Bridging predictions and experiments: some case examples

Ana M. Rojas-Mendoza
Sardinia, 30th June 2006

cnio
Centro Nacional de Investigaciones Oncológicas

cnio
Centro Nacional
de Investigaciones
Oncológicas

Biologist by training, Madrid, Spain.
Posdoc @ UC San Diego. (Dr. RF Doolittle)
Posdoc @ The Burnham Institute.
Researcher @ CNB, Spain.
Staff Scientist @ CNIO.

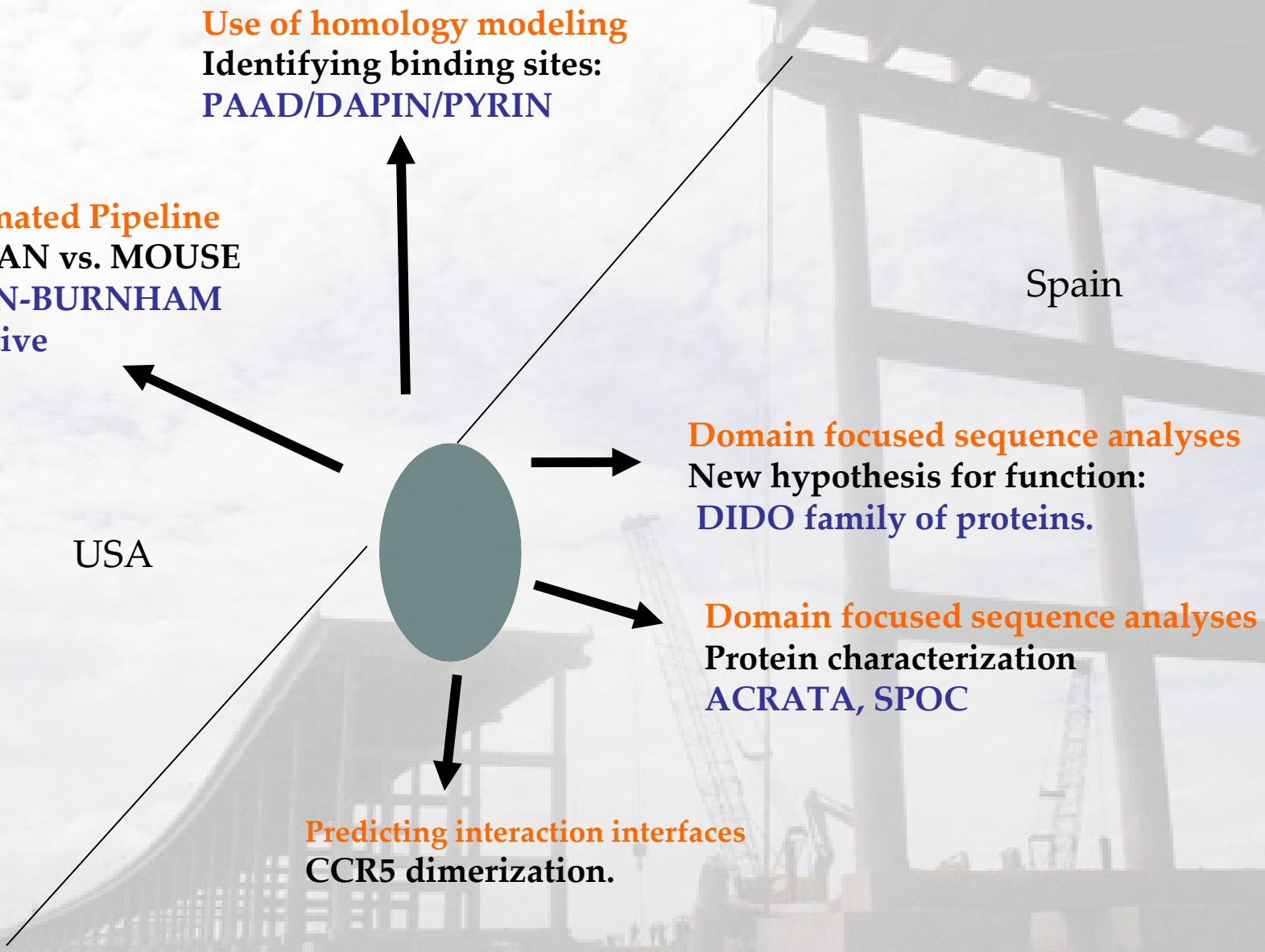
Apoptosis and Inflammation (Reed's lab).
Bioinformatics and Systems Biology (Adam Godzik's lab).

The Burnham Institute, la Jolla, CA (USA).

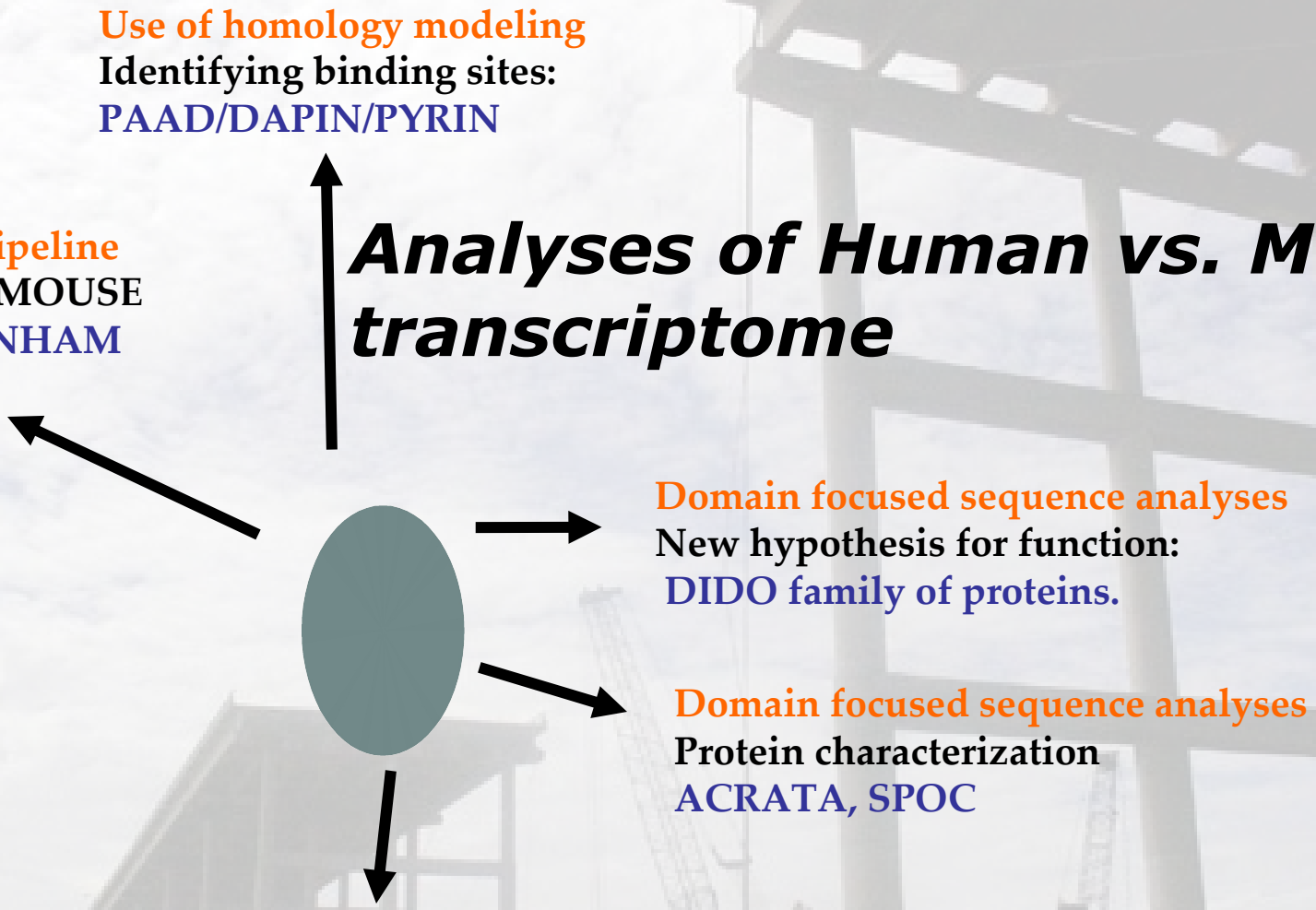
Cancer and apoptosis (Carlos Martinez-A's lab).
Protein Design Group (Alfonso Valencia's lab).

Spanish Centre for Biotechnology (CNB).

cnio
Centro Nacional de Investigaciones Oncológicas



cnio
Centro Nacional de Investigaciones Oncológicas



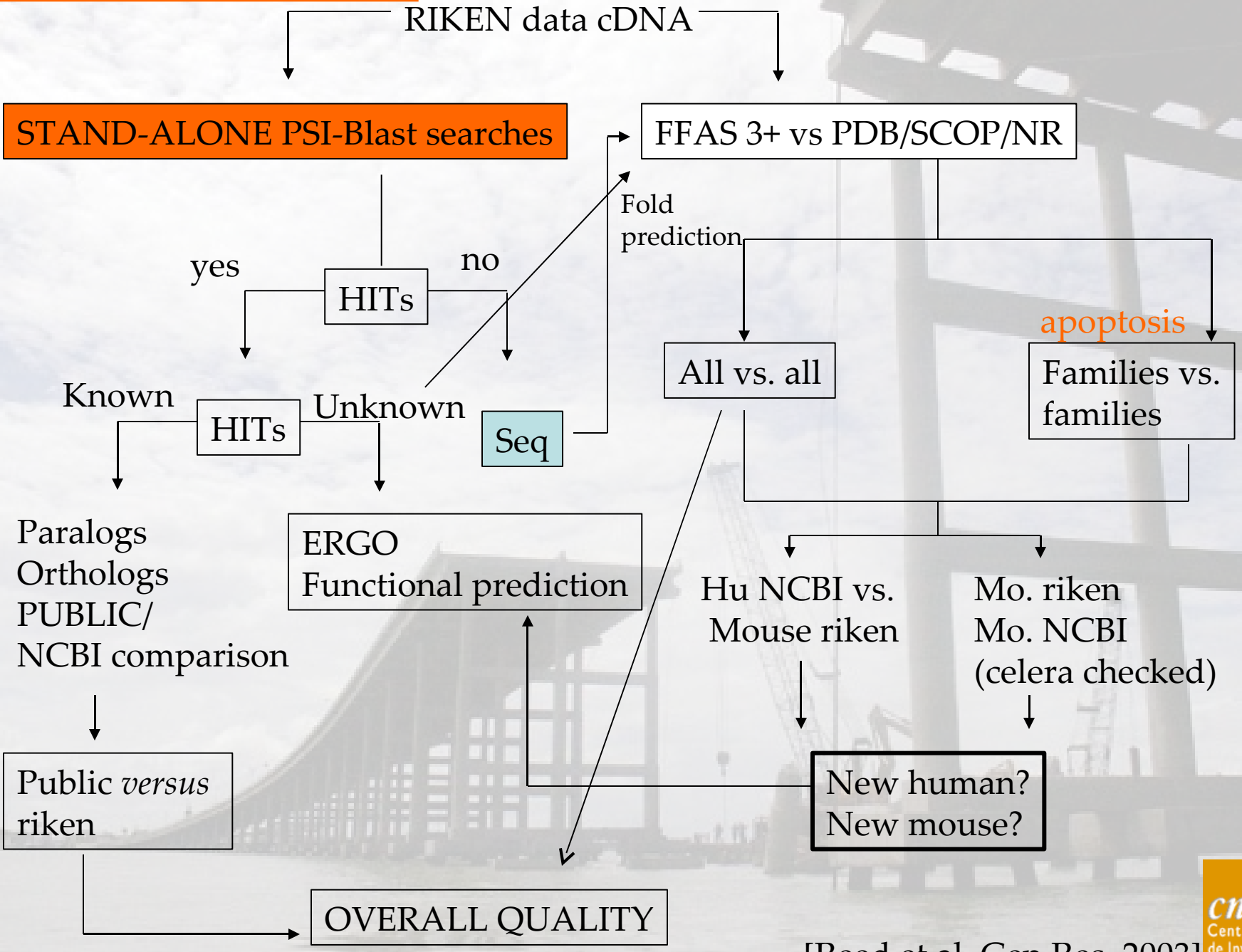
GOAL:

ANNOTATION OF PROTEIN INVOLVED IN APOPTOSIS USING THE MOUSE CDNA RIKEN-FANTOM2 COLLECTION (back to 2003).

[Reed et al, Gen Res, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

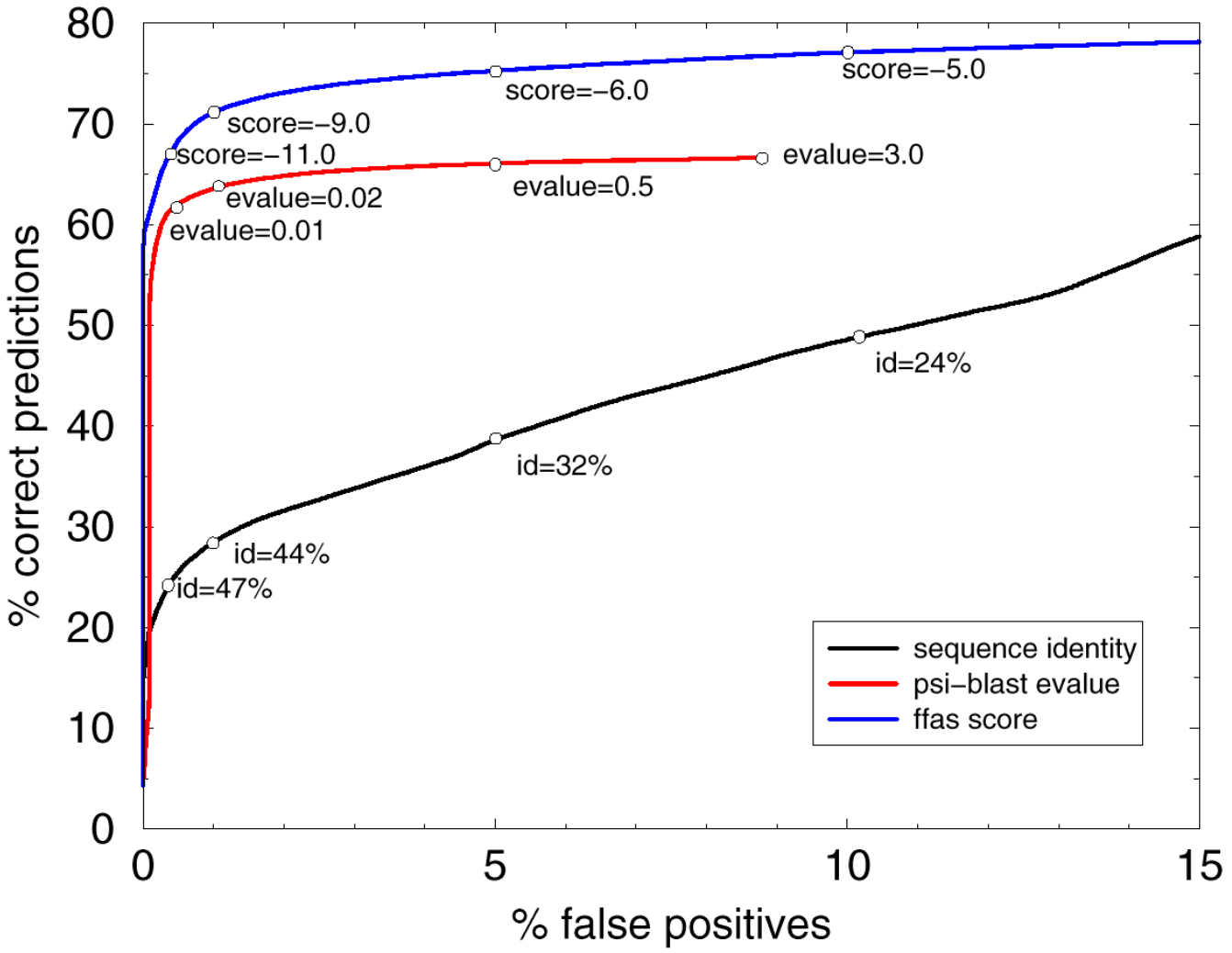
(1) PIPELINE CREATION



[Reed et al, Gen Res, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

Fold prediction validation



[Reed et al, Gen Res, 2003]

