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

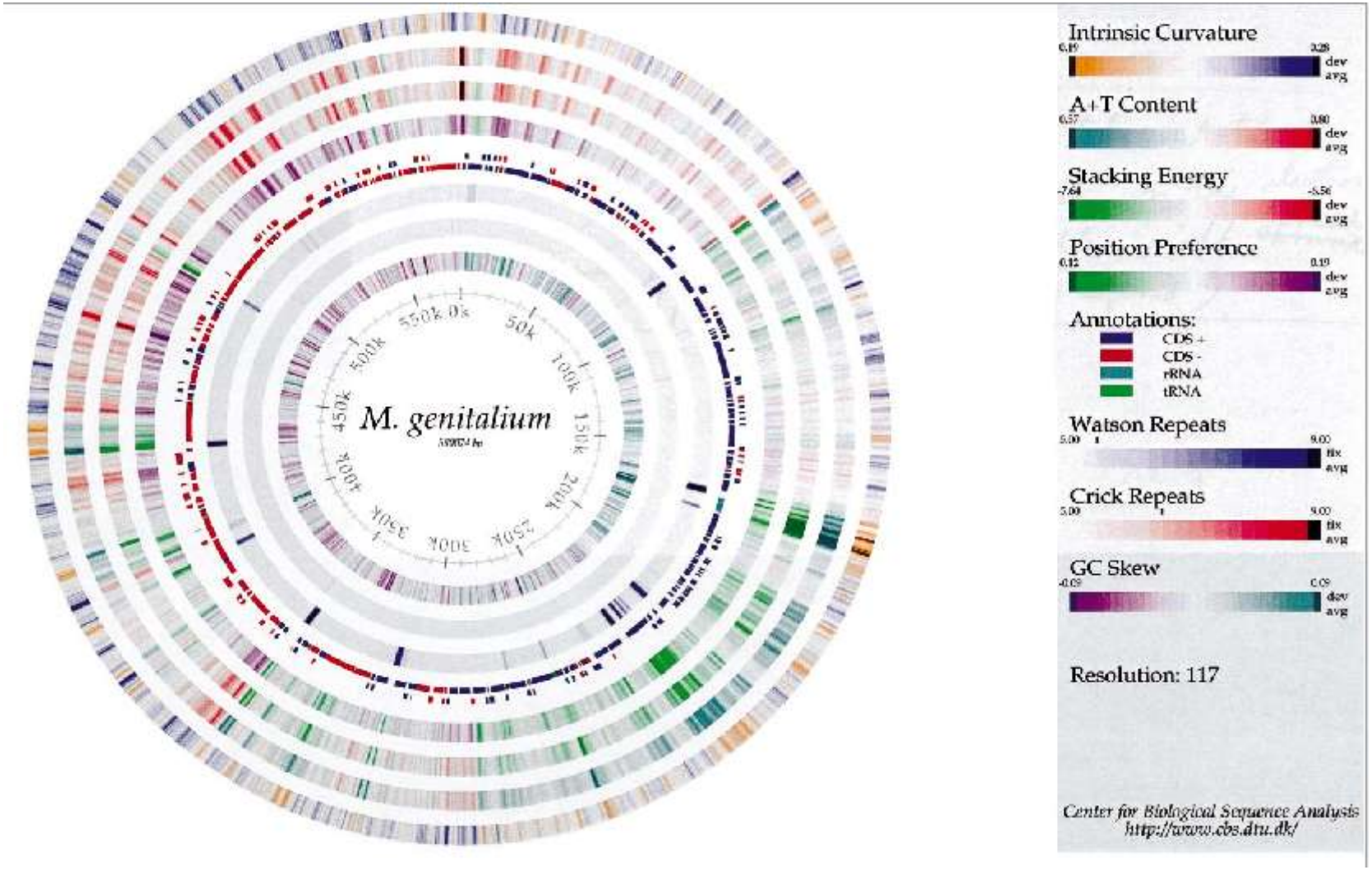
- Sekvence DNA a RNA
- Sekvence proteinů
- Struktura proteinů
- Údaje o aktivitě genů – DNA čip, „microarray“
- Údaje o expresi proteinů – 2-D gely + MS
- Údaje o struktuře DNA
- Mapy interakcí mezi proteiny a DNA
- Mapy interakcí mezi proteiny navzájem
- Literatura

Struktura DNA ?

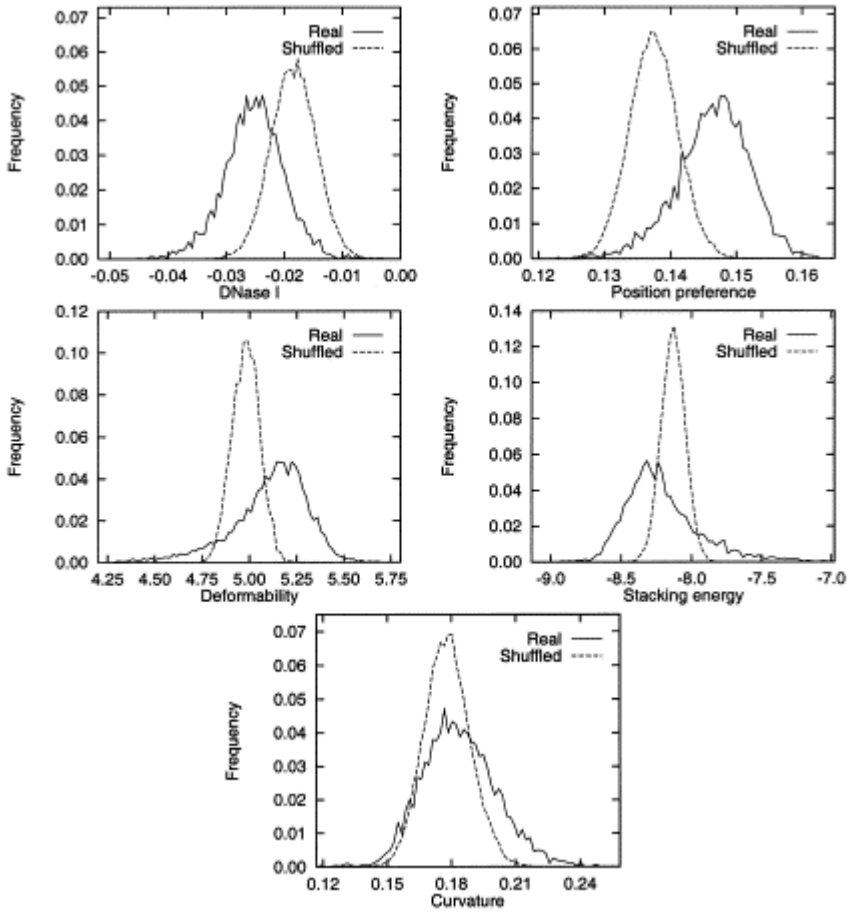
Specifické nukleotidy a jejich různá zvláštní opakování dávají molekule DNA různé vlastnosti.

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

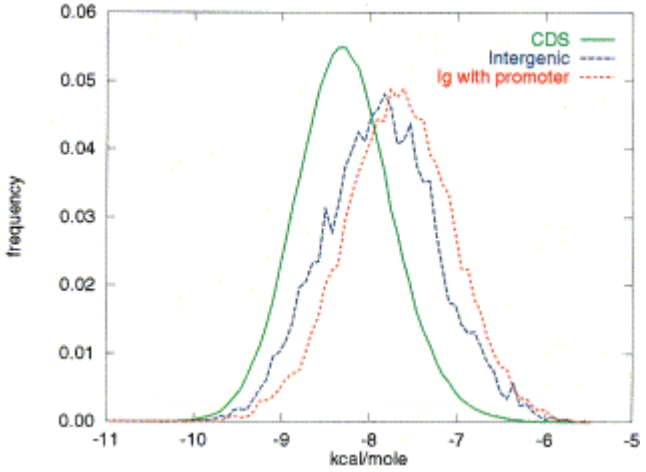
http://www.cbs.dtu.dk/services/GenomeAtlas/

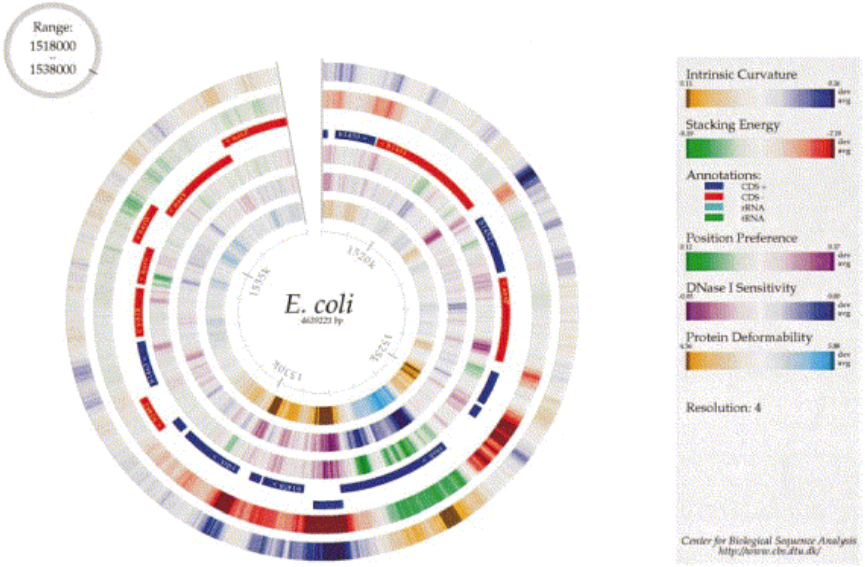


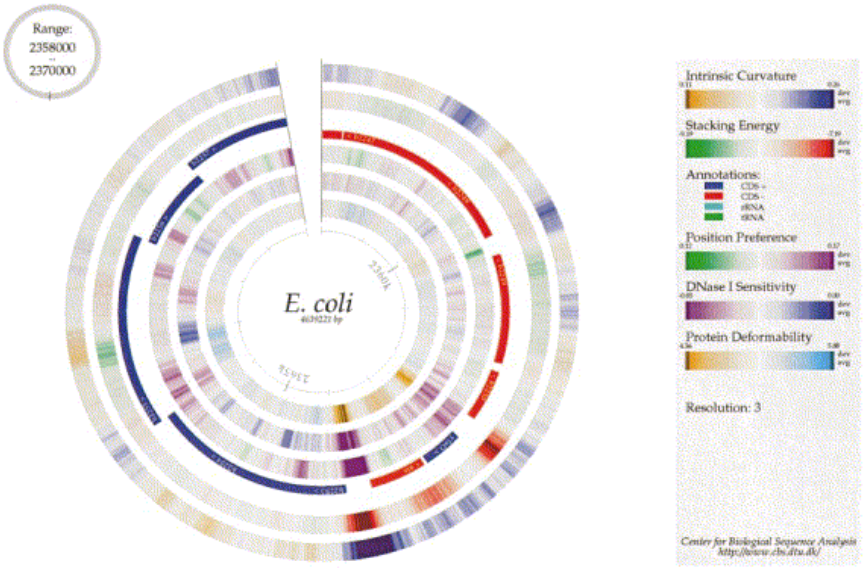
Rozdíly ve vlastnostech mezi DNA se skutečným a změněným pořadím nukleotidů

