



Vytáhl jsem si gen z genomové databáze, analyzoval jsem jeho sekvenci pomocí počítačového programu, odeslal jsem rukopis po internetu a článek mi vyšel v online časopisu. Celá tahle zkušenost ve mně zanechala jakýsi pocit prázdnoty.

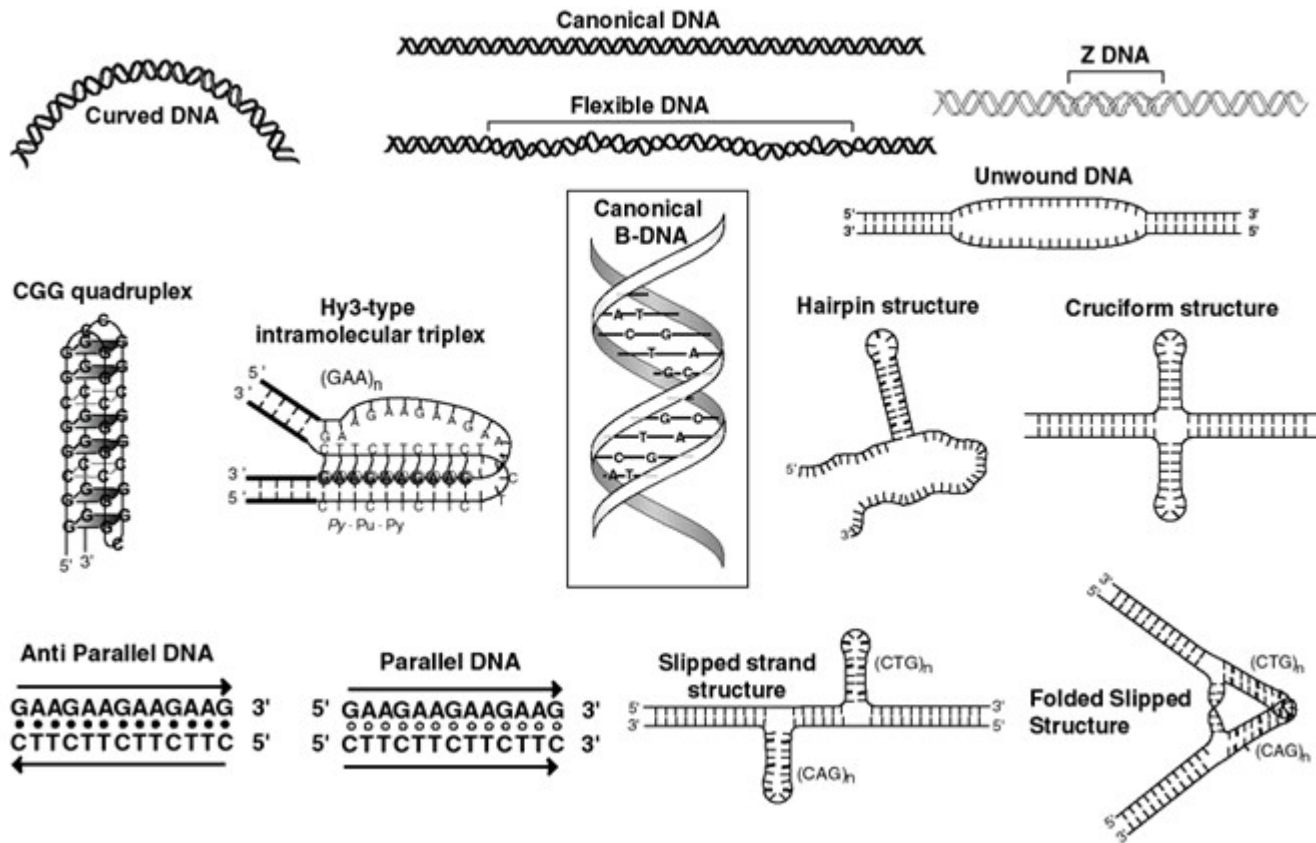
Bioinformatická data

- Sekvence DNA a RNA
- Sekvence proteinů
- Struktura proteinů
- Údaje o aktivitě genů – DNA čip, „microarray“
- Údaje o expresi proteinů – 2-D gely + MS
- Údaje o struktuře DNA
- Mapy interakcí mezi proteiny a DNA
- Mapy interakcí mezi proteiny navzájem
- Literatura

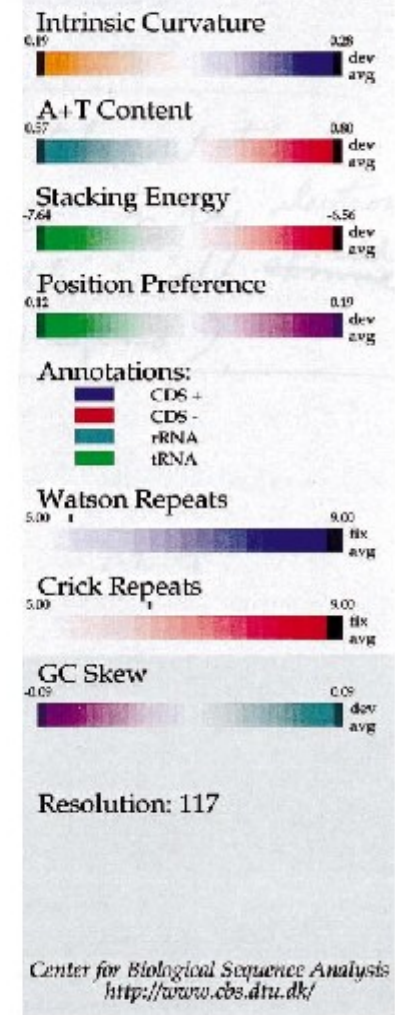
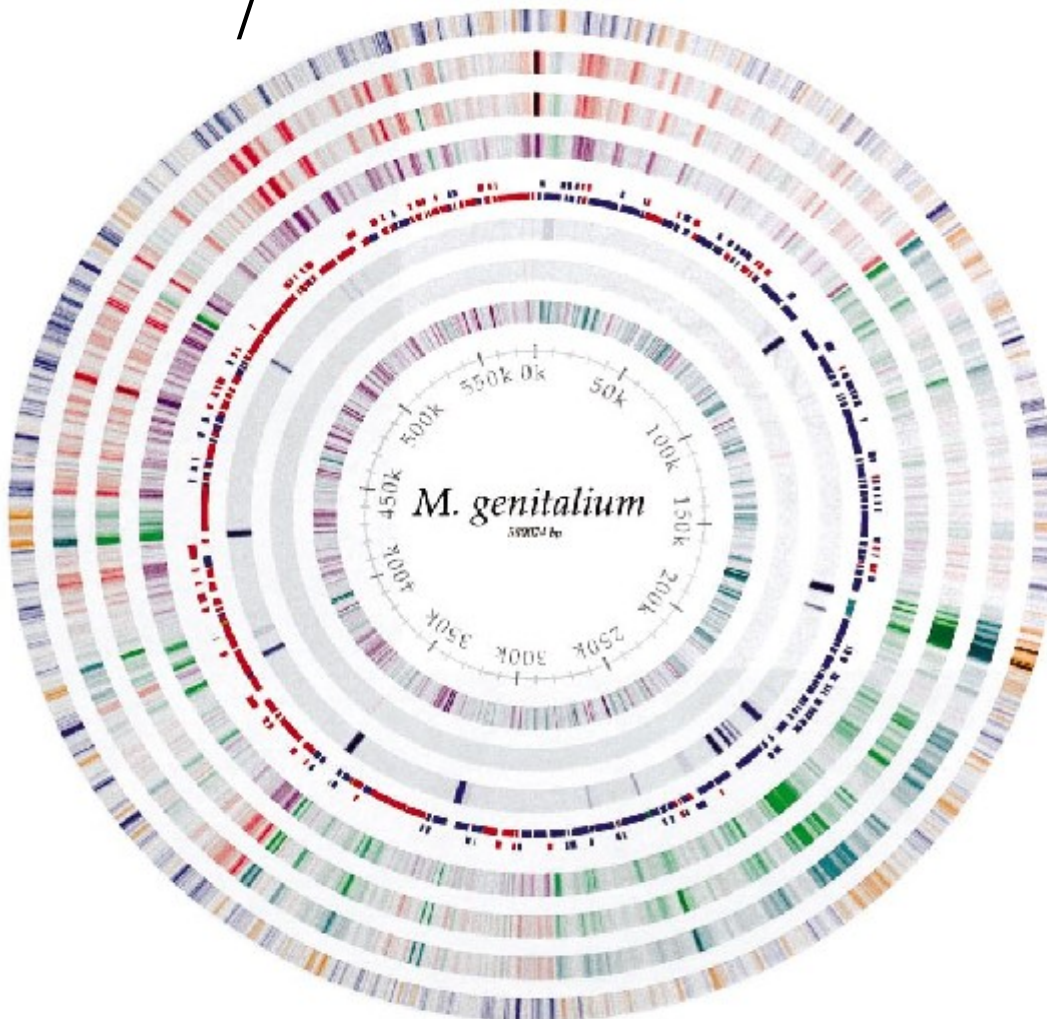
Struktura DNA ?

Specifické nukleotidy a jejich různá zvláštní opakování dávají molekule DNA různé vlastnosti.

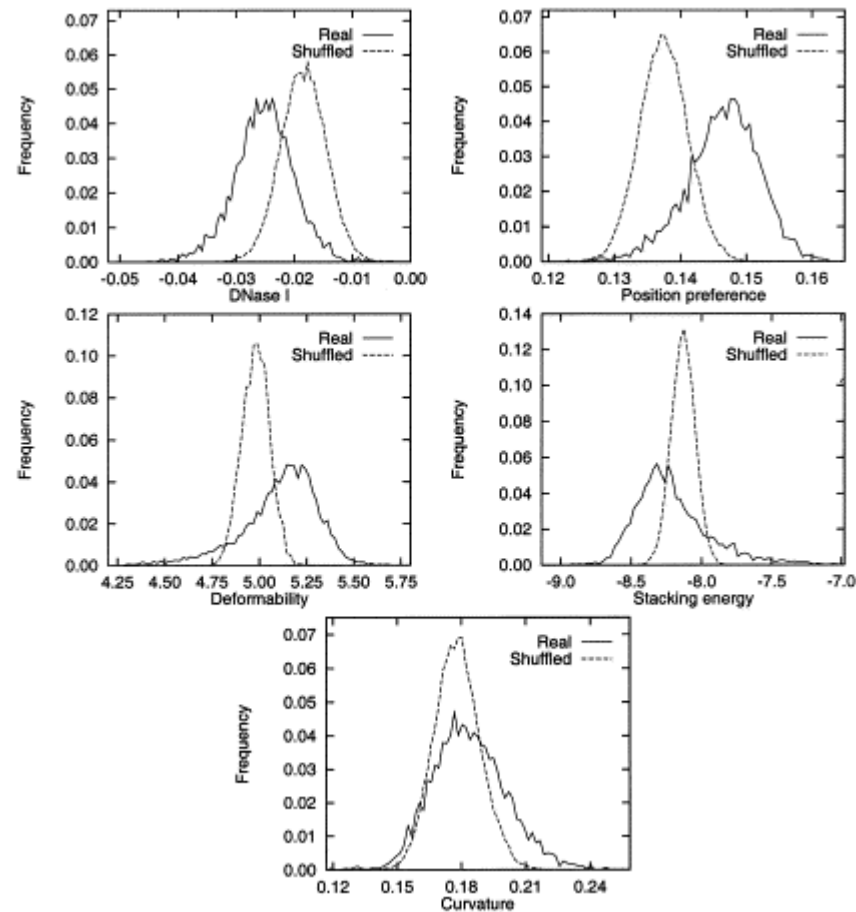
Například vyšší podíl bazí A a T snižuje počet vodíkových můstků mezi vlákny DNA – snadnější denaturace, ale i vazba některých proteinů, např. v procesu transkripce.



