Bioinformatics

Information networks

Bioinformatics - lectures

- Introduction
- Information networks
- Protein information resources
- Genome information resources
- DNA sequence analysis
- Pairwise sequence alignment
- Multiple sequence alignment
- Secondary database searching
- Analysis packages
- Protein structure modelling

Information networks

- what is the Internet?
- how do computers find each other?
- FTP and Telnet
- what is the Worl Wide Web?
- HTTP, HTML and URL
- EMBnet, EBI, NCBI
- SRS and ENTREZ

What is the Internet?

- Global network of computer networks that link government, academic and business institutions.
- communication by TCP/IP (Transmission Control Protocol/Internet Protocol)
- computers - nodes, data - packets
- packets may not be transferred directly from one computer to another

How do computers find each other?

- Each computer is assigned IP address

  147.251.28.2

  machine.site.domain

  bilbo.chemi.muni.cz

- FTP - File Transfer Protocol
- Telnet - remote connection

Example of Internet domains and subdomains

Country-based domains		*Other domains*		*Subdomains*	
Australia	.au	Educational	.edu	Academic	.ac
Denmark	.dk	Commercial	.com	Company	.co
Finland	.fi	Governmental	.gov	Other organisation	.org
France	.fr	Military	.mil	General	.gen
Germany	.de				
Greece	.gr				
Hungary	.hu				
Ireland	.ie				
Israel	.il				
Italy	.it				
Netherlands	.nl				
New Zealand	.nz				
Poland	.pl				
Portugal	.pt				
South Africa	.za				
Spain	.es				
Sweden	.se				
Switzerland	.ch				
United Kingdom	.uk				
USA	.us				

What is the World Wide Web?

- Developed at CERN - the European Laboratory of Particle Physics.
- The purpose was sharing of information.
- Hypermedia based information system.
- The most advanced information system found on the web.
- Very popular - almost synonymous with the Internet.

Web browsers

- Browser is the client communicating with servers using standard protocols.

- Home page is the first point of contact between browser and the server.

Lynx - academic, VT100 terminal
Mosaic - academic, X-windows
Netscape Navigator - commercial
Internet Explorer - commercial

HTTP, HTML and URL

- HTTP - HyperText Transport Protocol
  documents exploited by browsers are written in hypertext and transferred by HTTP

- HTML - HyperText markup Language
  standard language for writing a hypertext

- URL - Uniform Resourse Locator
  unique address for a document
  example: http://www.chemi.muni.cz/~jiri

EMBnet, EBI, NCBI

- 1988 established the network of European biocomputing and bioinformatics laboratories.
- Eliminates the need for multicopies of biology databases and retrieval software.
- Hinxton Hall = Sanger Centre + MRC Human Genome Mapping Project Resource Centre + European Bioinformatics Institute (EBI)
- National Center for Biotechnology Information (NCBI)

SRS, ENTREZ and LinkDB

SRS - The Sequence Retrieval System

- maintained by EBI
- network browser for databases in molecular biology
- allows indexation of flat-file databases
- allows customised search of selected databases
- link databanks: sequence, structure, bibliography, etc.

