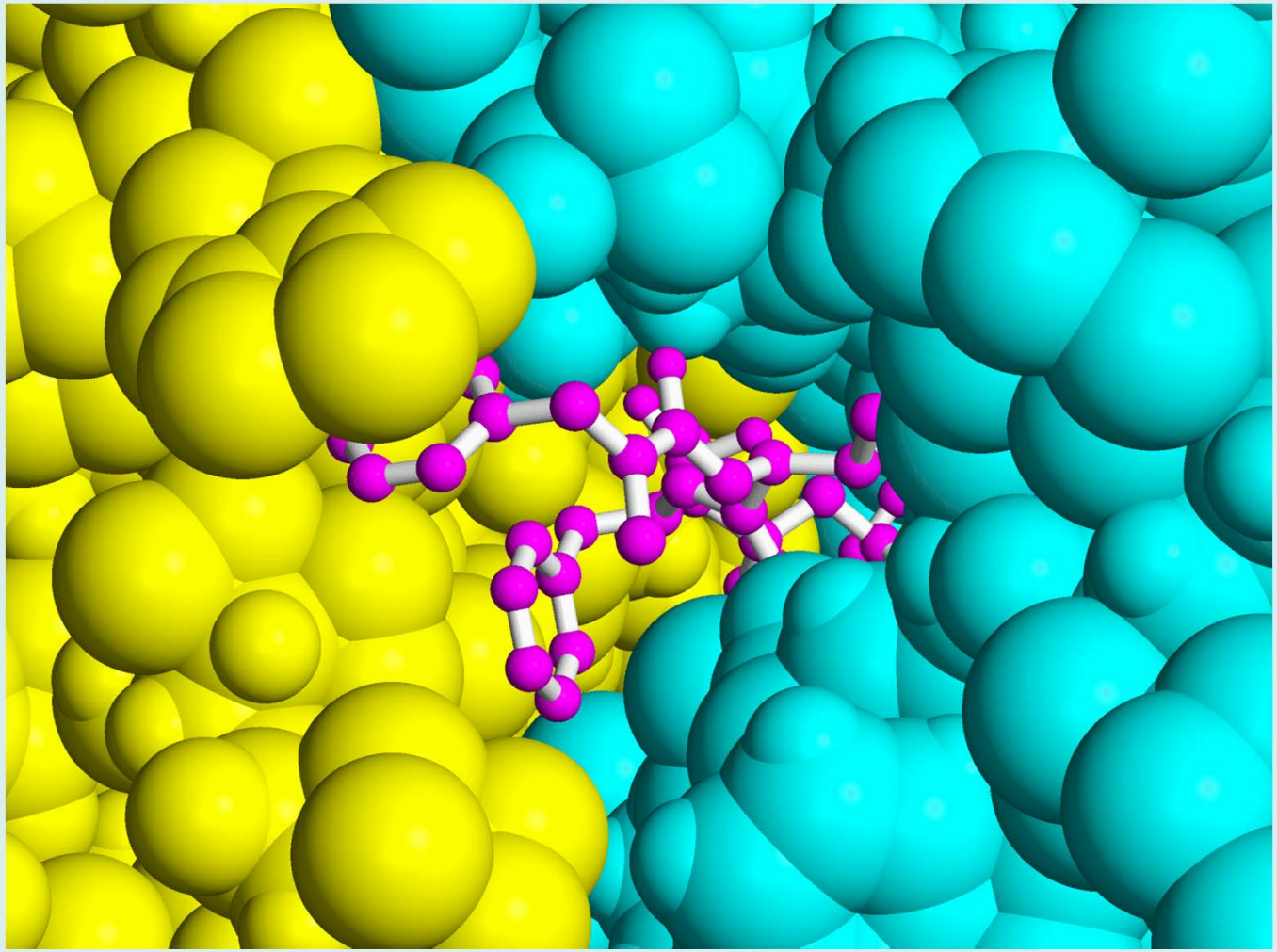
DATABASE INTEGRATION

Genes…..