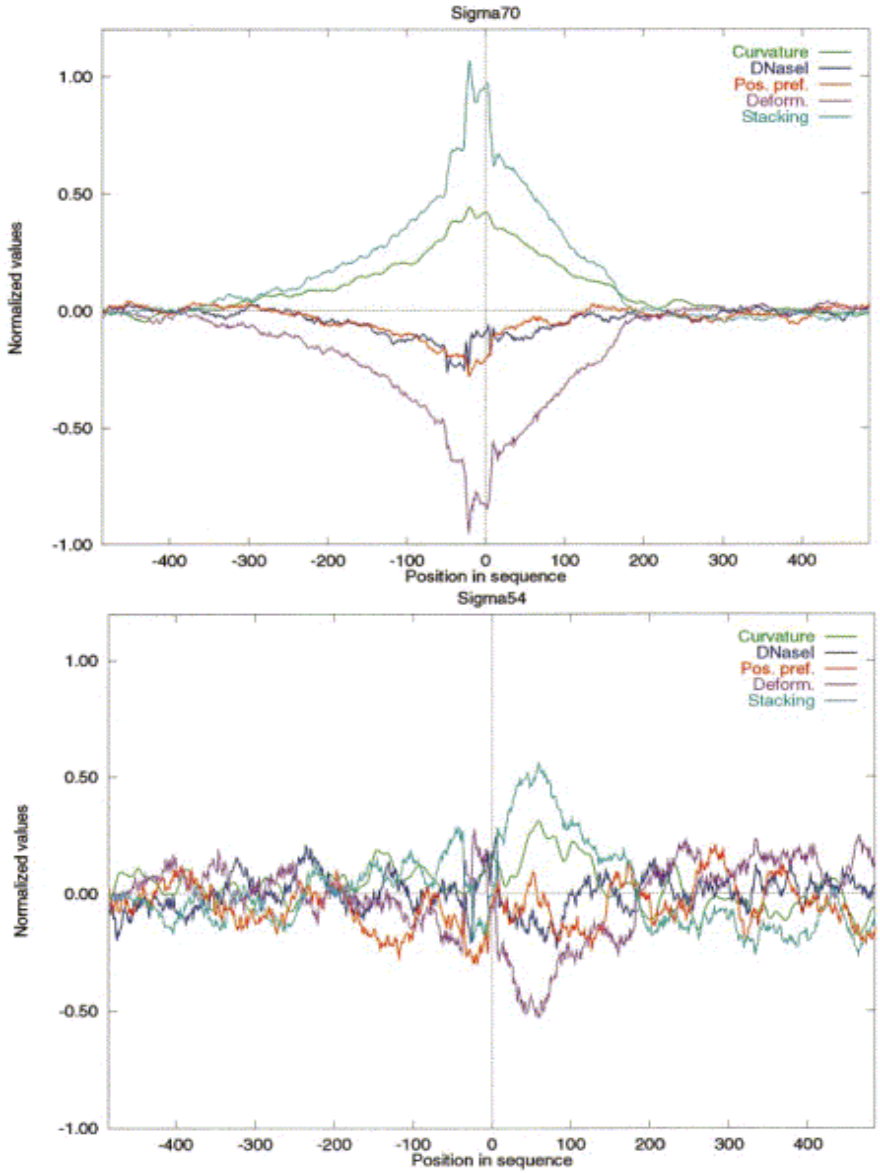


Rozdíly ve vlastnostech mezi DNA kódujících a nekódujících sekvencí



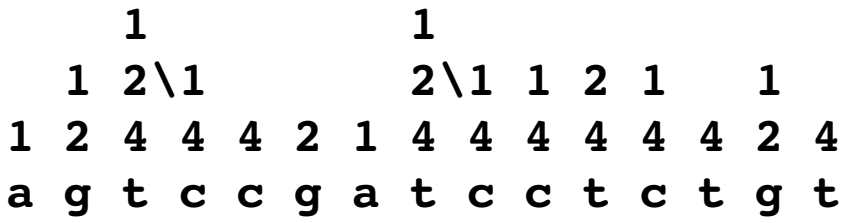

Range:
1518000
..
1538000
E. coli
4639221 bp
1538k
1524k
1510k
1496k
Intrinsic Curvature
Stacking Energy
Annotations:
CDS +
CDS -
rRNA
tRNA
Position Preference
DNase I Sensitivity
Protein Deformability
Resolution: 4
Center for Biological Sequence Analysis
http://www.cbs.dtu.dk/

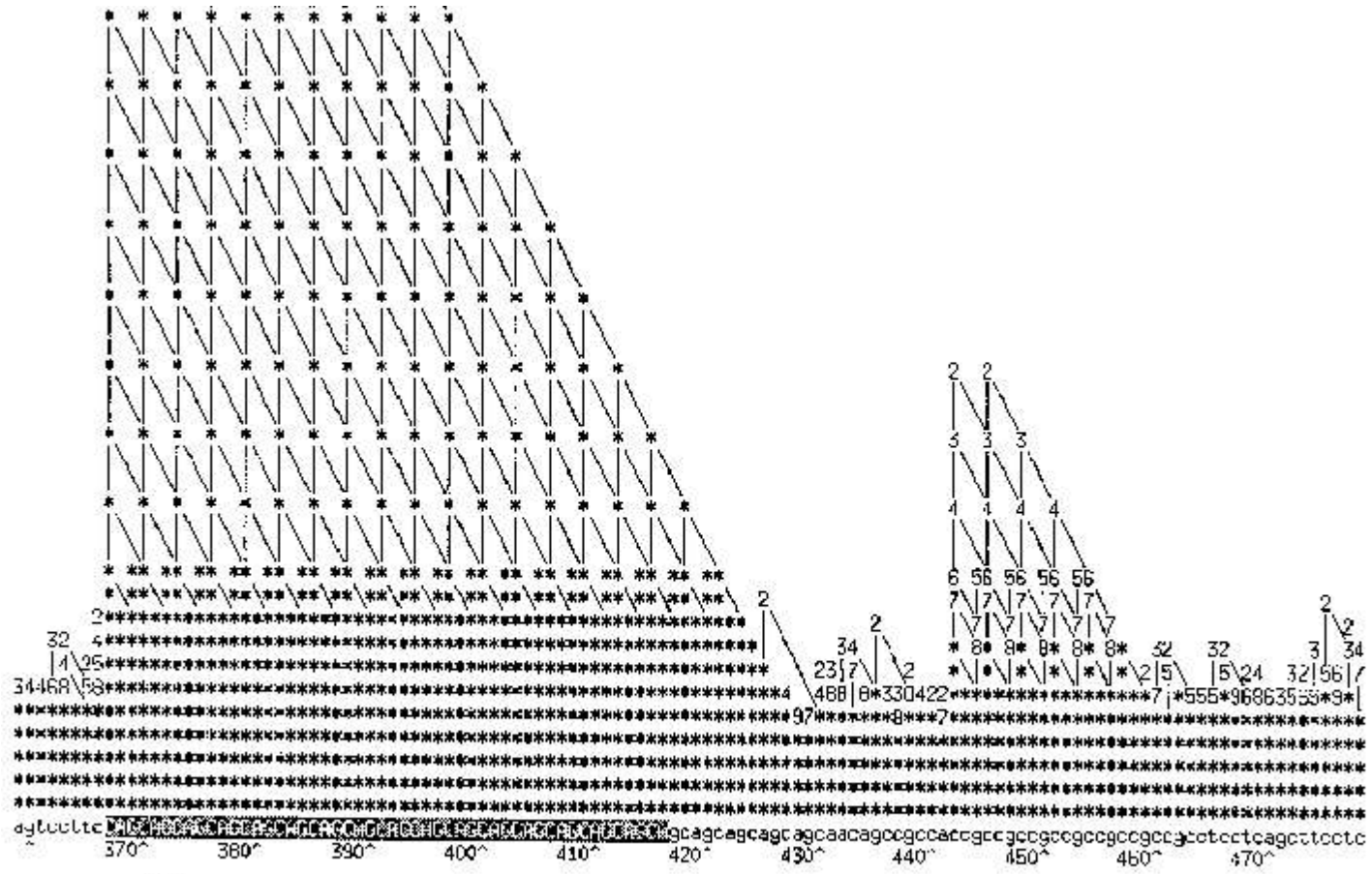

Range:
2358000
..
2370000
E. coli
4639221 bp
2370k
2358k
Intrinsic Curvature
Stacking Energy
Annotations:
CDS +
CDS -
rRNA
tRNA
Position Preference
DNase I Sensitivity
Protein Deformability
Resolution: 3
Center for Biological Sequence Analysis
http://www.cbs.dtu.dk/

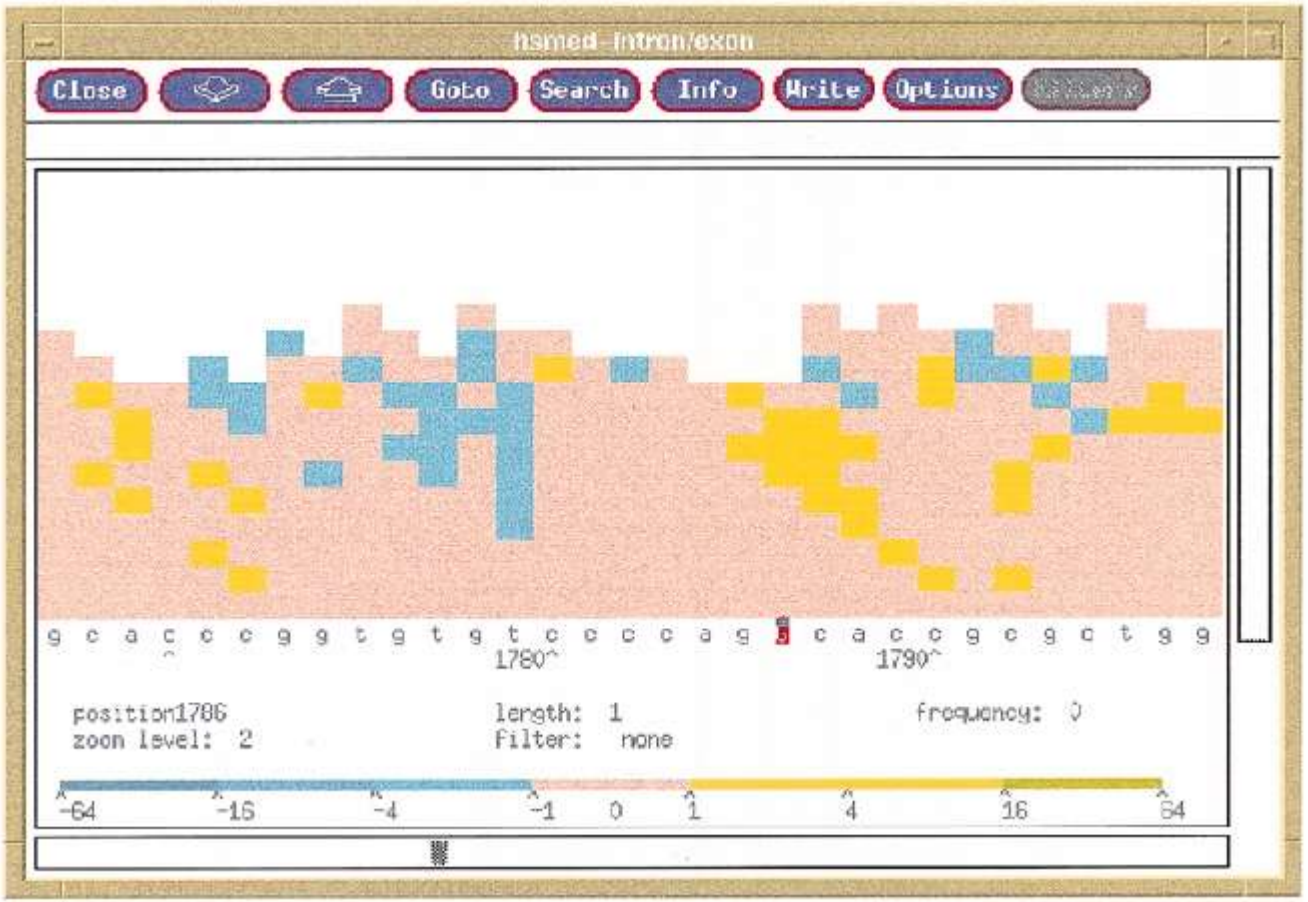

Sigma70
Curvature
DNasel
Pos. pref.
Deform.
Stacking
Normalized values
1.00
0.50
0.00
-0.50
-1.00
-400
-300
-200
-100
0
100
200
300
400
Position in sequence
Sigma54
Curvature
DNasel
Pos. pref.
Deform.
Stacking
Normalized values
1.00
0.50
0.00
-0.50
-1.00
-400
-300
-200
-100
0
100
200
300
400
Position in sequence

DNA Landscape (program Xlandscape)

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




370^
380^
390^
400^
410^
420^
430^
440^
450^
460^
470^
position:367
zoom level: 1
length: 50
filter: none
frequency: 6


hsmed_intron/exon
Close
Goto
Search
Info
Write
Options
g c a c c c g g t g t g t c c c c a g g c a c c g c g c t g g
1780^
1790^
position1786
zoon level: 2
length: 1
filter: none
frequency: 0
-64
-16
-4
-1
0
1
4
16
64