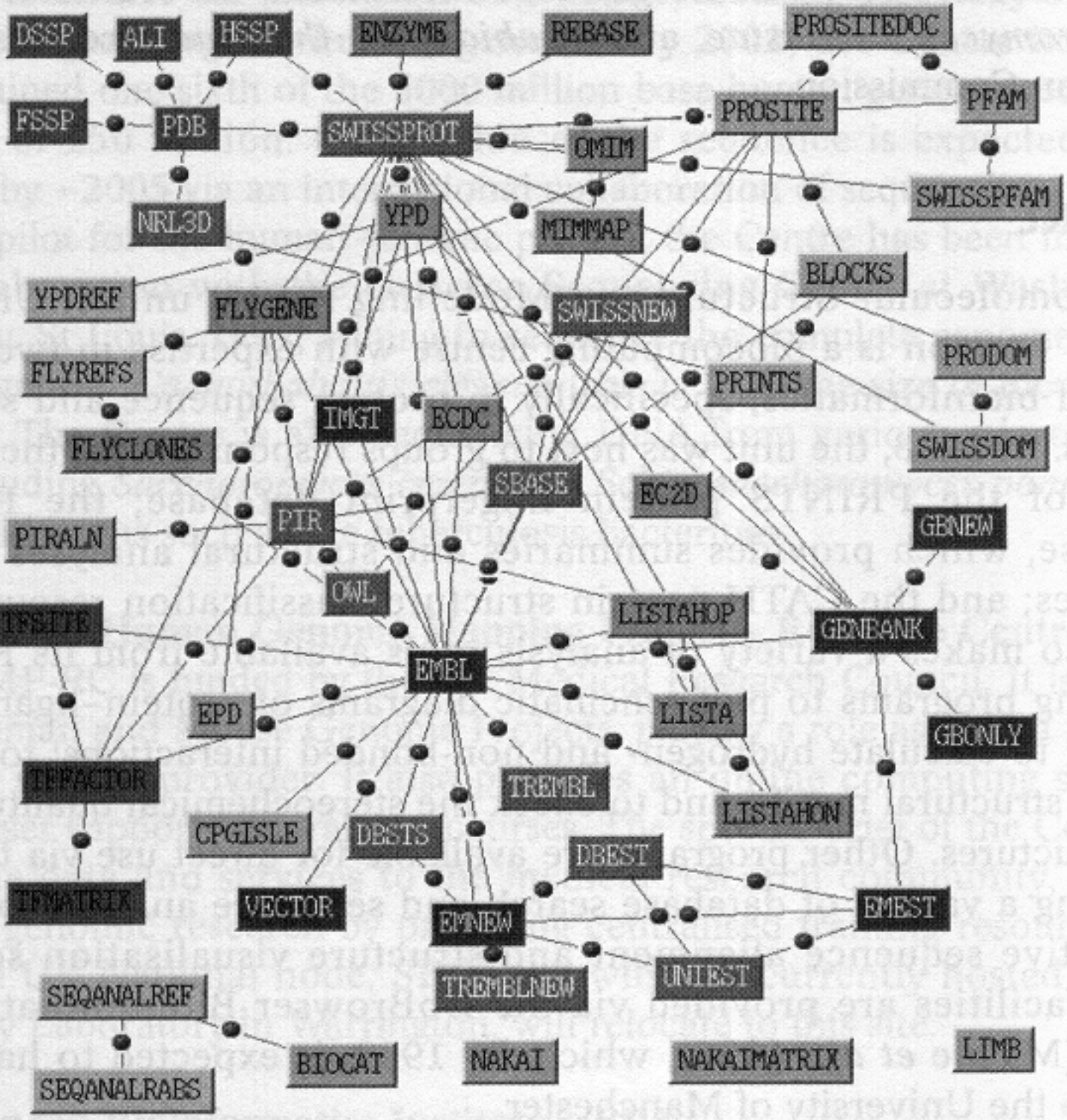
ENTREZ

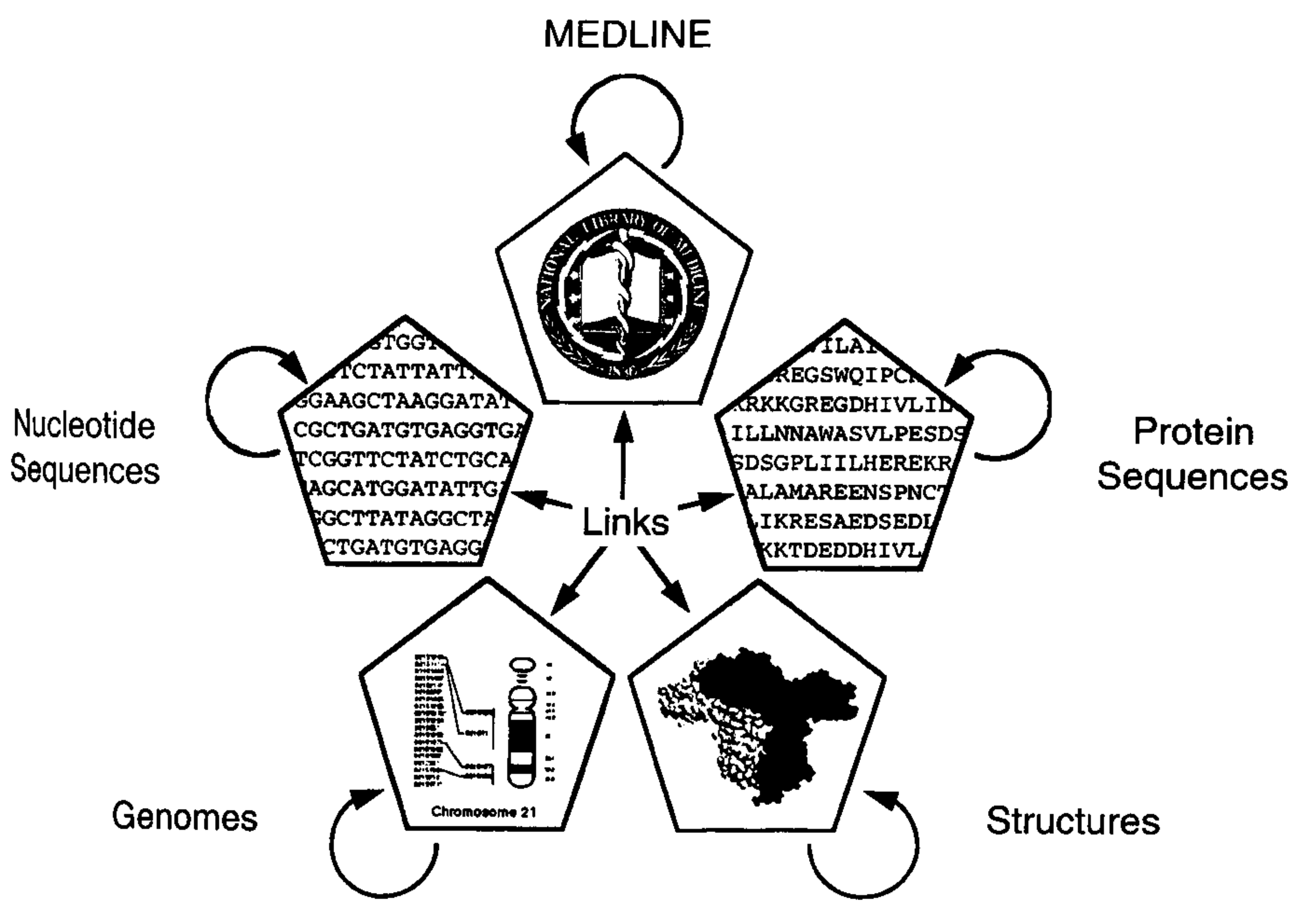
- integrates databases of NCBI
- less flexible then SRS
- valuable concept of neighbouring
- link databanks: DNA and protein sequences, genome data, structural data, PubMed bibliography