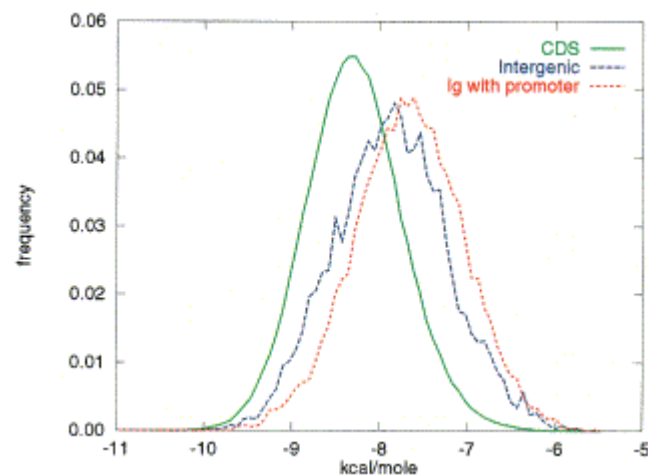
<http://www.cbs.dtu.dk/services/GenomeAtlas>



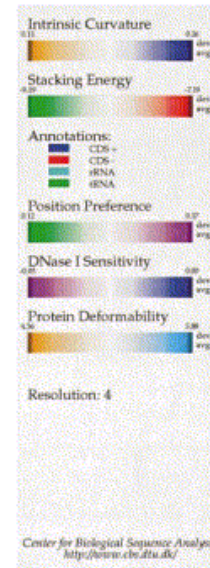
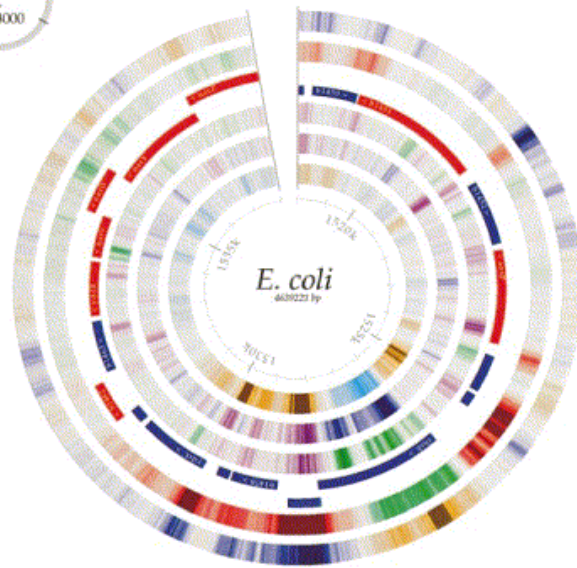
Rozdíly ve vlastnostech mezi DNA se skutečným a změněným pořadím nukleotidů



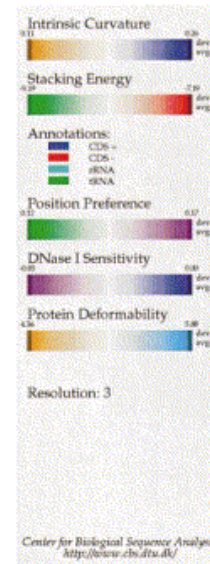
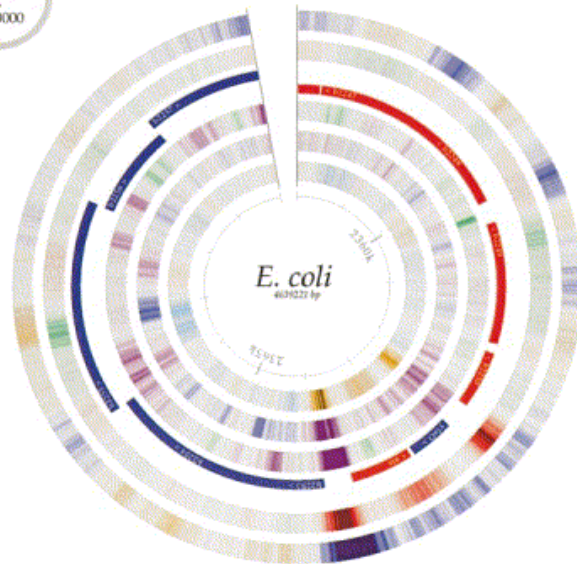
Rozdíly ve vlastnostech mezi DNA kódujících a nekódujících sekvencí

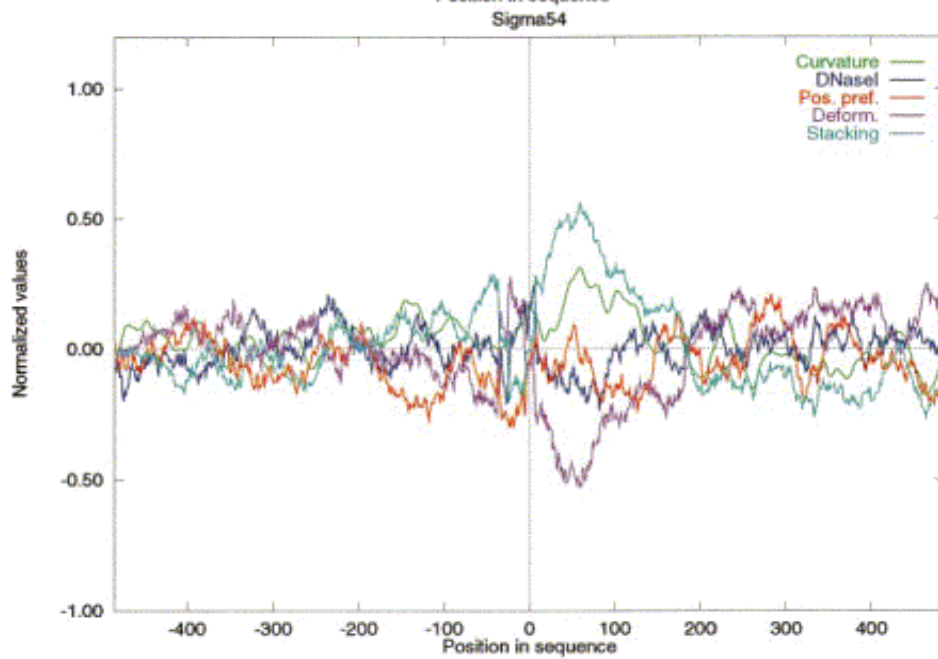
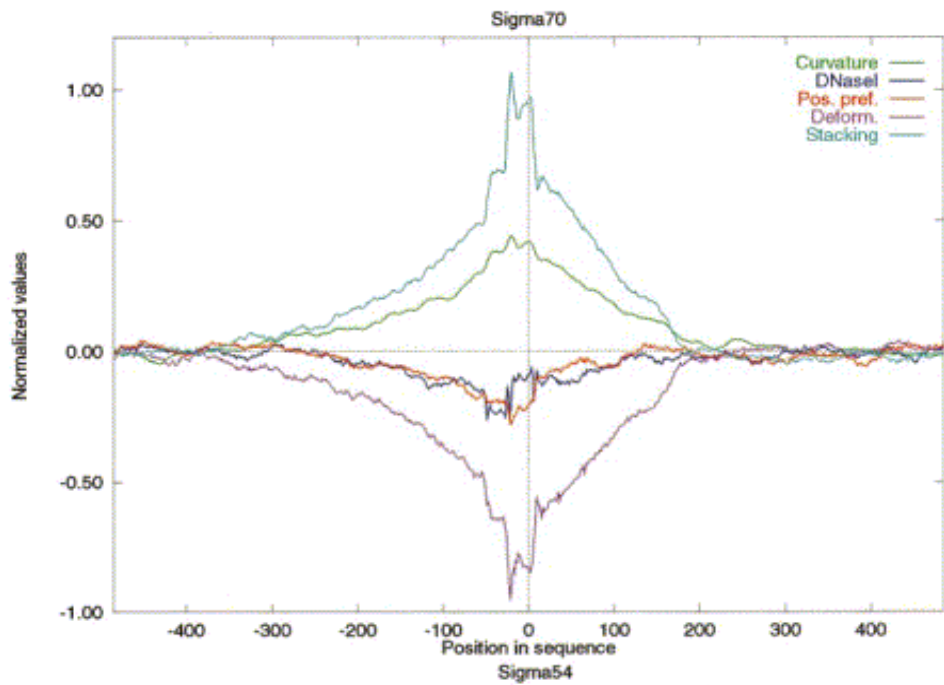