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

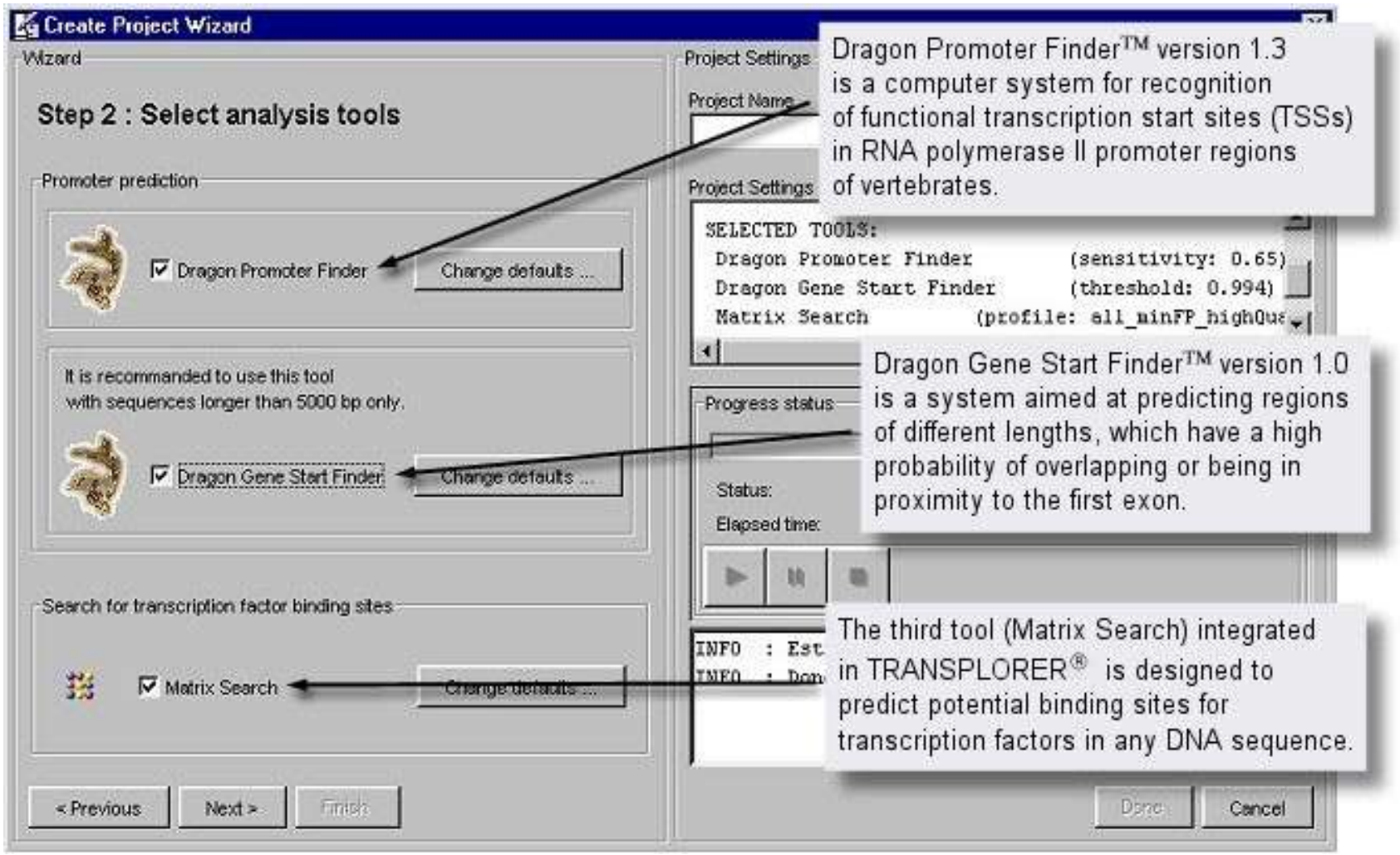
Factors T00306; GATA-1; Quality: 1; Species: human, Homo sapiens.
T00641; Oct-1; Quality: 4; Species: human, Homo sapiens.

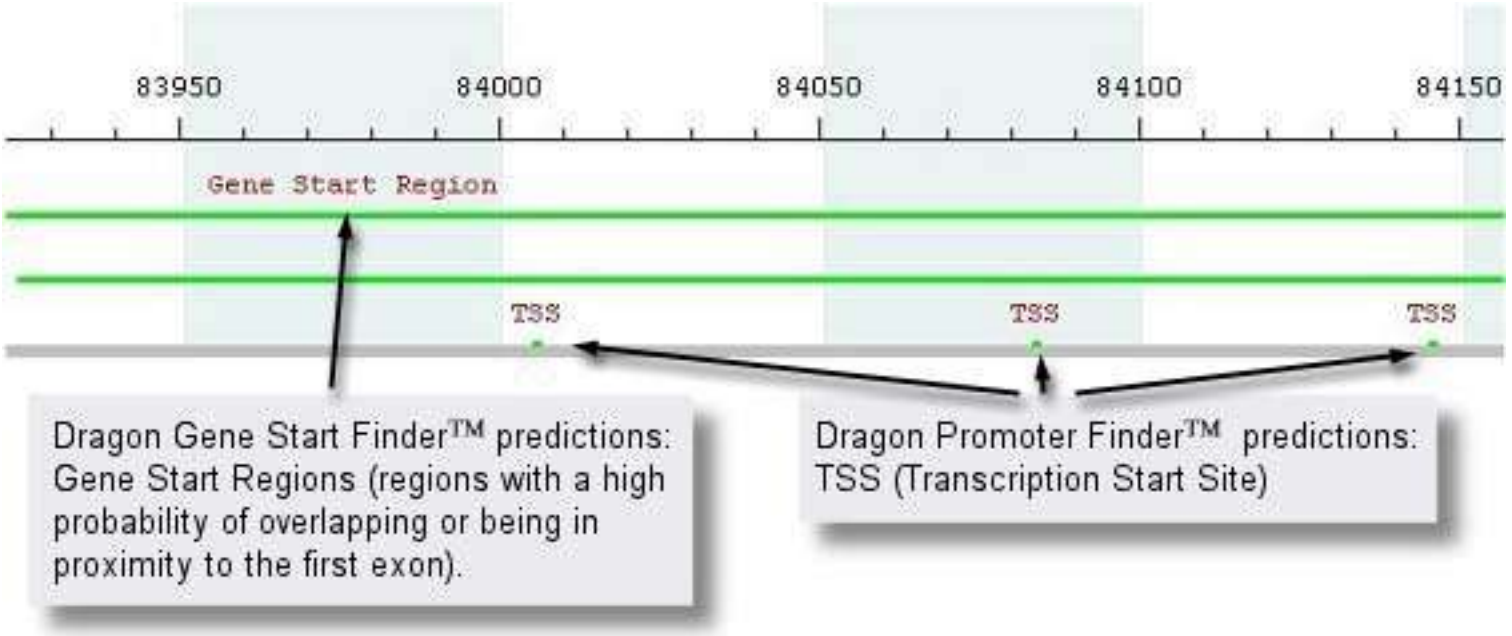
Species *human, Homo sapiens*

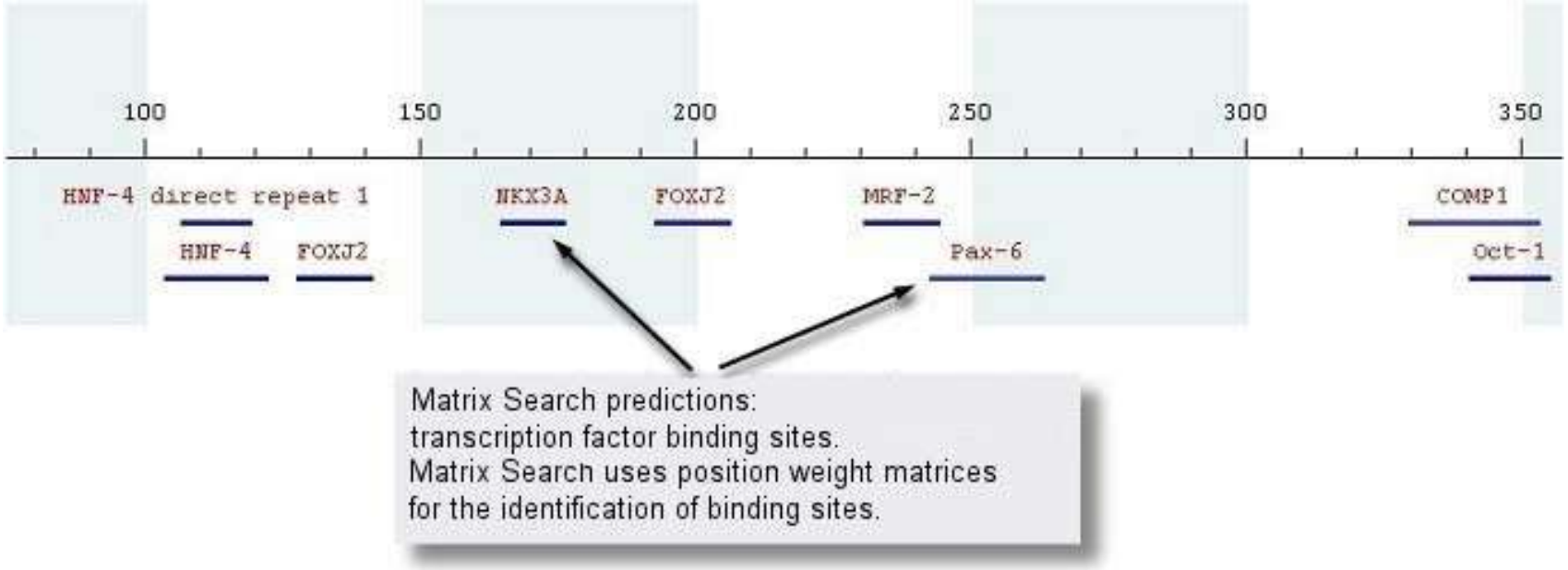
Classification *eukaryota; animalia; metazoa; chordata; vertebrata;*
tetrapoda; mammalia; eutheria; primates