agggcttggaaag
ctttttgctataag
tgggtggcaagtg
tcaaaacgtagta

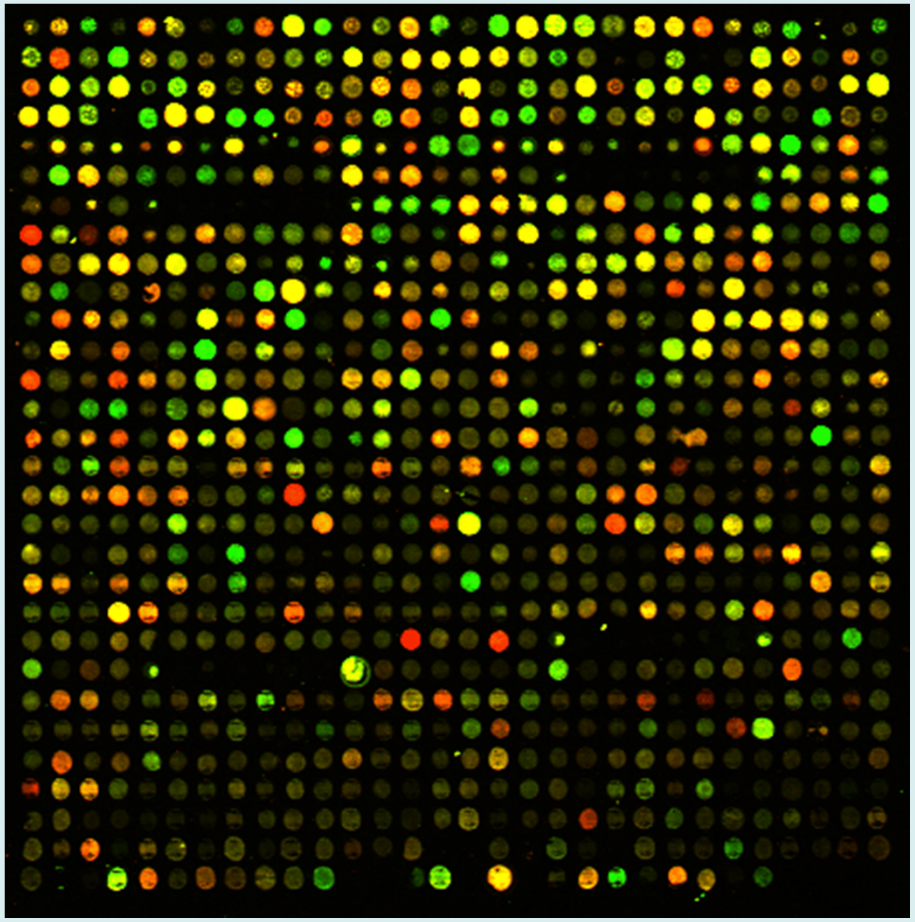
…….make proteins

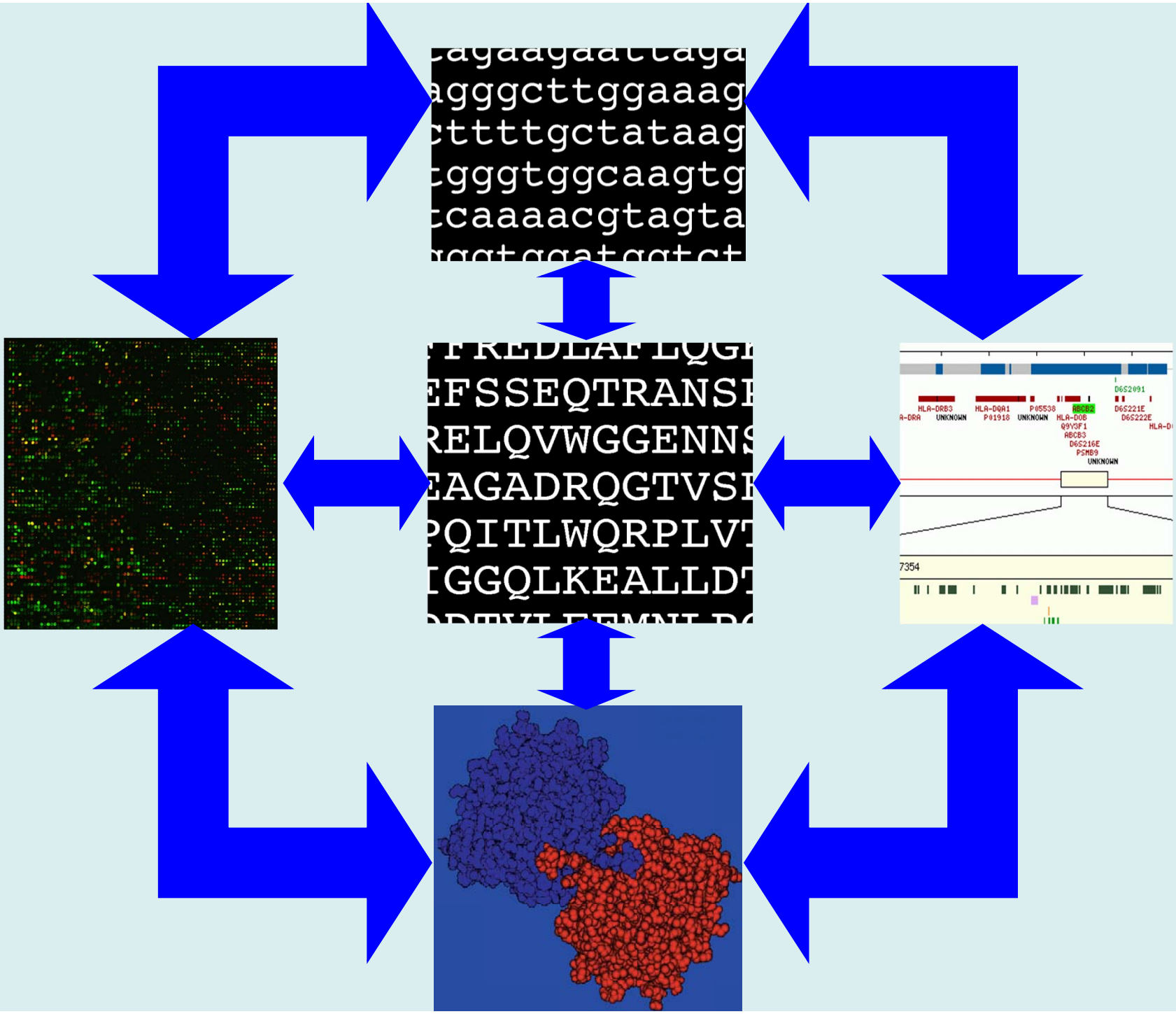
Proteins form complex 3D structures…

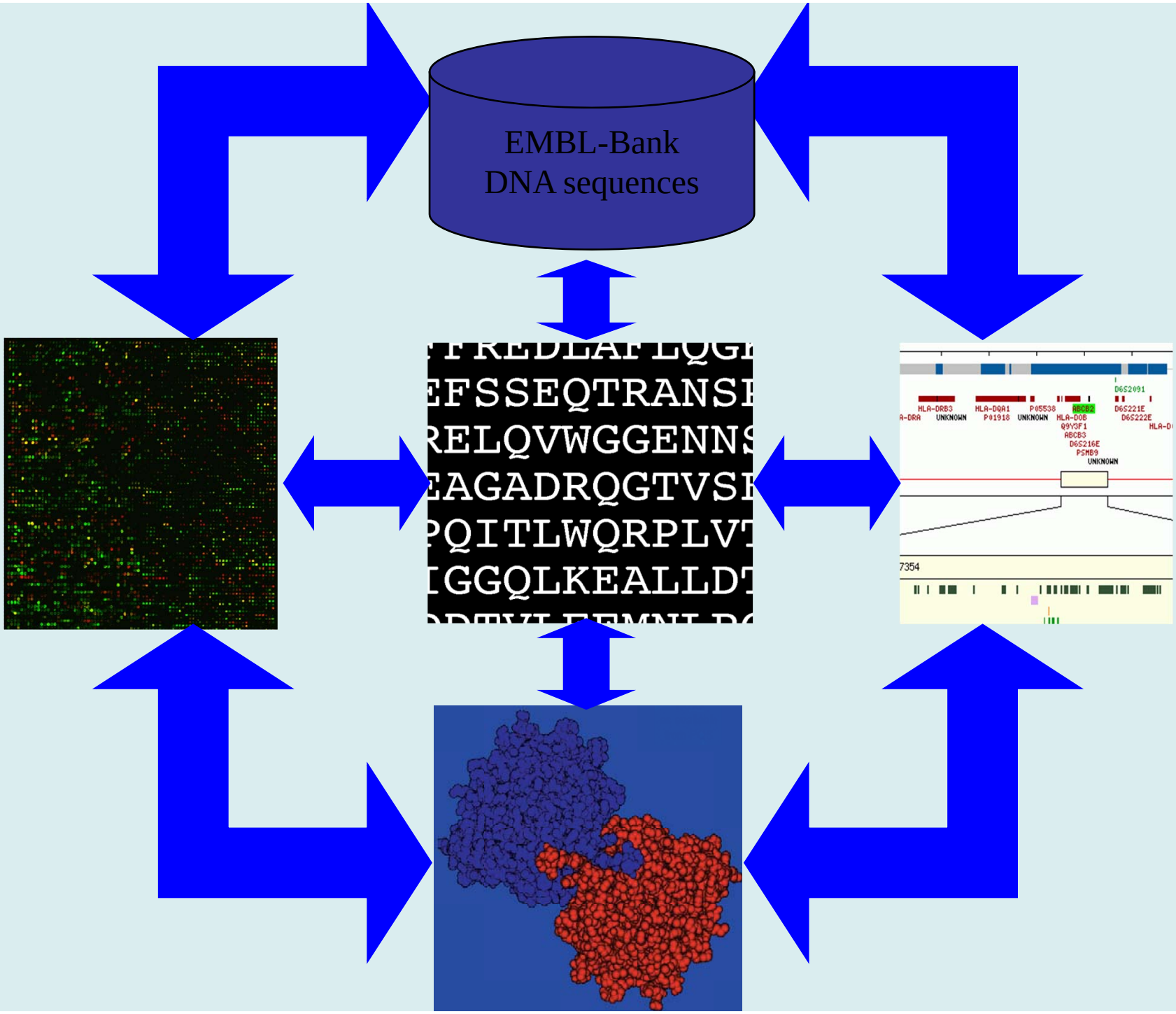


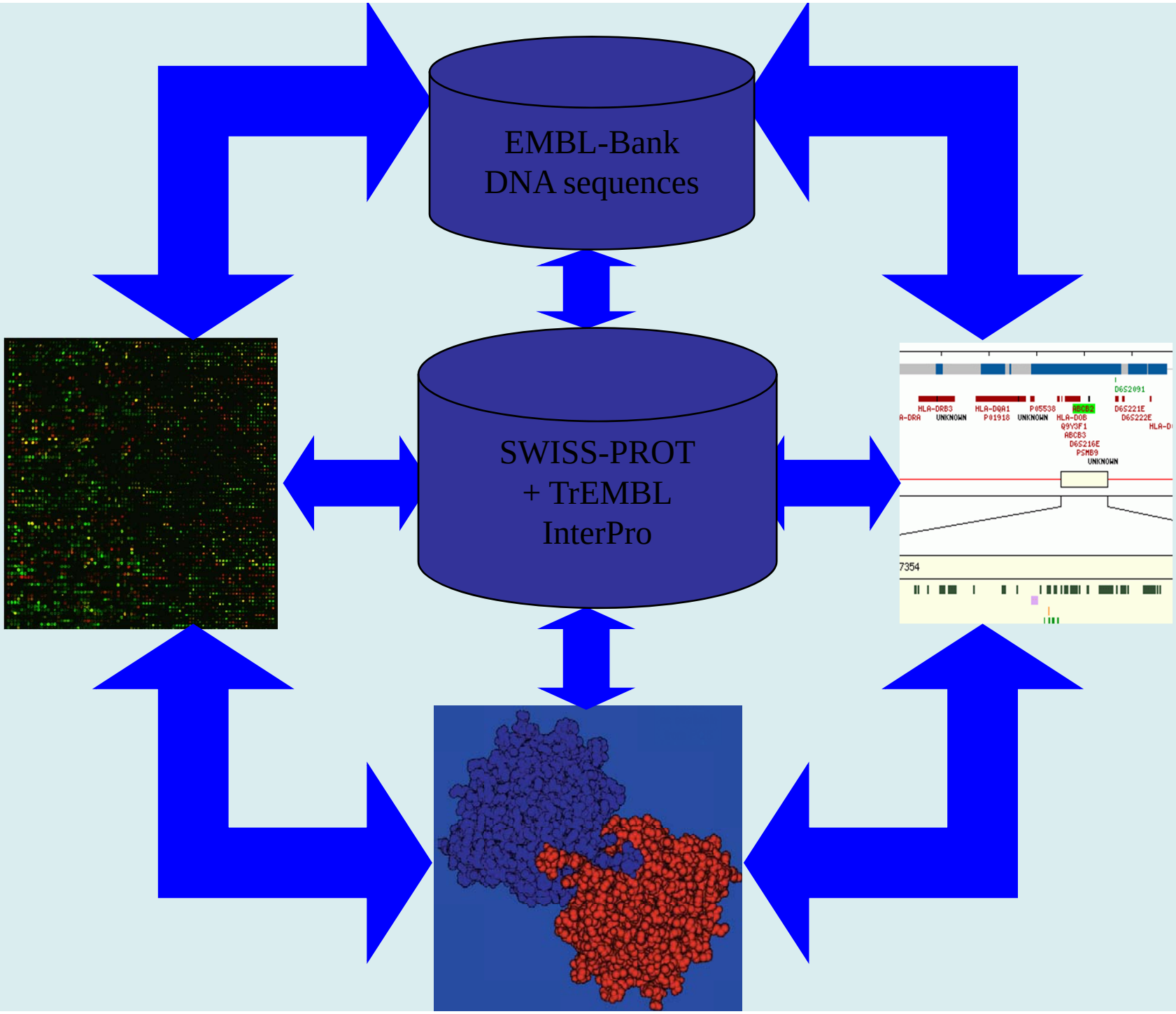
… and molecules interact

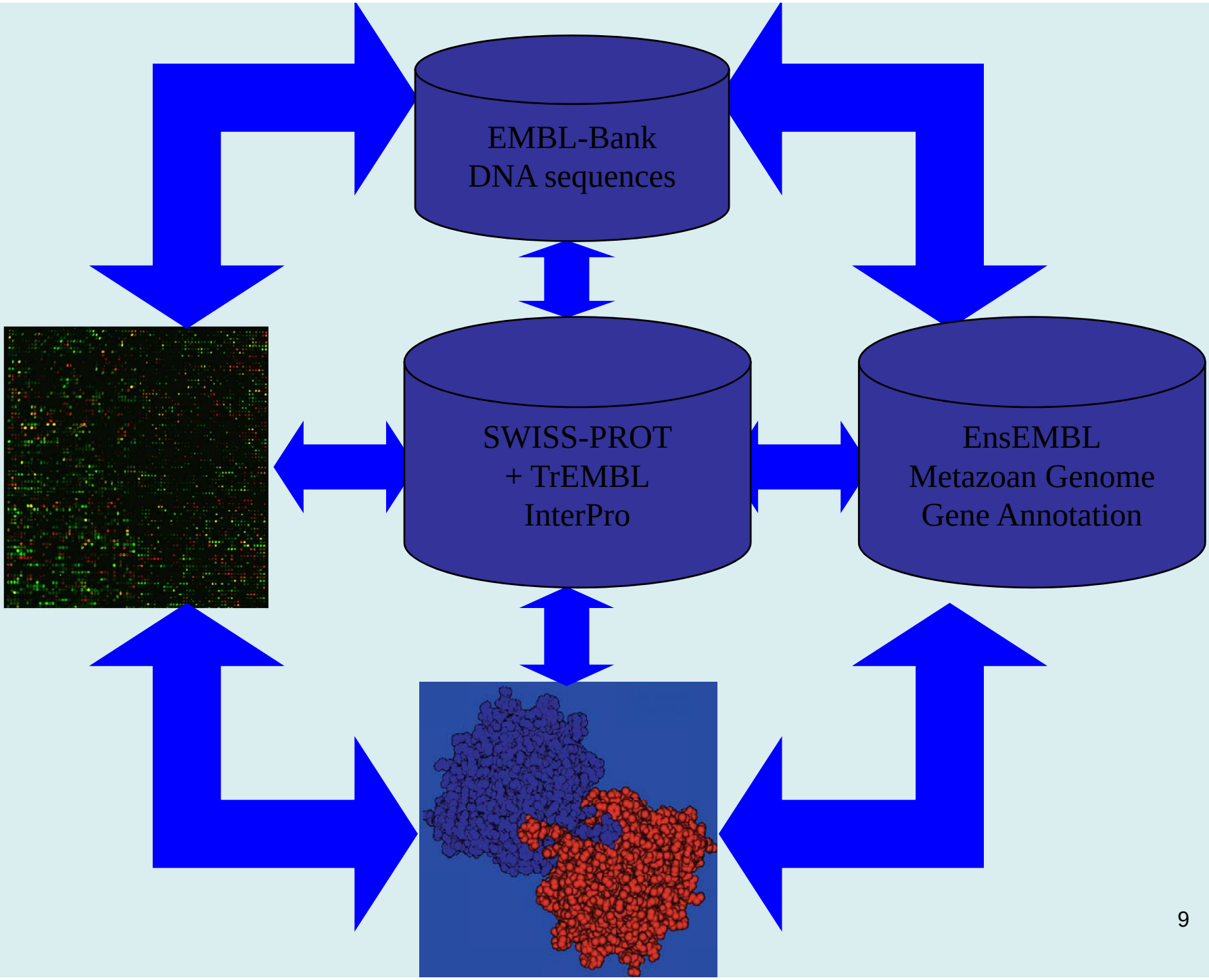
the right molecules need to be present at the right time