Range:
1518000
-
1538000



Range:
2358000
-
2370000

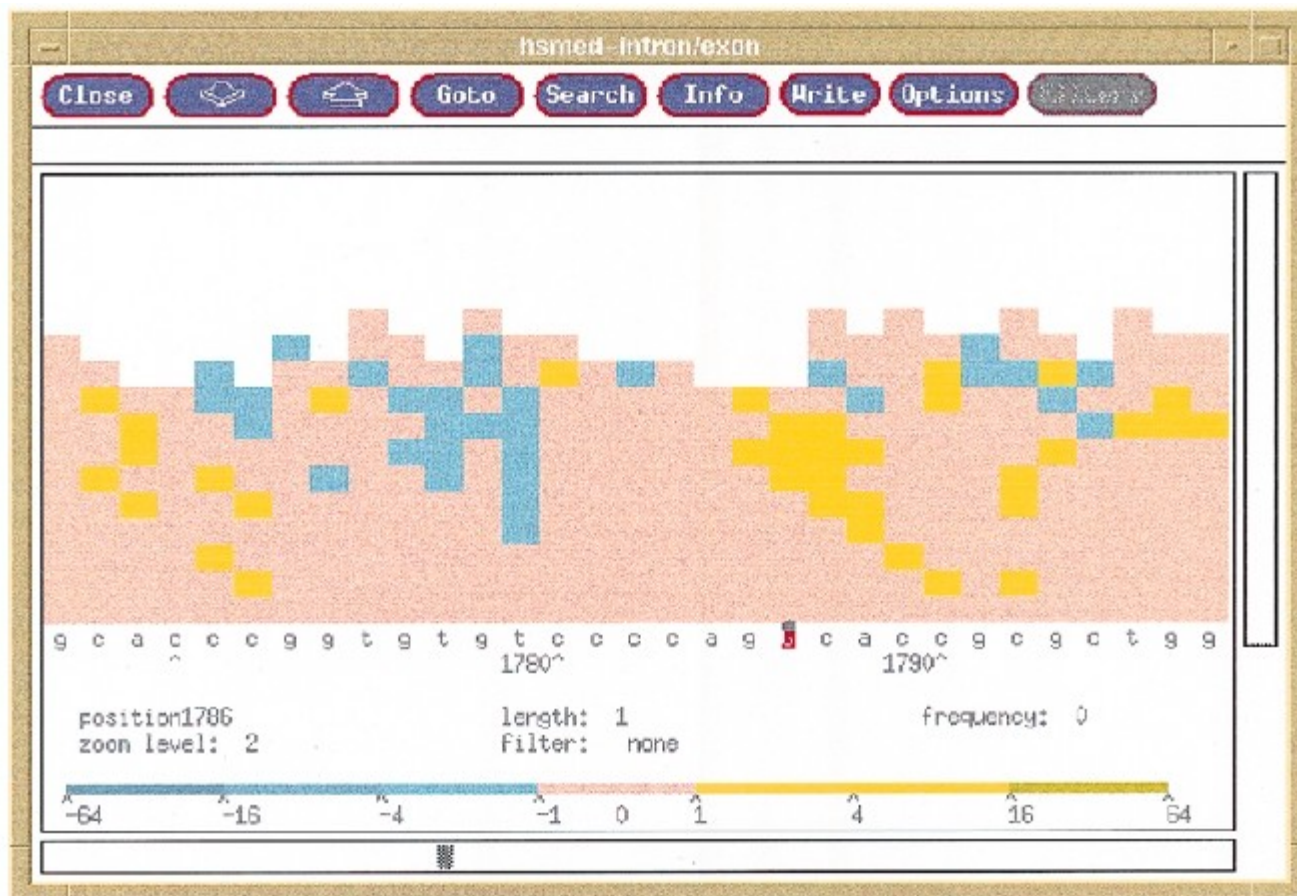




DNA Landscape (program Xlandscape)

<ftp://beagle.colorado.edu/pub/Landscape/xland.v.1.tar.Z>

```
          1          1
        1 2\1      2\1 1 2 1 1
1 2 4 4 4 2 1 4 4 4 4 4 2 4
a g t c c g a t c c t c t g t
```

DNA – protein

- transkripční faktory, strukturní proteiny

TESS <http://www.cbil.upenn.edu/tess/>

AccNo R00549

ID HS\$GG_03

History 20.06.1990 (created); ewi.
20.01.1993 (updated); thh; ewi.

Type D

Description gamma-globin; Gene: [G000261](#).

Sequences TATCTCaATGCAAATATCT.

Start -201

Stop -156

Factors [T00306](#); GATA-1; Quality: 1; Species: human, Homo sapiens.
[T00641](#); Oct-1; Quality: 4; Species: human, Homo sapiens.

Species *human, Homo sapiens*

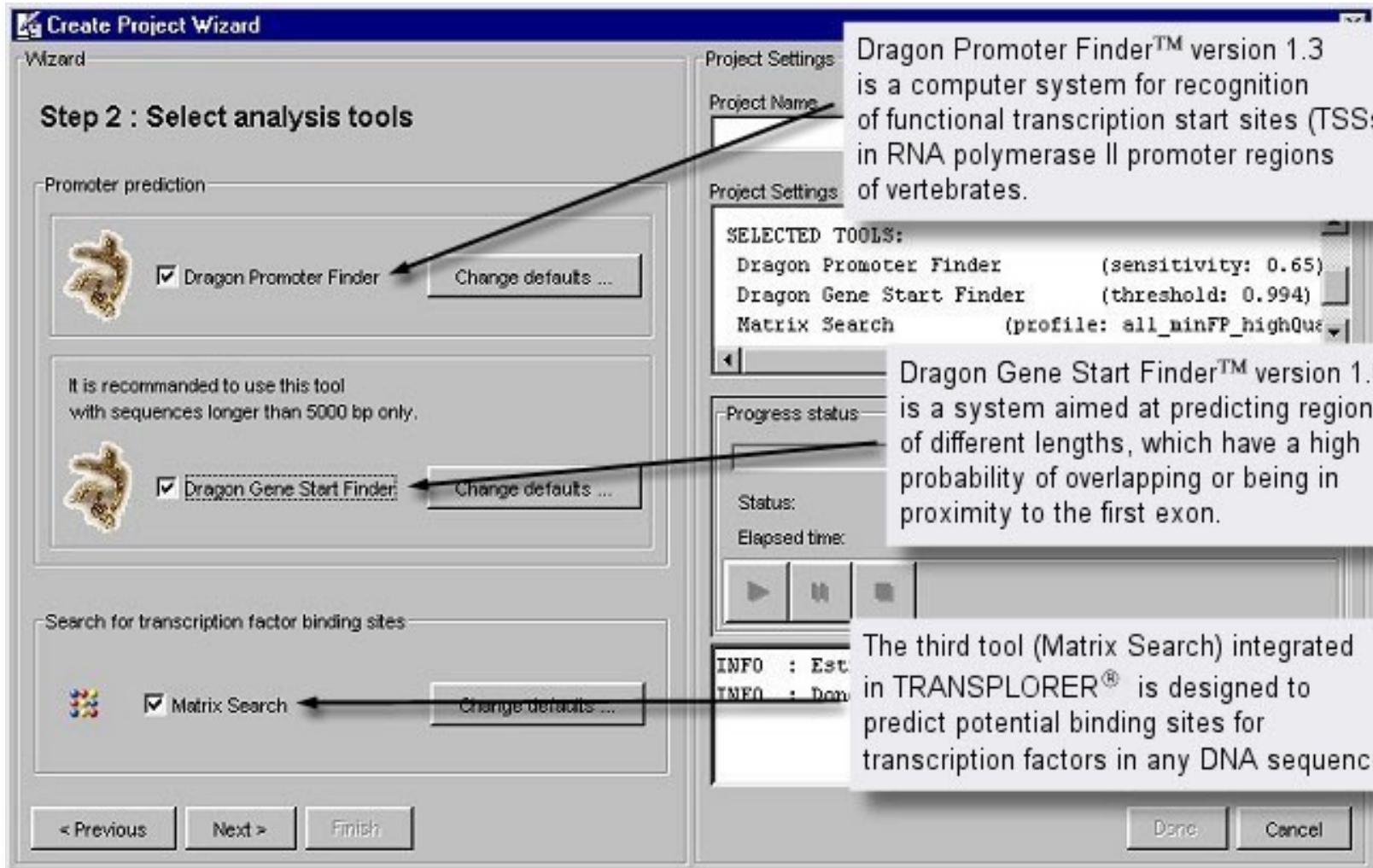
Classification *eukaryota; animalia; metazoa; chordata; vertebrata;*
tetrapoda; mammalia; eutheria; primates

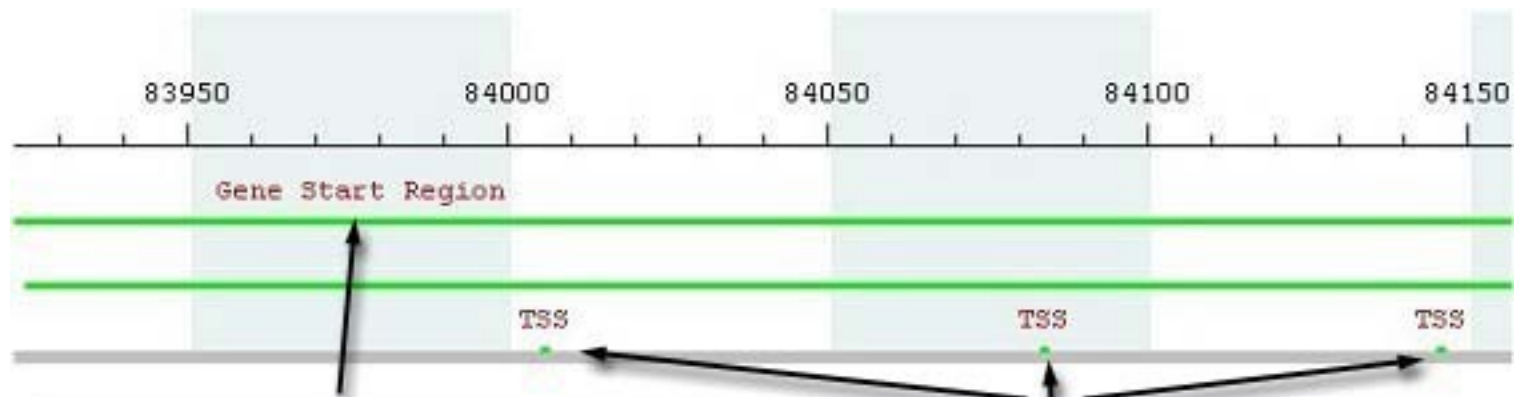
Source [0253](#); 32D cl3

[0126](#); K562

Method gel retardation

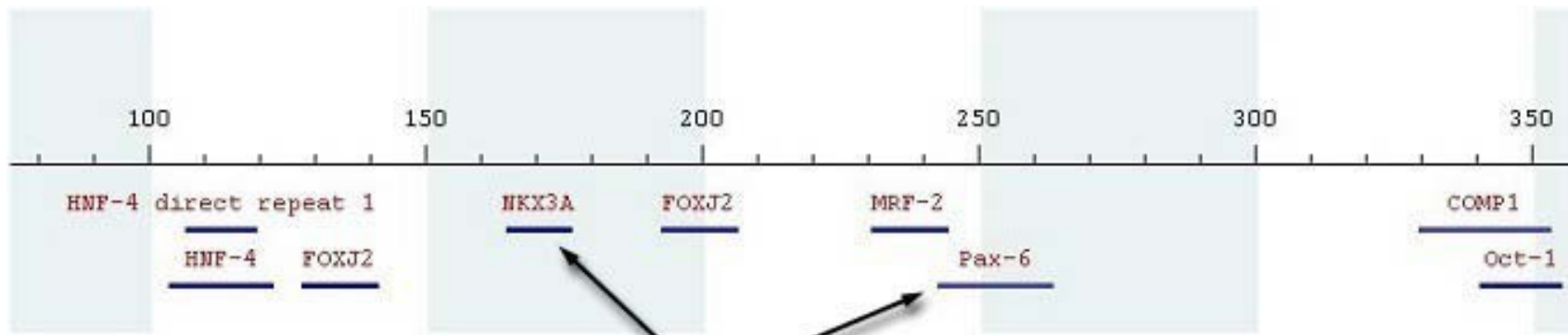
gel shift competition





Dragon Gene Start Finder™ predictions:
Gene Start Regions (regions with a high probability of overlapping or being in proximity to the first exon).

