structural problems with the help of biophysical and computer scientific methods and develops computer simulation methods for their modeling at the genomic, molecular and electronic levels. DKFZ enables access to: Genome Database, Biocomputing Service Group, Integrated Genome Database, Neural Networks and Genetic Algorithms in the Analysis of Protein and DNA Sequences, Applications of Quantum Mechanics on Biomolecules and Bioinformatics Seminars [2].

**Généthon (Evry, France)**

http://www.genethon.fr/

This Human Genome Research Centre offers in its server: The Collection of Families with Genetic Diseases (neuromuscular disorders, skin, neurologic and rheumatologic diseases), informations about Généthon's DNA and Cell Bank (at present the bank contains 22,500 specimens from 6,600 families affected by 200 different pathologies), The Map of Transcripts (a radiation hybrid map of the human whole-genome), The Search for Genes Which Cause Genetic Diseases [2].

**GenomeNet (Japan)**

http://www.genome.ad.jp/

GenomeNet is a Japanese computer network for genome research and related research areas in molecular and cellular biology. It is operated jointly by the Human Genome Center (HGC), Institute of Medical Science, the University of Tokyo and the Supercomputer Laboratory (SCL), Institute for Chemical Research (ICR), Kyoto University. This server gives access to: Genome Databases in Japan (BSORF: *Bacillus subtilis* — the international project of sequencing, *Escherichia coli* Databank, CyanoBase — database on unicellular cyanobacterium, Synechocystis sp. strain PCC6803, BodyMap — Expression Database of Human Genes, GENOTK — Human cDNA Database, SPAD — Signaling Pathway Database, Aberrant Splicing Database), KEGG — Kyoto Encyclopedia of Genes and Genomes, DBGET integrated Database Retrieval System, Sequence Interpretation Tools, and others [2].

**IBC (Institute for Biomedical Computing, Washington University, St. Louis, U.S.A.)**

http://ibc.wustl.edu/

The Institute for Biomedical Computing is a collaborative research organization between the Schools of Medicine and Engineering and Applied Sciences at Washington University. They specialize in using computational techniques for solving problems of sequence and genome analysis, microscopy, molecular modeling, biochemical databases and modeling, computational chemistry and image restoration [2].

**Jackson Laboratory (Bar Harbor, U.S.A.)**

http://www.jax.org/

The Jackson Laboratory has a threefold mission: to conduct research in basic genetics and the role of genes in health and disease; to educate the scientific community; and to provide genetically defined mice and other genetic resources to the world. This server offers access to The Mouse Genome Database, The Gene Expression Database Project, The Encyclopedia of the Mouse Genome, Resources, and to Training & Education [2].

**LBNL (Lawrence Berkeley National Laboratory, Berkeley, U.S.A.)**

http://genome.lbl.gov/

The Human Genome Center of the Lawrence Berkeley National Laboratory is oriented almost exclusively towards developing and implementing directed methodologies for cost--effective and accurate high throughput human DNA sequencing. Using the directed sequencing strategy they have sequenced 3.5 megabases of *Drosophila melanogaster* and 1.9 megabases of human sequence (mostly 5q31), and these sequences are available from their sequence archive [2].

**LLNL (Lawrence Livermore National Laboratory, Human Genome Center, Livermore, U.S.A.)**

http://www-bio.llnl.gov/bbrp/genome/

This Human Genome Center is organized into four broad areas of research and support: Resources, Physical Mapping, DNA Sequencing, and Enabling Technologies. Its server offers Maps of Human Chromosome 19, and Genomic sequencing (tool) [2].

**McGill University (PAHdb, Montreal, Canada)**

http://www.mcgill.ca/pahdb/

This site is for all who are interested in the PAH (phenylalanine hydroxylase) gene and alleles [2].

**NCGR (National Center for Genome Resources, U.S.A.)**

http://www.ncgr.org/

NCGR is a non-profit organization created to design, develop, and deliver resources in support of public and private genome research. The Center currently houses the Genome Sequence DataBase (GSDB), which is a collection of DNA sequence data and related information [2].

**Pasteur Institute (France)**

http://www.pasteur.fr/

This server offers in Biology part: Molecular Biology on the Web, Sequence analysis, Genotoxicity Database, Gene therapy, and others [2].