Analysis of RIKEN clone collection http://

Mouse genome annotation development site

@ The Burnham Institute
Godzik Lab

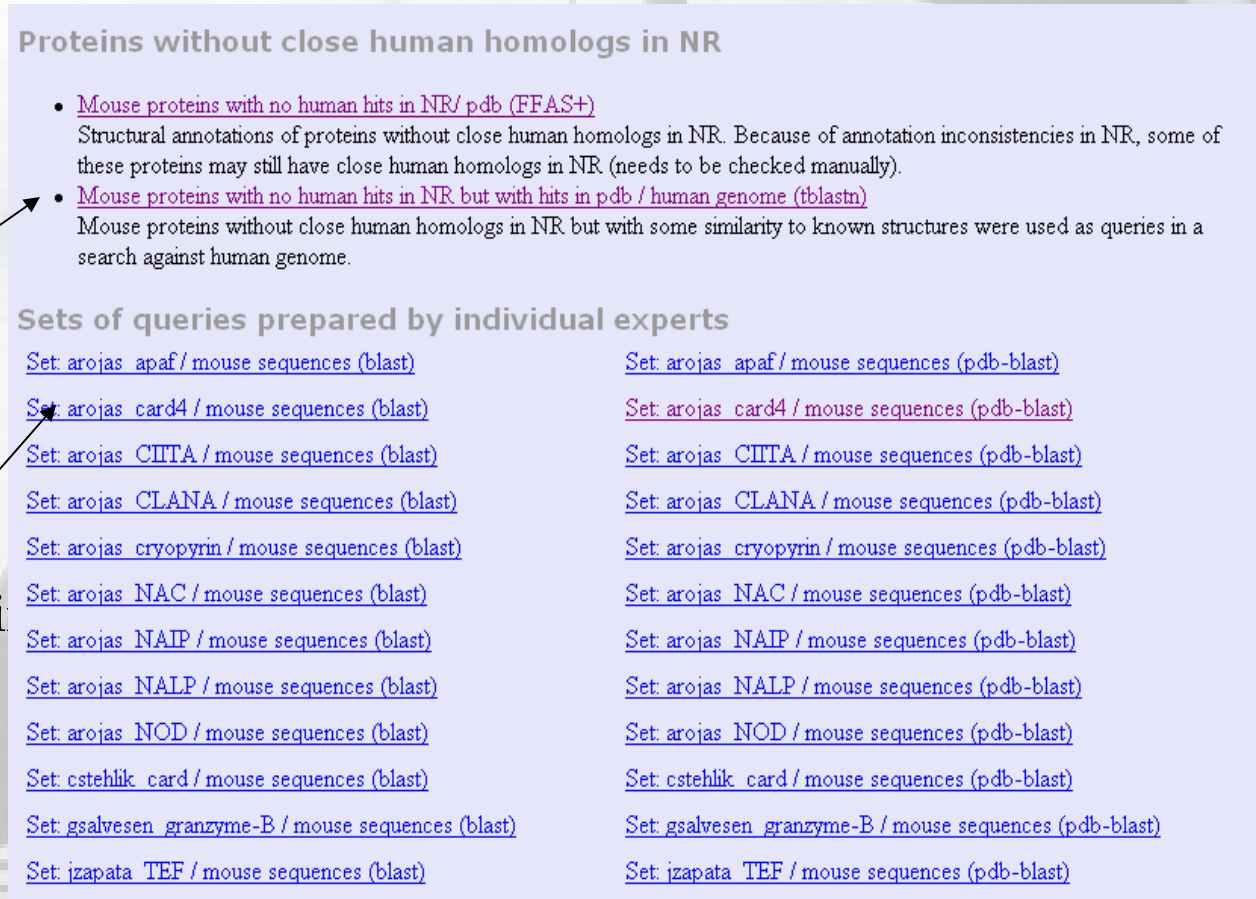
Annotation of apoptotic proteins

- Apoptotic proteins / mouse proteins (blast)
  These results contain strong matches between set of apoptotic proteins provided by the team of experts from The Burnham Institute and sequences of mouse proteins.
- Apoptotic proteins / mouse proteins (pdb-blast)
  Same as above but Pdb-Blast was used instead of Blast. Pdb-blast is more sensitive than Blast, but also more vulnerable to low complexity regions, which may yield false positives.
- Apoptotic Pfam domains / mouse proteins (blast)
  These results contain strong matches between set of sequences of apoptotic domains selected from PfamA and mouse proteins.
- Apoptotic Pfam domains / mouse proteins (pdb-blast)
  Same as above but Pdb-Blast was applied. In this case low-complexity problem is less important, since PfamA domain seqences usually don't contain such regions.
- Apoptotic Pfam domains / mouse proteins (FFAS+)
  Same as above but FFAS+ was applied. More interesting, "twilight zone" hits can be detected with this method.

[Reed et al, Gen Res, 2003]

Analysis of RIKEN FANTOM2 clone collection

- **Entire collection???**
- Double focus
  - New predictions
  - Support of analysis of apoptotic genes (in collaboration with Dr. John Reed's group i 15 domain families)



Proteins without close human homologs in NR

- Mouse proteins with no human hits in NR/ pdb (FFAS+)
  Structural annotations of proteins without close human homologs in NR. Because of annotation inconsistencies in NR, some of these proteins may still have close human homologs in NR (needs to be checked manually).
- Mouse proteins with no human hits in NR but with hits in pdb / human genome (tblastn)
  Mouse proteins without close human homologs in NR but with some similarity to known structures were used as queries in a search against human genome.

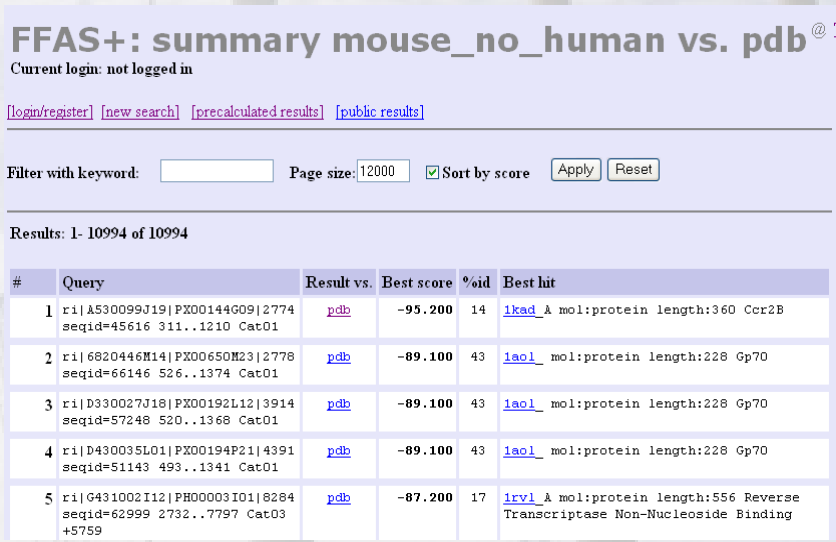
Sets of queries prepared by individual experts

Set: arojas_apaf / mouse sequences (blast)	Set: arojas_apaf / mouse sequences (pdb-blast)
Set: arojas_card4 / mouse sequences (blast)	Set: arojas_card4 / mouse sequences (pdb-blast)
Set: arojas_CIITA / mouse sequences (blast)	Set: arojas_CIITA / mouse sequences (pdb-blast)
Set: arojas_CLANA / mouse sequences (blast)	Set: arojas_CLANA / mouse sequences (pdb-blast)
Set: arojas_cryopyrin / mouse sequences (blast)	Set: arojas_cryopyrin / mouse sequences (pdb-blast)
Set: arojas_NAC / mouse sequences (blast)	Set: arojas_NAC / mouse sequences (pdb-blast)
Set: arojas_NAIP / mouse sequences (blast)	Set: arojas_NAIP / mouse sequences (pdb-blast)
Set: arojas_NALP / mouse sequences (blast)	Set: arojas_NALP / mouse sequences (pdb-blast)
Set: arojas_NOD / mouse sequences (blast)	Set: arojas_NOD / mouse sequences (pdb-blast)
Set: cstehlik_card / mouse sequences (blast)	Set: cstehlik_card / mouse sequences (pdb-blast)
Set: gsalvesen_granzyme-B / mouse sequences (blast)	Set: gsalvesen_granzyme-B / mouse sequences (pdb-blast)
Set: jzapata_TEF / mouse sequences (blast)	Set: jzapata_TEF / mouse sequences (pdb-blast)

[Reed et al, Gen Res, 2003]

Novel mouse proteins?

- ~10,000 clones with predicted amino acid sequence >100 aa. have **no** homologs in NR
- ~500 have statistically significant fold predictions (Z-score >9, estimated error rate <1%)
- ~1500 have a ~50% chance of having correct fold assigned



FFAS+: summary mouse_no_human vs. pdb

Current login: not logged in

[login/register] [new search] [precalculated results] [public results]

Filter with keyword: Page size: 12000 Sort by score Apply Reset

Results: 1- 10994 of 10994

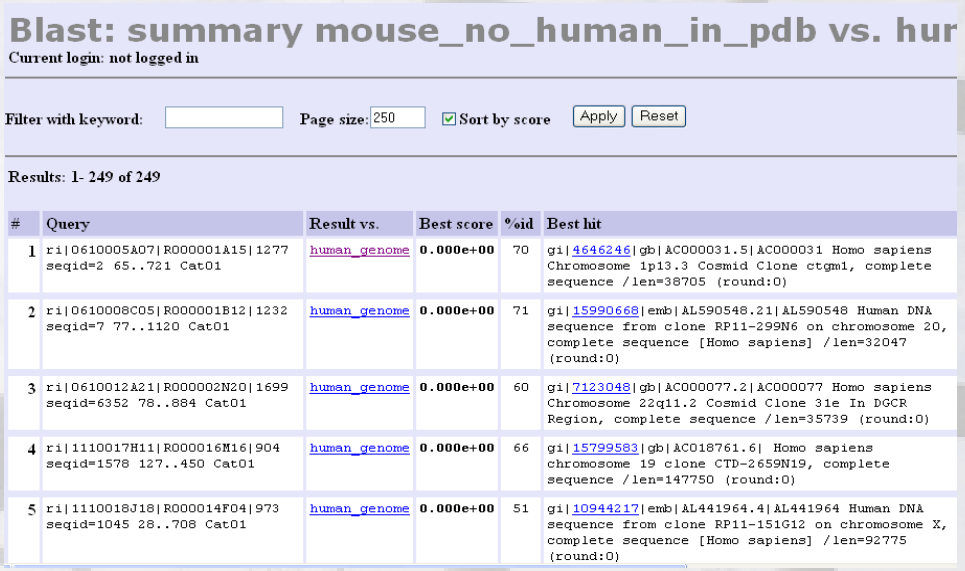
#	Query	Result vs.	Best score	%id	Best hit
1	ri|A530099J19|PX00144G09|2774 seqid=45616 311..1210 Cat01	pdb	-95.200	14	1kad_A mol:protein length:360 Ccr2B
2	ri|6820446M14|PX00650M23|2778 seqid=66146 526..1374 Cat01	pdb	-89.100	43	1aol_ mol:protein length:228 Gp70
3	ri|D330027J18|PX00192L12|3914 seqid=57248 520..1368 Cat01	pdb	-89.100	43	1aol_ mol:protein length:228 Gp70
4	ri|D430035L01|PX00194P21|4391 seqid=51143 493..1341 Cat01	pdb	-89.100	43	1aol_ mol:protein length:228 Gp70
5	ri|G431002I12|PH00003I01|8284 seqid=62999 2732..7797 Cat03 +5759	pdb	-87.200	17	1rvl_A mol:protein length:556 Reverse Transcriptase Non-Nucleoside Binding

[Reed et al, Gen Res, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

Novel human proteins?

- ~10,000 clones with predicted amino acid sequence >100 aa. have **no** homologs in NR
- ~250 have reliable hits in human genomic DNA (translated)



Blast: summary mouse_no_human_in_pdb vs. hur

Current login: not logged in

Filter with keyword: Page size: 250 Sort by score Apply Reset

Results: 1- 249 of 249

#	Query	Result vs.	Best score	%id	Best hit
1	ri|0610005A07|R000001A15|1277 seqid=2 65..721 Cat01	human_genome	0.000e+00	70	gi|4646246|gb|AC000031.5|AC000031 Homo sapiens Chromosome 1p13.3 Cosmid Clone ctgm1, complete sequence /len=38705 (round:0)
2	ri|0610008C05|R000001B12|1232 seqid=7 77..1120 Cat01	human_genome	0.000e+00	71	gi|15990668|emb|AL590548.21|AL590548 Human DNA sequence from clone RP11-299N6 on chromosome 20, complete sequence [Homo sapiens] /len=32047 (round:0)
3	ri|0610012A21|R000002N20|1699 seqid=6352 78..884 Cat01	human_genome	0.000e+00	60	gi|7123048|gb|AC000077.2|AC000077 Homo sapiens Chromosome 22q11.2 Cosmid Clone 31e In DGCR Region, complete sequence /len=35739 (round:0)
4	ri|1110017H11|R000016M16|904 seqid=1578 127..450 Cat01	human_genome	0.000e+00	66	gi|15799583|gb|AC018761.6| Homo sapiens chromosome 19 clone CTD-2659N19, complete sequence /len=147750 (round:0)
5	ri|1110018J18|R000014F04|973 seqid=1045 28..708 Cat01	human_genome	0.000e+00	51	gi|10944217|emb|AL441964.4|AL441964 Human DNA sequence from clone RP11-151G12 on chromosome X, complete sequence [Homo sapiens] /len=92775 (round:0)

[Reed et al, Gen Res, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

Summary of Annotations

- 219 mouse orthologs out of 227 human genes (96% coverage)

- Most of the difference due to PAAD and NACHT containing domain proteins (n = 23)
- 21 sequences of mouse from riken are absent in public databases.