Dragon Promoter Finder™ predictions:
TSS (Transcription Start Site)



Matrix Search predictions:
transcription factor binding sites.
Matrix Search uses position weight matrices
for the identification of binding sites.

TRANSPLORER - emb1_02_26_2

File Edit View Help

localhost

- Color schemes
- Matrices
- Profiles
- Projects
- Sequences

Overview Default Detailed Linked map Results table

426 852 1278 1704 19

...prim_tran
...intrc

1261 1281 1301 1321 1341 1361 1381 14

BR-C Z1 BR-C Z1

Lhx3 Oct-1

cttgatatgctatatataaaaaaaaactaattgaaattaattg
...intron...
cttgatatgctatatataaaaaaaaactaattgaaattaattg
298 1308 1318 1328 1338 1348
cttgatatgctatatataaaaaaaaactaattgaaattaattg
I\$BRCZ1_01 1310 1347 I\$BR

Ready

Filter View Options Regions Project info

Property	Value
Fit to width	<input checked="" type="checkbox"/>
Fit to height	<input type="checkbox"/>
Map options	
Max width	900000000
Interval	10

Above the sequence Map
Below the sequence

Legend Edit

weight & type.xml

weight	type
0.0-0.1	RNA
0.1-0.2	misc_
0.2-0.3	prim_
0.3-0.4	3'
0.4-0.5	3'
0.5-0.6	5'
0.6-0.7	5'
0.7-0.8	Cl
0.8-0.9	
0.9-1.0	
	ex
	inf

Color schemes

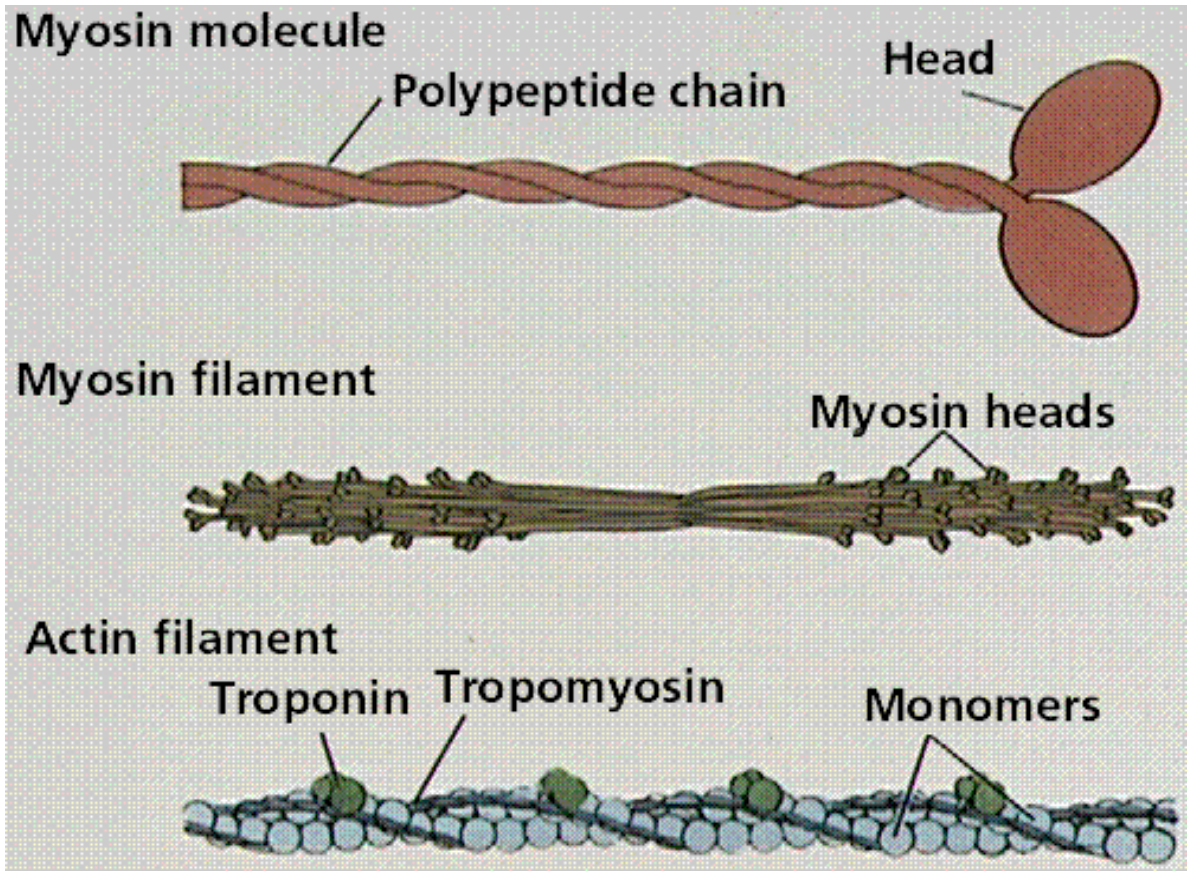
Different graphical representations of results, each displaying a different level of details.

Options for filtering results

Repository: List of all available data

protein – protein

důležité u většiny proteinů, protože jenom ojedinele fungují izolovaně



Metody určování interakcí

molekulární biologie a biochemie:

- two-hybrid test
- co-immunoprecipitation
- co-sedimentation
- in-vitro binding

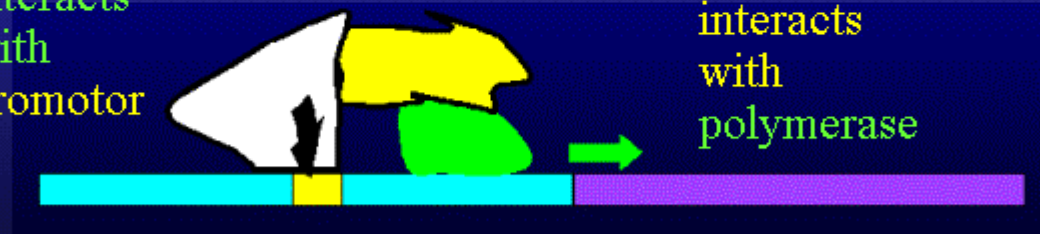
proteomika:

- microarrays
- MS of protein complexes
(hmotnostní spektrometrie)

Yeast two hybrid System 1

Gal4 protein: comprises DNA binding and activating domains

Binding domain
interacts
with
promotor



Activating
domain
interacts
with
polymerase

Measure reporter enzyme activity as e.g. blue colonies.

Yeast two hybrid System (3)

- This is achieved using gene fusions:
- Plasmids carrying different constructs can be expressed in yeast.



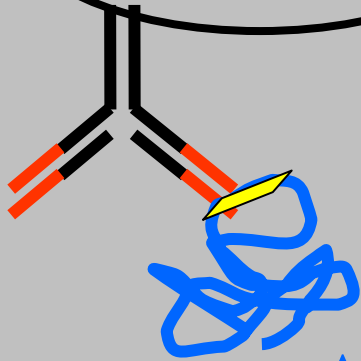
Binding domain as a translational fusion with the gene encoding another protein in one plasmid



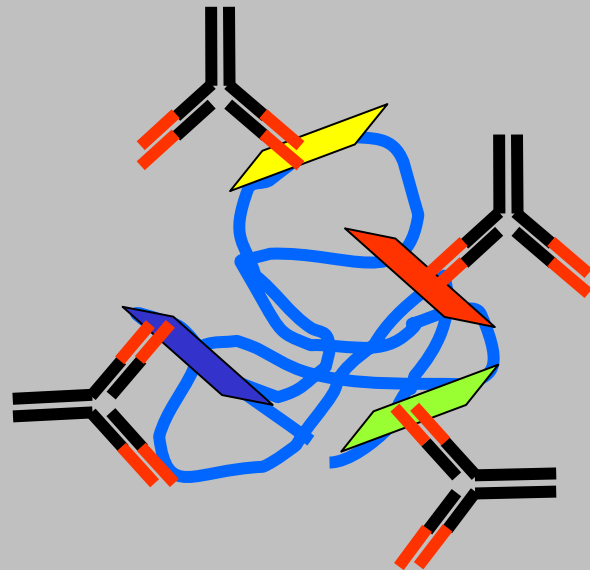
Activating domain as a translational fusion with the gene encoding a third protein in a second plasmid