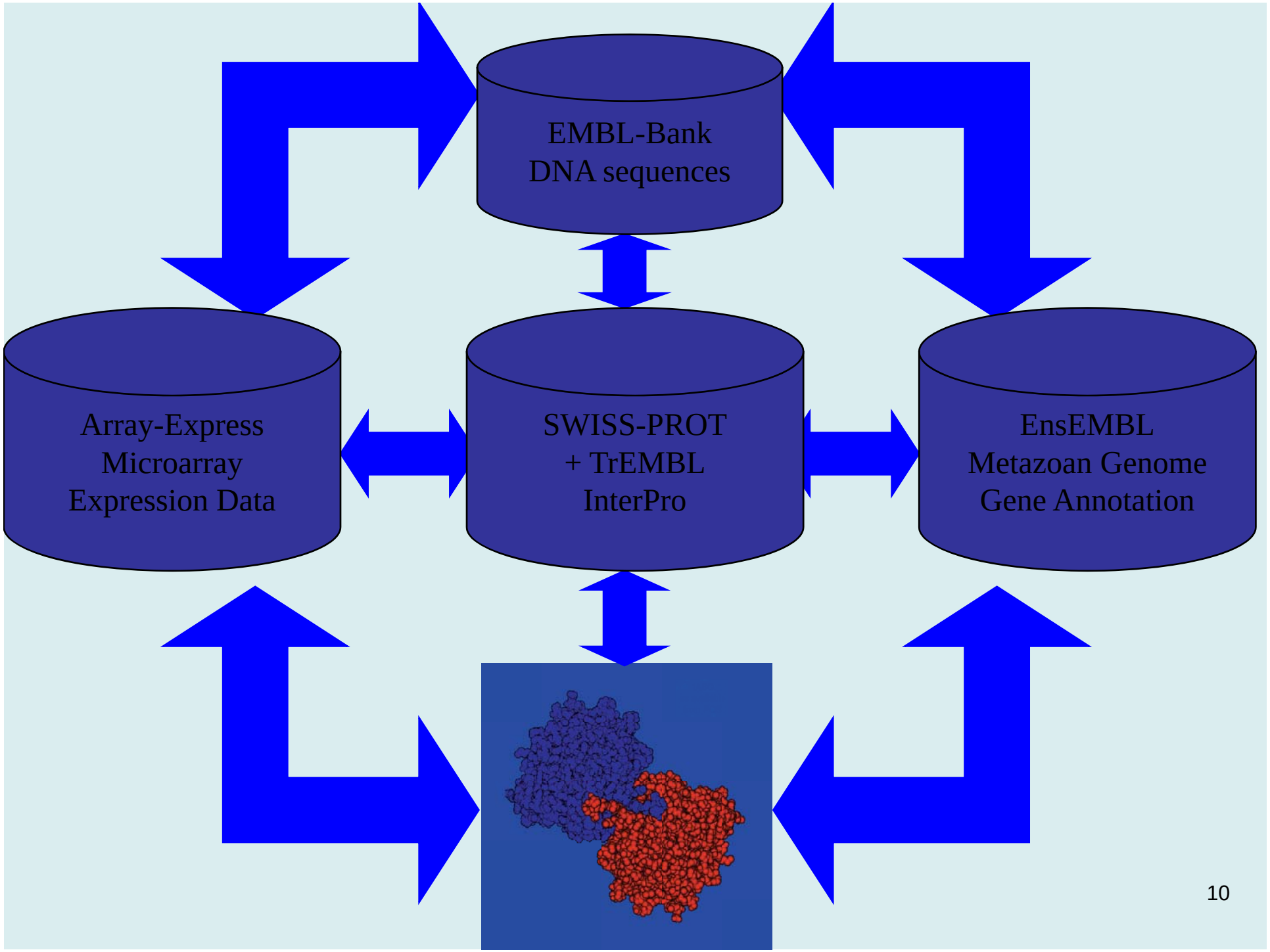


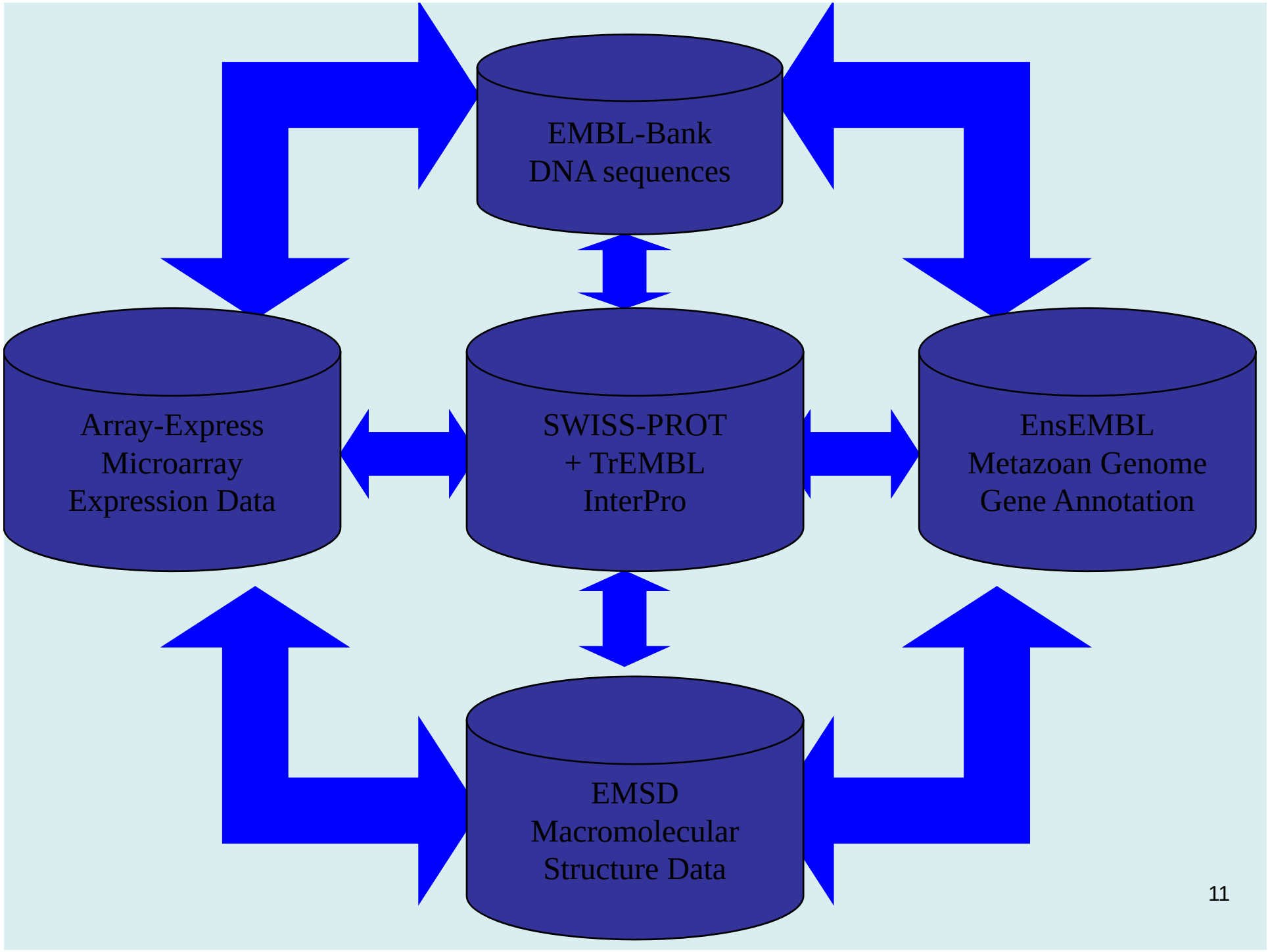

agggcttggaaag
ctttgctataag
tgggtggcaagtg
tcaaaacgtagta
EFSSEQTRANSI
RELQVWGGENNS
EAGADRQGTVSI
PQITLWQRPLVT
IGGQLKEALLDT
D6S2091
HLA-DRB3
UNKNOWN
HLA-DQA1
P01918
P05538
UNKNOWN
ABCB2
HLA-DOB
Q9Y3F1
ABCB3
D6S216E
PSMB9
UNKNOWN
D6S221E
D6S222E
7354