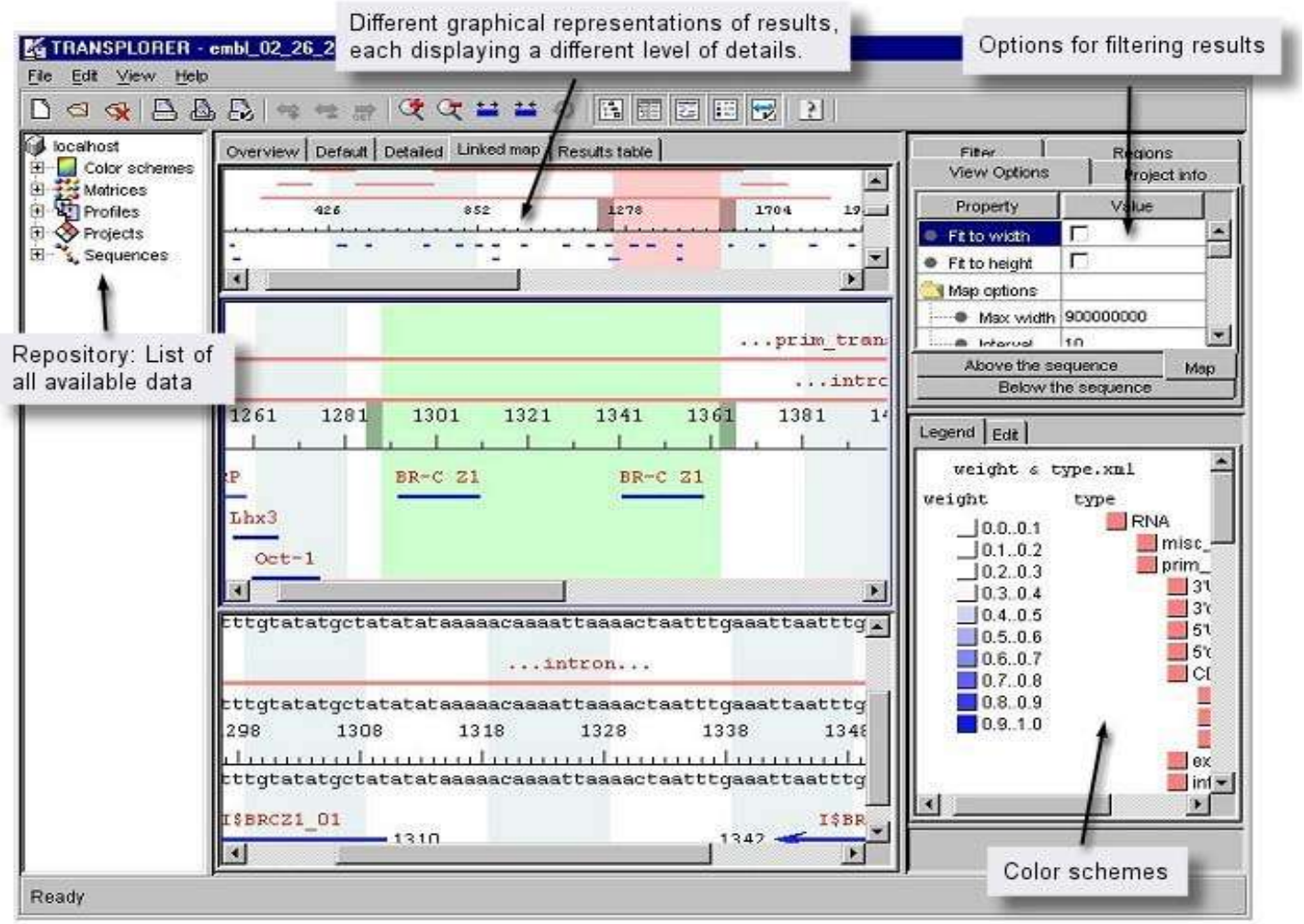
Source *0253; 32D cl3*
0126; K562

Method gel retardation
gel shift competition


Create Project Wizard
Wizard
Step 2 : Select analysis tools
Promoter prediction
Dragon Promoter Finder
Change defaults ...
It is recommanded to use this tool
with sequences longer than 5000 bp only.
Dragon Gene Start Finder
Change defaults ...
Search for transcription factor binding sites
Matrix Search
Change defaults ...
< Previous
Next >
Finish
Project Settings
Project Name
Project Settings
SELECTED TOOLS:
Dragon Promoter Finder (sensitivity: 0.65)
Dragon Gene Start Finder (threshold: 0.994)
Matrix Search (profile: all_minFP_highQua
Progress status
Status:
Elapsed time:
INFO : Est
Done
Cancel
Dragon Promoter Finder™ version 1.3 is a computer system for recognition of functional transcription start sites (TSSs) in RNA polymerase II promoter regions of vertebrates.
Dragon Gene Start Finder™ version 1.0 is a system aimed at predicting regions of different lengths, which have a high probability of overlapping or being in proximity to the first exon.
The third tool (Matrix Search) integrated in TRANSPLORER® is designed to predict potential binding sites for transcription factors in any DNA sequence.

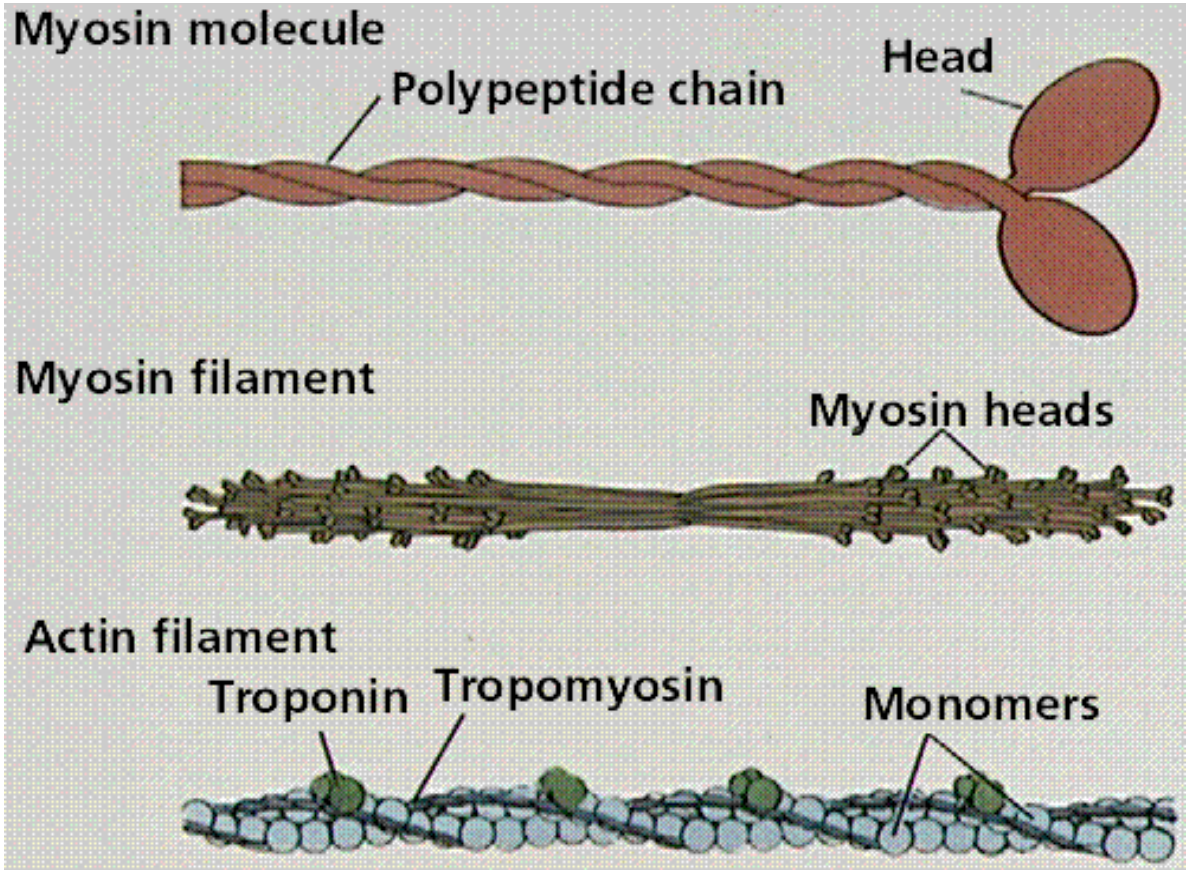

83950
84000
84050
84100
84150
Gene Start Region
TSS
TSS
TSS
Dragon Gene Start Finder™ predictions: Gene Start Regions (regions with a high probability of overlapping or being in proximity to the first exon).
Dragon Promoter Finder™ predictions: TSS (Transcription Start Site)


100
150
200
250
300
350
HNF-4 direct repeat 1
HNF-4
FOXJ2
NKX3A
FOXJ2
MRF-2
Pax-6
COMP1
Oct-1
Matrix Search predictions:
transcription factor binding sites.
Matrix Search uses position weight matrices
for the identification of binding sites.


Different graphical representations of results, each displaying a different level of details.
Options for filtering results
TRANSPLORER - embl_02_26_2
File Edit View Help
localhost
Color schemes
Matrices
Profiles
Projects
Sequences
Repository: List of all available data
Overview Default Detailed Linked map Results table
426 852 1278 1704
...prim_tran
...intrc
1261 1281 1301 1321 1341 1361 1381
BR-C Z1
BR-C Z1
Lhx3
Oct-1
...intron...
1298 1308 1318 1328 1338
I$BRCZ1_01
1310
1342
Filter Regions
View Options Project info
Property Value
Fit to width
Fit to height
Map options
Max width 900000000
Above the sequence
Below the sequence
Map
Legend Edit
weight & type.xml
weight
type
RNA
0.0..0.1
0.1..0.2
0.2..0.3
0.3..0.4
0.4..0.5
0.5..0.6
0.6..0.7
0.7..0.8
0.8..0.9
0.9..1.0
Color schemes
Ready

protein – protein

důležité u většiny proteinů, protože jenom ojediněle fungují izolovaně


Myosin molecule
Polypeptide chain
Head
Myosin filament
Myosin heads
Actin filament
Troponin
Tropomyosin
Monomers

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

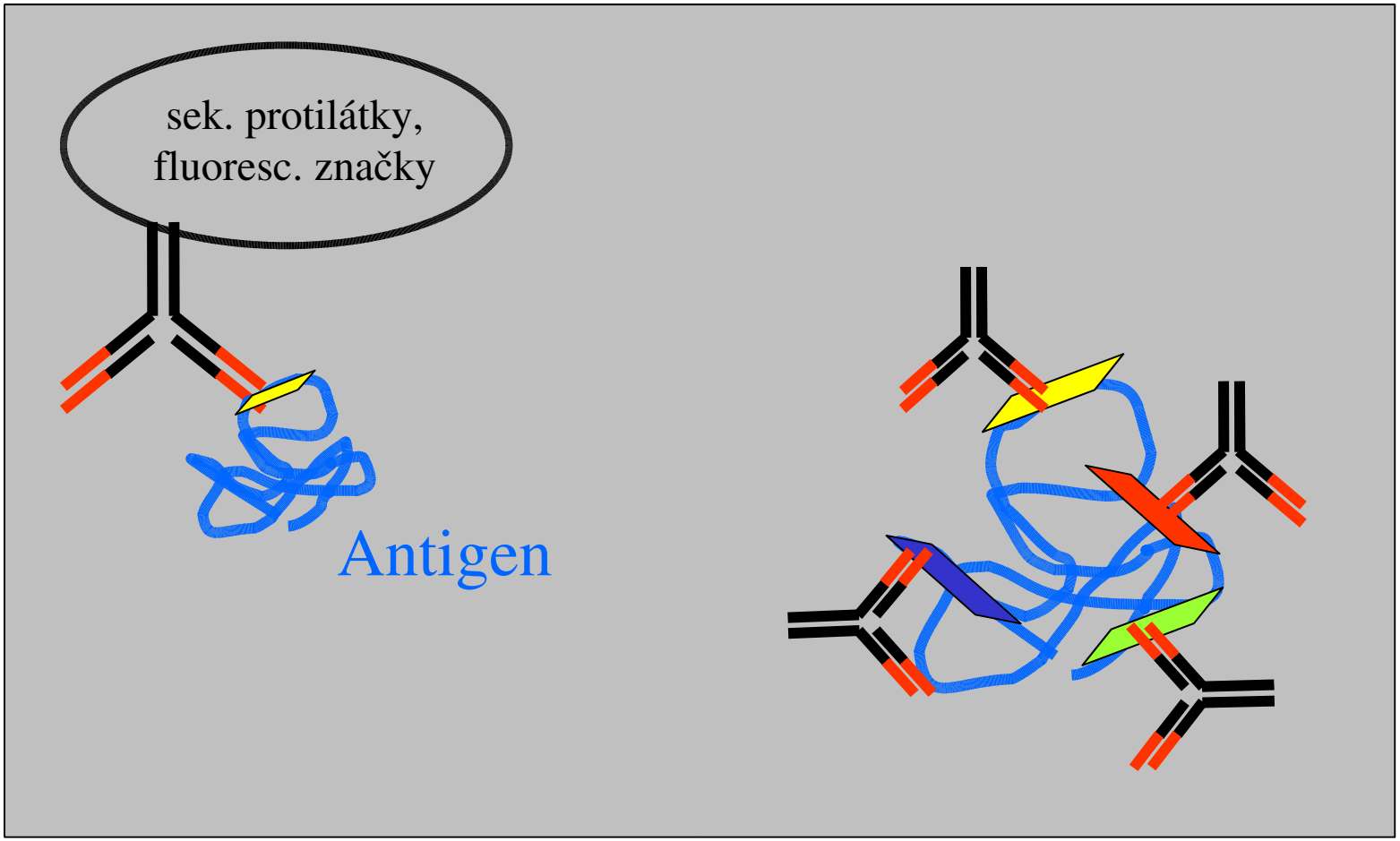
Measure reporter enzyme activity as e.g. blue colonies.