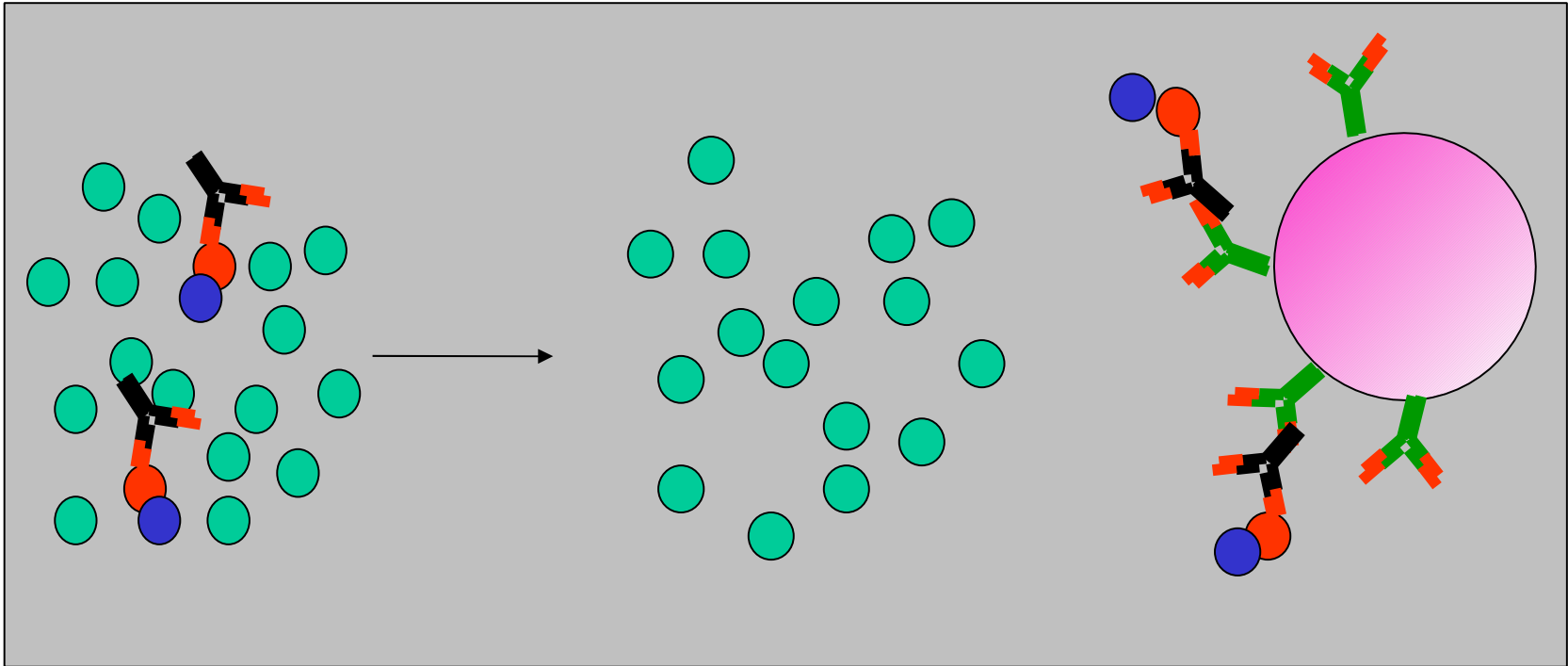


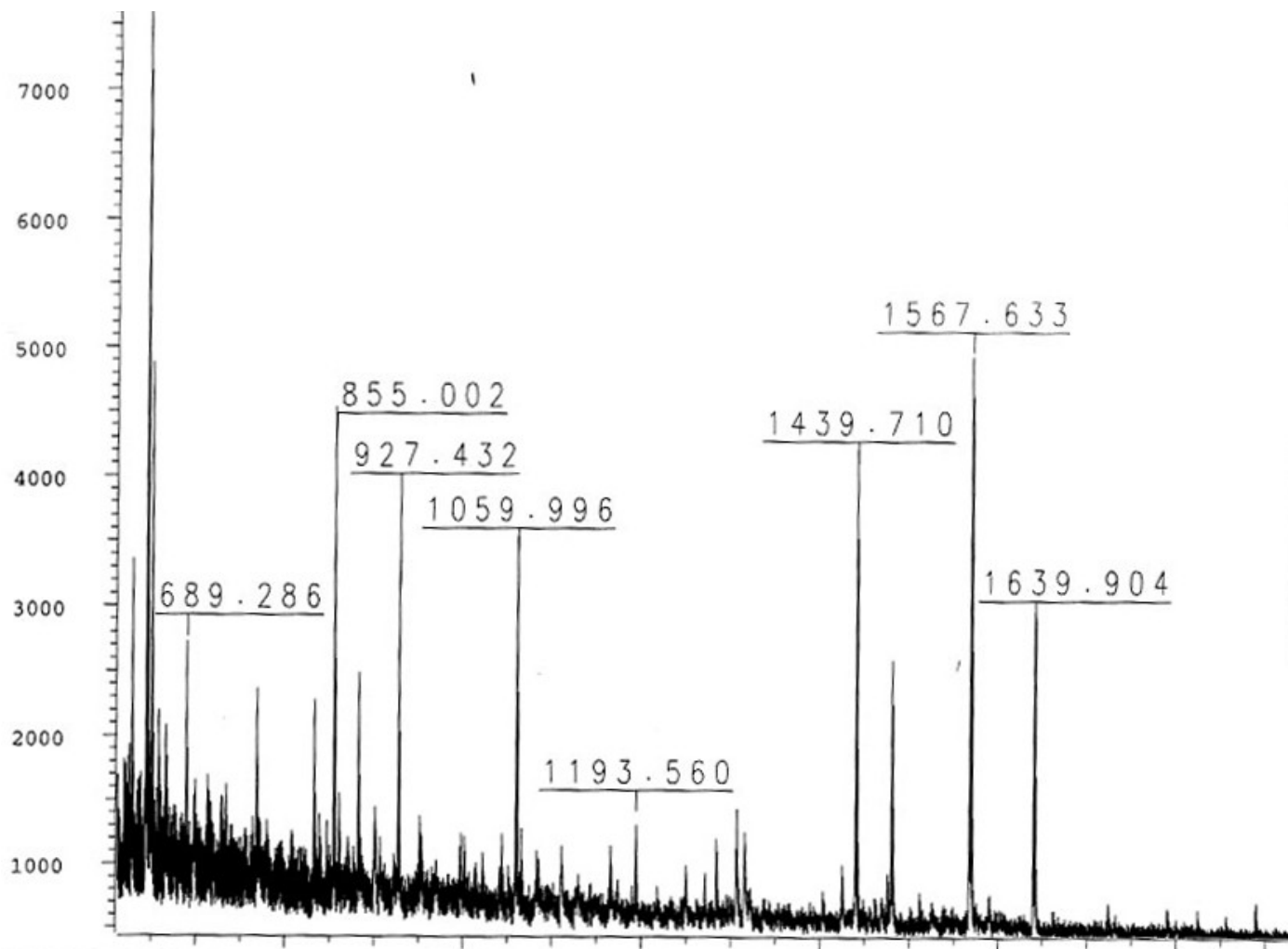
sek. protilátky,
fluoresc. značky



Antigen



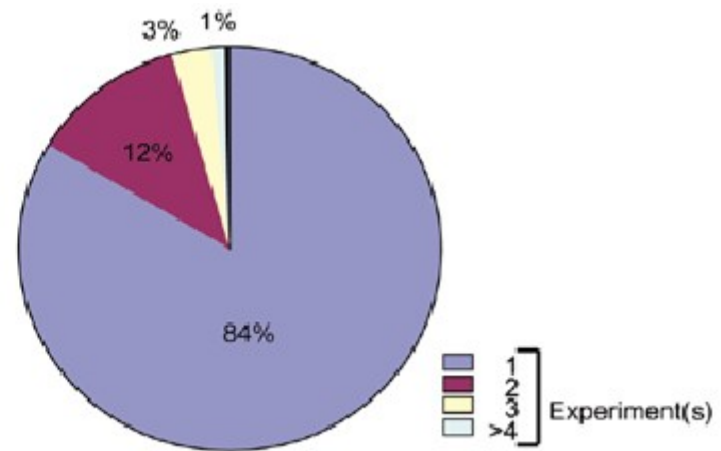
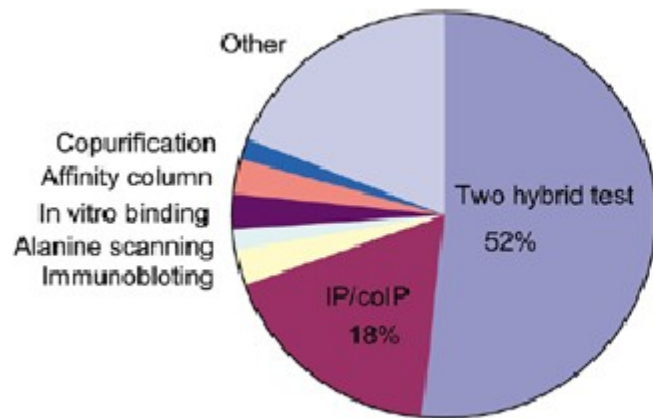






DIP <http://dip.doe-mbi.ucla.edu/>

Number of proteins	17556
Number of organisms	109
Number of interactions	46463
Number of distinct experiments describing an interaction	51915
Number of data sources (articles)	2884
Number of data sources (other)	34



1
 2
 3
 >4

} Experiment(s)

JDIP2D

File Edit Render Tools Windows Help

Protein Information Data

PIR	S51380
SWISSPROT	ST11_YEAST
Genpept (G)	gi_1078309
LENGTH	738
ORGANISM	Saccharomyces cerevisiae
DESCRIPTION	protein kinase STE11
FUNCTION	
EC	2.7.1.-
LOCALIZATION	
SUPERFAMILY	protein kinase Gyr2, protein kinase homology, SAM homology
PDB	-
KEYWORD	ATP, phosphotransferase; protein kinase
YPD:	STE11

Nodes: 25
Edges: 78

JDIP2D NODE: ST11_YEAST

Name	ST11_YEAST
Source	DIP
ID	861

Name	Value
Attributes	
Ecol	
Nuclear	<input type="checkbox"/>
Cytoplasmic	<input type="checkbox"/>
Membrane	<input type="checkbox"/>
Mitochondrial	<input type="checkbox"/>
Scalar	<input type="checkbox"/>
Text	<input type="checkbox"/>

Close

JDIP2D EDGE:

From	11
To	5
Source	DIP
ID	1010

Name	Value
Attributes	
Ecol	
Sucrose gradient se	<input type="checkbox"/>
Two hybrid test	<input checked="" type="checkbox"/>
Copurification	<input type="checkbox"/>
Co-sedimentation	<input type="checkbox"/>
Immunoprecipitatio	<input type="checkbox"/>
In vitro binding	<input checked="" type="checkbox"/>
Genetic	<input type="checkbox"/>
Co-immunoprecipit	<input checked="" type="checkbox"/>
Affinity Column	<input type="checkbox"/>
Immunolocalization	<input type="checkbox"/>