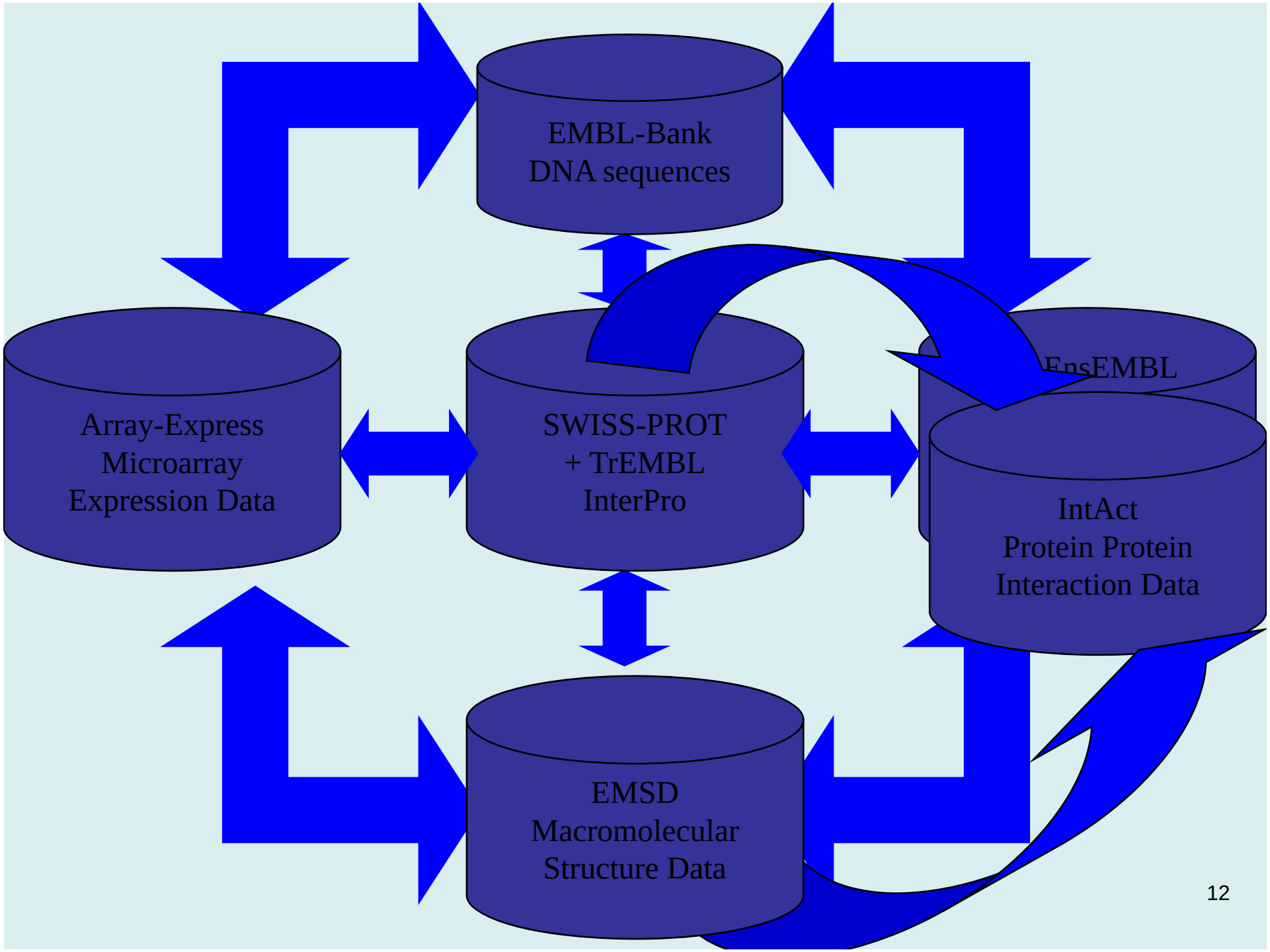

EMBL-Bank
DNA sequences
EFSSEQTRANSI
RELQVWGGENNS
EAGADRQGTVSI
PQITLWQRPLVT
IGGQLKEALLDT
D6S2091
HLA-DRB3
HLA-DQA1
P05538
ABCB2
D6S221E
UNKNOWN
P01918
UNKNOWN
HLA-DOB
D6S222E
Q9Y3F1
ABCB3
D6S216E
PSMB9
UNKNOWN
7354


EMBL-Bank
DNA sequences
SWISS-PROT
+ TrEMBL
InterPro
D6S2091
HLA-DRB3
HLA-DQA1
P05538
ABCB2
D6S221E
UNKNOWN
P01918
UNKNOWN
HLA-DOB
D6S222E
Q9Y3F1
ABCB3
D6S216E
PSMB9
UNKNOWN
7354


EMBL-Bank
DNA sequences
SWISS-PROT
+ TrEMBL
InterPro
EnsEMBL
Metazoan Genome
Gene Annotation


EMBL-Bank
DNA sequences
Array-Express
Microarray
Expression Data
SWISS-PROT
+ TrEMBL
InterPro
EnsEMBL
Metazoan Genome
Gene Annotation


EMBL-Bank
DNA sequences
Array-Express
Microarray
Expression Data
SWISS-PROT
+ TrEMBL
InterPro
EnsEMBL
Metazoan Genome
Gene Annotation
EMSD
Macromolecular
Structure Data


EMBL-Bank
DNA sequences
Array-Express
Microarray
Expression Data
SWISS-PROT
+ TrEMBL
InterPro
EnsEMBL
IntAct
Protein Protein
Interaction Data
EMSD
Macromolecular
Structure Data

EMBnet

Developed in the 80s to link European laboratories which used bioinformatics as a research tool in Biology.

Takes advantage of the Internet potential for global communication and resources centralization.

EMBnet Members

Argentina
Australia
Austria
Belgium
Brasil
Canada
Chile
China
Colombia
Cuba
Denmark
Finland
France
Germany

Greece
Hungary
India
Ireland
Israel
Italy
Mexico
Netherlands
Norway
Poland
Portugal
Russia
Slovakia
South Africa

Spain
Sweden
Switzerland
UK

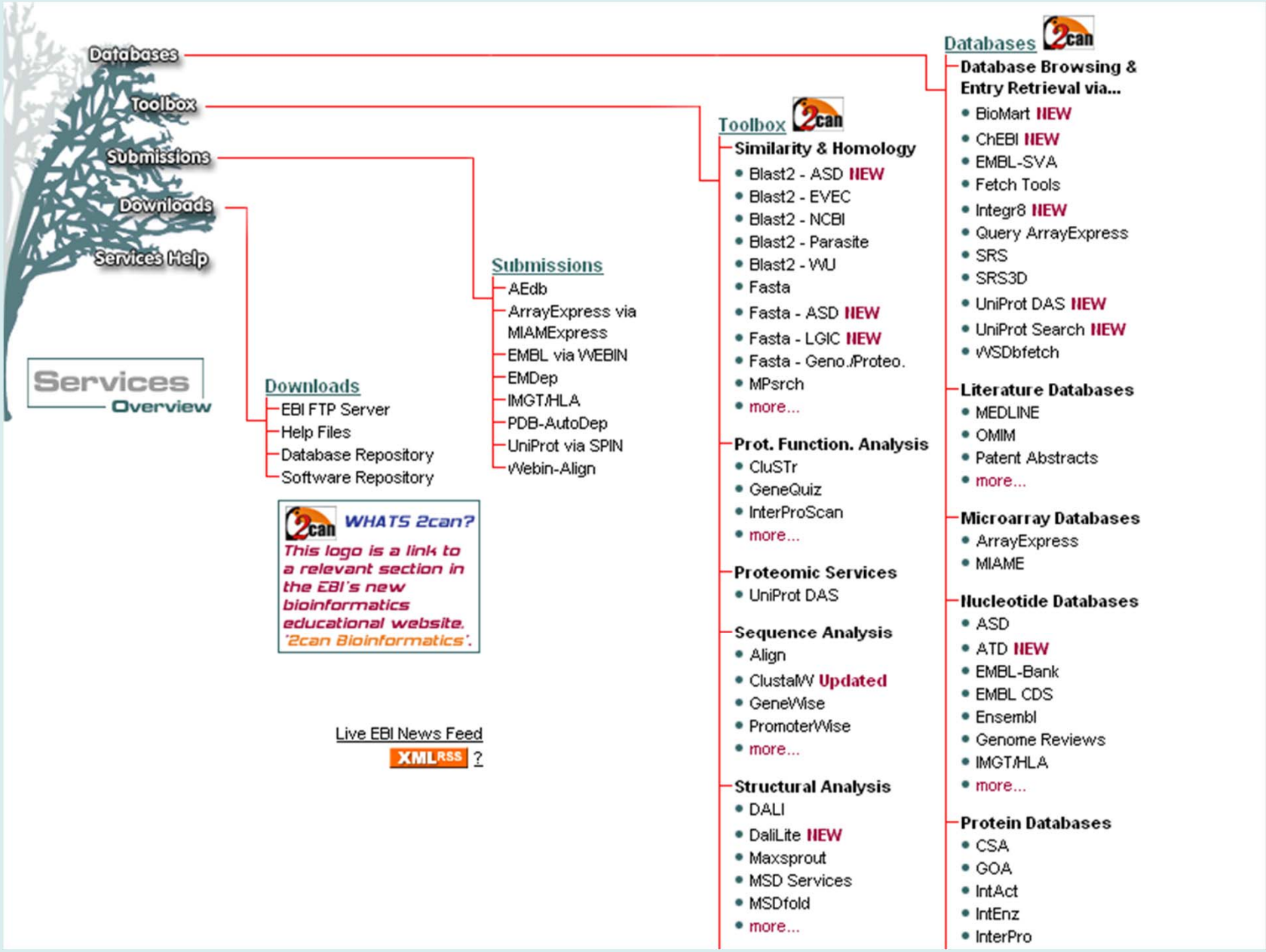
EBI
ETI
ICGEB
MIPS
UMBER
Hoffman-La Roche
LION Bioscience

The National Nodes

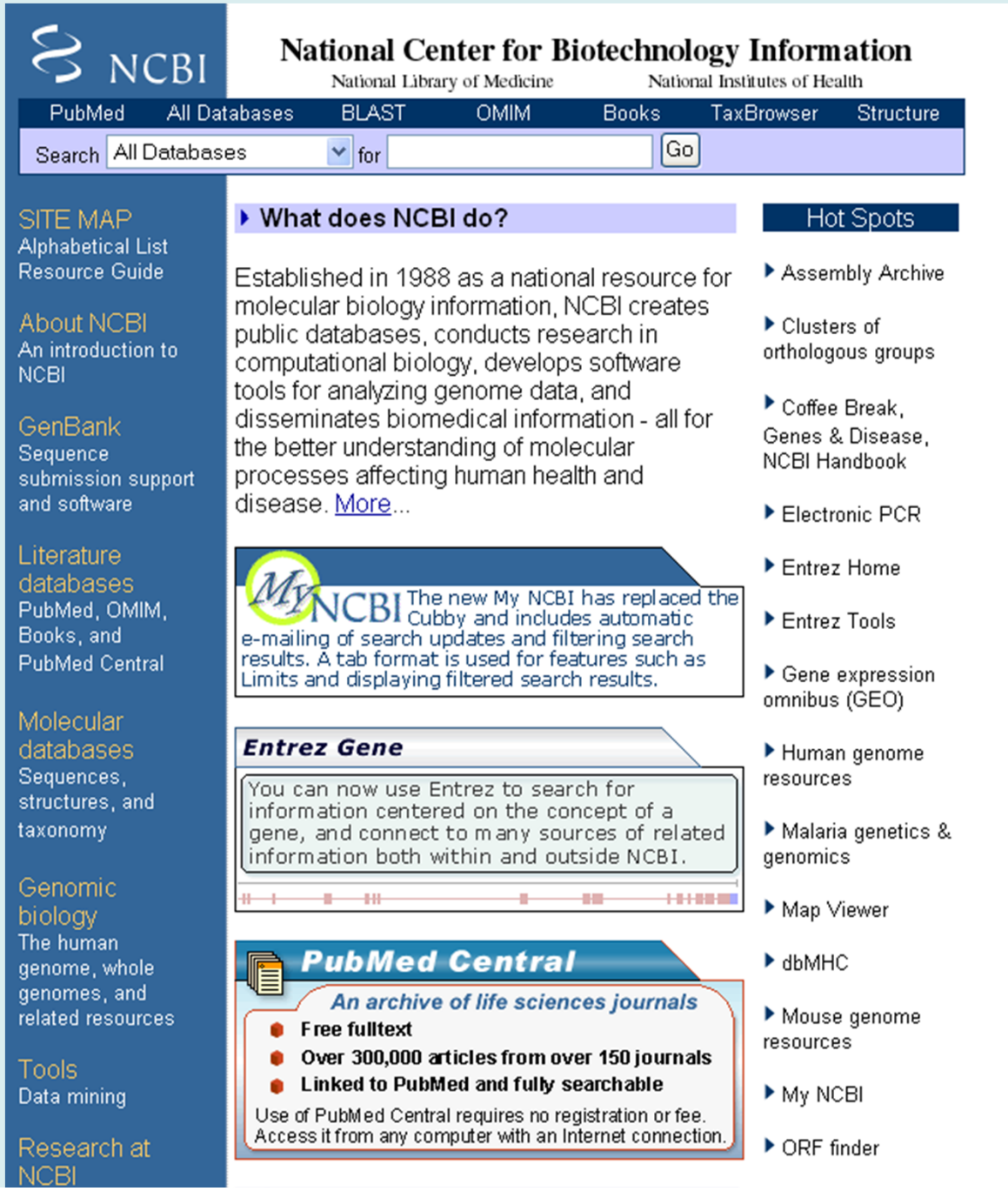
Argentina	Node info		
Australia	Node info	Israel	Node info
Austria	Node info	Italy	Node info
Belgium	Node info	Mexico	Node info
Brazil	Node info	Netherlands (The)	Node info
Canada	Node info	Norway	Node info
Chile	Node info	Poland	Node info
China	Node info	Portugal	Node info
Colombia	Node info	Russia	Node info
Cuba	Node info	Slovakia	Node info
Finland	Node info	South Africa	Node info
France	Node info	Spain	Node info
Germany	Node info	Sweden	Node info
Hungary	Node info	Switzerland	Node info
India	Node info	United Kingdom	Node info