- Overall, 29 additi
orthologs/paralogs r

- 0 novel mouse ge
signature domains, c

Table 1. Summary of Protein Domain Family Comp
for Humans and Mice

Protein domain	Human	Mouse	Only
Caspase	11	10	
CARD[a]	23	18	
DED[a]	11	11	
DD	33	33	
BIR	8	7[b]	
Bcl-2	24	27	
TNF-ligands	18	17	
TNF-Rs	29	27	
TIR	14	16	
TRAF/TEF	14	18	
PAAD	19	12	
NACHT	20	16	
REL	5	5	
IκB	8	7	

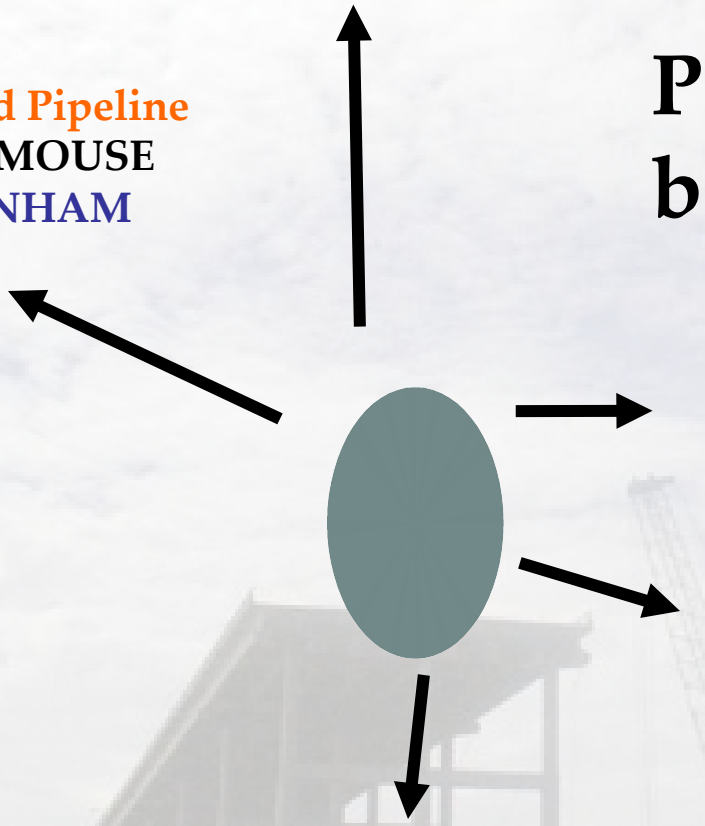
Comparative Analysis of Apoptosis and Inflammation Genes of Mice and Humans

John C. Reed,[1,3,4] Kutbuddin Doctor,[1] Ana Rojas,[1] Juan M. Zapata,[1] Christian Stehlik,[1] Loredana Fiorentino,[1] Jason Damiano,[1] Wilfried Roth,[1] Shu-ichi Matsuzawa,[1] Ruchi Newman,[1] Shinichi Takayama,[1] Hiroyuki Marusawa,[1] Famming Xu,[1] Guy Salvesen,[1] RIKEN GER Group[2] and GSL Members,[3,5] and Adam Godzik[1]

[1]The Burnham Institute, La Jolla, California 92037, USA; [2]Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan; [3]Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan

ANT2, ANT3, VDAC1, VDAC2, VDAC3, Beclin, BI-1, RT
Aven, Calcineurin-A, Calcineurin-B, Nip1, and Nip2 and

[Reed et al; Gen Res 2003]



(2) Use of homology modeling
Identifying binding sites:
PAAD/DAPIN/PYRIN

(1) Automated Pipeline
HUMAN vs. MOUSE
RIKEN-BURNHAM
initiative

Prediction of binding mode.

(3)Domain focused sequence analyses
New hypothesis for function:
DIDO family of proteins.

*Domain focused sequence analyses
Protein characterization
ACRATA, SPOC

(4) Predicting interaction interfaces
CCR5 dimerization.

GOAL:

Use bionformatics tools, homology modeling to predict binding Interfaces.

[Liu,T et al, Prot Sci, 2003]

cnio
Centro Nacional de Investigaciones Oncológicas

(2) USE OF HOMOLOGY MODELING

Pyrin, Aim (absent in melanoma), Asc (apoptosis associated speck-like protein containing a Caspase recrutiment domain) and a Death domain-like (DD)

WHERE IS THE PAAD DOMAIN?

Nacht family: PAN/NALPs/DEFCAP/PYCARD, CATERPILLER
(Tschopp et al, Nature, 2003)

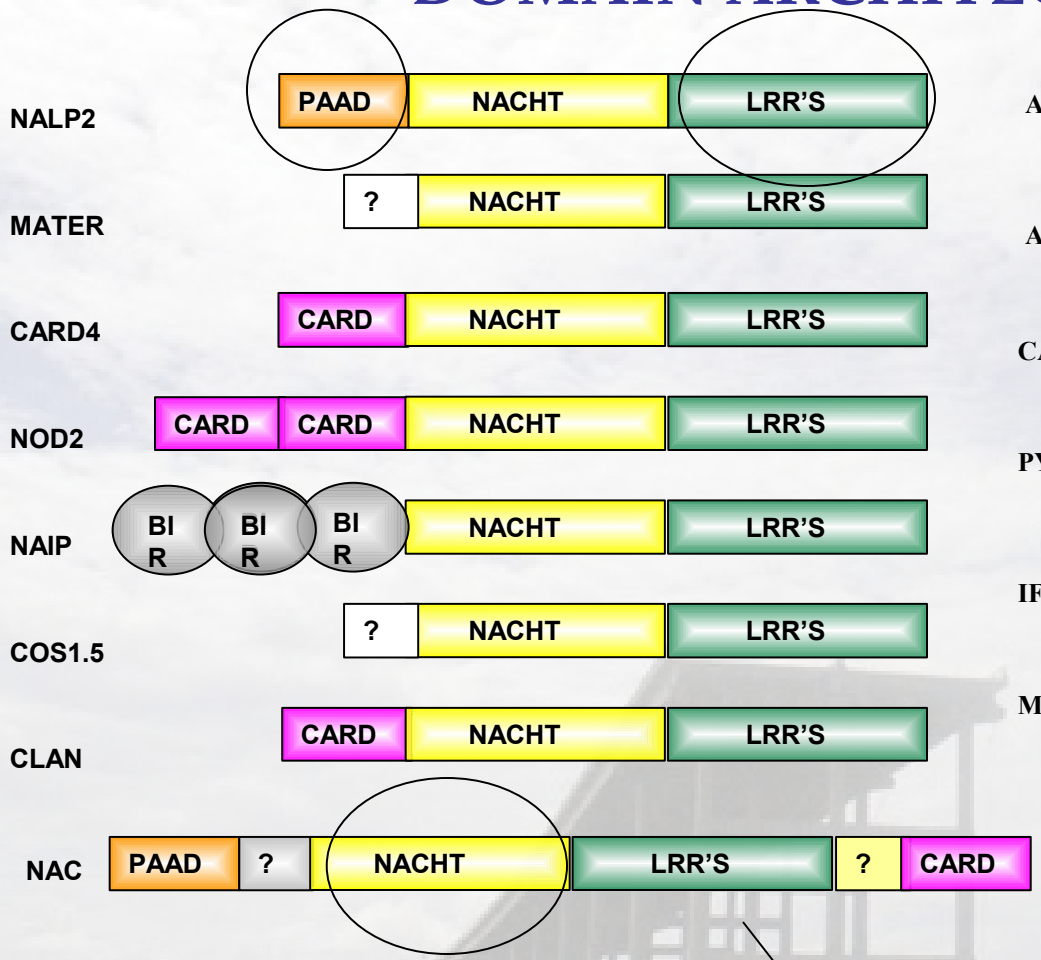
PAAD family: MEFV/PYRIN (Pawlowski, et.al., 2001 , others)

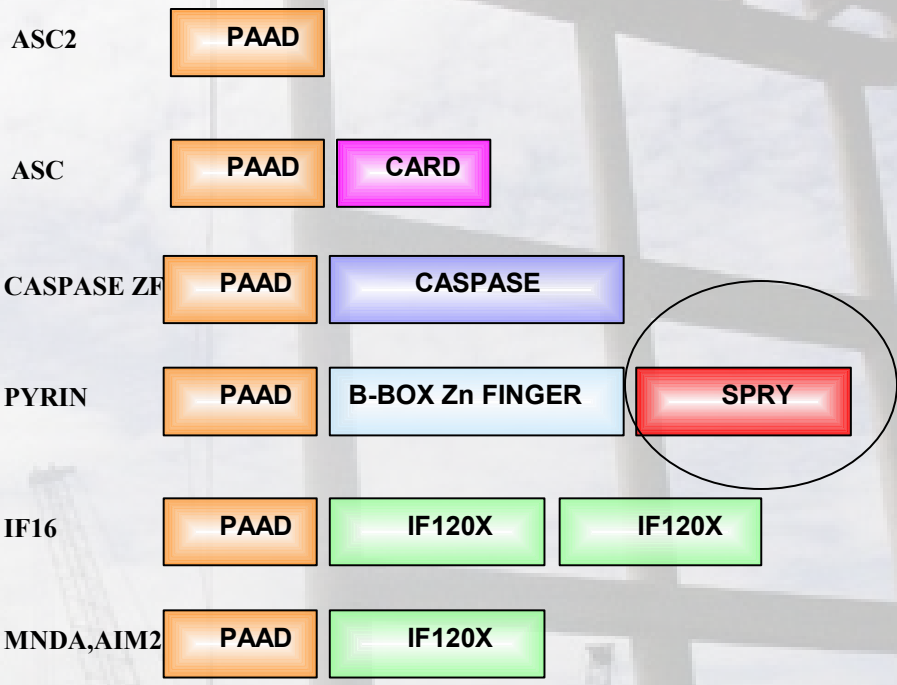
[Liu,T et al, Prot Sci, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

(2) USE OF HOMOLOGY MODELING

DOMAIN ARCHITECTURES





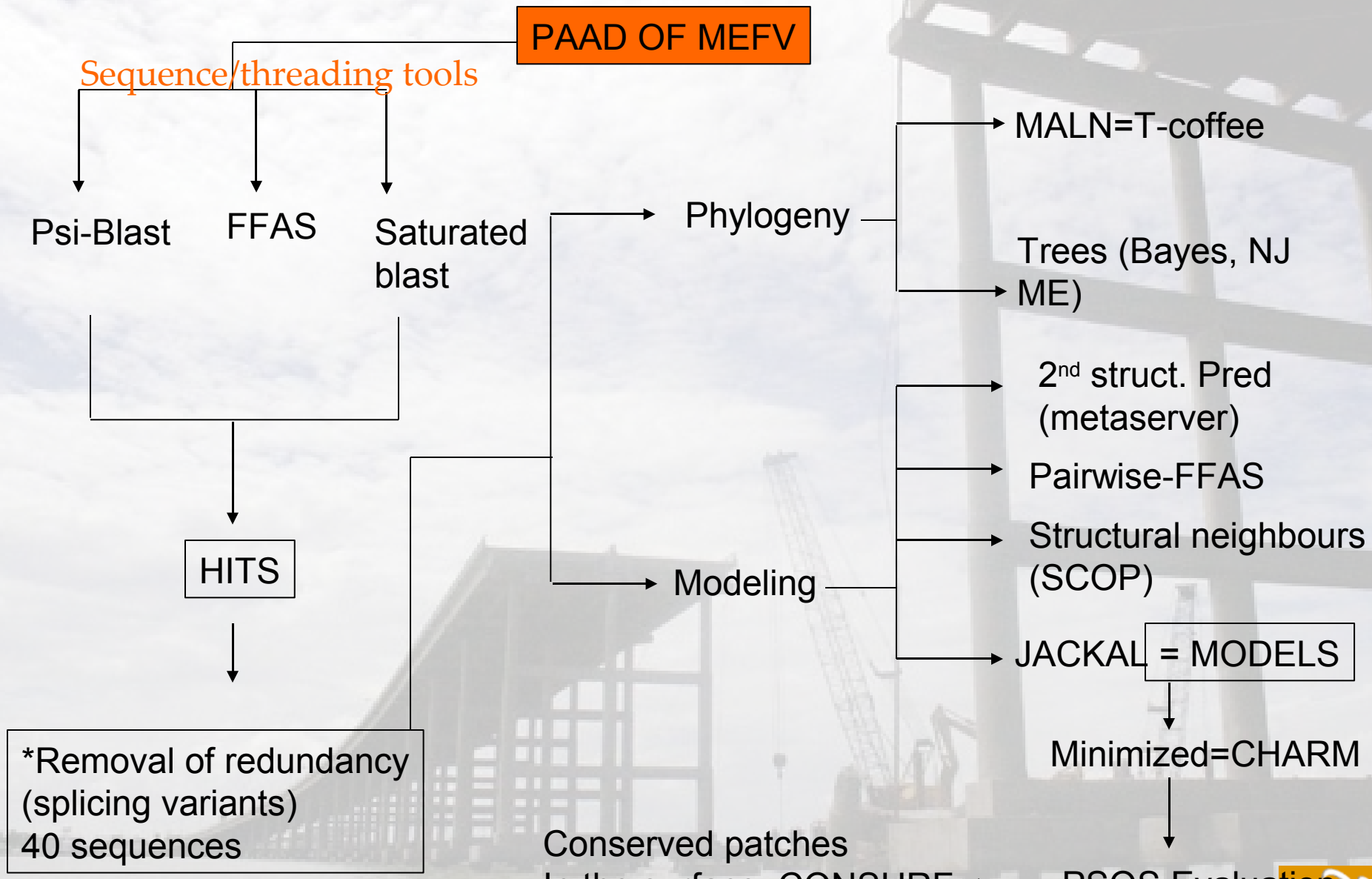
Sensors!

They connect different pathways!

[Liu,T et al, Prot Sci, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

(2) USE OF HOMOLOGY MODELING

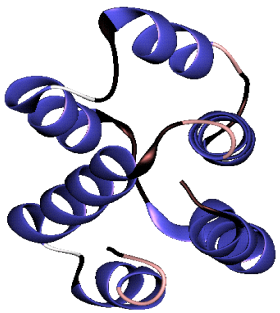


[Liu,T et al, Prot Sci, 2003]

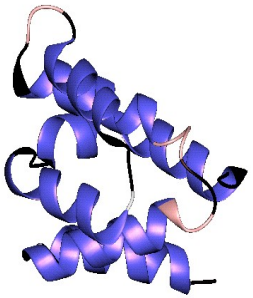
cnio Centro Nacional de Investigaciones Oncológicas