Close

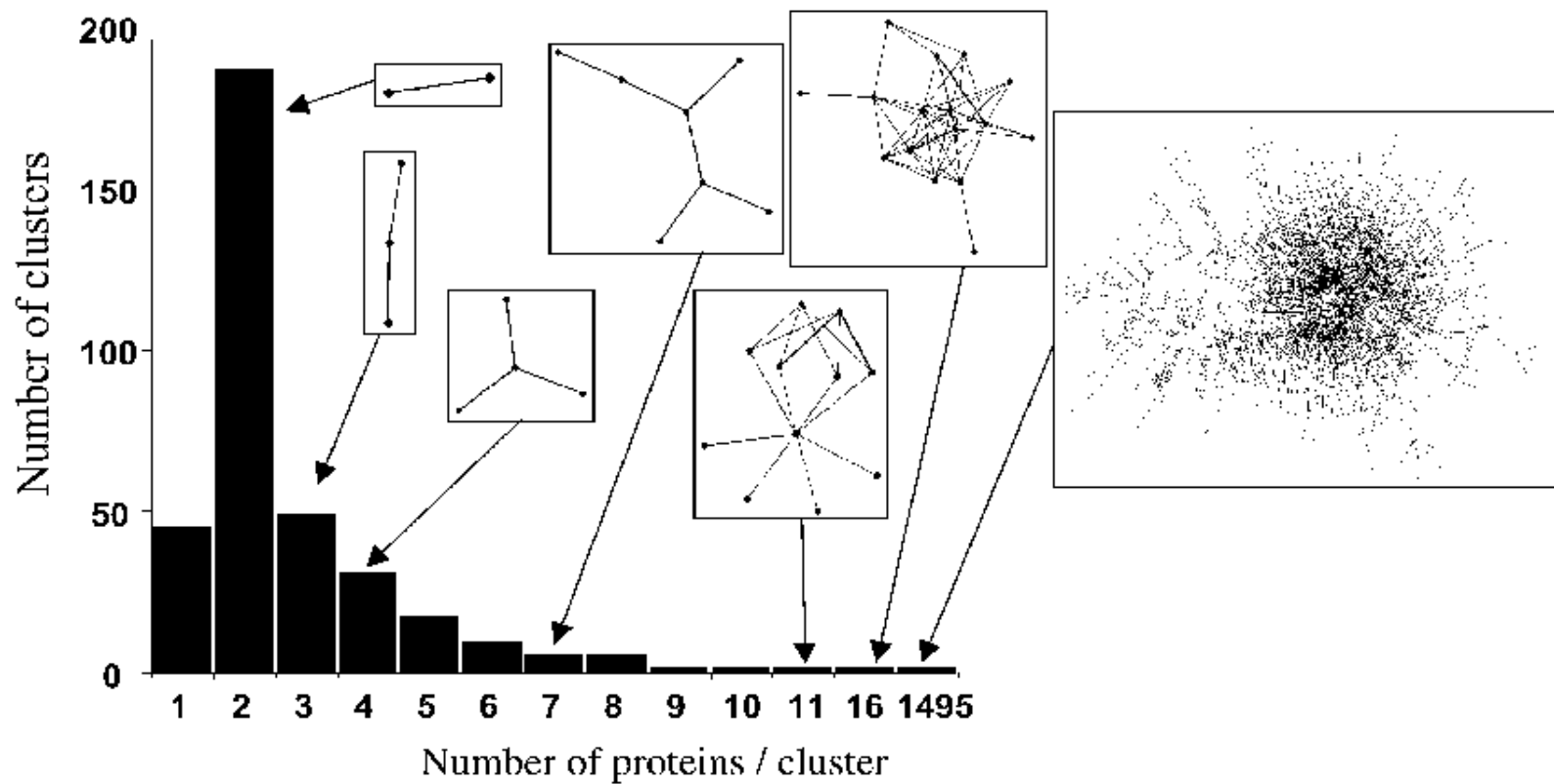
DIP ENTRY #1015

PROTEIN A

PIR	S51254
SWISSPROT	STES_YEAST
Genpept (G)	gi_1071352

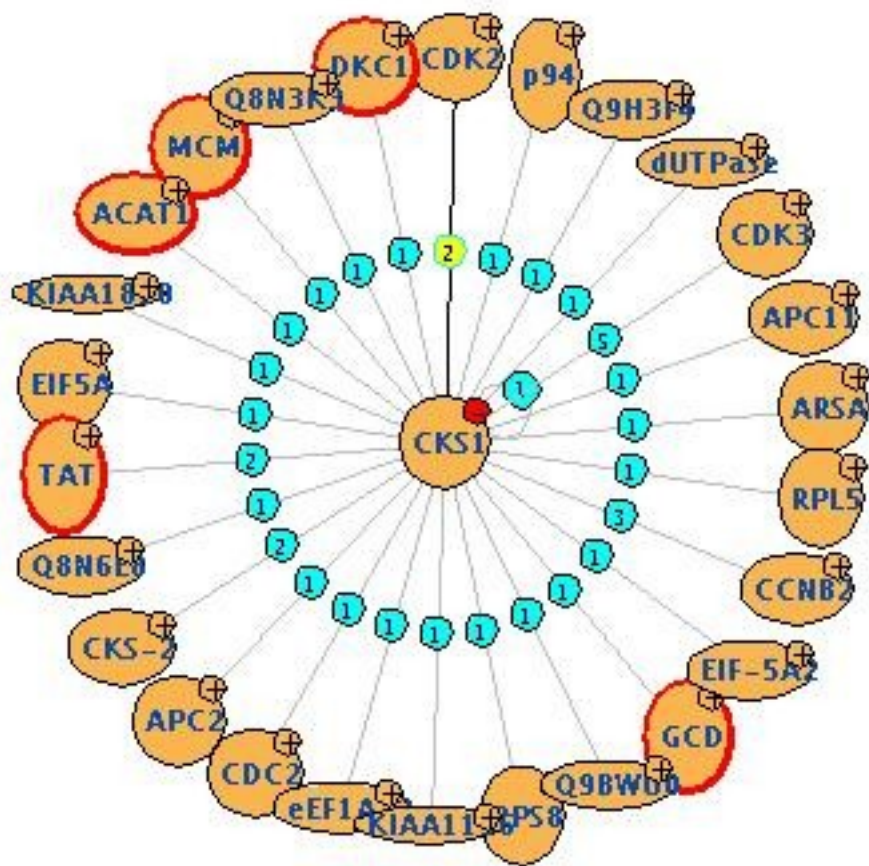
Freeze

<click> node to select; shift-<click> to get node options; press H for help



MINT

<http://mint.bio.uniroma2.it/mint/>



Other databases:

BIND: <http://www.blueprint.org/bind/bind.php>

DIP: <http://dip.doe-mbi.ucla.edu/>

Intact: <http://www.ebi.ac.uk/intact/>

mips: <http://mips.gsf.de/>

PPID: http://www.anc.ed.ac.uk/mscs/PPID/cgi-bin/ppid_search.pl

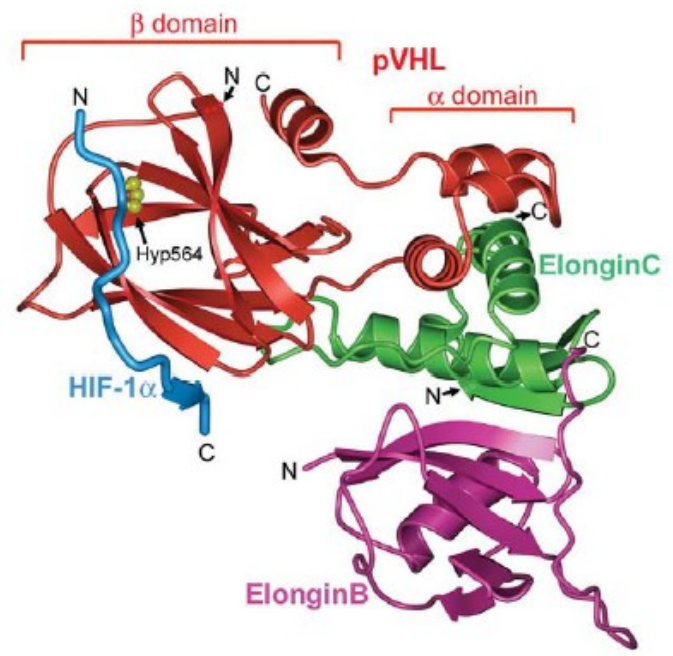
grid: <http://biodata.mshri.on.ca/grid/servlet/Index>

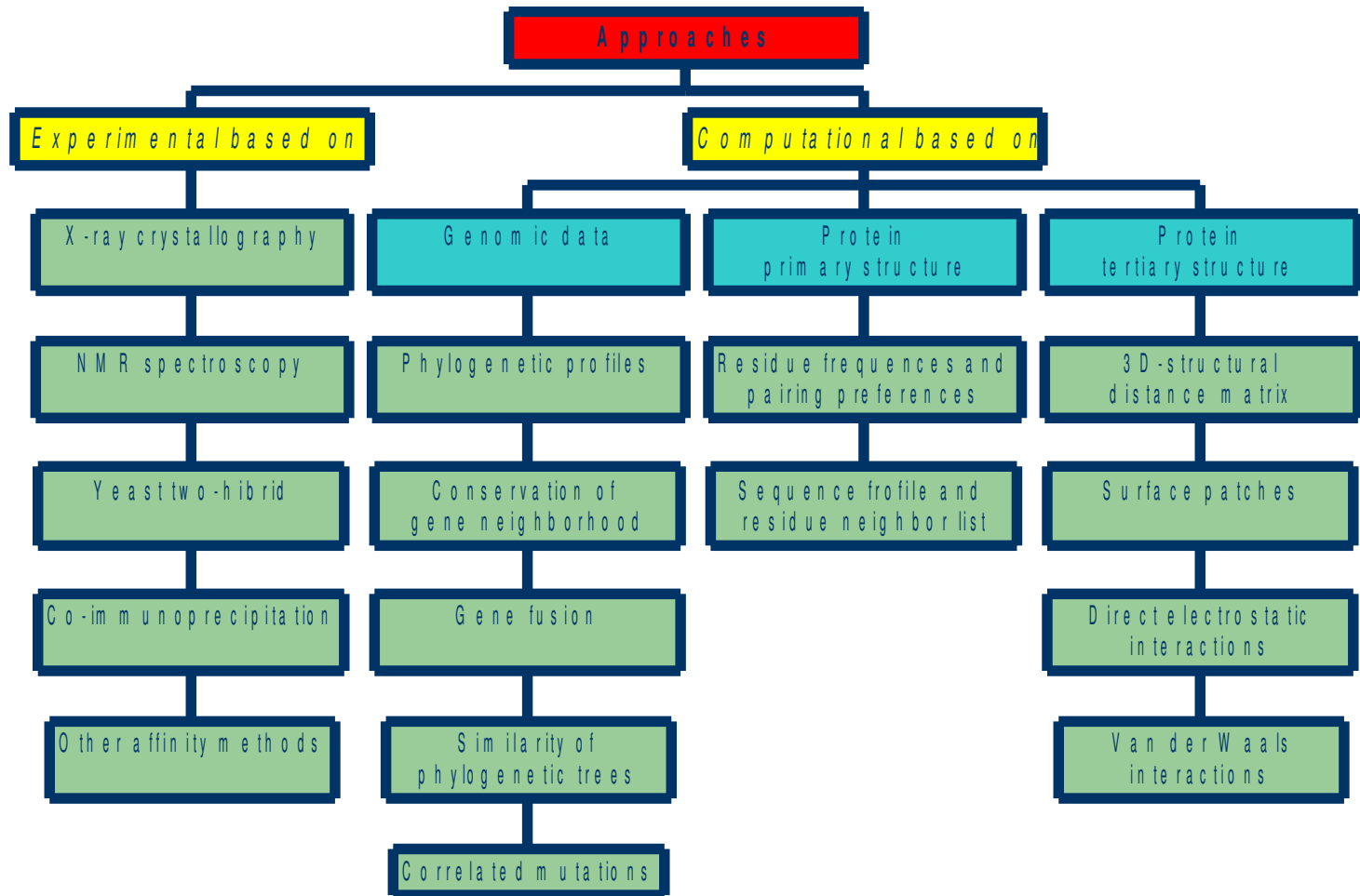
Human protein reference database: <http://www.hprd.org/>