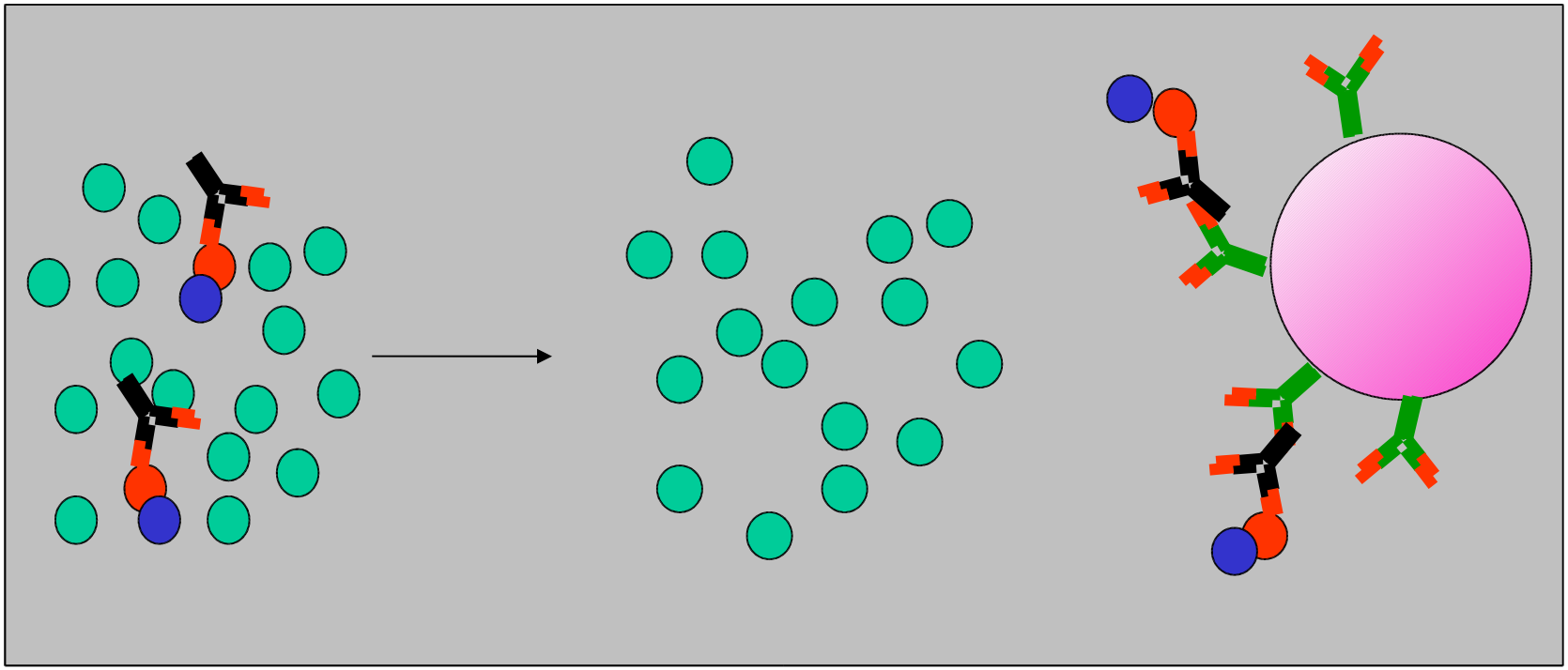
Yeast two hybrid System (3)

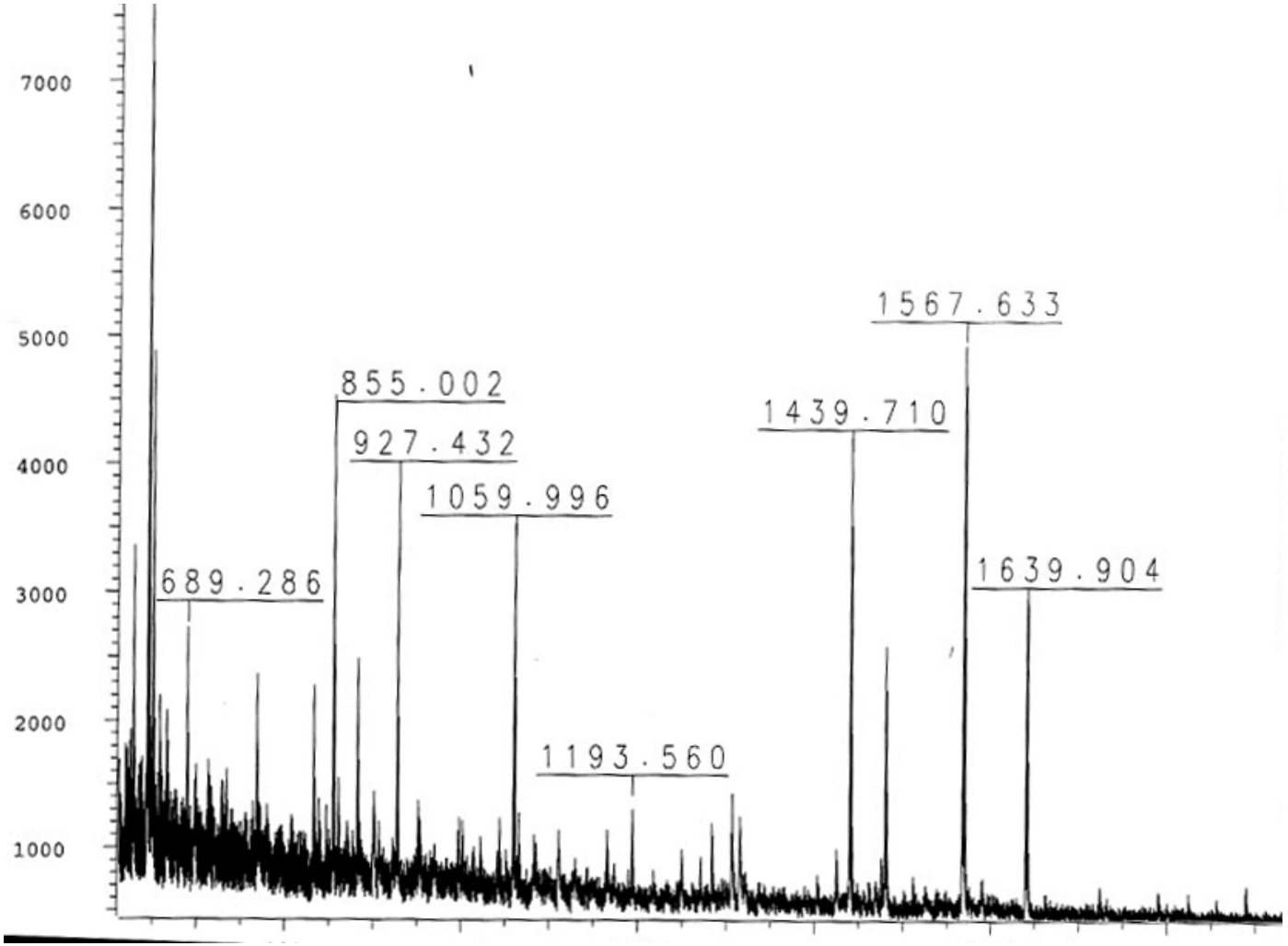
- This is achieved using gene fusions:
- Plasmids carrying different constructs can be expressed in yeast.