**Sanger Centre (Hinxton, U.K.)**

http://www.sanger.ac.uk/

The Sanger Centre is a research center established jointly by the Wellcome Trust and the British Medical Research Council to provide a major focus in the U.K. for mapping and sequencing the human genome, and genomes of other organisms [2].

**Stanford Human Genome Center (U.S.A.)**

http://shgc.stanford.edu/

The two primary research goals of SHGC are the construction of high resolution Radiation Hybrid maps of the human genome and the sequencing of large, contiguous genomic regions [2].

**TIGR (The Institute for Genomic Research, Rockville, U.S.A.)**

http://www.tigr.org/

This Institute is a not-for-profit research institute with interests in structural, functional, and comparative analysis of genomes and gene products in viruses, eubacteria, pathogenic bacteria, archaea, and eukaryotes, both plant and animal, including humans. This server offers TIGR Database (TDB) — a collection of molecular biology databases containing DNA and protein sequence, gene expression, cellular role, isology class, taxonomic and sample collection data for a variety of species. TDB currently comprises the Microbial Database (MDB, containing the genome sequences of *Haemophilus influenzae* and *Mycoplasma genitalium*), the Human cDNA Database (HCD), the Expressed Gene Anatomy Database (EGAD), Sequences, Sources, Taxa (SST), and the *Arabidopsis thaliana* Database [2].

**University of Michigan (Human Genome Center, U.S.A.)**

http://mendel.hgp.med.umich.edu/

This server offers: UM Genome Center Resources and Core Services, Biology and Genome Informatics Resources, General UM and Internet Resources, and Computer and Software Documentation [2].

**VGC (Virtual Genome Center, University of Minnesota Medical School, U.S.A.)**

http://alces.med.umn.edu/

The Virtual Genome Center is a place where molecular biologists have an access to tools and databases for their work. The VGC has no director, staff, grant support, or corporate sponsors. It simply exists. Its server gives access to: Programs you can run at the VGC, *Saccharomyces* and *Candida* databases, Molecular Biology databases and tables [2].

**Washington University School of Medicine (Department of Genetics, U.S.A.)**

http://genome.wustl.edu/

Research interests in the department include the basic molecular and developmental genetics of model eukaryotic systems such as the *Drosophila melanogaster*, the *Caenorhabditis elegans*, and the *Saccharomyces cerevisiae*, as well as mapping and identification of genes responsible for important hereditary diseases [2].

**Whitehead Institute for Biomedical Research (MIT Center for Genome Research, Massachusetts, U.S.A.)**

http://www-genome.wi.mit.edu/

This server is run by the Center for Genome Research at the Whitehead Institute for Biomedical Research in Cambridge. It contains information on map and sequence releases (Genome Maps, Lab Protocols, Automation — The Sequatron, Sequencing Informatics, Sequence Data, and Genome Center Software) [2].

## NUCLEIC ACIDS

**DNA Data Bank of Japan (National Institute of Genetics, Japan)**

http://www.ddbj.nig.ac.jp/

This is the Center for Information Biology, National Institute of Genetics server [2].

**rRNA (University of Antwerp, Department of Biochemistry, Belgium)**

http://rrna.uia.ac.be/

This server offers: The LSU rRNA Database (aligned sequences with secondary structure information from the large subunit ribosomal RNA in a number of formats), The SSU rRNA Database (aligned sequences with secondary structure information from the small subunit ribosomal RNA in a number of formats), CARD (a DOS program to draw RNA structure models) [2].

**University of Utah (Departments of Medicinal Chemistry and Biochemistry, Salt Lake City, U.S.A.)**

http://www-medlib.med.utah.edu/RNAmods/

The RNA modification database provides a comprehensive listing of posttranscriptionally modified nucleosides from RNA [2, 16].

## PROTEINS

**BLOCKS (Fred Hutchinson Cancer Research Center, Seattle, U.S.A.)**

http://blocks.fhcrc.org/

A Fred Hutchinson Cancer Research Center server is useful for the detection and verification of protein sequence homology. Blocks offered by this place are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins [2, 17].

**IMGT ( Immunogenetics, Montpellier, France)**

http://Imgt.cnusc.fr/

IMGT is an integrated database specializing in immunogenetics. It combines nucleotide and protein sequences of immunoglobulins, T-cell receptors (LIGM-DB) and HLA molecules (HLA-DB) [2].