(2) USE OF HOMOLOGY MODELING

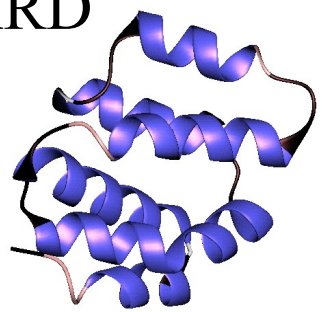
ANCESTORAL DOMAIN



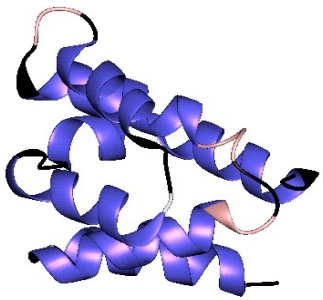
PAAD



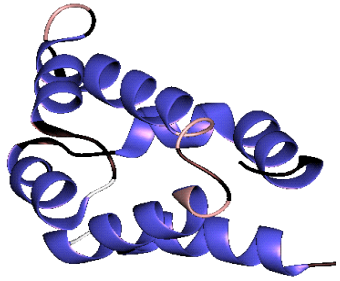
CARD



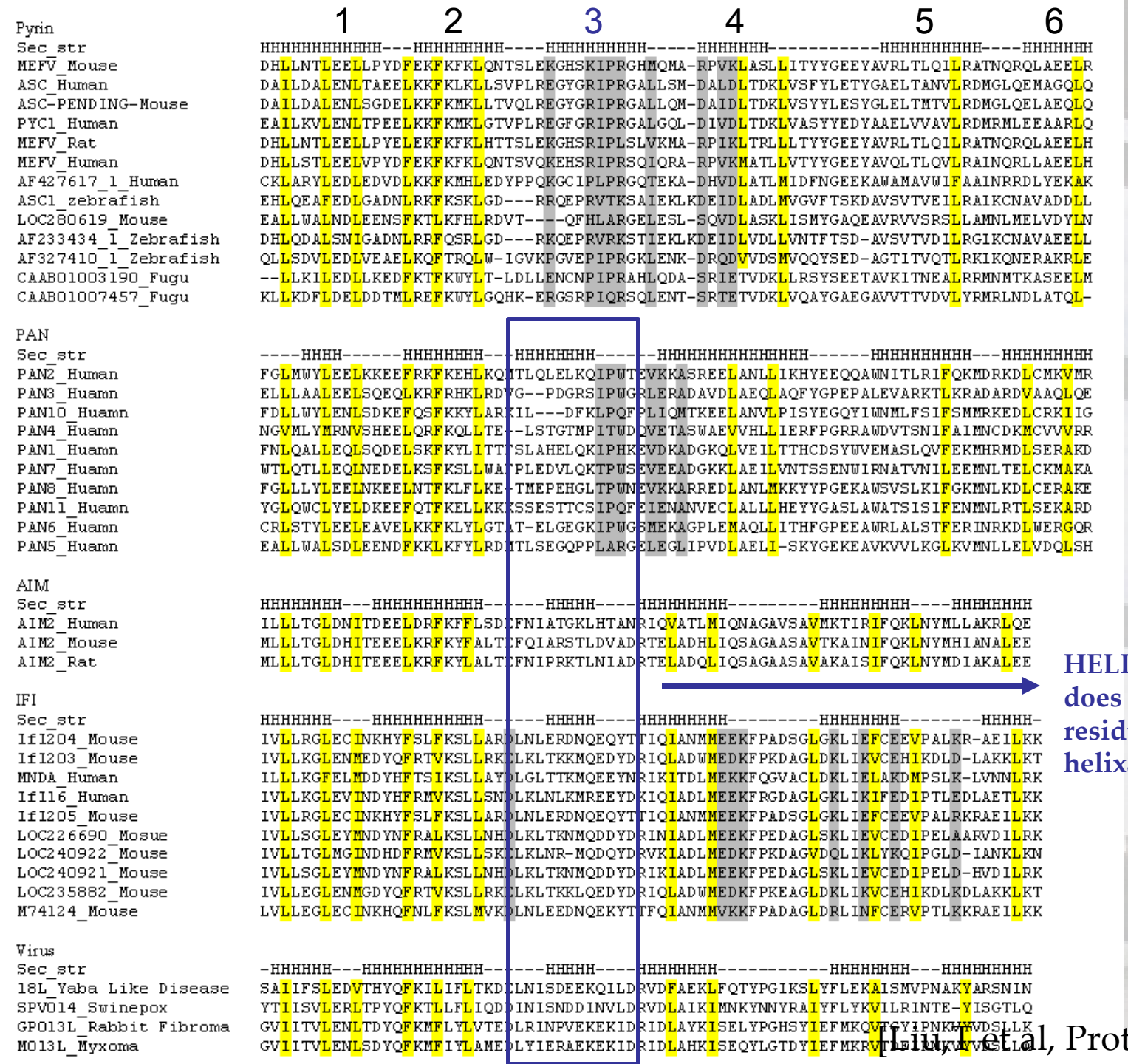
DD



DED



[Liu,T et al, Prot Sci, 2003]



Hydrophobic core (sol. acc. area <10% maximum solv. area)

Sec. Structure Prediction

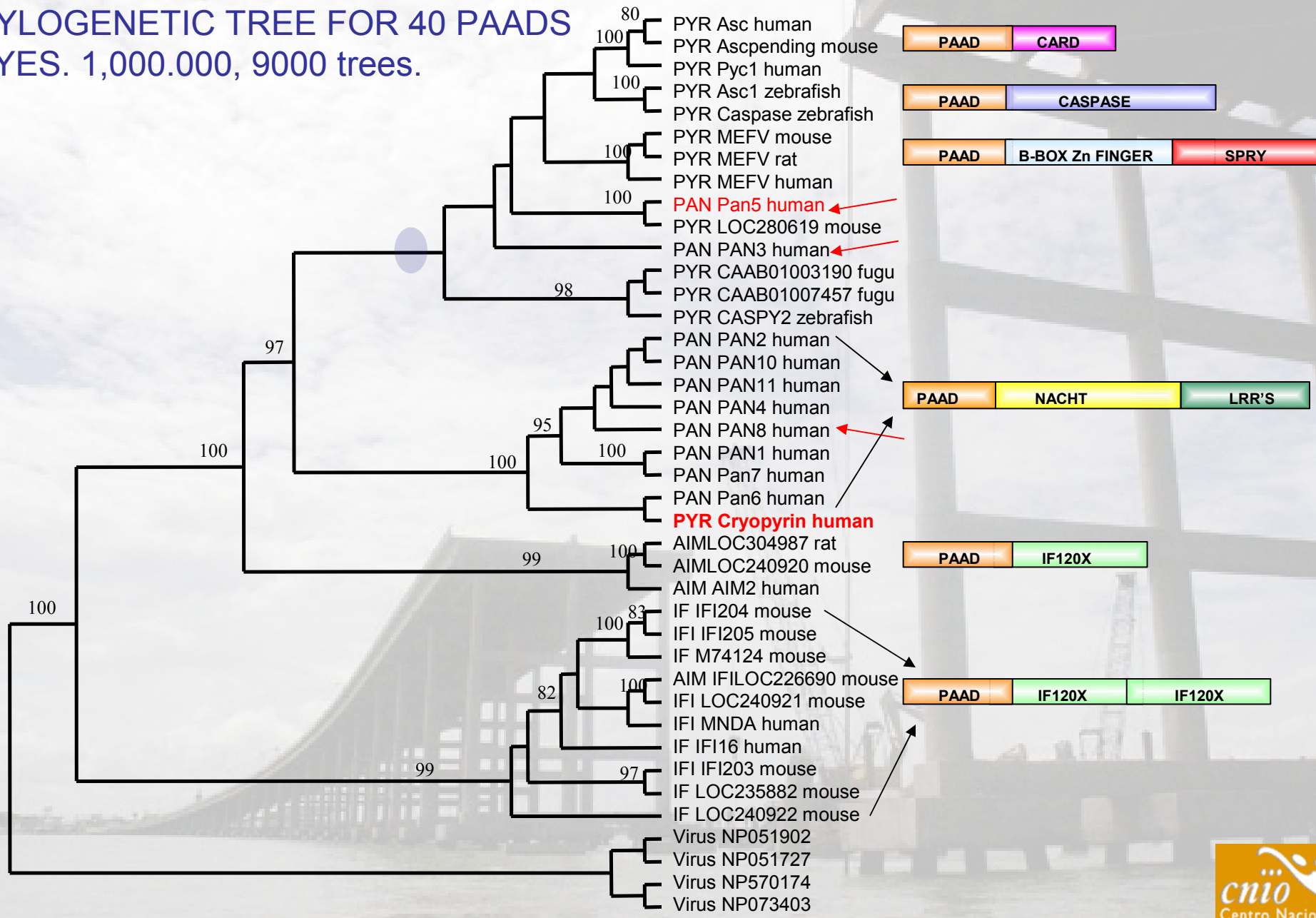
HELIX 3 does not have core residues. In DD, and others helix3 doesn't pack too well

[Liu, T. et al, Prot Sci, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

(2) USE OF HOMOLOGY MODELING

HYLOGENETIC TREE FOR 40 PAADS
AYES. 1,000.000, 9000 trees.

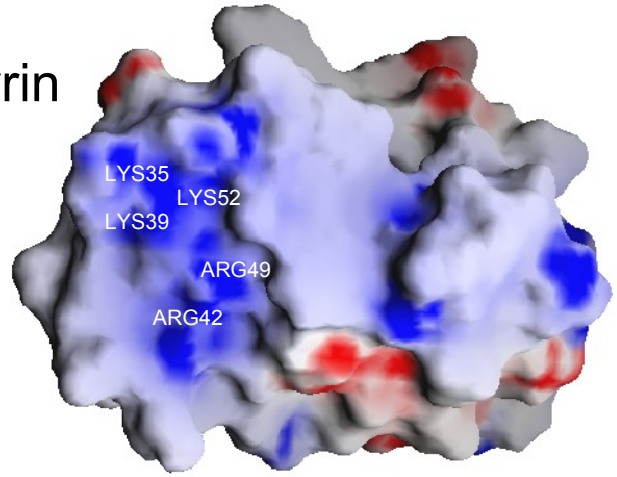


[Liu,T et al, Prot Sci, 2003]

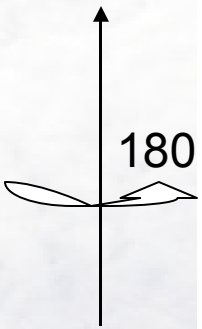
cnio Centro Nacional de Investigaciones Oncológicas

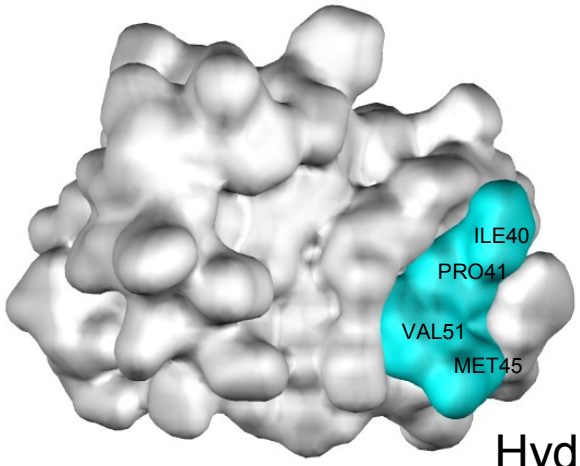
(2) USE OF HOMOLOGY MODELING

pyrin



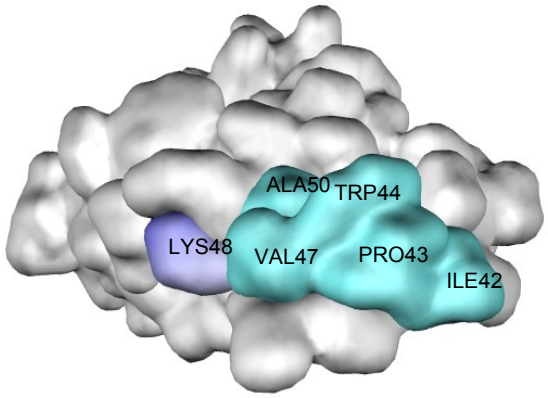
Charged patch





Hydrophobic patch

Pan2/NALP4

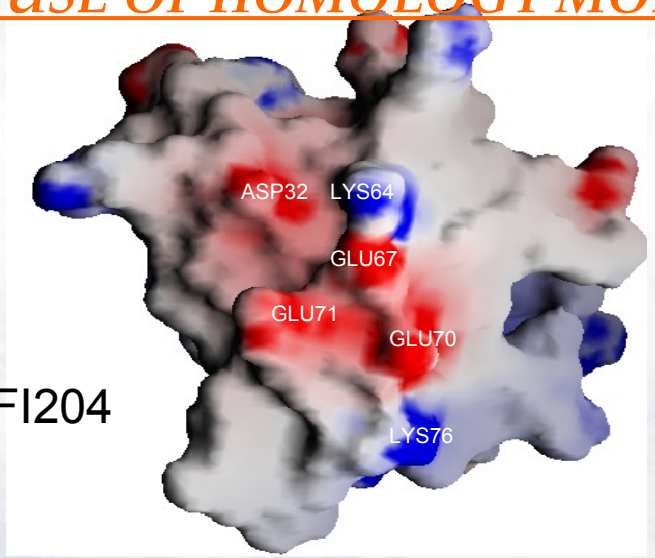


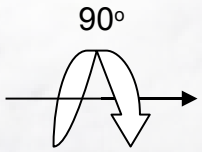
[Liu,T et al, Prot Sci, 2003]

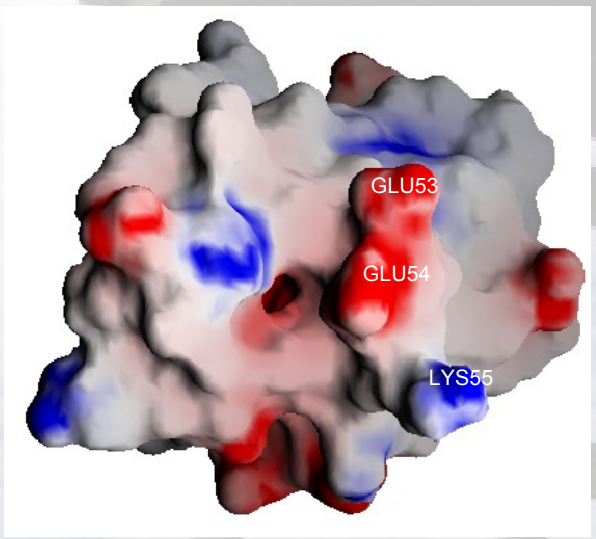
cnio Centro Nacional de Investigaciones Oncológicas

(2) USE OF HOMOLOGY MODELING

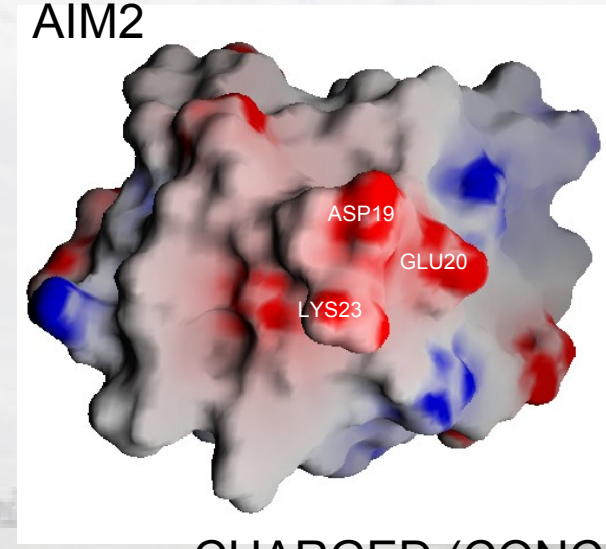
IFI204

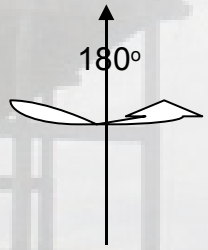


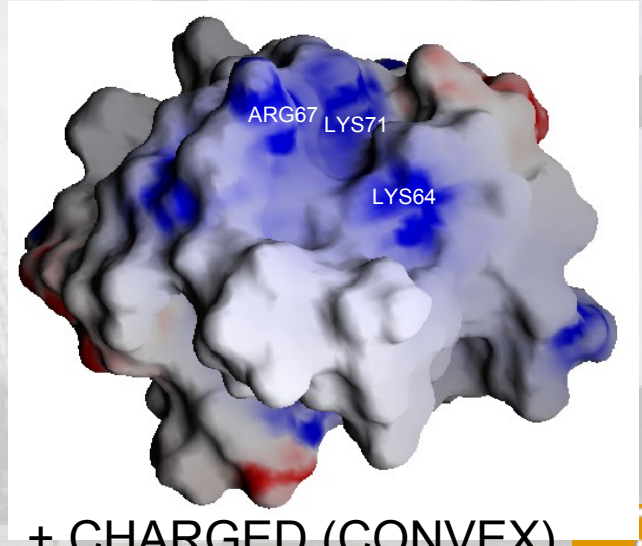




AIM2







- CHARGED (CONCAVE)

+ CHARGED (CONVEX)

[Liu,T et al, Prot Sci, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

SUMMARY

PAAD_DAPIN is a vertebrate-specific domain

PAAD from MEFV genes are the ancestral ones, sucesive duplications of the PAAD-PYR group yielded the mammalian pool

Viral PAAD's might mimic IFI/AIM family

Helix3 is disordered in DD/DED/CARD structures.
Needs partner interaction to fold properly .