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

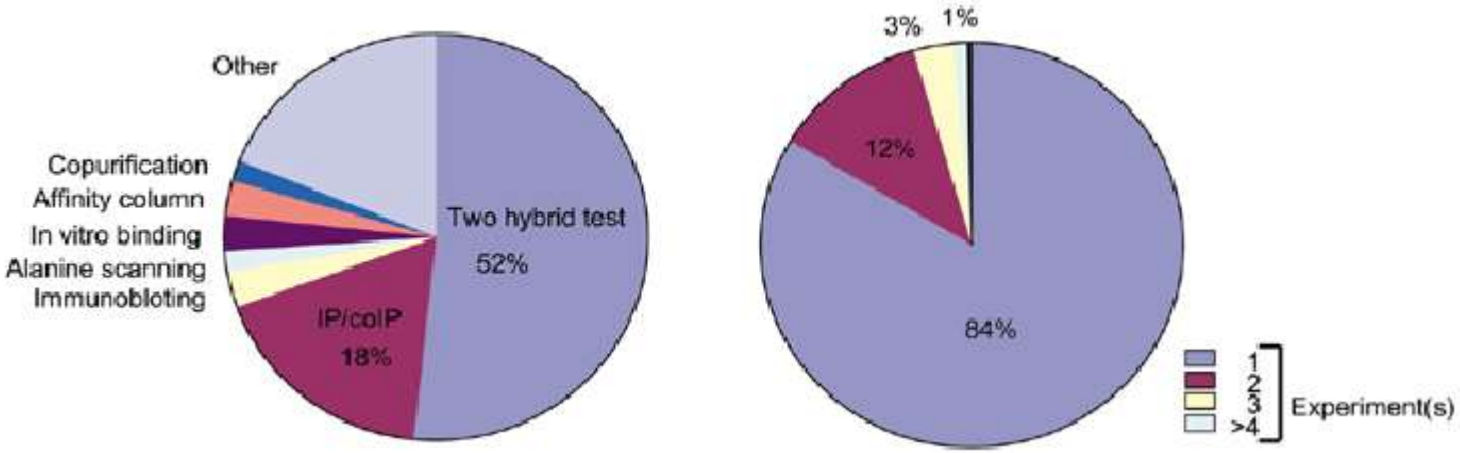

sek. protilátky,
fluoresc. značky
Antigen

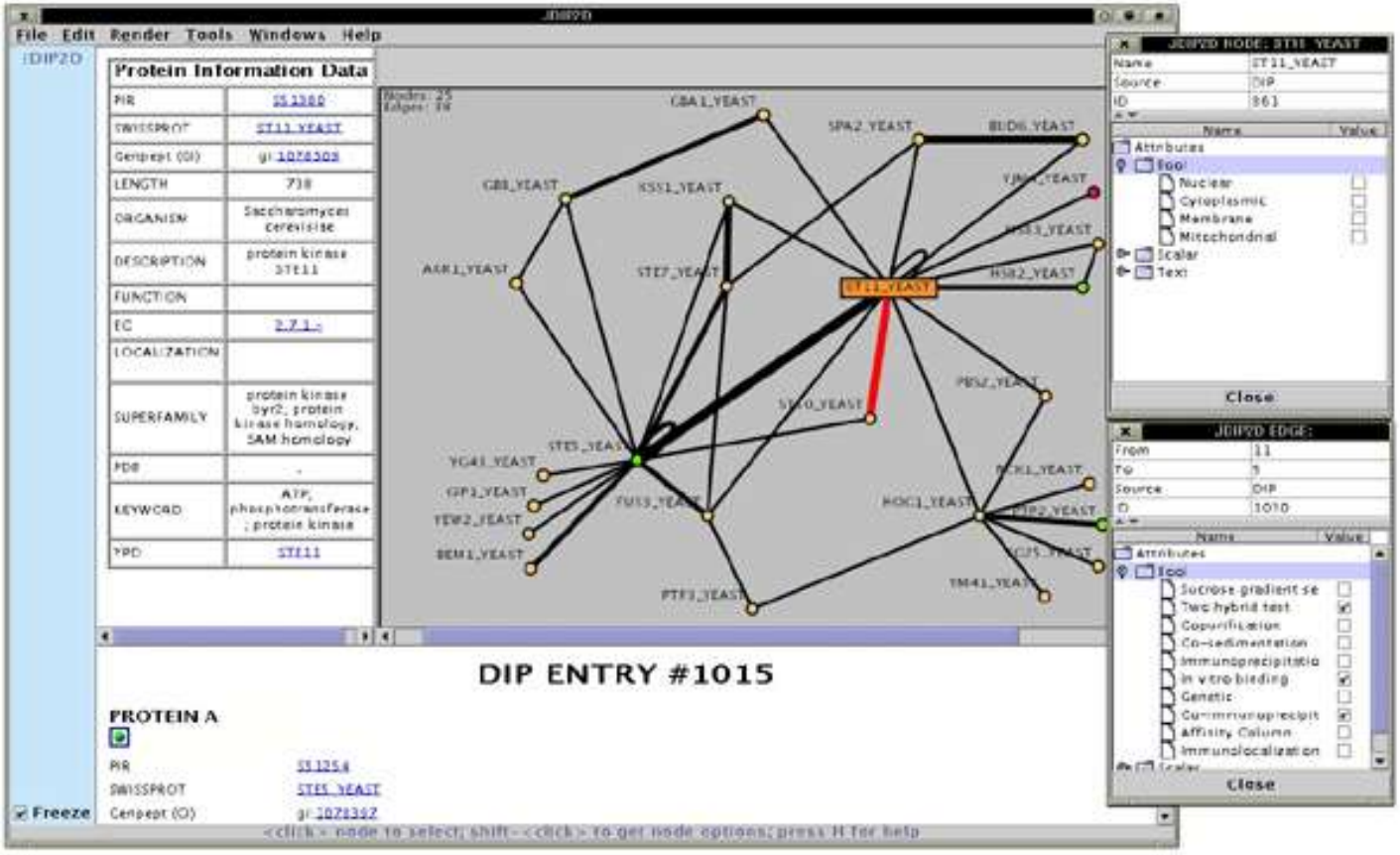


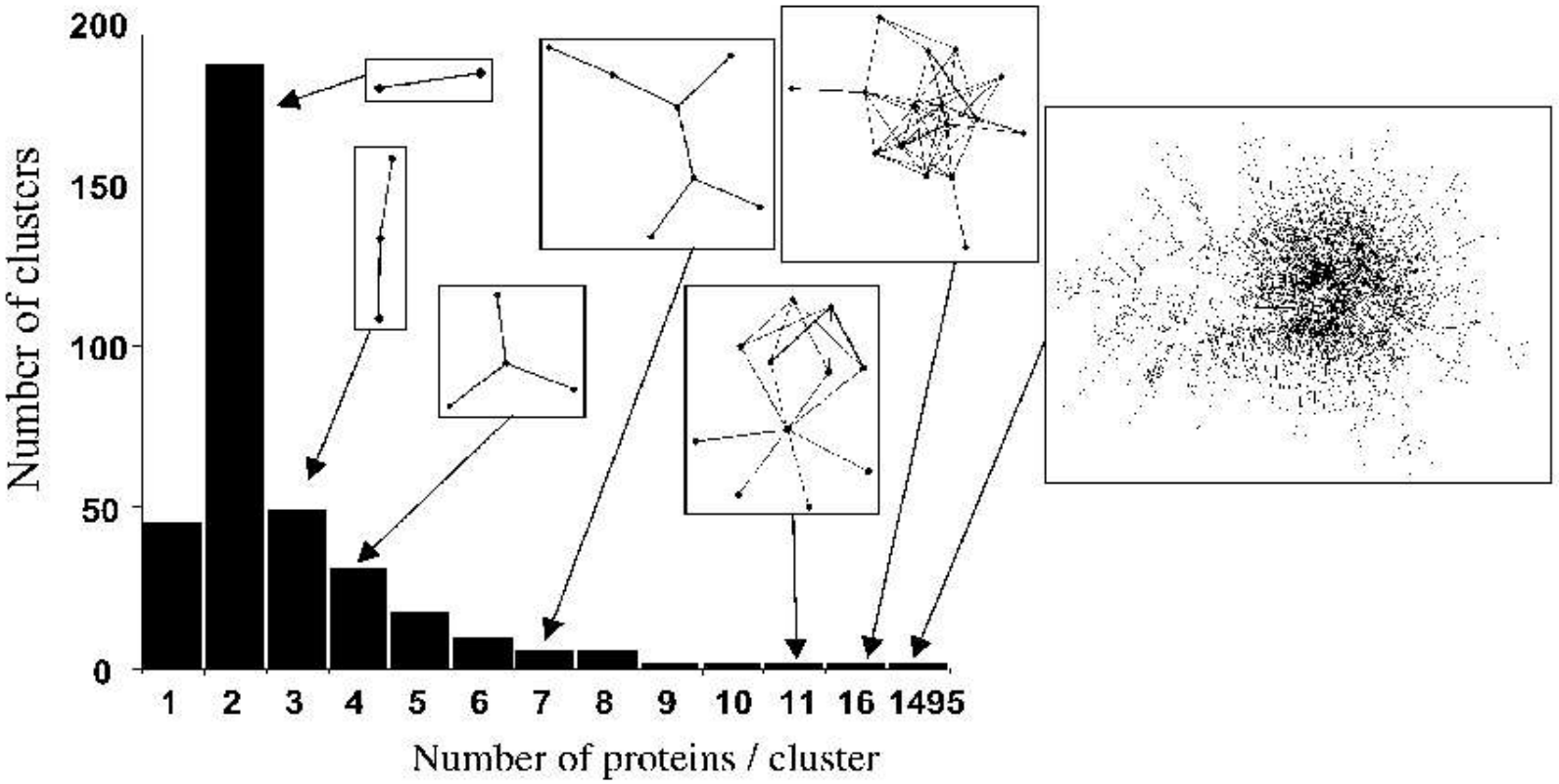

7000
6000
5000
4000
3000
2000
1000
1567.633
855.002
1439.710
927.432
1059.996
689.286
1639.904
1193.560

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

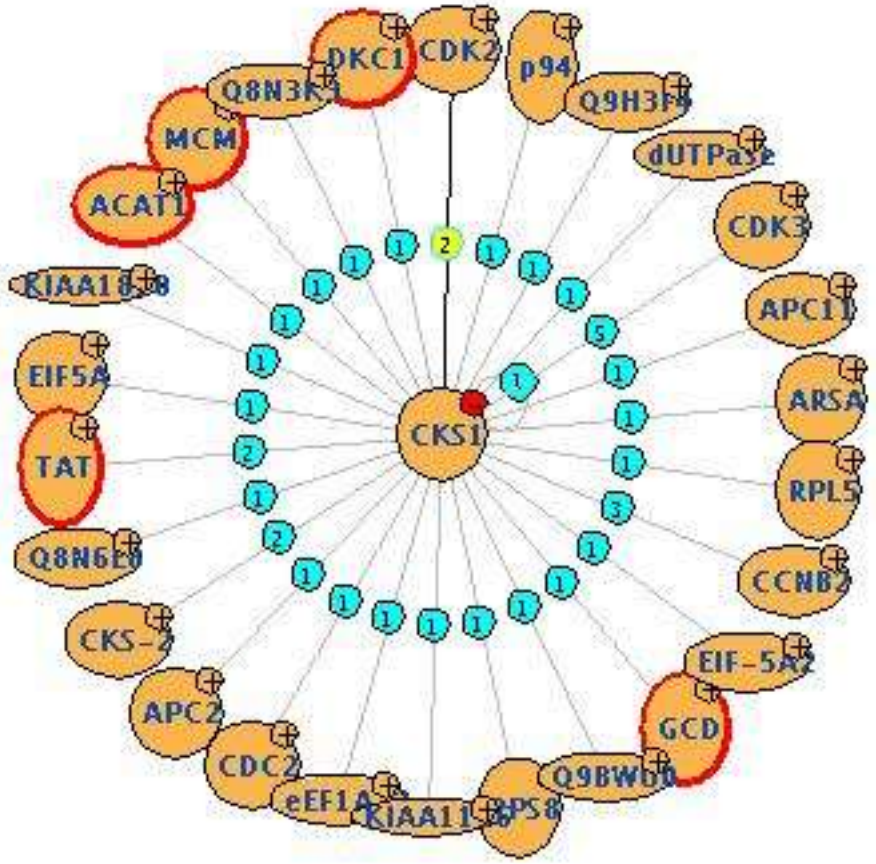
Number of proteins	**17556**
Number of organisms	**109**
Number of interactions	**46463**
Number of distinct experiments describing an interaction	**51915**
Number of data sources (articles)	**2884**
Number of data sources (other)	**34**


Other
Copurification
Affinity column
In vitro binding
Alanine scanning
Immunobloting
Two hybrid test
52%
IP/coIP
18%
3%
1%
12%
84%
1
2
3
>4
Experiment(s)


File Edit Render Tools Windows Help
DIP2D
Protein Information Data
PIR
SWISSPROT
Genpept (GI)
LENGTH
738
ORGANISM
Saccharomyces cerevisiae
DESCRIPTION
protein kinase STE11
FUNCTION
EC
LOCALIZATION
SUPERFAMILY
protein kinase byr2, protein kinase homology, SAM homology
PDB
KEYWORD
ATP; phosphotransferase; protein kinase
YPD
STE11
Nodes: 25
Edges: 34
GBA1_YEAST
SPA2_YEAST
BUD6_YEAST
GB8_YEAST
KSS1_YEAST
AKR1_YEAST
STE7_YEAST
STE11_YEAST
HSH2_YEAST
PBS2_YEAST
STE5_YEAST
YG41_YEAST
GP3_YEAST
YEW2_YEAST
BEM1_YEAST
FUS3_YEAST
BCK1_YEAST
ROC1_YEAST
PTP2_YEAST
PTP3_YEAST
YM41_YEAST
DIP ENTRY #1015
PROTEIN A
PIR
SWISSPROT
STE5_YEAST
Genpept (GI)
Freeze
<click> node to select; shift-<click> to get node options; press H for help
JDIP2D NODE: STE11_YEAST
Name
ST11_YEAST
Source
DIP
ID
861
Name
Value
Attributes
Too
Nuclear
Cytoplasmic
Membrane
Mitochondrial
Scalar
Text
Close
JDIP2D EDGE:
From
11
To
5
Source
DIP
ID
1010
Name
Value
Attributes
Too
Sucrose gradient se
Two hybrid test
Copurification
Co-sedimentation
Immunoprecipitatio
In vitro binding
Genetic
Co-immunoprecipit
Affinity Column
Immunolocalization
Close


200
150
100
50
0
Number of clusters
1
2
3
4
5
6
7
8
9
10
11
16
1495
Number of proteins / cluster

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


DKC1
CDK2
p94
MCM
ACAT1
CDK3
EIF5A
TAT
CKS1
ARSA
RPL5
CCNB2
CKS-2
APC2
CDC2
GCD
eEF1A

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

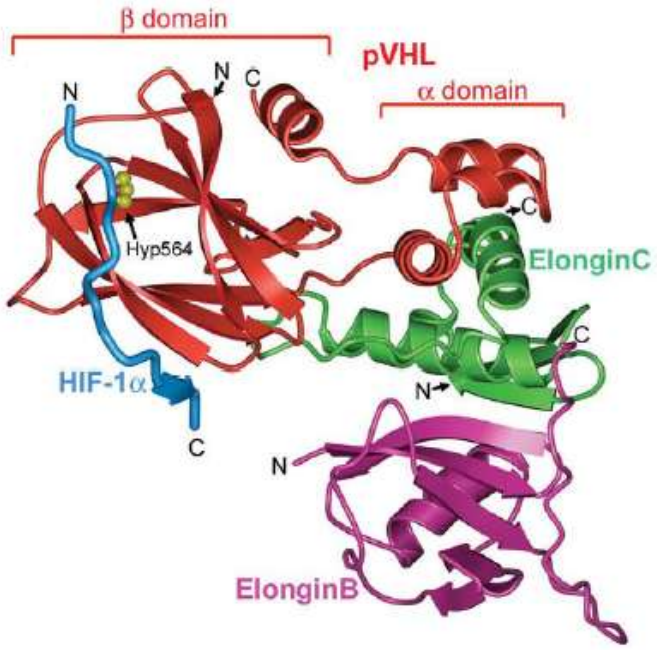
grid: http://biodata.mshri.on.ca/grid/servlet/Index

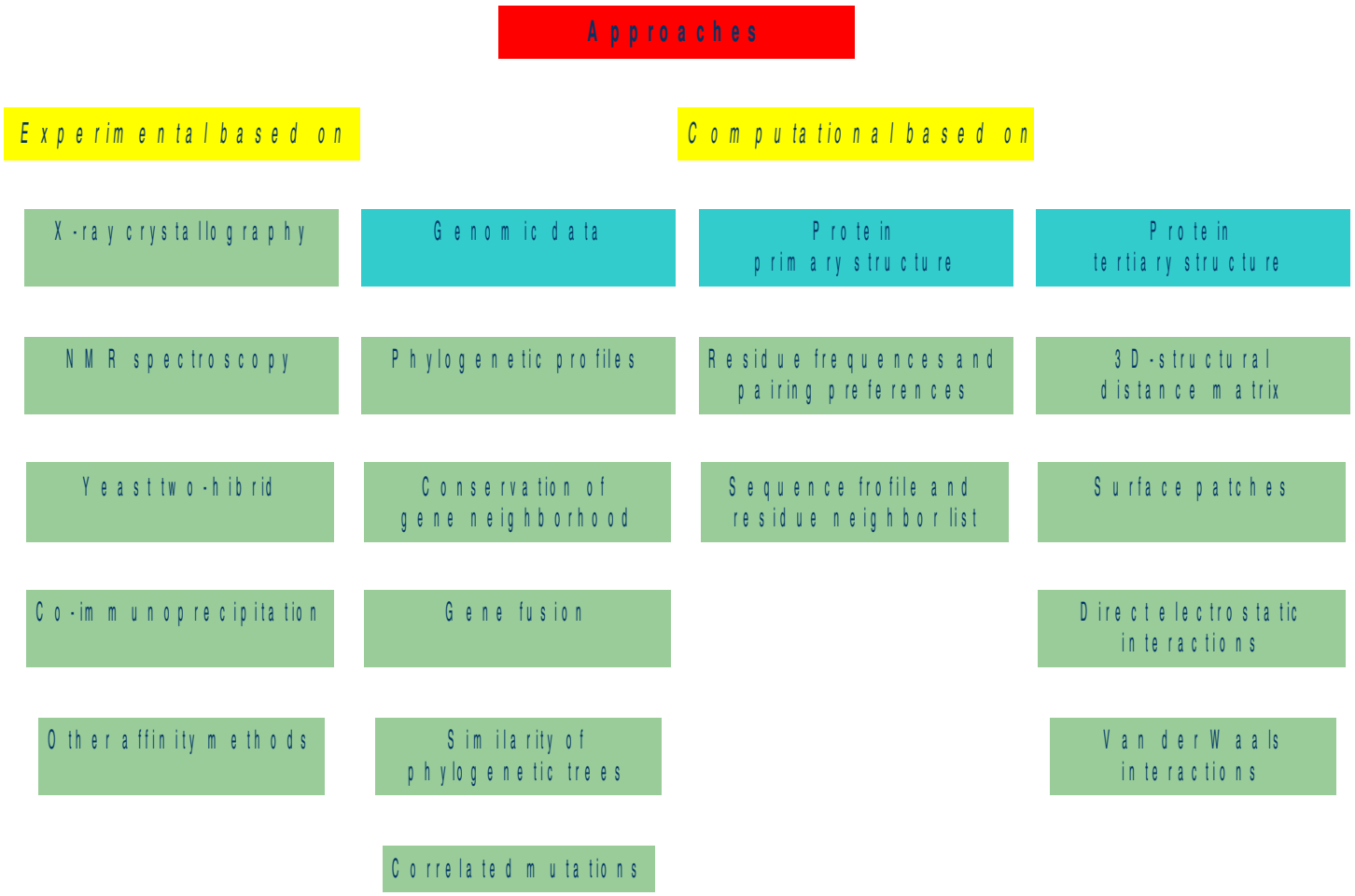
Human protein reference database: http://www.hprd.org/