**PDB (Protein Data Bank, the Brookhaven National Laboratory, U.S.A.)**

http://www.pdb.bnl.gov/

The Protein Data Bank is an archive of experimentally determined three-dimensional structures of biological macromolecules [2].

**PIR (Protein Information Resource)**

http://watson.gmu.edu/pirwww/

The PIR is one of three collaborating data centers responsible for collecting, organizing, and distributing the PIR-International Protein Sequence Database. The database contains information concerning all naturally occurring, wild-type proteins whose primary structure (the sequence) is known [2, 18].

**UCL'S Database Browser (University College London, Department of Biochemistry and Molecular Biology, U.K.)**

http://www.biochem.ucl.ac.uk/bsm/dbbrowser/

UCL'S Database Browser provides facilities for interrogating the OWL composite sequence database and its derived database of fingerprints, i.e. PRINTS. It also provides links to a variety of other data sources: e.g. SWISS-PROT, PIR, GenBank, NRL-3D, PROSITE, ProDom, BLOCKS, SBASE, etc.; and to numerous sites and analysis tools: e.g. ExPASy, GenomeNet, SEQNET, EBI, etc., including a unique fingerprint search facility [2, 19].

## REFERENCES

1. *Genome Digest*, October (1995), 10–13.
2. Internet.
3. Rodriguez-Tomé, P., Stoehr, P.J., Cameron, G.N. & Flores, T.P. (1996) The European Bioinformatics Institute (EBI) databases. *Nucleic Acids Res.* **24,** 6–12.
4. Bairoch, A. & Apweiler, R. (1996) The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res.* **24,** 21–25.
5. Holm, L. & Sander, C. (1996) The FSSP database: Fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res.* **24,** 206–209.
6. Schneider, R. & Sander, C. (1996) The HSSP database of protein structure-sequence alignments. *Nucleic Acids Res.* **24,** 201–205.
7. Bairoch, A., Bucher, P. & Hofmann, K. (1996) The PROSITE database, its status in 1995. *Nucleic Acids Res.* **24,** 189–196.
8. Roberts, R. & Macelis, D. (1996) REBASE-restriction enzymes and methylases. *Nucleic Acids Res.* **24,** 223–235.
9. Appel, R.D., Sanchez, J.-C., Bairoch, A., Golaz, O., Miu, M., Vargas, J.R. & Hochstrasser, D.F. (1993) SWISS-2DPAGE: a database of two-dimensional gel electrophoresis images. *Electrophoresis* **14,** 1232–1238.
10. Appel, R.D., Sanchez, J.-C., Bairoch, A., Golaz, O., Ravier, F., Pasquali, C., Hughes, G.J. & Hochstrasser, D.F. (1996) The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis, its status in 1995. *Nucleic Acids Res.* **24,** 180–181.
11. Bairoch, A. (1996) The ENZYME data bank in 1995. *Nucleic Acids Res.* **24,** 221–222.
12. Fasman, K.H., Letovsky, S.I., Cottingham, R.W. & Kingsbury, D.T. (1996) Improvements to the GDB™ Human Genome Data Base. *Nucleic Acids Res.* **24,** 57–63.
13. Benson, D.A., Boguski, M., Lipman, D.J. & Ostell, J. (1996) GenBank. *Nucleic Acids Res.* **24,** 1–5.
14. Boguski, M.S. (1995) Hunting for genes in computer data bases. *N. Engl. J. Med.* 333, 645–647.
15. Brusic, V., Rudy, G., Kyne, A.P. & Harrison, L.C. (1996) MHCPEP-a database of MHC-binding peptides: update 1995. *Nucleic Acids Res.* **24,** 242–244.
16. Crain, P.F. & McCloskey, J.A. (1996) The RNA modification database. *Nucleic Acids Res.* **24,** 98–99.
17. Pietrokovski, S., Henikoff, J.G. & Henikoff, S. (1996) The Blocks database — a system for protein classification. *Nucleic Acids Res.* **24,** 197–200.
18. George, D.G., Barker, W.C., Mewes, H.-W., Pfeiffer, F. & Tsugita, A. (1996) The PIR —International Protein Sequence Database. *Nucleic Acids Res.* **24,** 17–20.
19. Attwood, T.K., Beck, M.E., Bleasby, A.J., Degtyarenko, K. & Smith, D.J.P. (1996) Progress with the PRINTS protein fingerprint database. *Nucleic Acids Res.* **24,** 182–188.