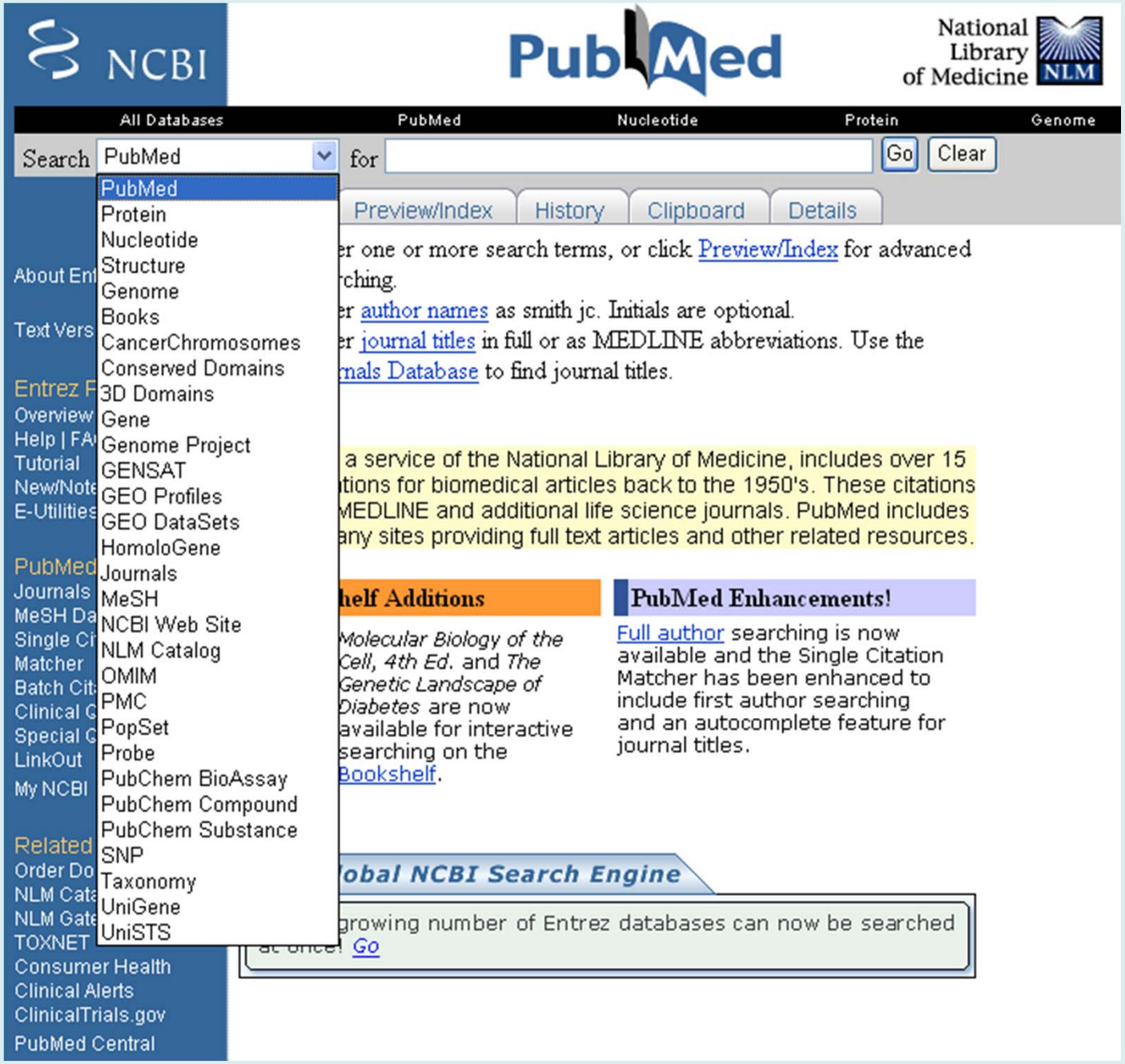
EBI


Database Search for
in
Nucleotide sequences
Go
Search EBI Website
Go
European Bioinformatics Institute
Groups at the European Bioinformatics Institute
Bioinformatics Products and Services
FLASH Intro
European Bioinformatics Institute
a part of the European Molecular Biology Laboratory
the path to knowledge
About EBI | Funding | Whats New | Groups | Services | Toolbox | EBI Databases | Downloads | Submissions
BioMart Database Queries | SRS Database Queries | Site Search | Site Map | Services Map | Contact Us | Terms of Use
2can
2can Bioinformatics - Training and Education at the EBI
XML RSS
?
© Copyright European Bioinformatics Institute 2002-2004. All Rights and Trademarks Reserved.


Databases
Toolbox
Submissions
Downloads
Services Help
Services
Overview
Downloads
EBI FTP Server
Help Files
Database Repository
Software Repository
Submissions
AEdb
ArrayExpress via MIAMExpress
EMBL via WEBIN
EMDep
IMGT/HLA
PDB-AutoDep
UniProt via SPIN
Webin-Align
WHATS 2can?
This logo is a link to a relevant section in the EBI's new bioinformatics educational website, '2can Bioinformatics'.
Live EBI News Feed
XMLRSS ?
Toolbox
Similarity & Homology
Blast2 - ASD NEW
Blast2 - EVEC
Blast2 - NCBI
Blast2 - Parasite
Blast2 - WU
Fasta
Fasta - ASD NEW
Fasta - LGIC NEW
Fasta - Geno./Proteo.
MPsrch
more...
Prot. Function. Analysis
CluSTr
GeneQuiz
InterProScan
more...
Proteomic Services
UniProt DAS
Sequence Analysis
Align
ClustalW Updated
GeneWise
PromoterWise
more...
Structural Analysis
DALI
DaliLite NEW
Maxsprout
MSD Services
MSDfold
more...
Databases
Database Browsing & Entry Retrieval via...
BioMart NEW
ChEBI NEW
EMBL-SVA
Fetch Tools
Integr8 NEW
Query ArrayExpress
SRS
SRS3D
UniProt DAS NEW
UniProt Search NEW
WSDbfetch
Literature Databases
MEDLINE
OMIM
Patent Abstracts
more...
Microarray Databases
ArrayExpress
MIAME
Nucleotide Databases
ASD
ATD NEW
EMBL-Bank
EMBL CDS
Ensembl
Genome Reviews
IMGT/HLA
more...
Protein Databases
CSA
GOA
IntAct
IntEnz
InterPro

NCBI

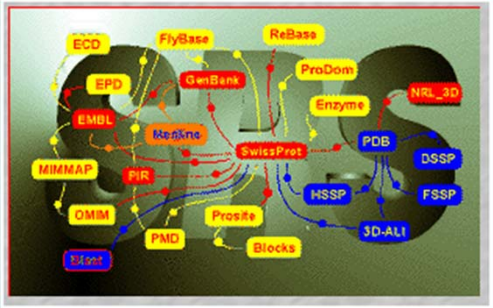

NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed
All Databases
BLAST
OMIM
Books
TaxBrowser
Structure
Search
All Databases
for
Go
SITE MAP
Alphabetical List
Resource Guide
About NCBI
An introduction to NCBI
GenBank
Sequence submission support and software
Literature databases
PubMed, OMIM, Books, and PubMed Central
Molecular databases
Sequences, structures, and taxonomy
Genomic biology
The human genome, whole genomes, and related resources
Tools
Data mining
Research at NCBI
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
MyNCBI
The new My NCBI has replaced the Cubby and includes automatic e-mailing of search updates and filtering search results. A tab format is used for features such as Limits and displaying filtered search results.
Entrez Gene
You can now use Entrez to search for information centered on the concept of a gene, and connect to many sources of related information both within and outside NCBI.
PubMed Central
An archive of life sciences journals
Free fulltext
Over 300,000 articles from over 150 journals
Linked to PubMed and fully searchable
Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.
Hot Spots
Assembly Archive
Clusters of orthologous groups
Coffee Break, Genes & Disease, NCBI Handbook
Electronic PCR
Entrez Home
Entrez Tools
Gene expression omnibus (GEO)
Human genome resources
Malaria genetics & genomics
Map Viewer
dbMHC
Mouse genome resources
My NCBI
ORF finder


NCBI
PubMed
National Library of Medicine
NLM
All Databases
PubMed
Nucleotide
Protein
Genome
Search PubMed for Go Clear
PubMed
Protein
Nucleotide
Structure
Genome
Books
CancerChromosomes
Conserved Domains
3D Domains
Gene
Genome Project
GENSAT
GEO Profiles
GEO DataSets
HomoloGene
Journals
MeSH
NCBI Web Site
NLM Catalog
OMIM
PMC
PopSet
Probe
PubChem BioAssay
PubChem Compound
PubChem Substance
SNP
Taxonomy
UniGene
UniSTS
Preview/Index
History
Clipboard
Details
PubMed Enhancements!
Full author searching is now available and the Single Citation Matcher has been enhanced to include first author searching and an autocomplete feature for journal titles.
Text Vers
Overview
Tutorial
Journals
Matcher
LinkOut
My NCBI
Order Do
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

How to get to data

There are several systems to retrieve resources:

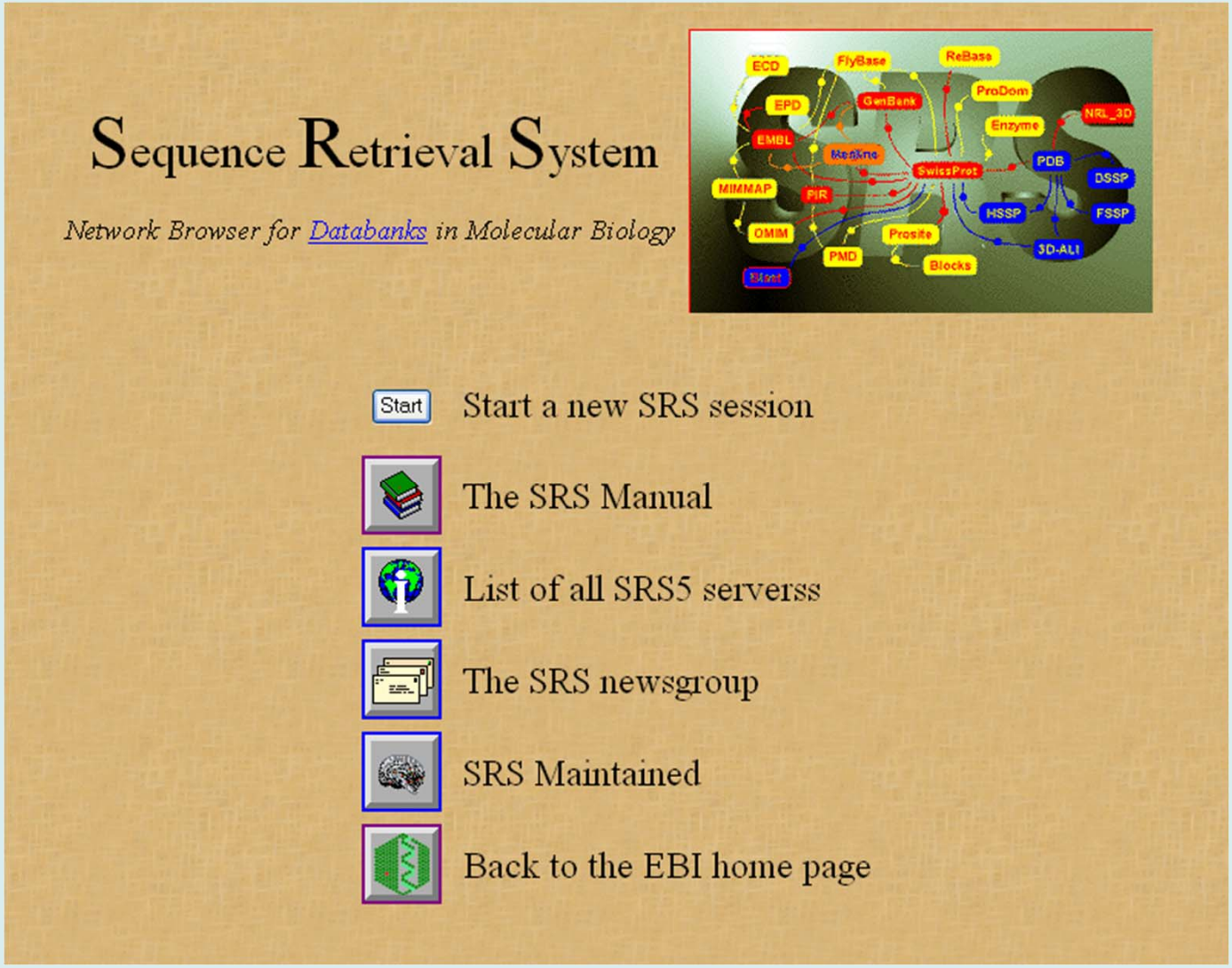
SRS



ENTREZ

They are **integrated systems to retrieve sequences**, based on text search, to be used with the main databases, including PubMed, Nucleotide and Protein Sequences, Protein Structures, Whole Genomes, Taxonomy, etc.

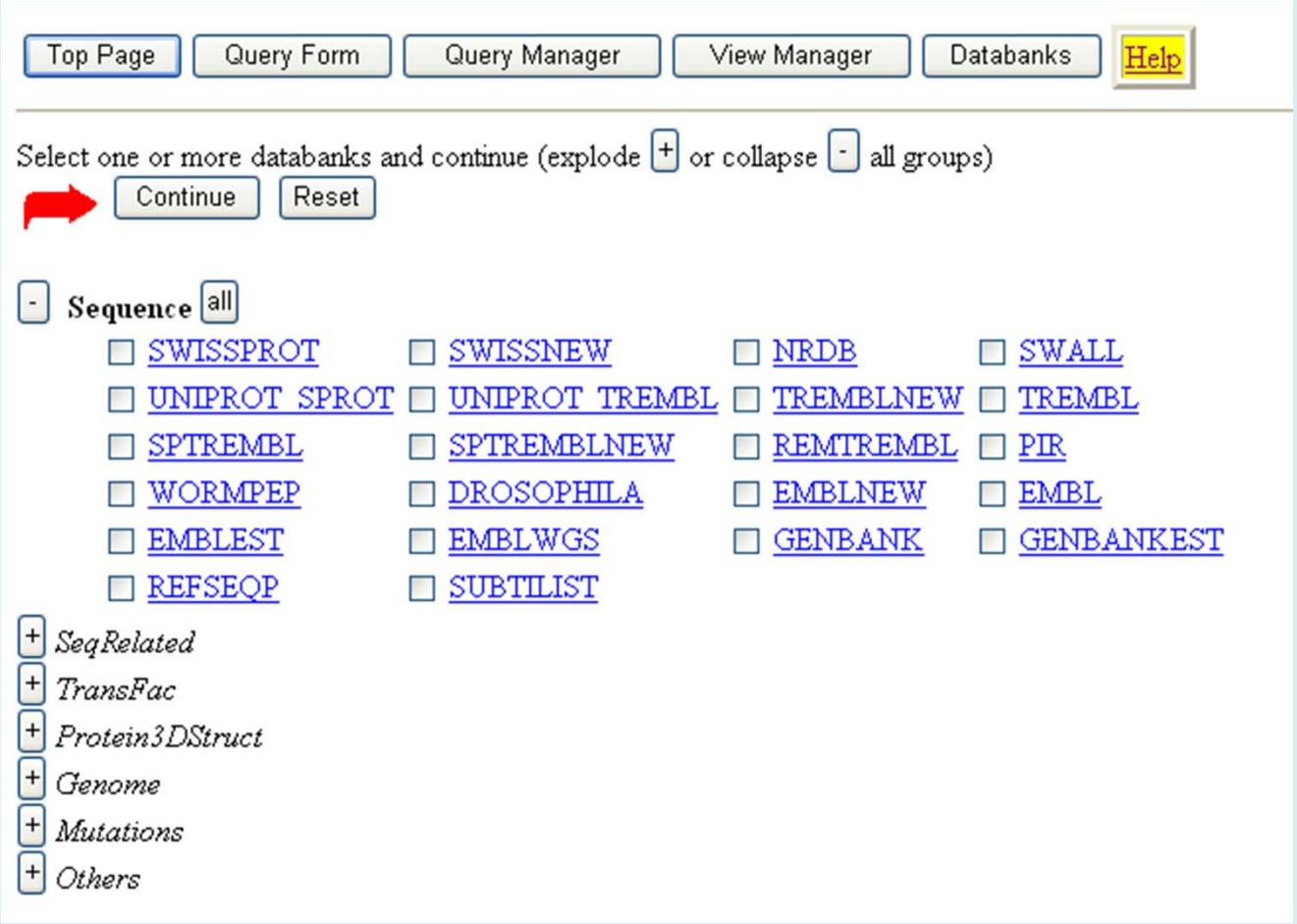
SRS is an **integrated system of sequence retrieval**, created by EBI, to be used with the main databases.



SRSWWW

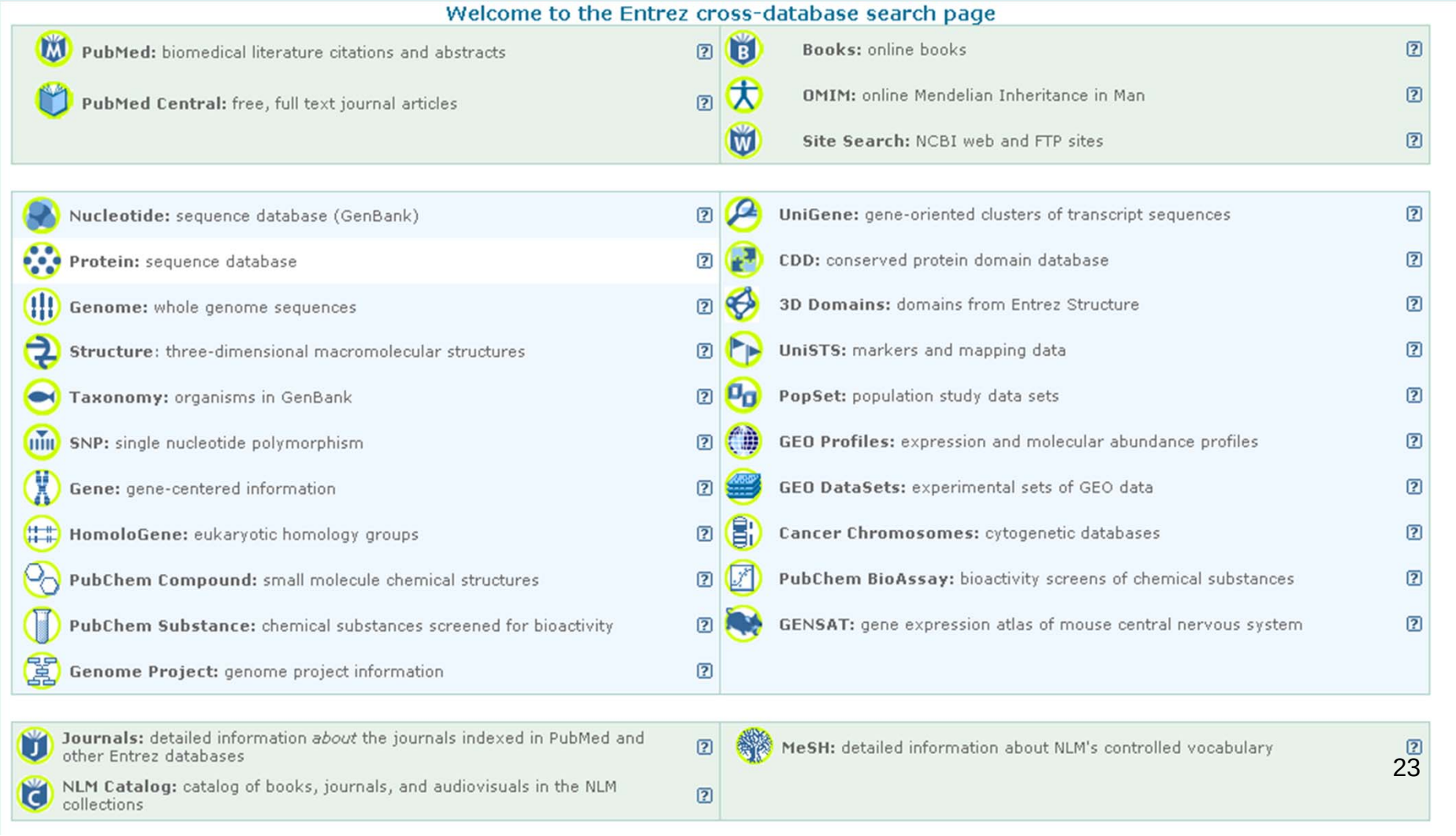
Easy to use

Allows recording of “sessions”