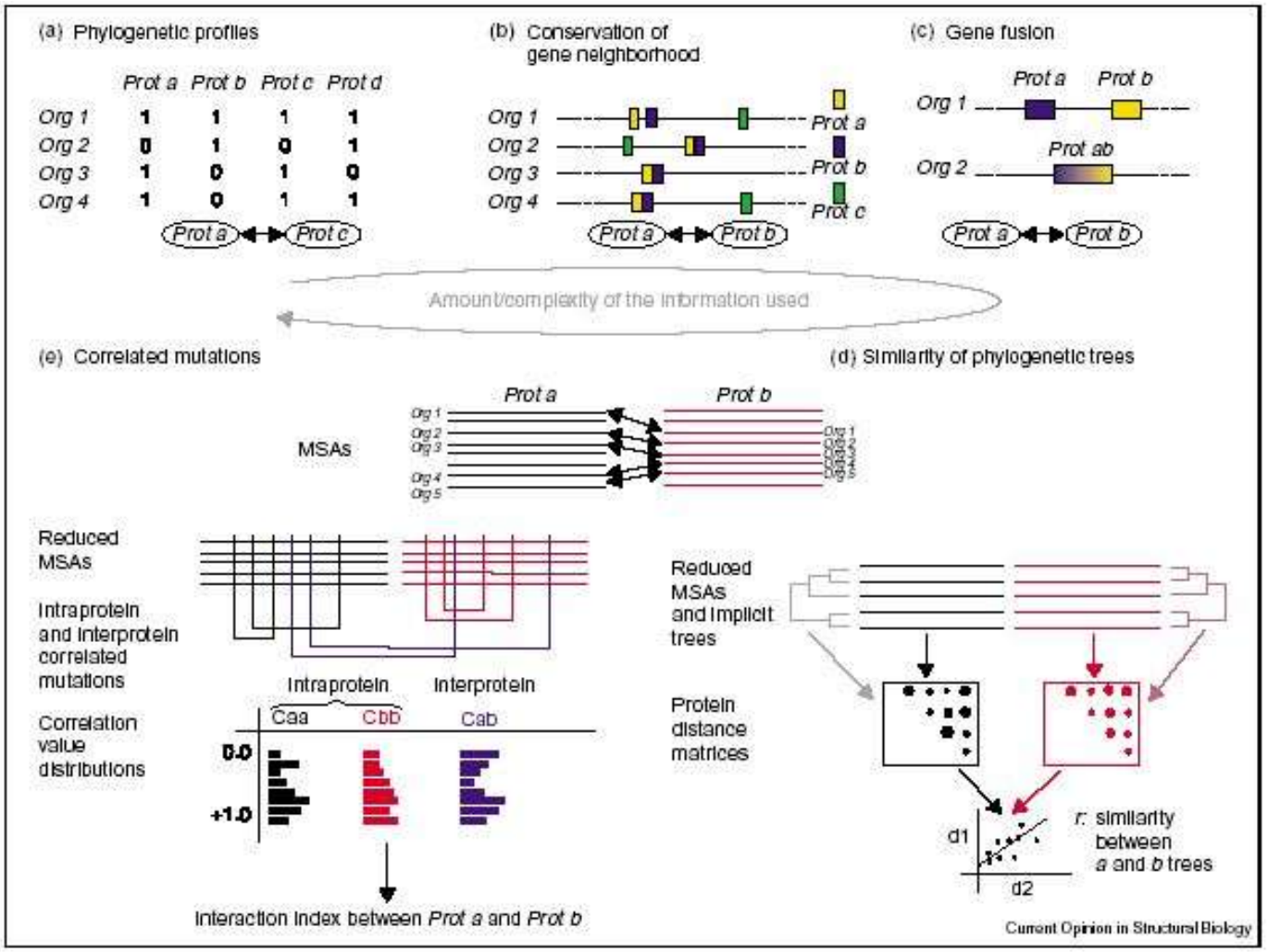
Visualization tools:

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

Cytoscape: http://www.cytoscape.org

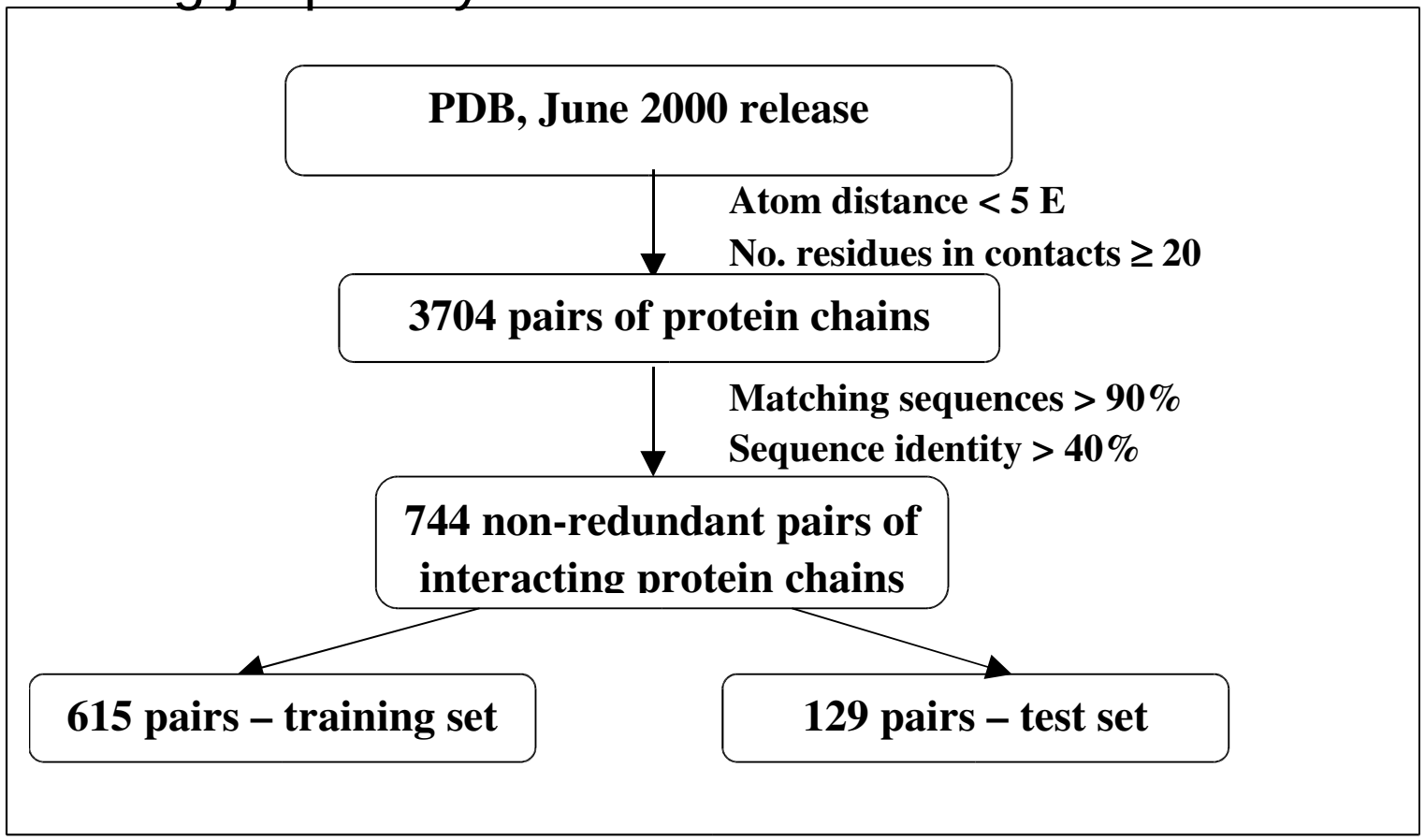

β domain
pVHL
α domain
N
N
C
C
Hyp564
ElonginC
C
HIF-1α
N
C
N
ElonginB


Approaches
Experimental based on
Computational based on
X-ray crystallography
Genomic data
Protein primary structure
Protein tertiary structure
NMR spectroscopy
Phylogenetic profiles
Residue frequences and pairing preferences
3D-structural distance matrix
Yeast two-hibrid
Conservation of gene neighborhood
Sequence frofile and residue neighbor list
Surface patches
Co-immunoprecipitation
Gene fusion
Direct electrostatic interactions
Other affinity methods
Similarity of phylogenetic trees
Van der Waals interactions
Correlated mutations