The binding interface contains at least 10 hydrophobic residues. By analogy with CARD domains, electrostatic forces are also important.

id, character and conserved patches are as divergent within PAAD, as PAAD with DED/DD/CARD=> suggest specialization for not "cross-talking"

[Liu,T et al, Prot Sci, 2003]

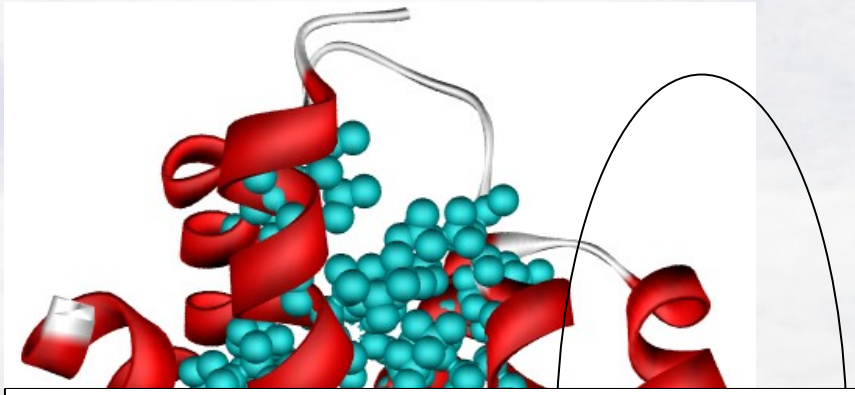
(3) USE OF HOMOLOGY MODELING

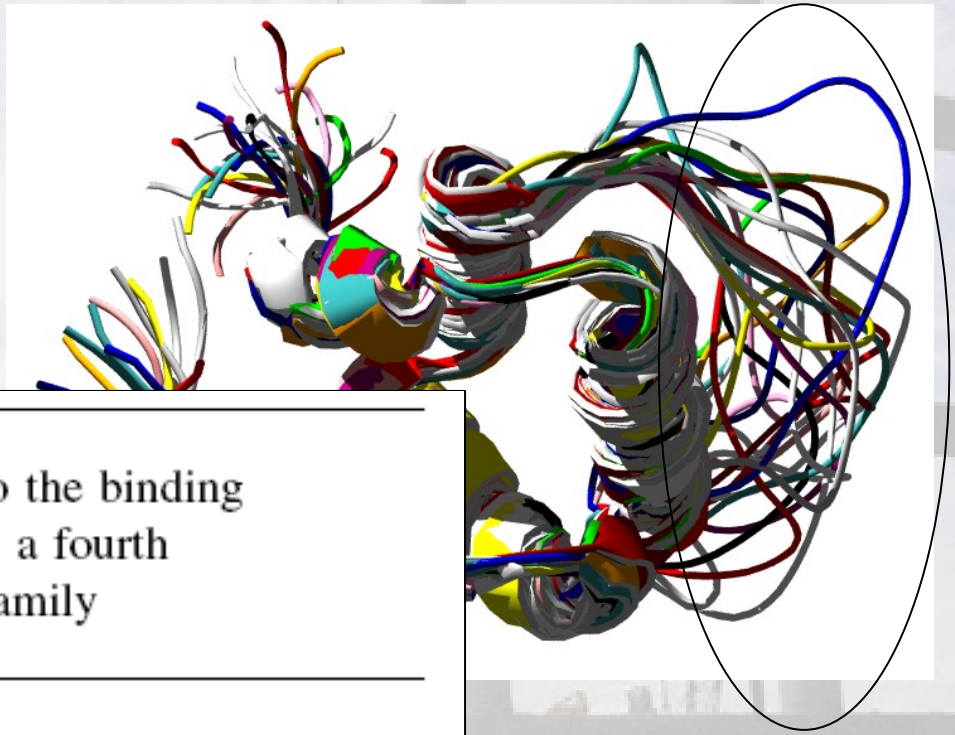
PAAD is a 6 alpha helical bundle

Helix 3 is disordered

Binding patches correctly predicted

Real structure 1PN5
Released October 2003





Homology modeling provides insights into the binding mode of the PAAD/DAPIN/pyrin domain, a fourth member of the CARD/DD/DED domain family

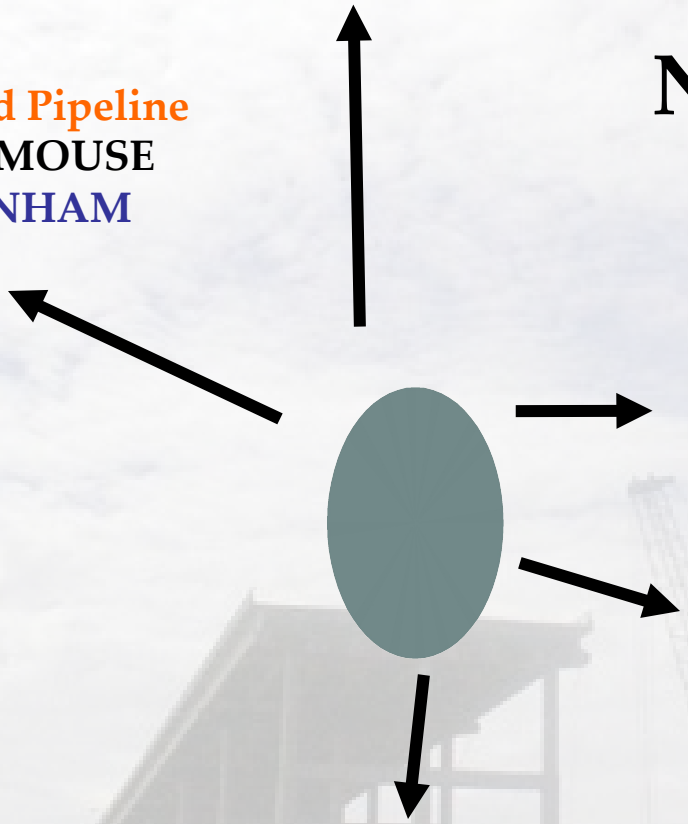
TONG LIU,[1] ANA ROJAS,[1] YUZHEN YE, AND ADAM GODZIK
The Burnham Institute, La Jolla, California 92037, USA

(RECEIVED March 5, 2003; FINAL REVISION May 23, 2003; ACCEPTED May 27, 2003)

[Liu,T et al, Prot Sci, 2003]

cnio Centro Nacional de Investigaciones Oncológicas

Novel Hypothesis.



(2) Use of homology modeling
Identifying binding sites:
PAAD/DAPIN/PYRIN

(1) Automated Pipeline
HUMAN vs. MOUSE
RIKEN-BURNHAM
initiative

Domain focused sequence analyses
New hypothesis for function:
(3) DIDO family of proteins.

Domain focused sequence analyses
Protein characterization
ACRATA, SPOC

Predicting interaction interfaces
CCR5 dimerization.

GOAL:

Use domain focused sequence searching to get insights into novel Function.

[Rojas, et al, FEBS J, 2005]

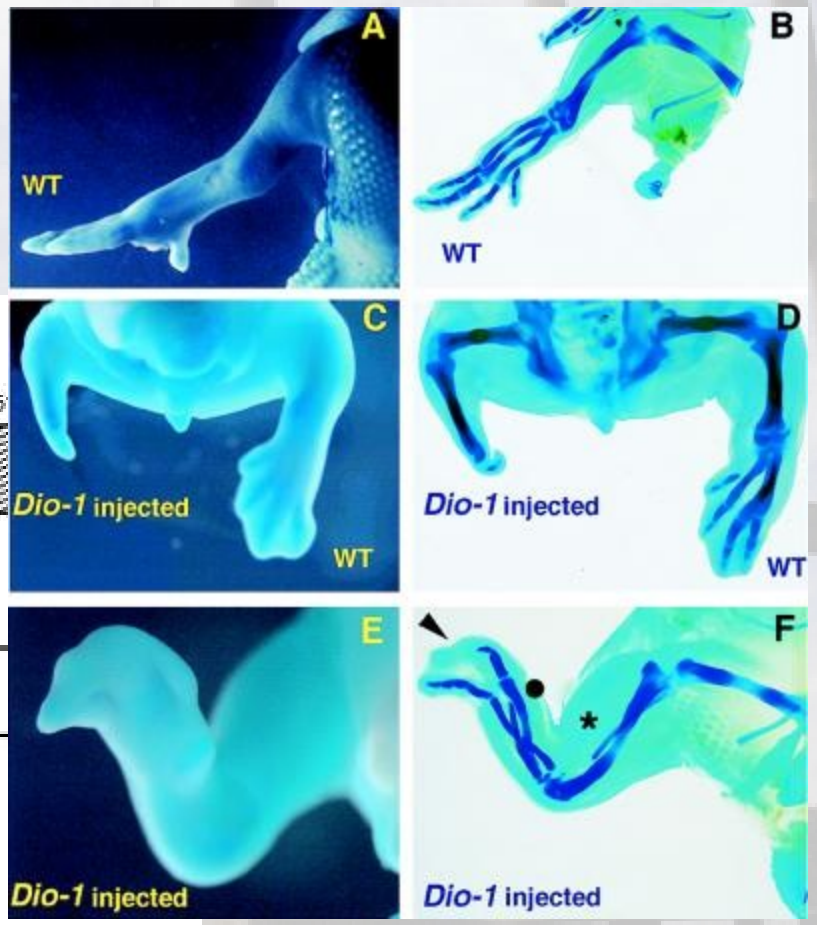
cnio
Centro Nacional de Investigaciones Oncológicas

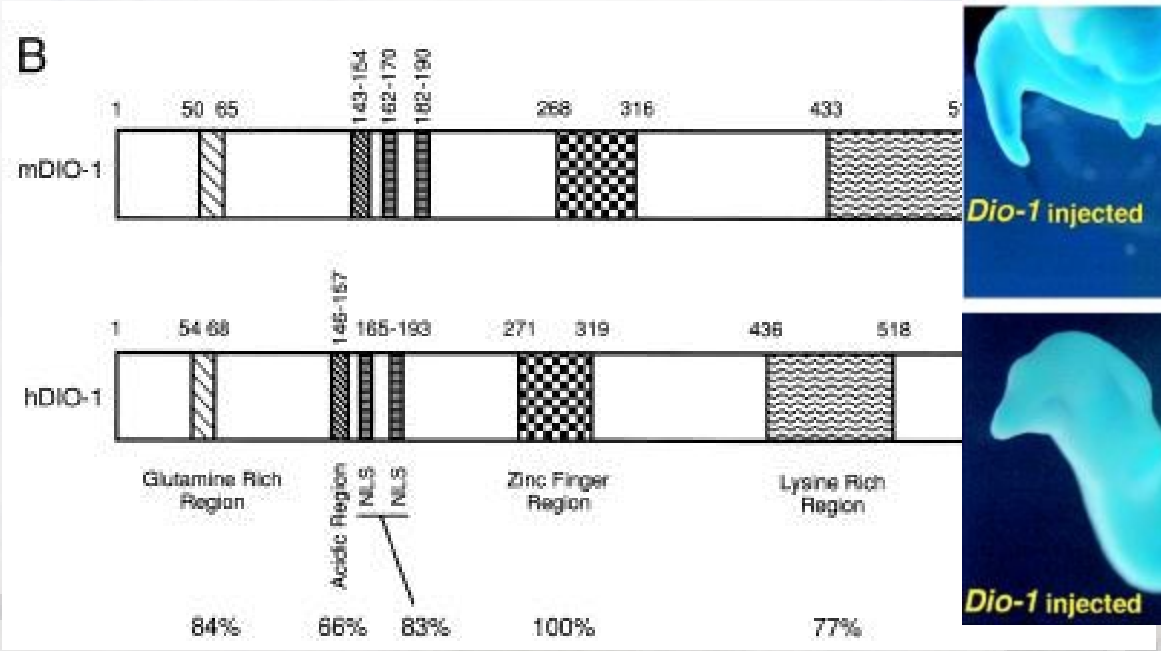
(3) DIDO-1 PROTEIN ANALYSES

DEATH ASSOCIATED TRANSCRIPTION GENE (DAFT), also known as DIO.

DISRUPTS LIMB DEVELOPMENT (Garcia-Domingo et al,. 1999)

- **Is Present in All Tissues and Its Levels Are Up-Regulated During Apoptosis.**
- **Alteration of Limb Development by *DIO-1* Overexpression**





Suggest that the gene is a putative transcription factor

[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

WHAT IS KNOWN: DEATH ASSOCIATED FACTOR GENE

INVOLVED IN APOPTOSIS (Garcia-Domingo et al,. 2003)

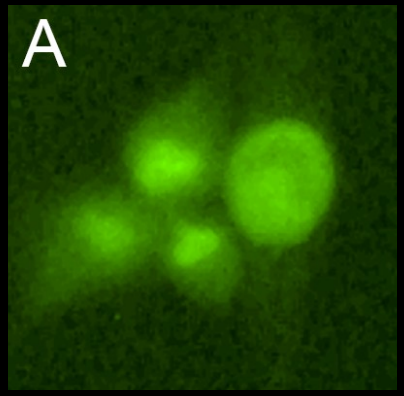
- **DIO-1 nuclear translocation following apoptotic stimulation requires the NLS.**
- **DIO-1 forms oligomers.**
- **DIO-1 is present in multiple forms with distinct subcellular localizations.**
- **DIO-1 overexpression upregulates procaspase levels, leading to increased caspase activity.**
- **DIO-1ΔNLS is a dominant negative mutant that protects cells from apoptosis.**

[Rojas, et al, FEBS J, 2005]

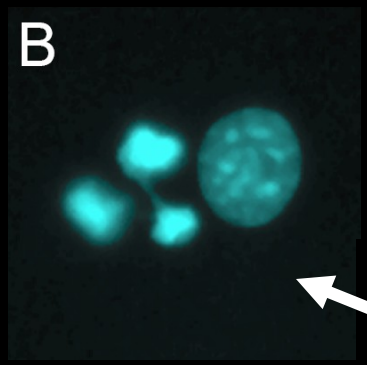
cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

ADDITIONAL EXPERIMENTAL DATA

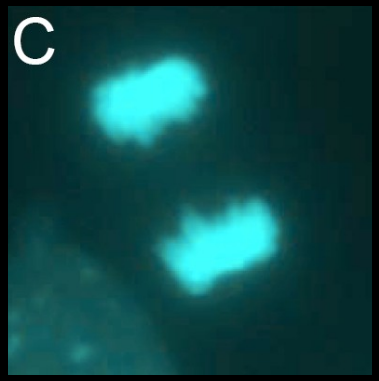


DIO-1 is present in mitotic chromosomes

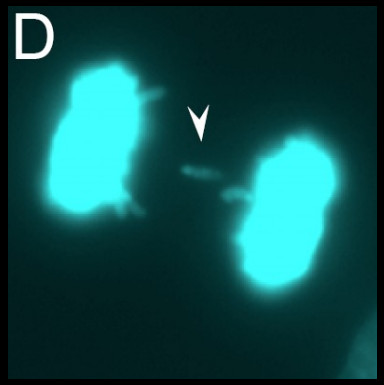


Mitosis on DIO overexpressed-cells

Asymmetric divisions!