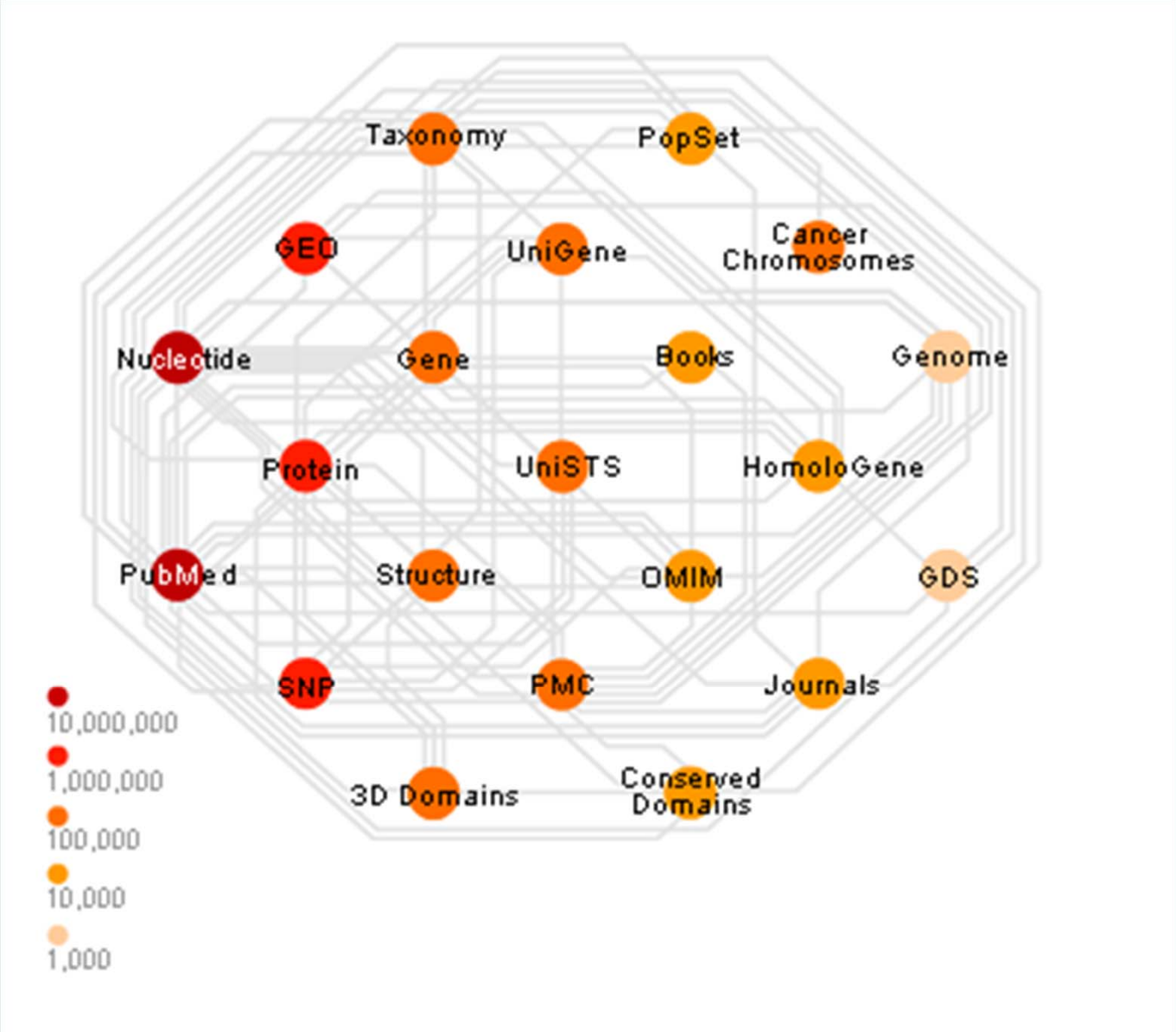


ENTREZ

ENTREZ is an **integrated system of sequence retrieval**, created by NCBI, to be used with the main databases.



DATABASE INTEGRATION WITH ENTREZ



Software Tools For Database Analysis

Software Tools for Sequence Analysis

General Packages:

Packages that offer a comprehensive range of bioinformatics tools for sequence analysis.

Most researchers would expect to use such packages at some time.

Specialised Packages

Packages that offer tools for a particular type of analysis.

Used intensely by researchers in the relevant area, not at all by everyone else.

WWW Resources

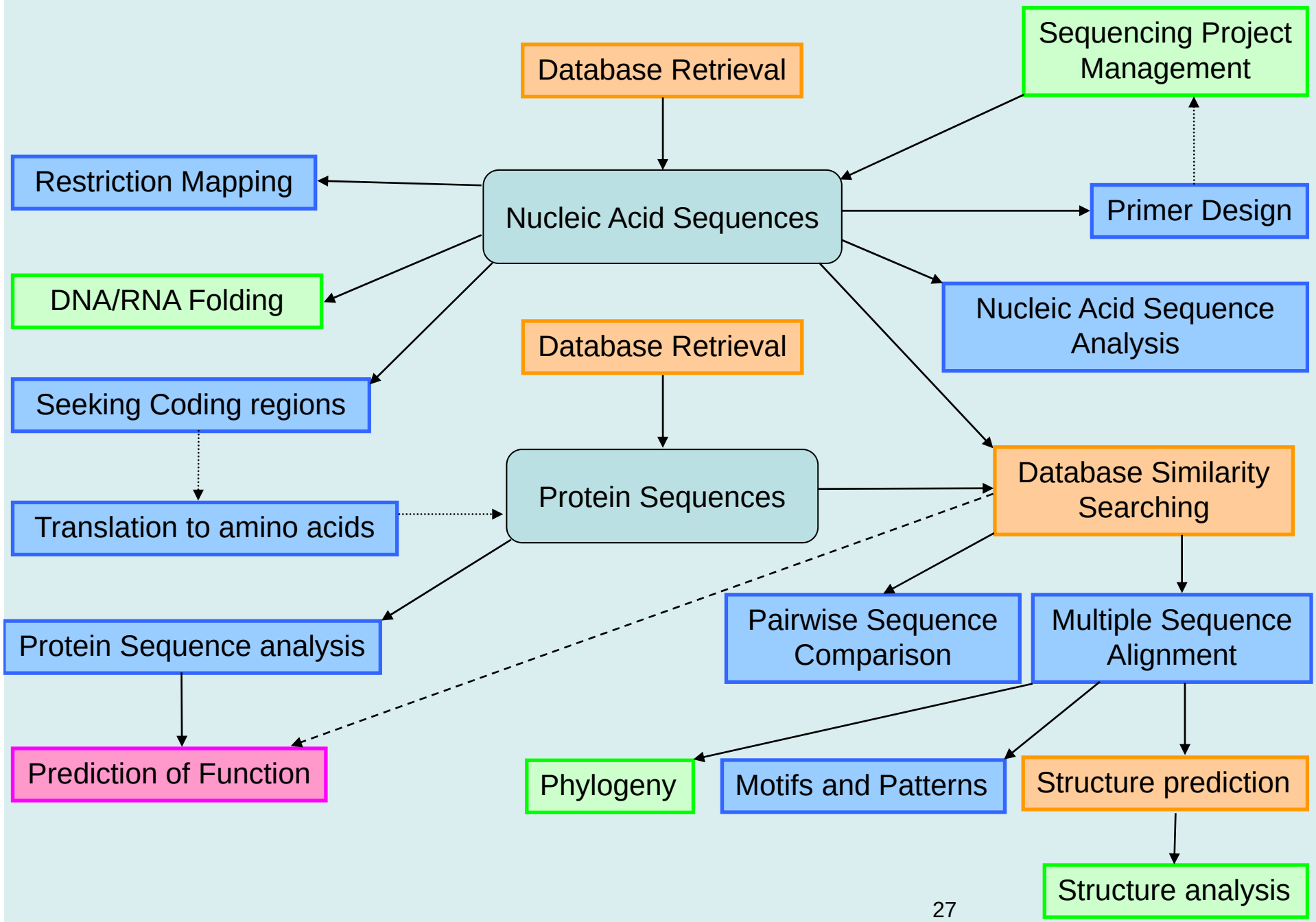
Tools whose nature inclines them to be primarily accessed over the network.

These categorisations are very general

Many specialist programs are incorporated into the general packages.

Most things can be done at a web site somewhere.