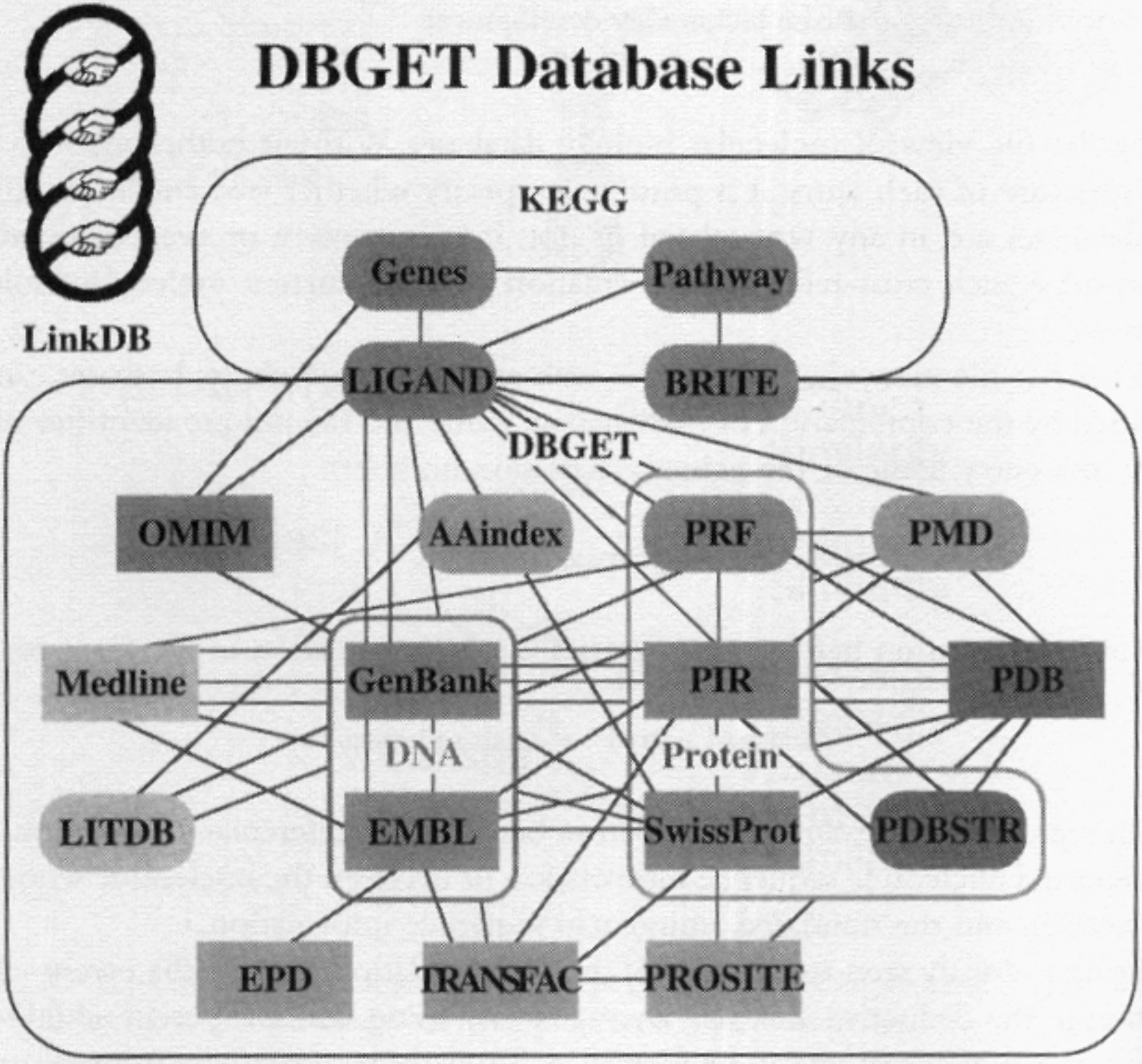
SRS, ENTREZ and LinkDB

LinkDB

- maintained by Institute for Chemical Reseach, Japan
- network browser for databases in DBGET and KEGG (Kyoto encyclopedia of genes and genomes)
- link databanks: sequence, motifs, structure, amino acid properties, ligands, metabolic pathways


DSSP
ALI
HSSP
ENZYME
REBASE
PROSITEDOC
FSSP
PDB
SWISSPROT
PROSITE
PFAM
OMIM
SWISSPFAM
NRL3D
YPD
MIMMAP
BLOCKS
YPDREF
FLYGENE
SWISSNEW
PRODOM
FLYREFS
PRINTS
IMGT
ECDC
FLYCLONES
SWISSDOM
SBASE
EC2D
PIR
GBNEW
PIRALN
OWL
LISTAHOP
TFSITE
GENBANK
EMBL
EPD
LISTA
GBONLY
TFFACTOR
TREMBL
LISTAHON
CPGISLE
DBSTS
DBEST
TFMATRIX
VECTOR
EMNEW
EMEST
TREMBLNEW
UNIEST
SEQANALREF
BIOCAT
NAKAI
NAKAIMATRIX
LIMB
SEQANALRABS


MEDLINE
Nucleotide Sequences
Protein Sequences
Links
Genomes
Structures
Chromosome 21


DBGET Database Links
LinkDB
KEGG
Genes
Pathway
LIGAND
BRITE
DBGET
OMIM
AAindex
PRF
PMD
Medline
GenBank
PIR
PDB
DNA
Protein
LITDB
EMBL
SwissProt
PDBSTR
EPD
TRANSFAC
PROSITE