Normal anaphase



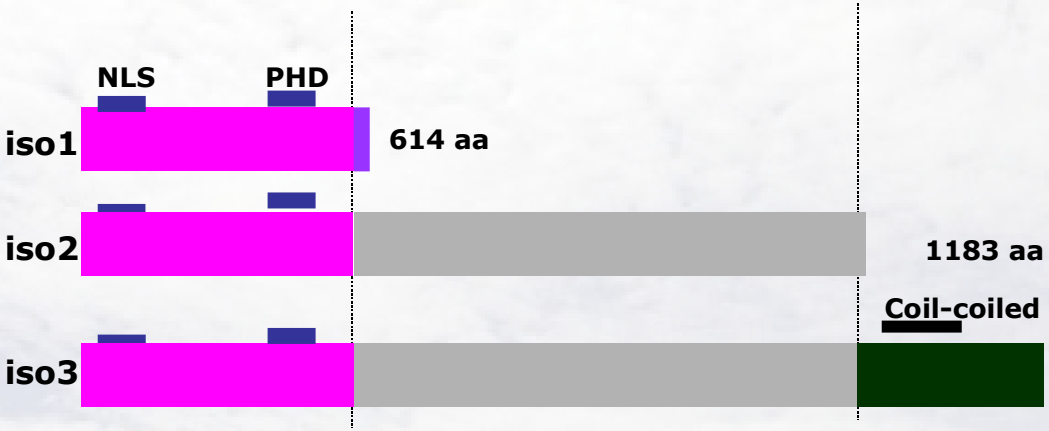
DIO-targeted cells show abnormal anaphases: lagging chromosomes

TARGETED MICE SHOW SEVERE SUB-FERTILITY!!

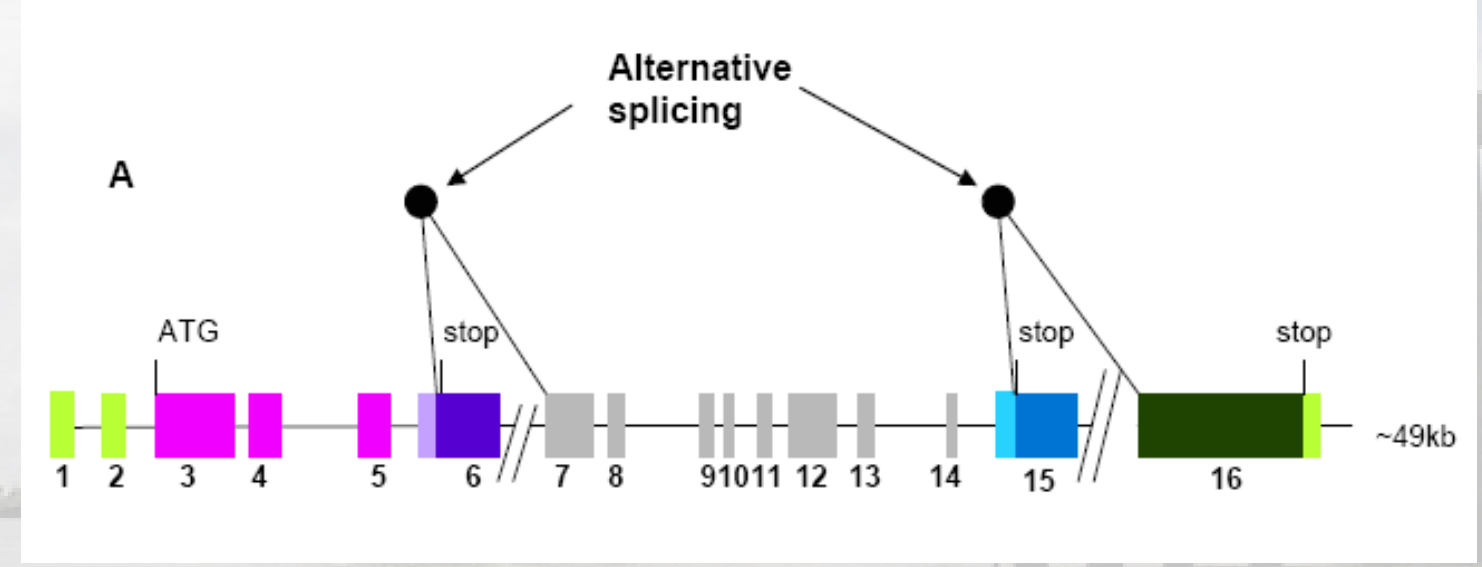
[Rojas, et al, FEBS J, 2005]

(3) DIDO-1 PROTEIN ANALYSES

The gene contains 3 splicing variants



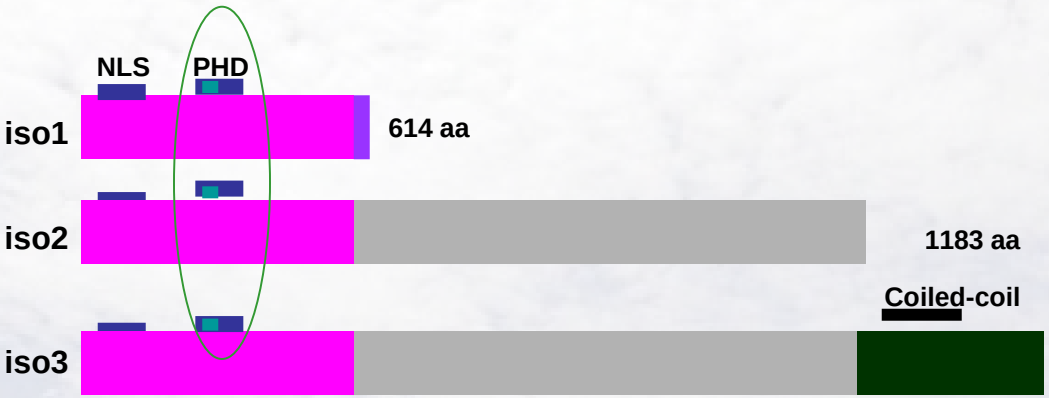
LONG PARTS OF THE PROTEIN REMAIN UNCOVERED!

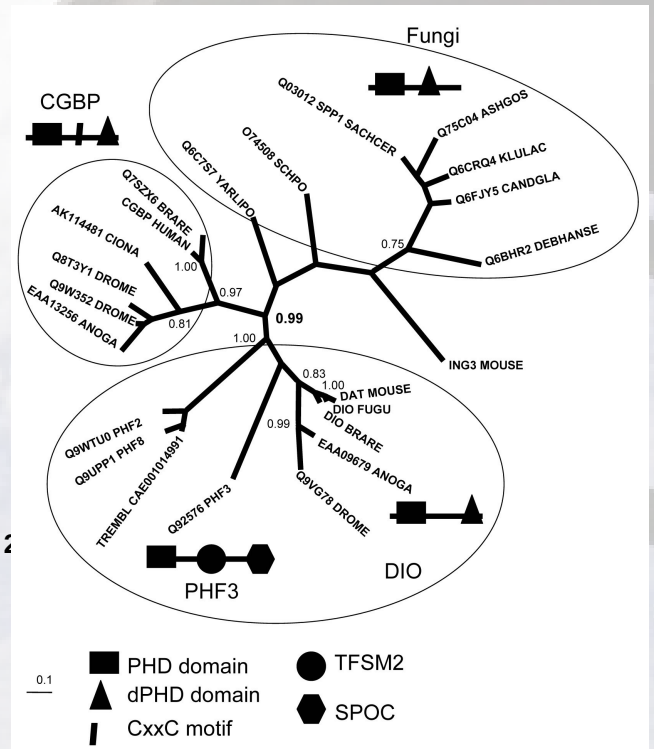


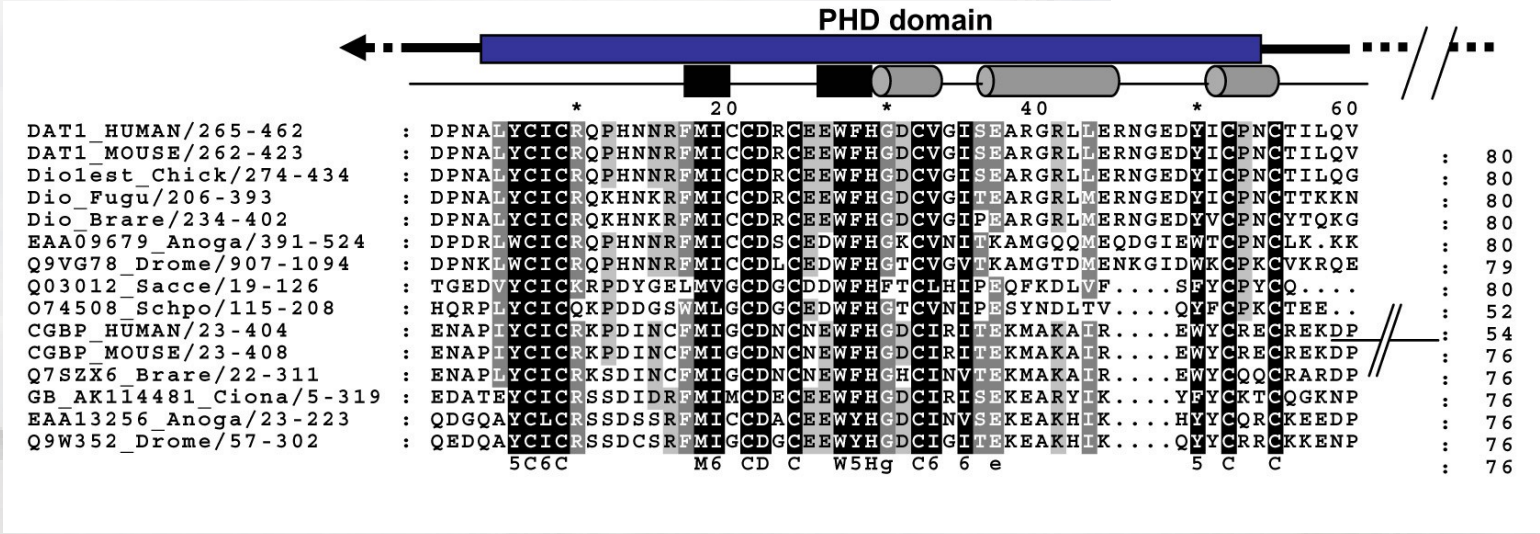
[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES





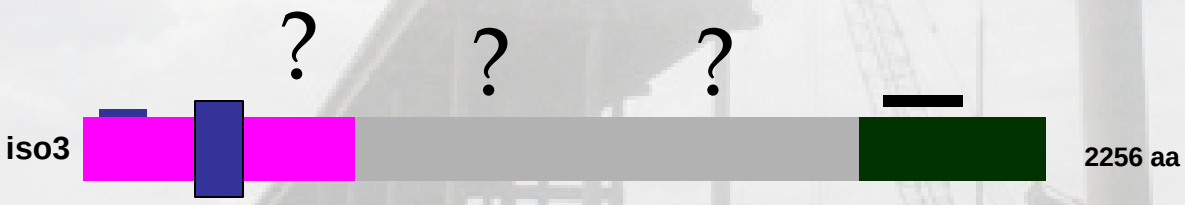


[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

Then…

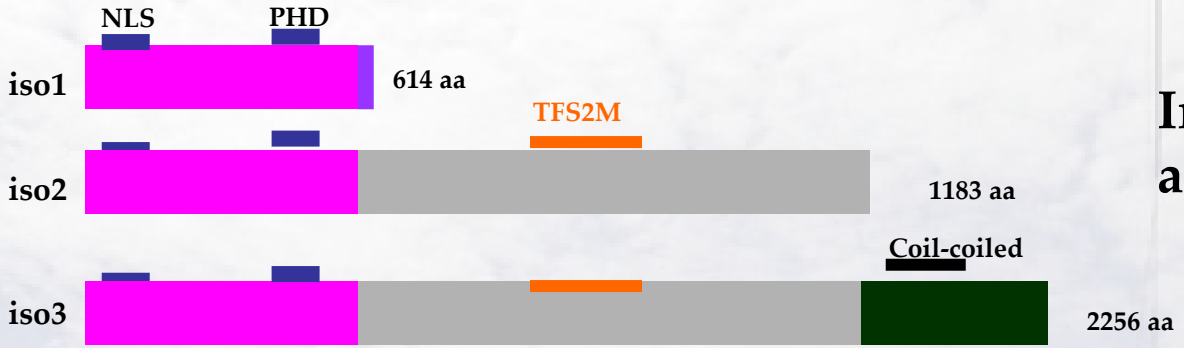
WHAT ELSE CAN BE FOUND IN THESE UNCOVERED REGIONS?



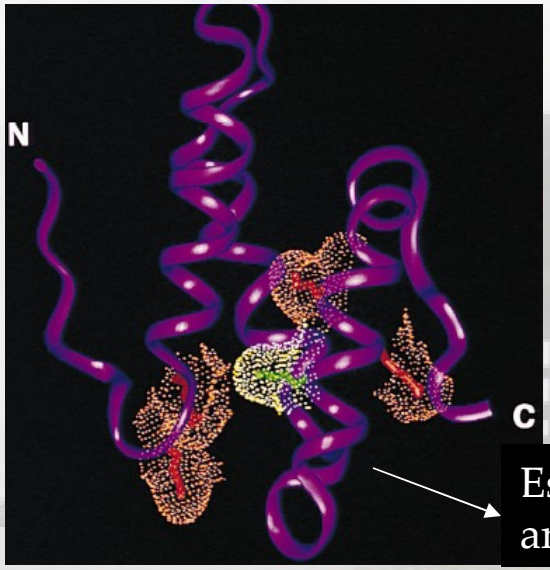
[Rojas, et al, FEBS J, 2005]

(3) DIDO-1 PROTEIN ANALYSES

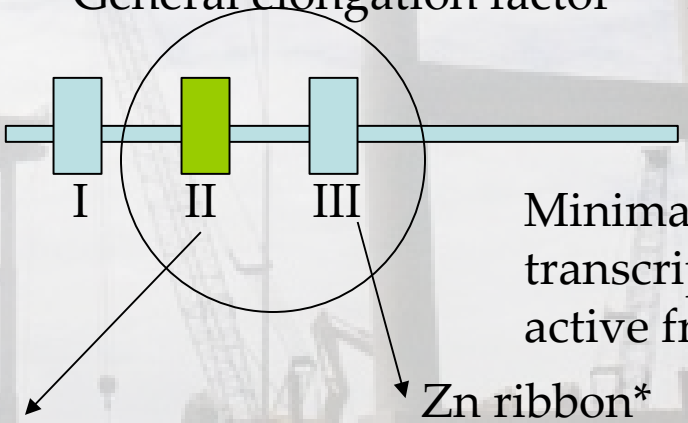
METHODS | AUTOMATIC SEARCHES



Involved in DNA Binding!
a regulatory domain



General elongation factor TFS2M



Minimal transcriptionally active fragment!

Essential for Pol II to read in pause sites and transcripts cleavage.