Sequence Analysis – an Overview



Software Tools for Sequence Analysis

General Packages:

GCG Wisconsin Package

accelrys

Commercial

WWW and X GUIs

Widely available

UNIX only

Comprehensive

emboss

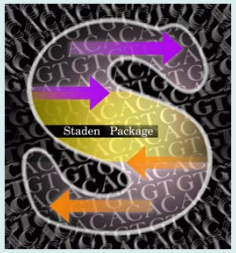
Open source

Several GUIs (java, WWW, X)

Similar structure to the GCG package

UNIX only

Comprehensive



Open source

Excellent GUI including interactive graphical output

Not comprehensive but allows access to EMBOSS

Windows, MacOS X, UNIX

Software Tools for Sequence Analysis

General Packages:

Commercial — Expensive

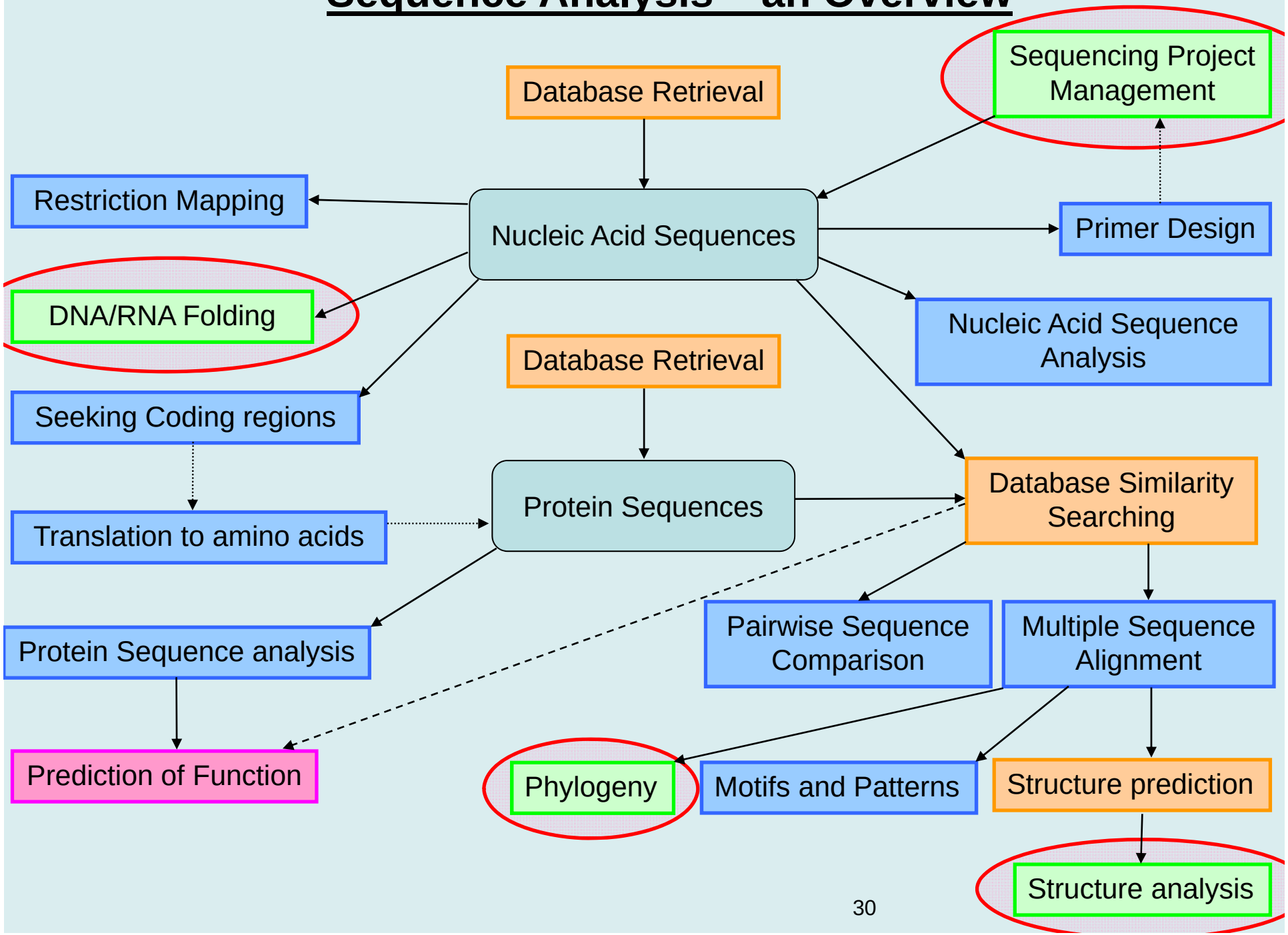
Other options

Windows PCs or Macintoshes — Good GUIs

Public Domain — Windows, Macintosh, UNIX

Modern intuitive GUI — Access remote databases

Sequence Analysis – an Overview

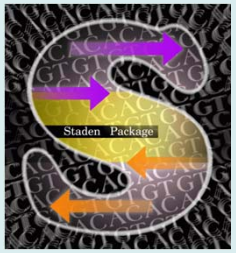


Software Tools for Sequence Analysis

Specialised Packages

Sequencing Project Management

"The Phred - Phrap Package"
By Phil Green et al



Free academic licence

Excellent base call confidence estimation (phred)

Excellent large scale contig assembler (phrap)

Available by anonymous ftp

Excellent GUI

Excellent contig editor

Excellent finishing tools

Simple confidence estimation
Contig assembler – not good for big projects
BUT
phred and phrap can be accessed from Staden GUI

Software Tools for Sequence Analysis

Specialised Packages

DNA/RNA Folding

Michael Zuker`s Programs

Free for academic use

Can be installed locally or run via a WWW page

Incorporated into the GCG general package

Protein Structure Analysis

Whatif by Gert Vriend

Nominal fee for academic use

LINUX, IRIX, Windows

Software Tools for Sequence Analysis

Specialised Packages

Protein Structure Analysis – for very rich people

SYBYL

Tripos

IRIX, HP-UX, LINUX

Insight II

accelrys

IRIX, AIX, LINUX

Both systems are **very** impressive @ **very** expensive

Software Tools for Sequence Analysis

Specialised Packages

Phylogeny

PHYLIP

Available by anonymous ftp

Windows, Macintosh, UNIX

Incorporated into the EMBOSS general package

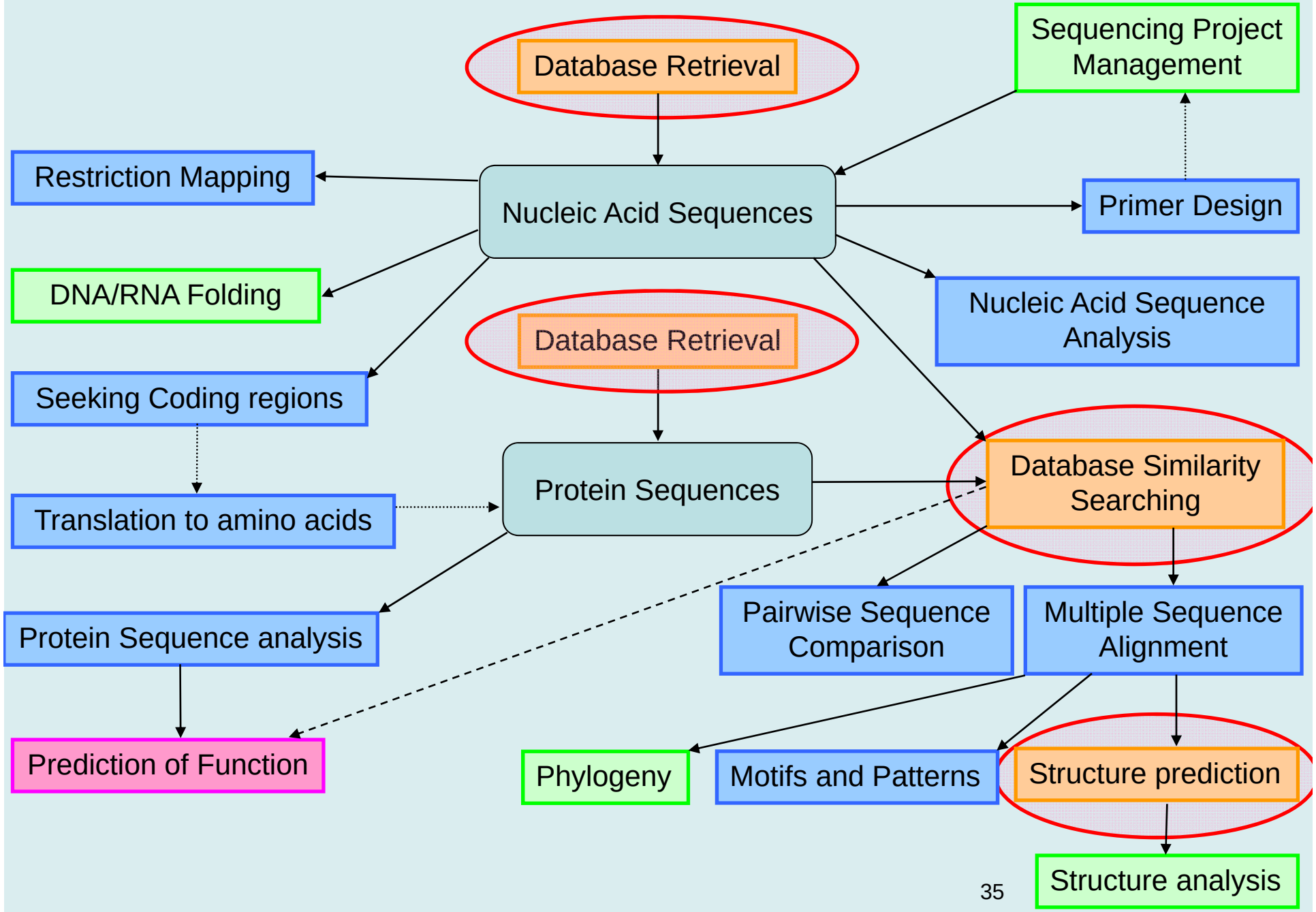
PAUP*

Commercial, but reasonable

UNIX, VMS, DOS and windows

Incorporated into the GCG general package

Sequence Analysis – an Overview



Software Tools for Sequence Analysis

WWW Resources

Database Retrieval

Sequence Retrieval System

Retrieves MUCH more than sequences

SRS

Core elements free to academic sites

LION **Bioscience AG**

Implemented in many places

It is possible to integrate analysis tools

Elements of SRS are incorporated into EMBOSS

Software Tools for Sequence Analysis

WWW Resources

Database Retrieval

Retrieves MUCH more than sequences

Access to NCBI databases only

Entrez client software available by anonymous ftp

Most general packages include tools to access local sequence databases

EMBOSS programs can access sequences from remote **SRS** servers

Software Tools for Sequence Analysis

Database Similarity Searching

WWW Resources

NCBI BLAST

Very popular, very widely available

Not sensitive – But extremely fast

UNIVERSITY of VIRGINIA HEALTH SYSTEM **FASTA**

Popular, widely available

Not sensitive – much slower than blast

Can be installed locally or run via a WWW page

Available by anonymous ftp (blast, fasta)

BOTH blast & fasta

DNA/Protein query **V** DNA/Protein database

Incorporated into the GCG general package

Software Tools for Sequence Analysis

WWW Resources

Structure prediction

The Barton Group

School of Life Sciences, University of Dundee, Dow St, Dundee DD1 5EH, Scotland, UK

JPRED: ING

Was consensus service now **JNet** only

JNet available by anonymous ftp

Burkhard Rost

Older service, similar approach to **JNet**

Main element is called PHD

Both **JPred** and **PHD** work best from aligned protein families

Simpler methods predicting from single sequences in most general packages

Software Tools for Sequence Analysis

WWW Resources

Other WWW services

General Services:

EBI

Pasteur Institute

And many more

Protein sequence analysis

Expasy

Gene finding

genscan at the MIT (Free academic license)

Simple gene finding in most general packages

Primer design

primer3 at the MIT(Available by anonymous ftp)

Primer design in most general packages

Primer design in EMBOSS **is** primer3