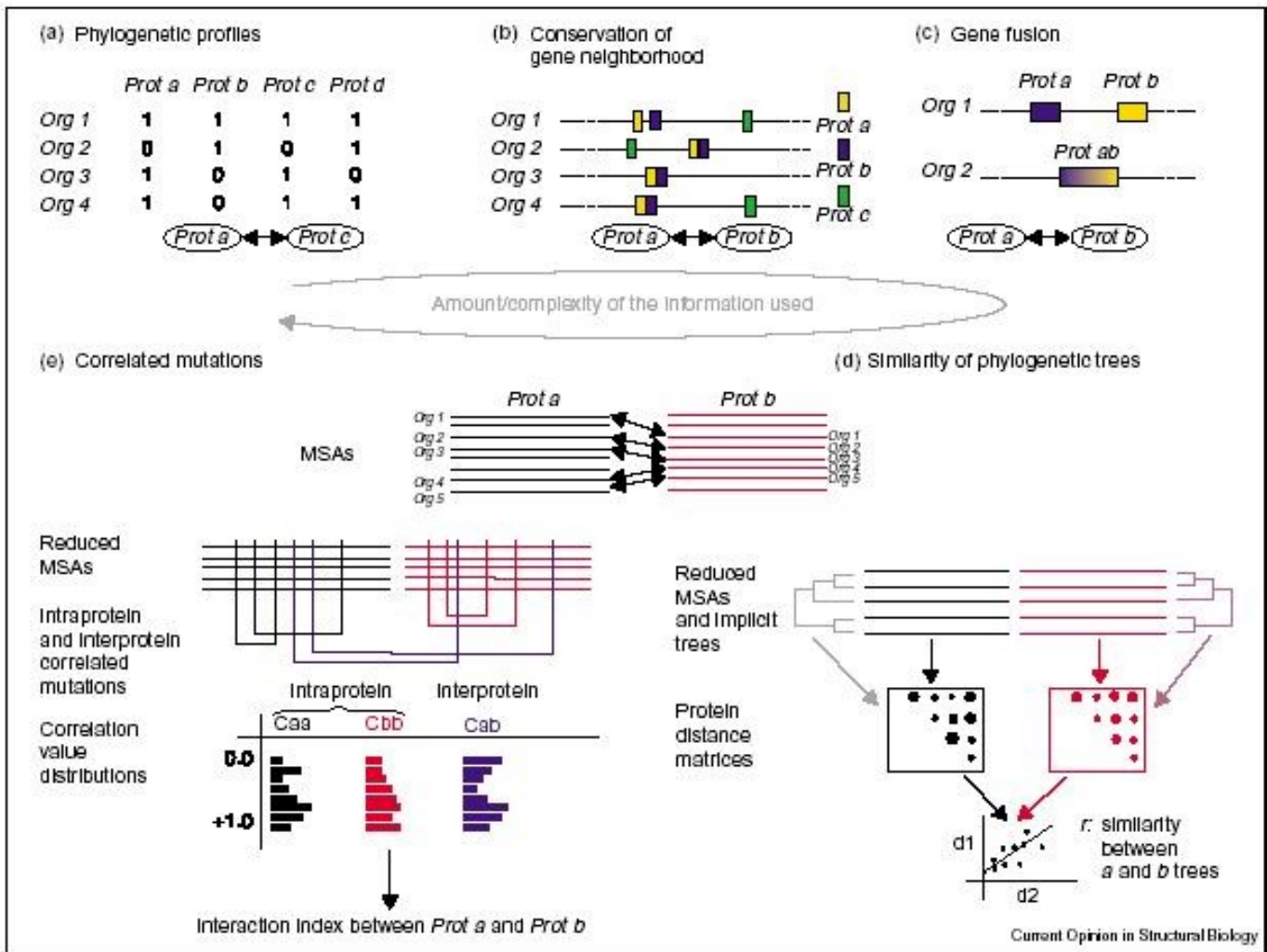
Visualization tools:

Osprey: <http://biodata.mshri.on.ca/osprey/servlet/Index>

Cytoscape: <http://www.cytoscape.org>

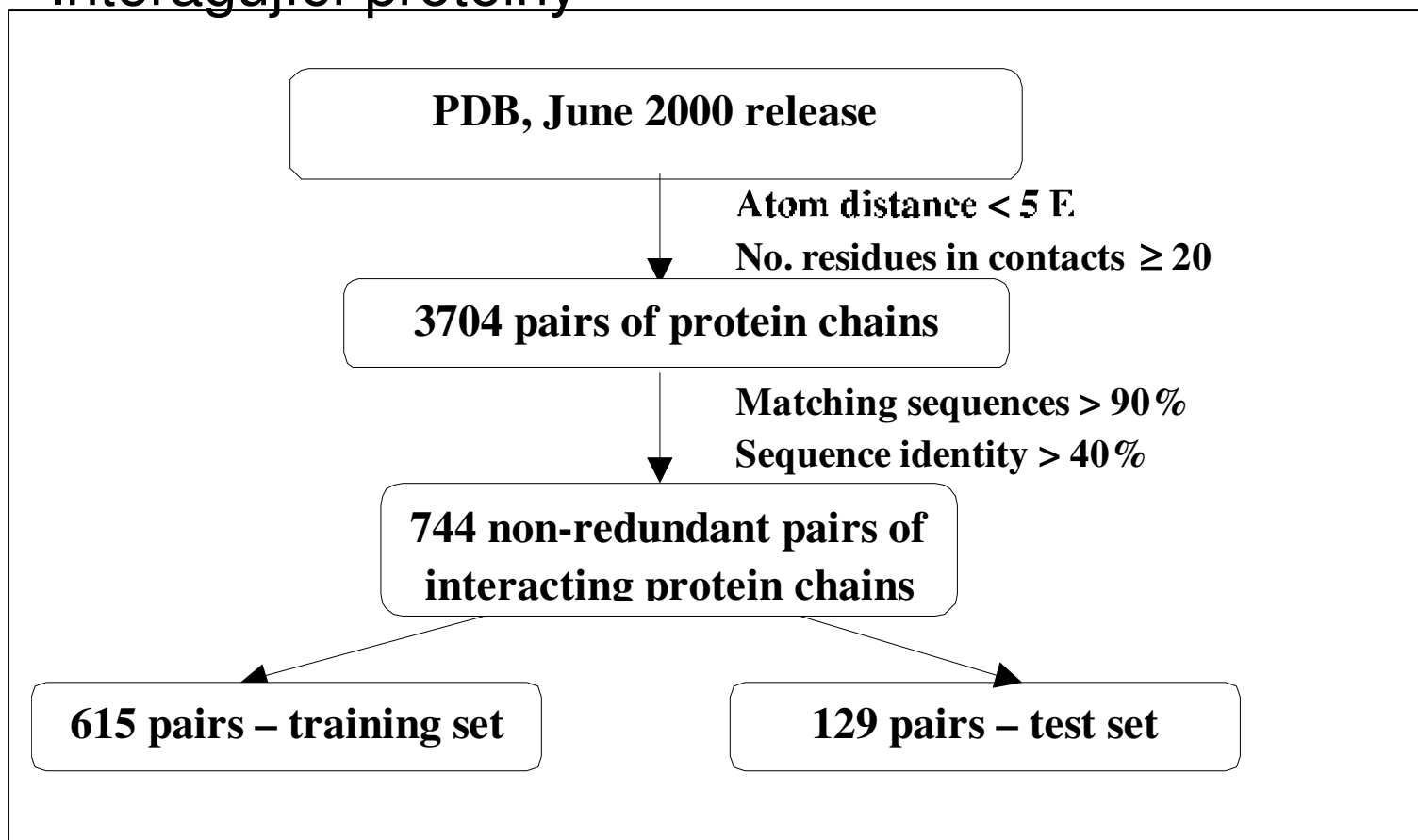




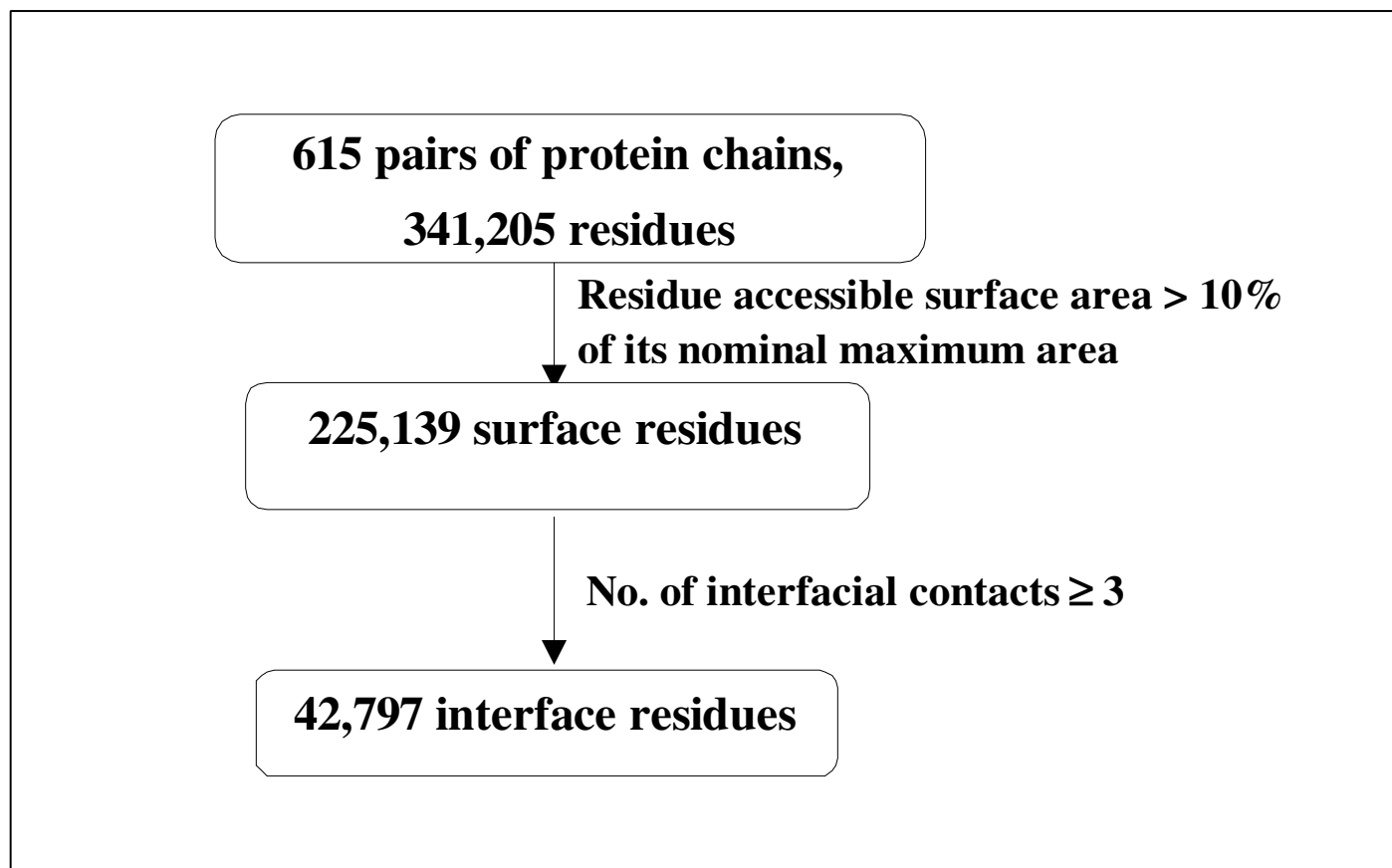


- Četnost v rozhraních
- Hydrofobicita
- Planarita
- Vyčnívavost
- Exponovaný povrch

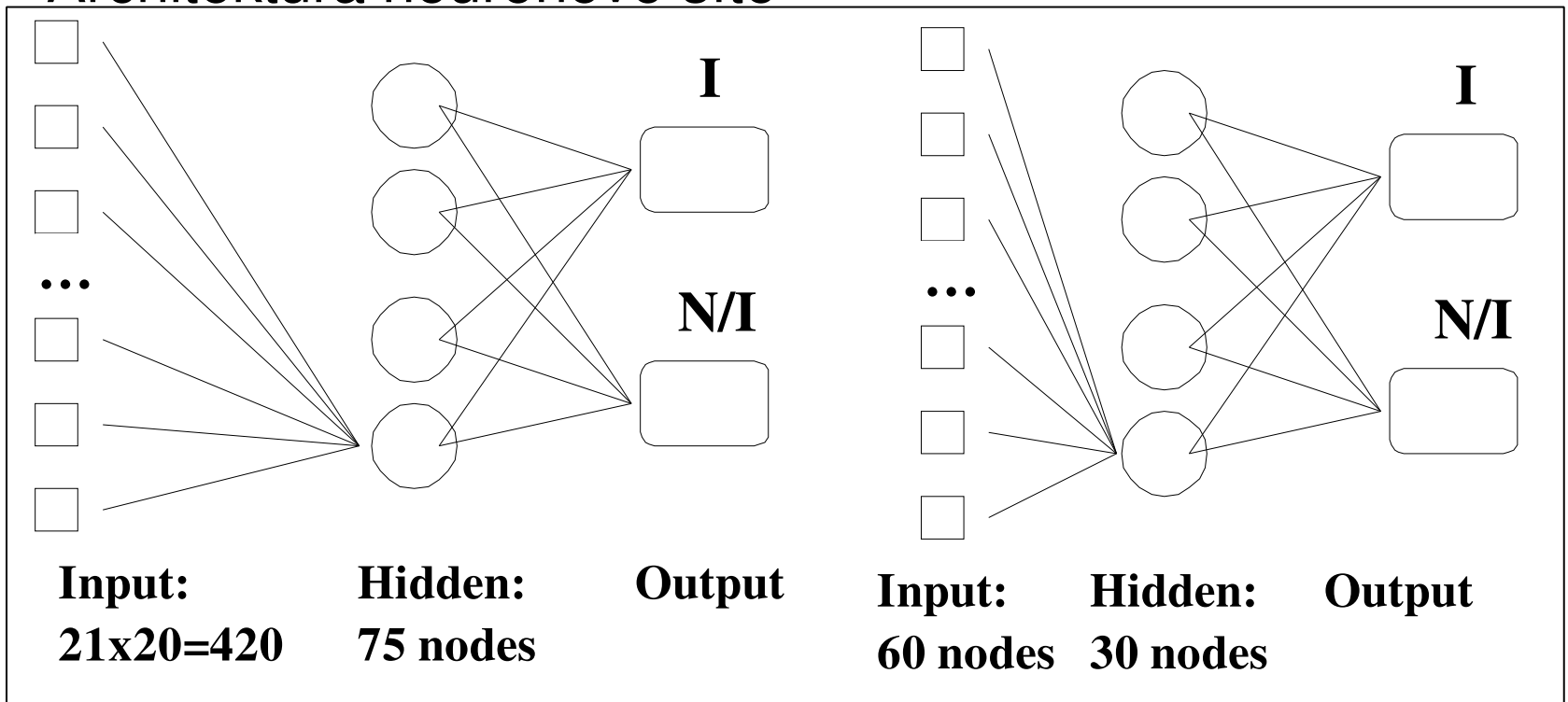
Interagující proteiny



Aminokyseliny na povrchu PDB proteinů



Architektura neuronové sítě



Sloupec z PSI-BLAST PSSM + 1 SAA (exponovaná plocha) pro každou z 20 AA (Daná AA + 19 prostorově nejbližších)

EPR Expression Profile Reliability

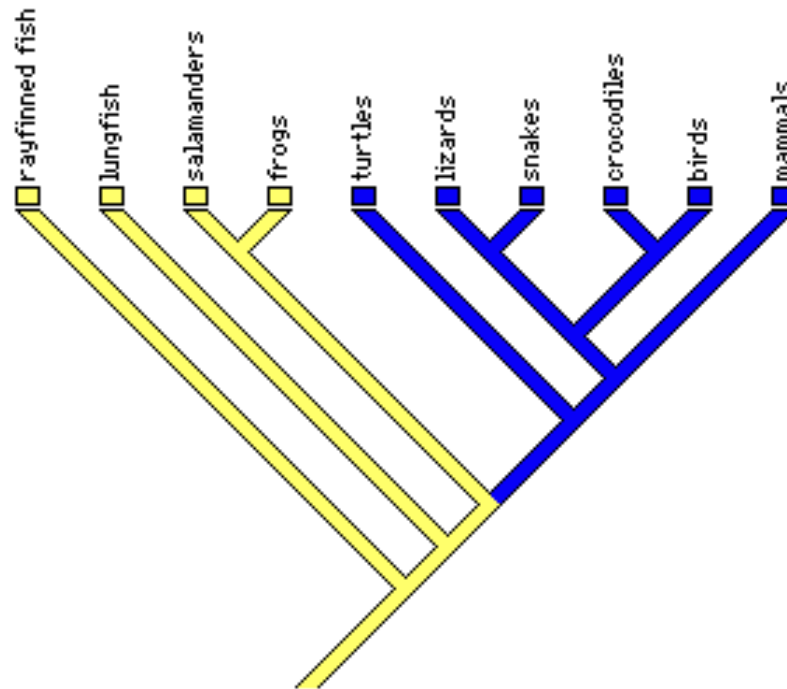
PVM Paralogous Verification

DPV Domain Pair Verification

<http://www.russell.embl.de/interprets/>



Fylogenetické stromy



Většina metod pro konstrukci stromů je založena na informacích o vztazích mezi dvěma prvky (organizmy, sekvencemi a pod.)

- UPGMA (metoda postupného združování párů)
- TDM (metoda transformované vzdálenosti)
- NRM, NJM (metody hodnotící sousedy)
- metody založeny na pravděpodobnosti substitucí (v podstatě vícenásobné zarovnání sekvencí)

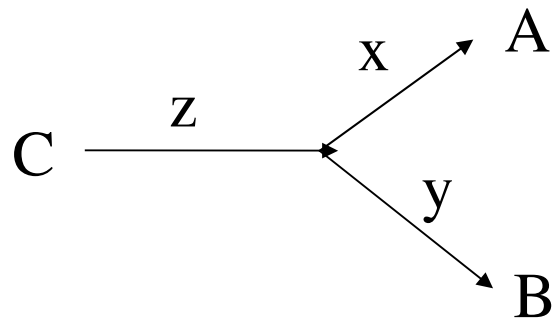
UPGMA

postupně se redukuje vzdálenostní matice obsahující prvky $d(XY)$ tak, že nejbližší dva prvky X a Y se nahradí prvkem W , který je od nejbližšího společného souseda Z vzdálen

$$d(WZ) = (d(XZ) + d(YZ))/2$$

Rozměr matice se zmenší o 1.

Délka větví fylogenetického stromu sestrojeného metodou UPGMA se počítá vždy pro hodnocenou dvojici a společného předka všech ostatních členů stromu.



$$d(AC) = x + z$$

$$d(AB) = x + y$$

$$d(BC) = y + z$$

$$x = (d(AC) + d(AB) - d(BC))/2$$



COG <http://www.ncbi.nlm.nih.gov/COG/>

“a(A) je ortologem proteinu b(B)”

\Leftrightarrow

Protein a je homologem proteinu b, který plní v organizmu A obdobnou funkci, jakou plní protein b v organizmu B

- má podobnou sekvenci
- má podobnou regulaci
- má podobné partnery (substrát, ligand,...)

COG je nástrojem, který podobně jako např. Gene Ontology dokáže propojovat na vyšší úrovni a tím zhodnocovat informace přicházející ze sekvenace různých genomů.

Čím víc ortologů známe, tím větší máme naději, že dokážeme popsat dosud neznámý vztah mezi proteiny. Vztahy v jednom organizmu můžeme často převést do jiných organizmů.

- kolokace v genomu
- interakce v metabolismu

Dokážeme usuzovat na fylogenetické závislosti a příčinu absence-přítomnosti proteinů v organizmech

- ztráta genů
- horizontální transfer

[K] [COG2771](#) DNA-binding HTH domains

36 COGs