[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

METHODS: manual | Selecting regions

Query seq

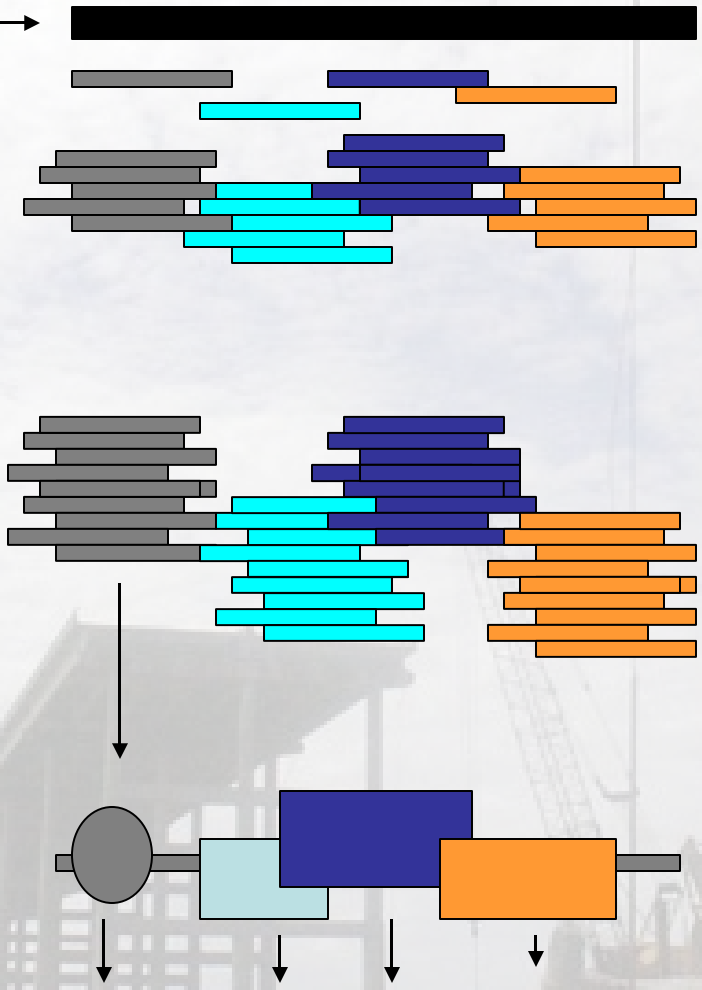
Blast to nr/uniprot90

Blast to EST's & unfinished genomes

TO ENRICH PROFILE!

Multiple alignment
T-COFFEE,
MUSCLE, etc

PROFILE BUILDING



HMMER/PSI-BLAST SEARCHES in Uniprot90

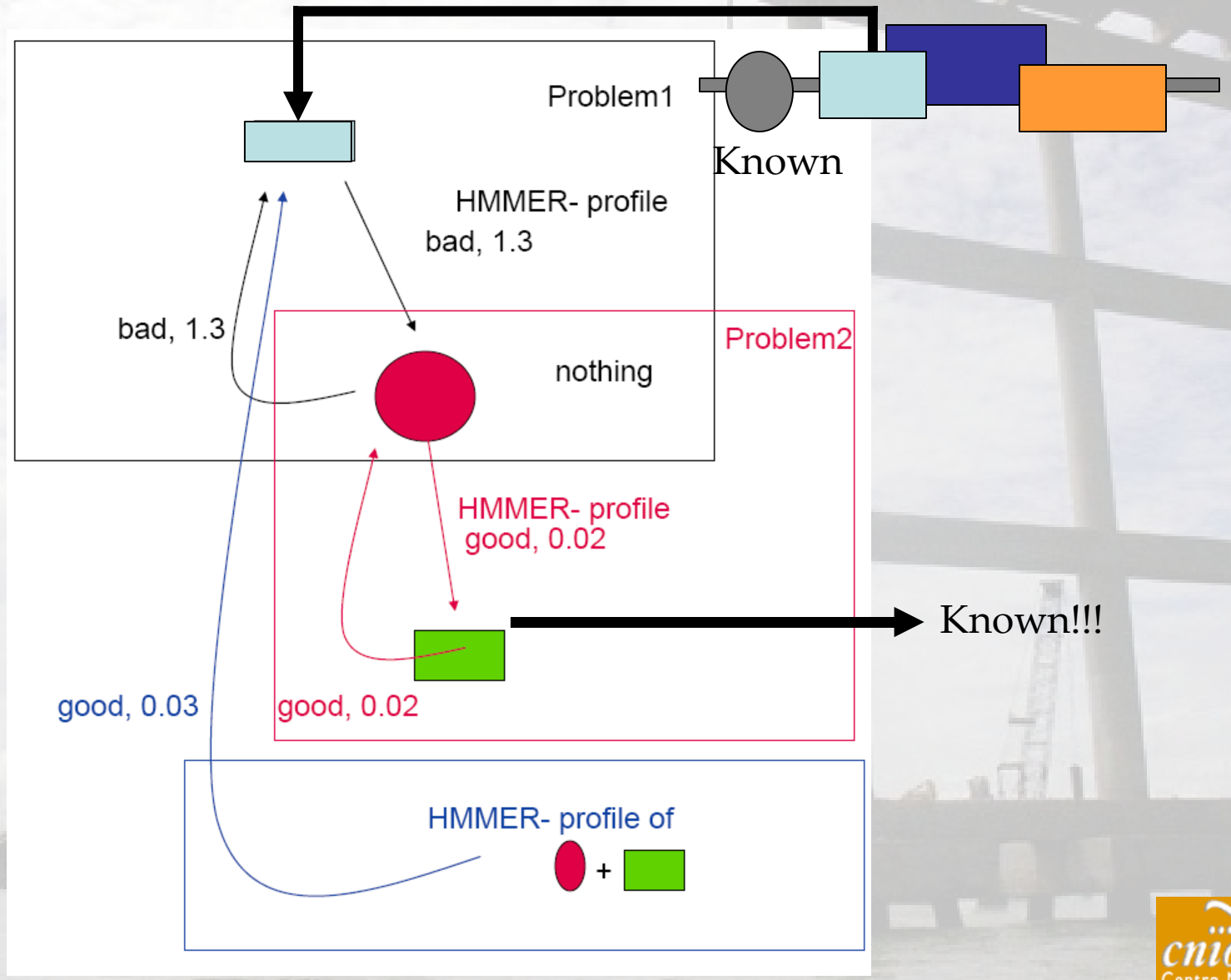
[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

METHODS: HMMER strategy

Intermediate profile searching!

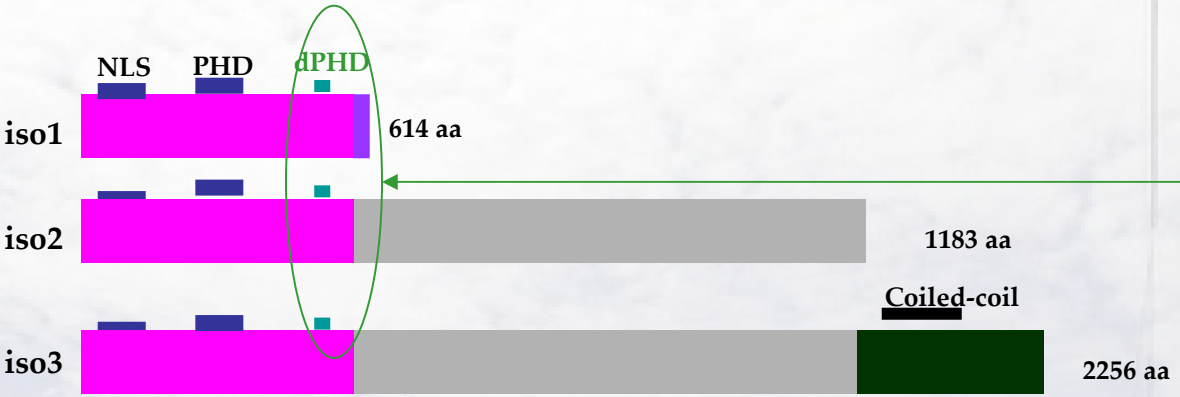


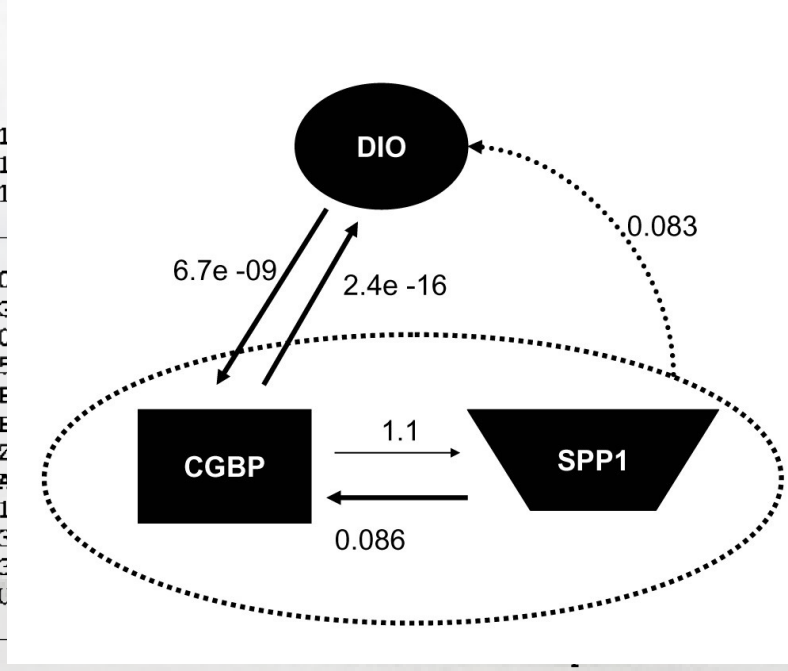
[Rojas, et al, FEBS J, 2005]

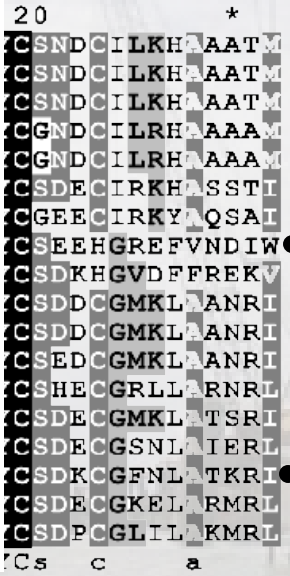
cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

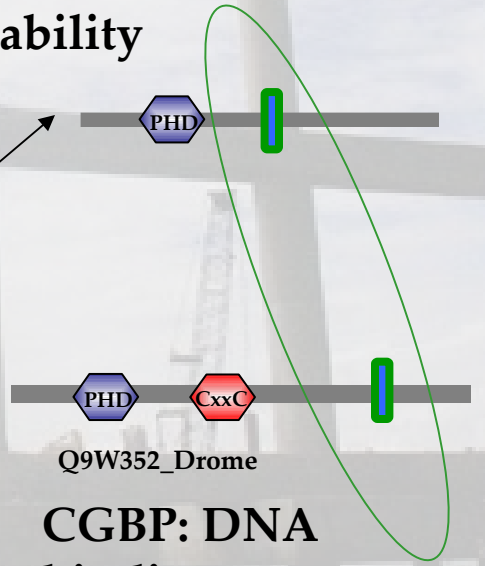
METHODS HMMER ANALYSES I







SPP1: Chromatin stability



CGBP: DNA binding!

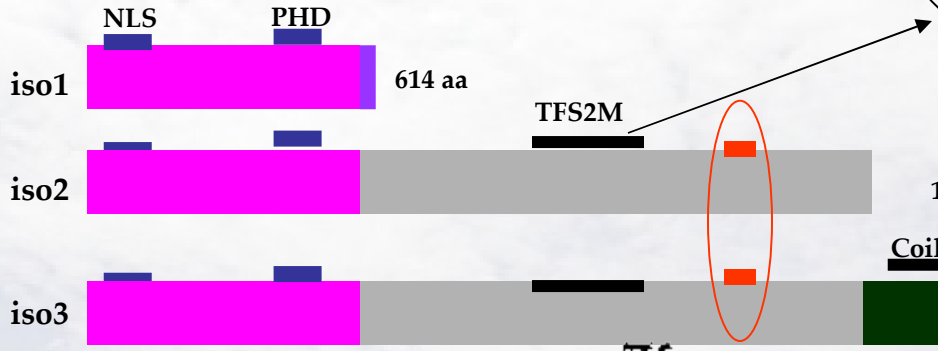
[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

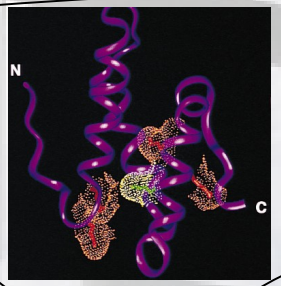
(3) DIDO-1 PROTEIN ANALYSES

METHODS

HMMER ANALYSES II

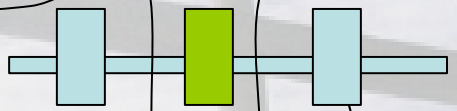


General elongation factor

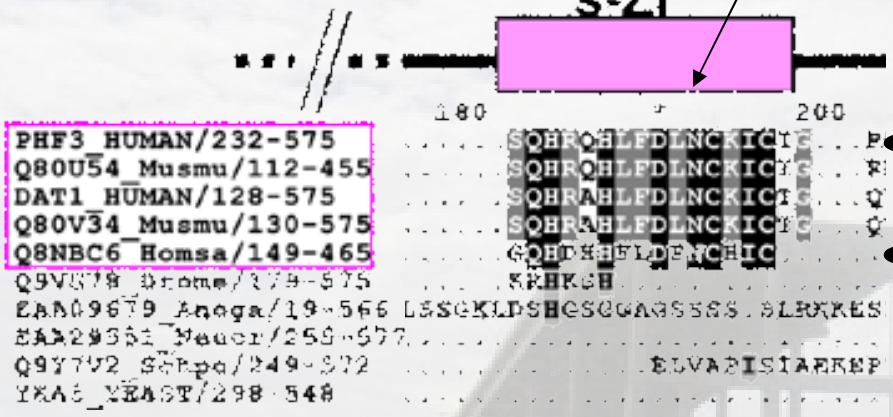


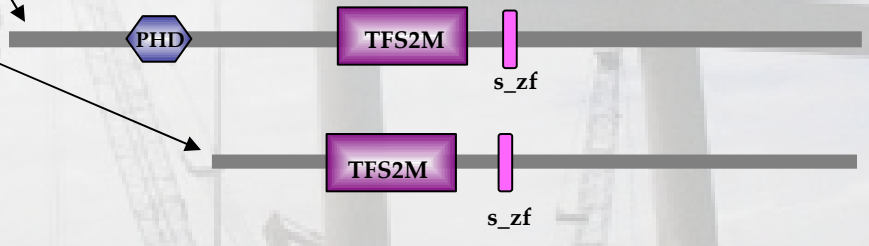
Involved in DNA Binding!
a regulatory domain

TFS2M
Minimal transcriptionally active fragment!