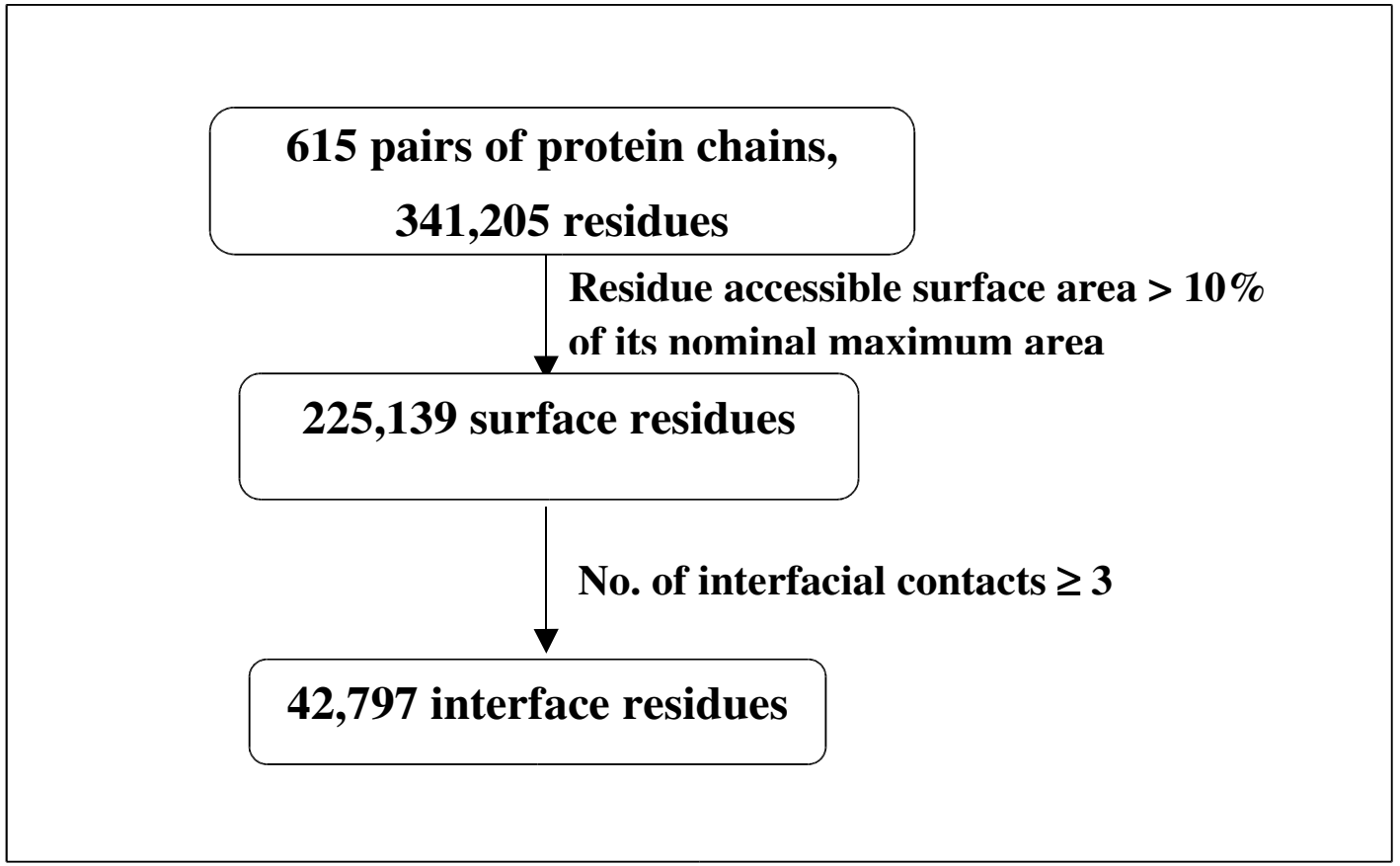

(a) Phylogenetic profiles
Prot a Prot b Prot c Prot d
Org 1 1 1 1 1
Org 2 0 1 0 1
Org 3 1 0 1 0
Org 4 1 0 1 1
Prot a
Prot c
(b) Conservation of gene neighborhood
Org 1
Org 2
Org 3
Org 4
Prot a
Prot b
Prot c
Prot a
Prot b
(c) Gene fusion
Prot a
Prot b
Org 1
Prot ab
Org 2
Prot a
Prot b
Amount/complexity of the information used
(e) Correlated mutations
(d) Similarity of phylogenetic trees
MSAs
Prot a
Prot b
Org 1
Org 2
Org 3
Org 4
Org 5
Reduced MSAs
Intraprotein and interprotein correlated mutations
Correlation value distributions
Intraprotein
Interprotein
Caa
Cbb
Cab
0.0
+1.0
Interaction index between Prot a and Prot b
Reduced MSAs and implicit trees
Protein distance matrices
d1
d2
r: similarity between a and b trees
Current Opinion in Structural Biology

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

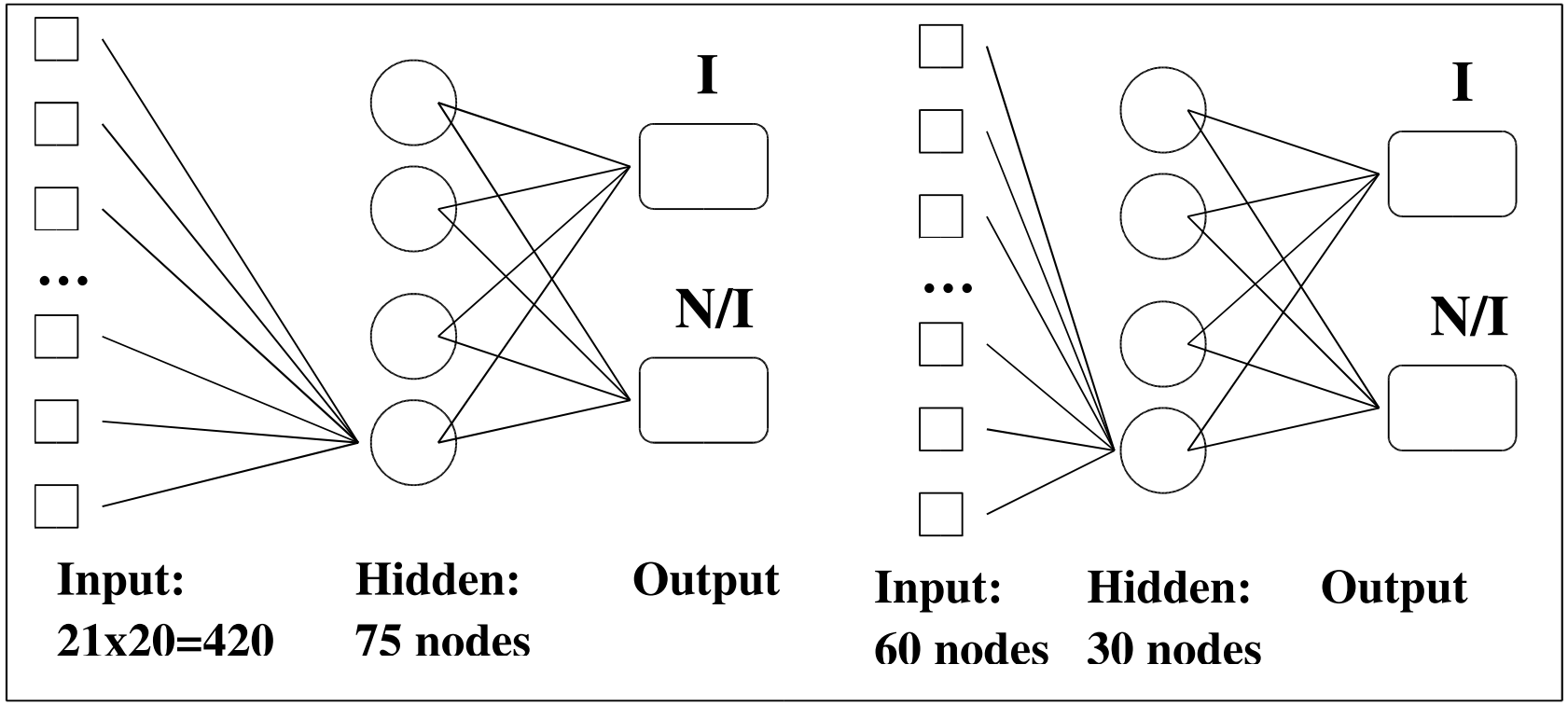
Interagující proteiny



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)

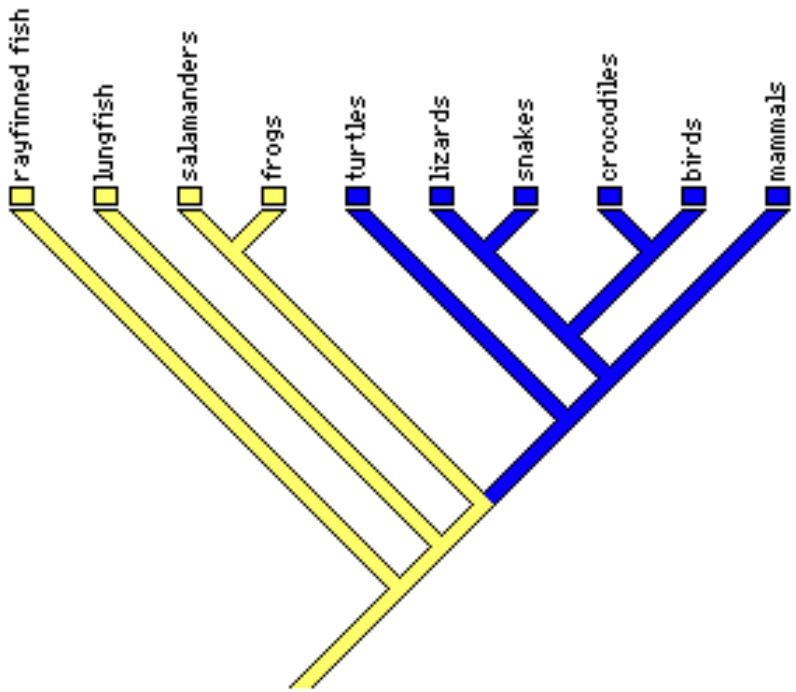
EPR Expression Profile Reliability

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)

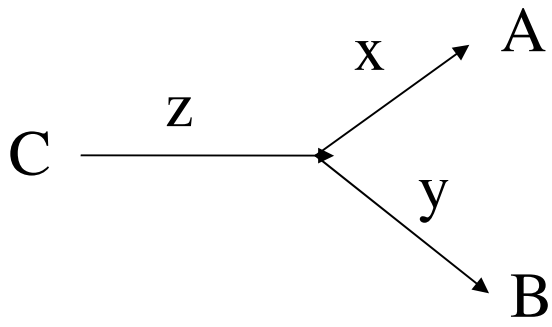
UPGMA

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)”

<=>

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,...)

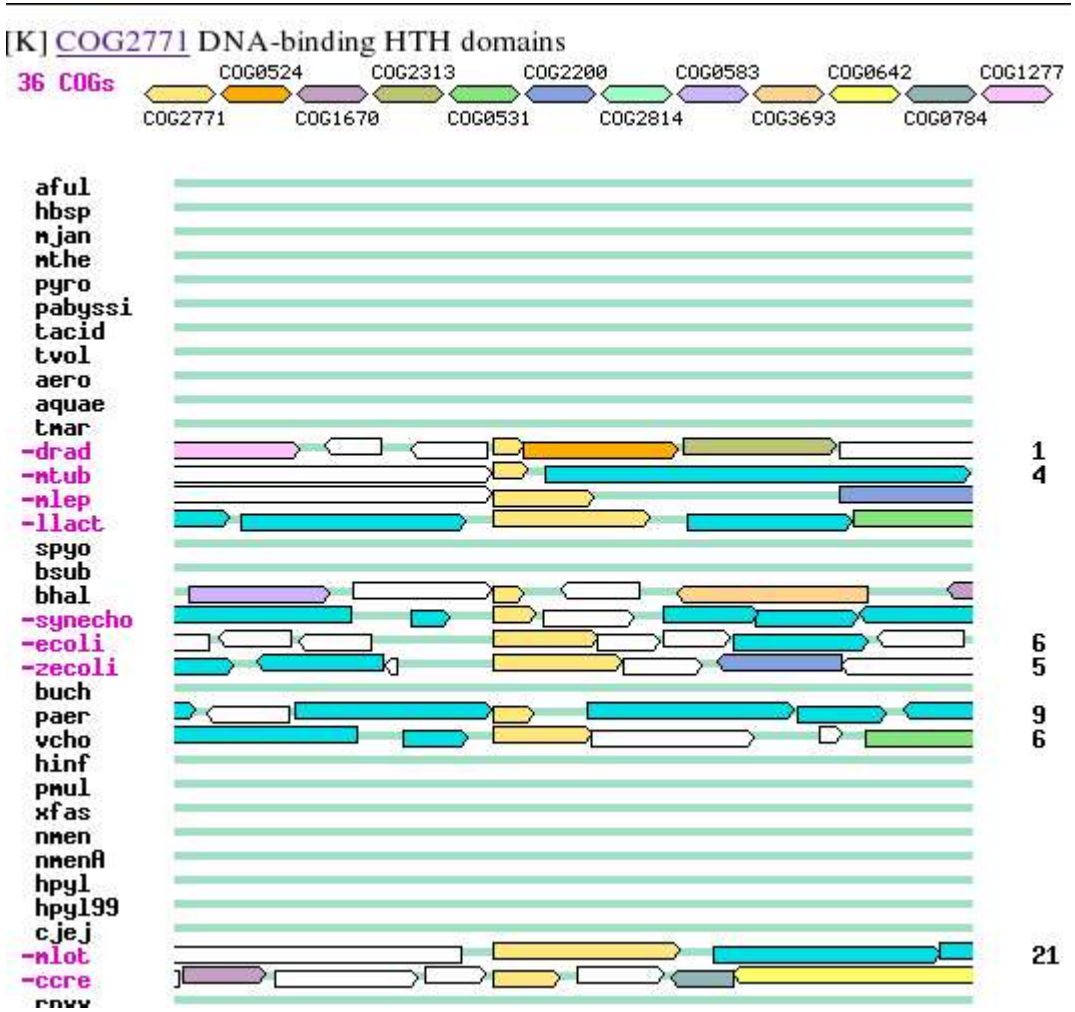
COG je nástrojem, který podobně jako např.
Gene Ontology dokáže propojovat na vyšší úrovni
a tím zhodnocovat informace přicházející ze
sekvenace různých genomů.

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 metabolizmu

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

- ztráta genů
- horizontální transfer


[K] COG2771 DNA-binding HTH domains
36 COGs
COG0524
COG2313
COG2200
COG0583
COG0642
COG1277
COG2771
COG1670
COG0531
COG2814
COG3693
COG0784
aful
hbsp
mjan
mthe
pyro
pabyssi
tacid
tvol
aero
aquae
tmar
-drad
-mtub
-mlep
-llact
spyo
bsub
bhal
-synecho
-ecoli
-zecoli
buch
paer
vcho
hinf
pmul
xfas
nmen
nmenA
hpyl
hpyl99
cjej
-mlot
-ccre
1
4
6
5
9
6
21