Zn ribbon*



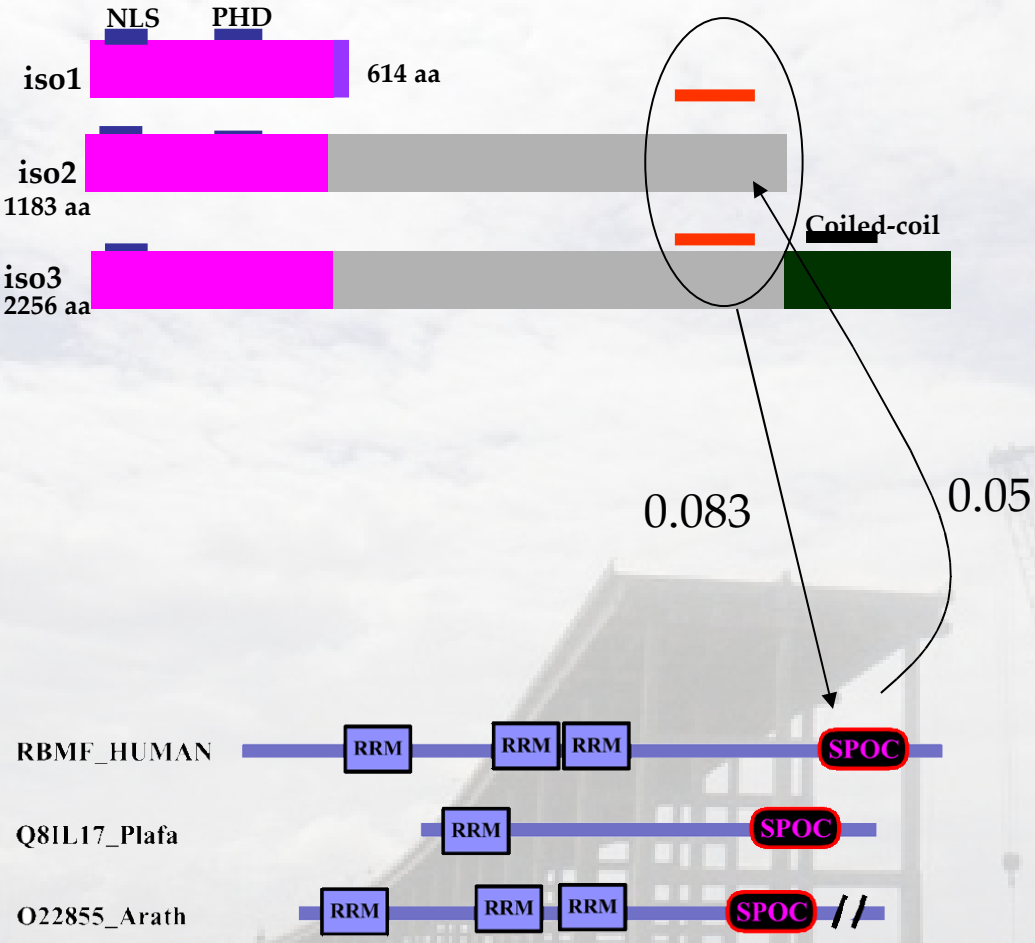


[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

METHODS HMMER ANALYSES III



SPOC: Protein-protein interaction

[Sanchez-Pulido, et al, BMC Bioinformatics 2004]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

METHODS

HMME

BMC Bioinformatics

BioMed Central

Research article

Open Access

SPOC: A widely distributed domain associated with cancer, apoptosis and transcription

Luis Sánchez-Pulido*[1], Ana M Rojas[1], Karel H van Wely[2], Carlos Martinez-A[2] and Alfonso Valencia[1]

Address: [1]Protein Design Group, Centro Nacional de Biotecnología (CNB-CSIC). Cantoblanco, E-28049 Madrid, Spain and [2]Department of Immunology and Oncology, Centro Nacional de Biotecnología (CNB-CSIC). Cantoblanco, E-28049 Madrid, Spain

Email: Luis Sánchez-Pulido* - sanchez@cnb.uam.es; Ana M Rojas - arojas@cnb.uam.es; Karel H van Wely - kvanwely@cnb.uam.es; Carlos Martinez-A - cmartineza@cnb.uam.es; Alfonso Valencia - valencia@cnb.uam.es

* Corresponding author

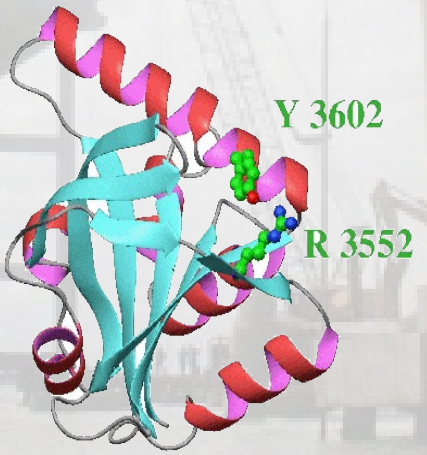
iso2

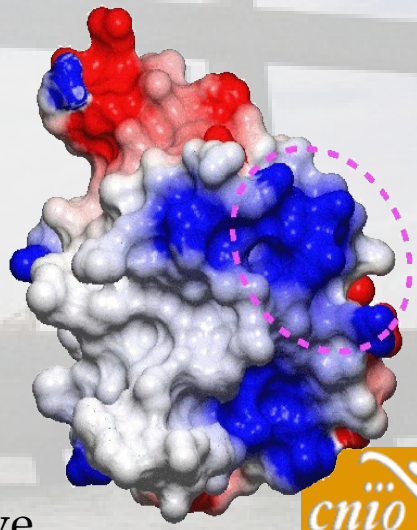
Model

SPOC: Protein-protein interaction

SPOC

RBMF_HUMAN



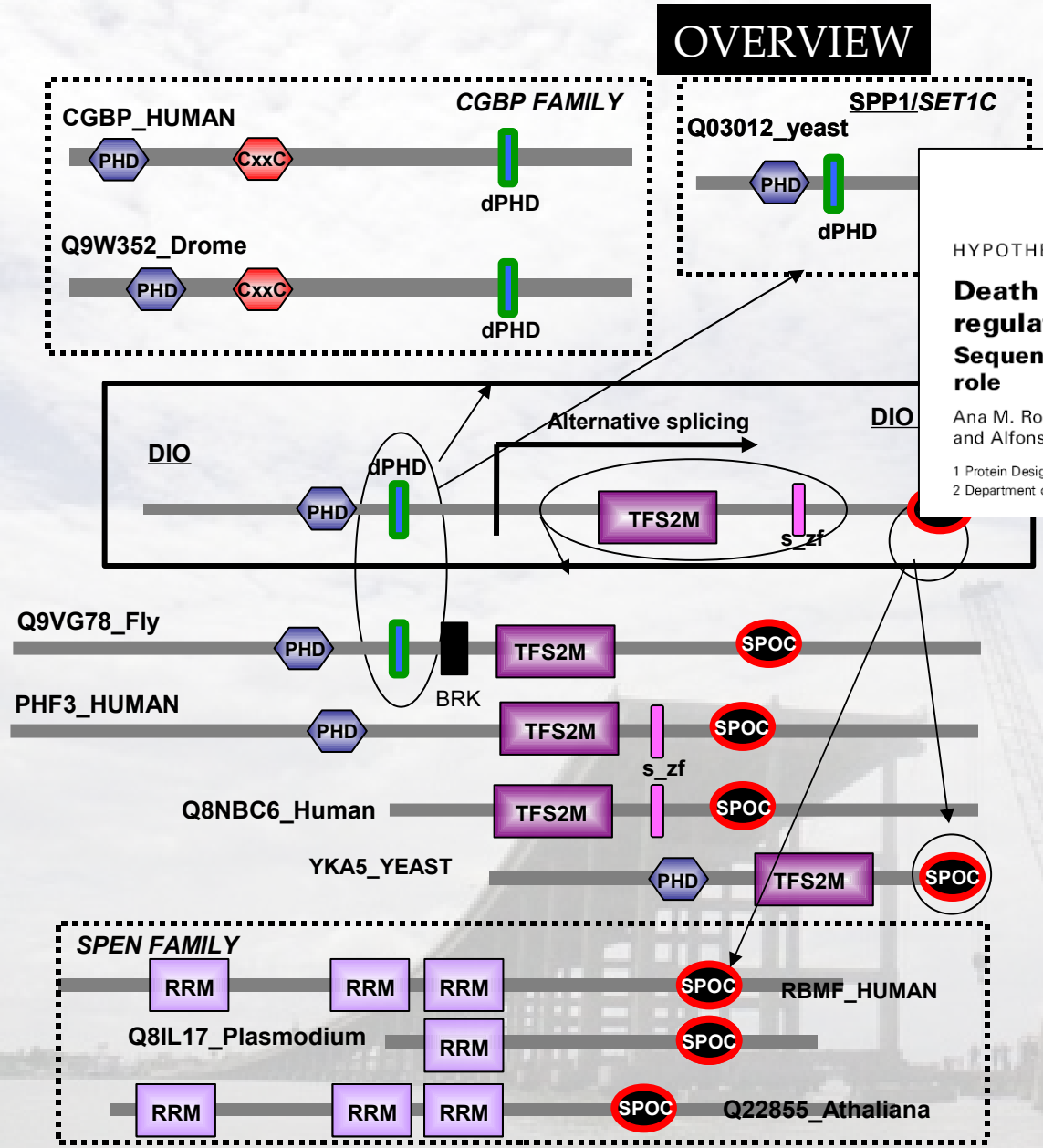


Comparative modeling

[Sanchez-Pulido, et al, BMC Bioinformatics 2004]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES



HYPOTHESIS

Death inducer obliterator protein 1 in the context of DN regulation

Sequence analyses of distant homologues point to a novel functio role

Ana M. Rojas[1,2], Luis Sanchez-Pulido[1], Agnes Fütterer[2], Karel H. M. van Wely[2], Carlos Martine and Alfonso Valencia[1]

1 Protein Design Group, CNB/CSIC, Madrid, Spain
2 Department of Immunology and Oncology, CNB/CSIC, Madrid, Spain

DIO : apoptosis

chromatin stability

SPEN: transcriptional repression

[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(3) DIDO-1 PROTEIN ANALYSES

PERSPECTIVES

Hypothesis: DIO's main role involves chromatin stability/recombination rather than to apoptosis or cancer.

Pitfalls (many): Role of isoforms is not well established in meiosis.
This protein can be located everywhere (nucleous-citopl)…

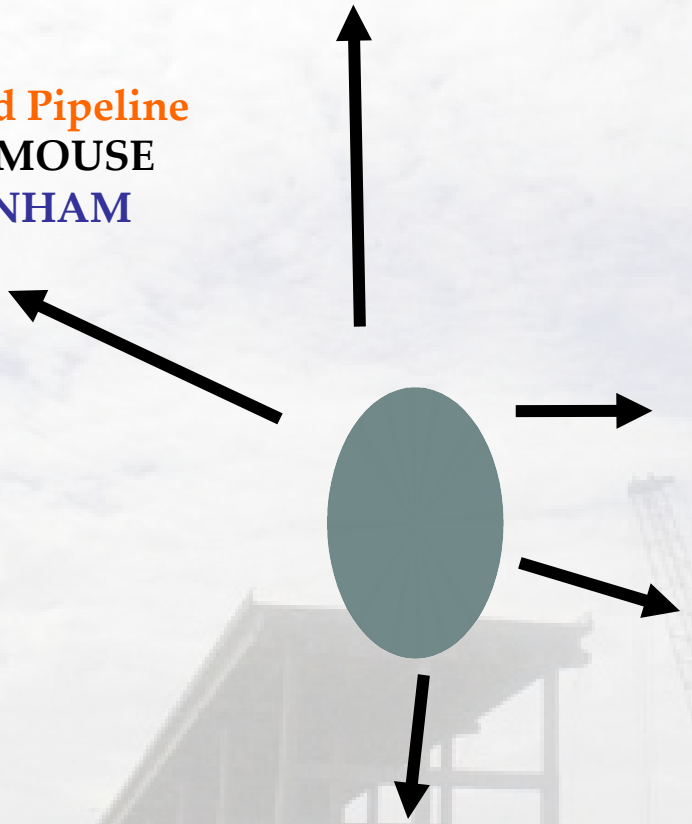
QUESTIONS: How is the exact mechanism?
Why this protein is so unusually rich on cys?
Might be weird-metal regulated, i.e.: Molib.?

ON-GOING: 3D structure of dPHD and SPOC domains.

[Rojas, et al, FEBS J, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

CCR5 Interaction.



(2) Use of homology modeling
Identifying binding sites:
PAAD/DAPIN/PYRIN

(1) Automated Pipeline
HUMAN vs. MOUSE
RIKEN-BURNHAM initiative

(3)Domain focused sequence analyses
New hypothesis for function:
DIDO family of proteins.

*Domain focused sequence analyses
Protein characterization
ACRATA, SPOC

(4) Predicting interaction interfaces
CCR5 dimerization.

GOAL:

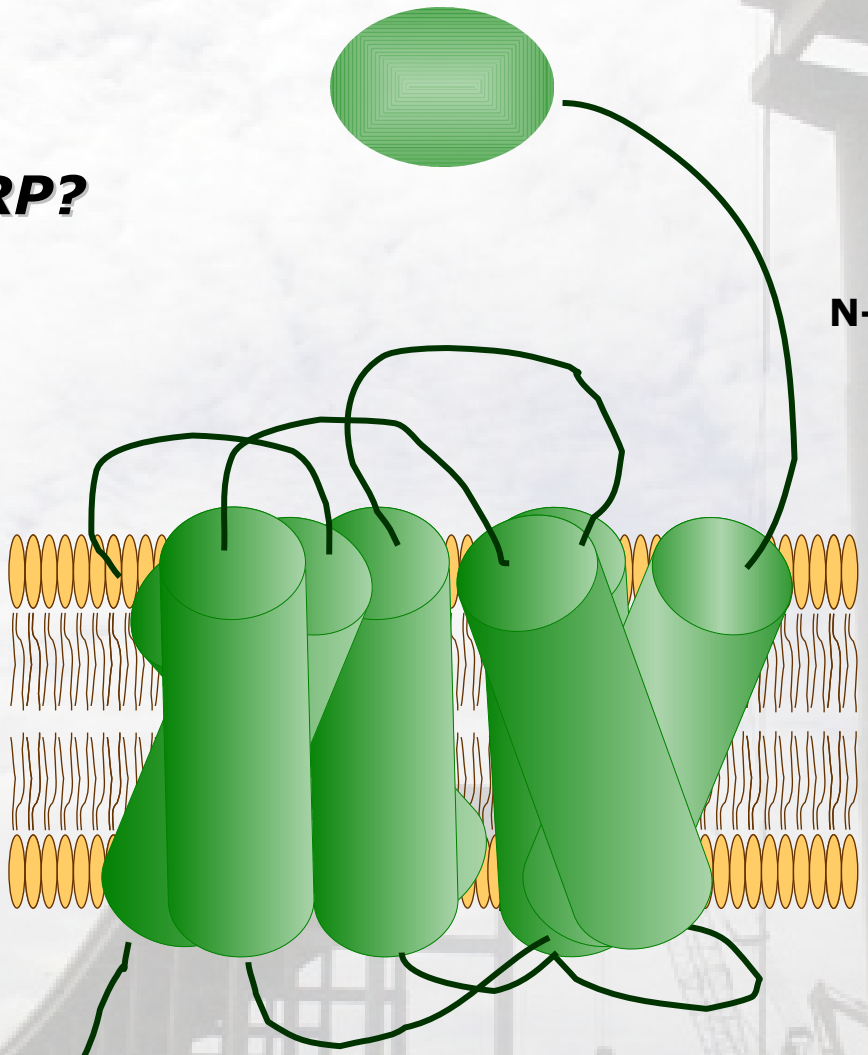
Use bionformatics tools to predict residues involved in binding.

[de Juan et al, Bioinformatics, 2005]

cnio Centro Nacional de Investigaciones Oncológicas

(4) CCR5: dimerization

What are GCRP?



cnio
Centro Nacional de Investigaciones Oncológicas

(4) CCR5

Why are important?

- They are extremely diverse and transduce very different messages!

(photons, odors, nucleotides, peptides, lipids….)

- They are involved in:

Inflammation, pain response, etc…= Pharmacological targets!

MEPNETTM (Membrane Protein Network): ~ 100 GPCR Xtals!!
Consortium with 40 pharma co's.

cnio
Centro Nacional de Investigaciones Oncológicas

GPCRDB home page - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.gpcr.org/7tm/ Search

Mail Home Radio Netscape Search Bookmarks The Linux Docu...

GPCRDB home page

Last sequence update: 8-Mar-05 ; Swiss-Prot rel. 47.0, TrEMBL rel. 29

Where are the GCRP?

GPCRDB: Information system for G protein-coupled receptors (GPCRs)

March 2005 release (9.0) - Spring release

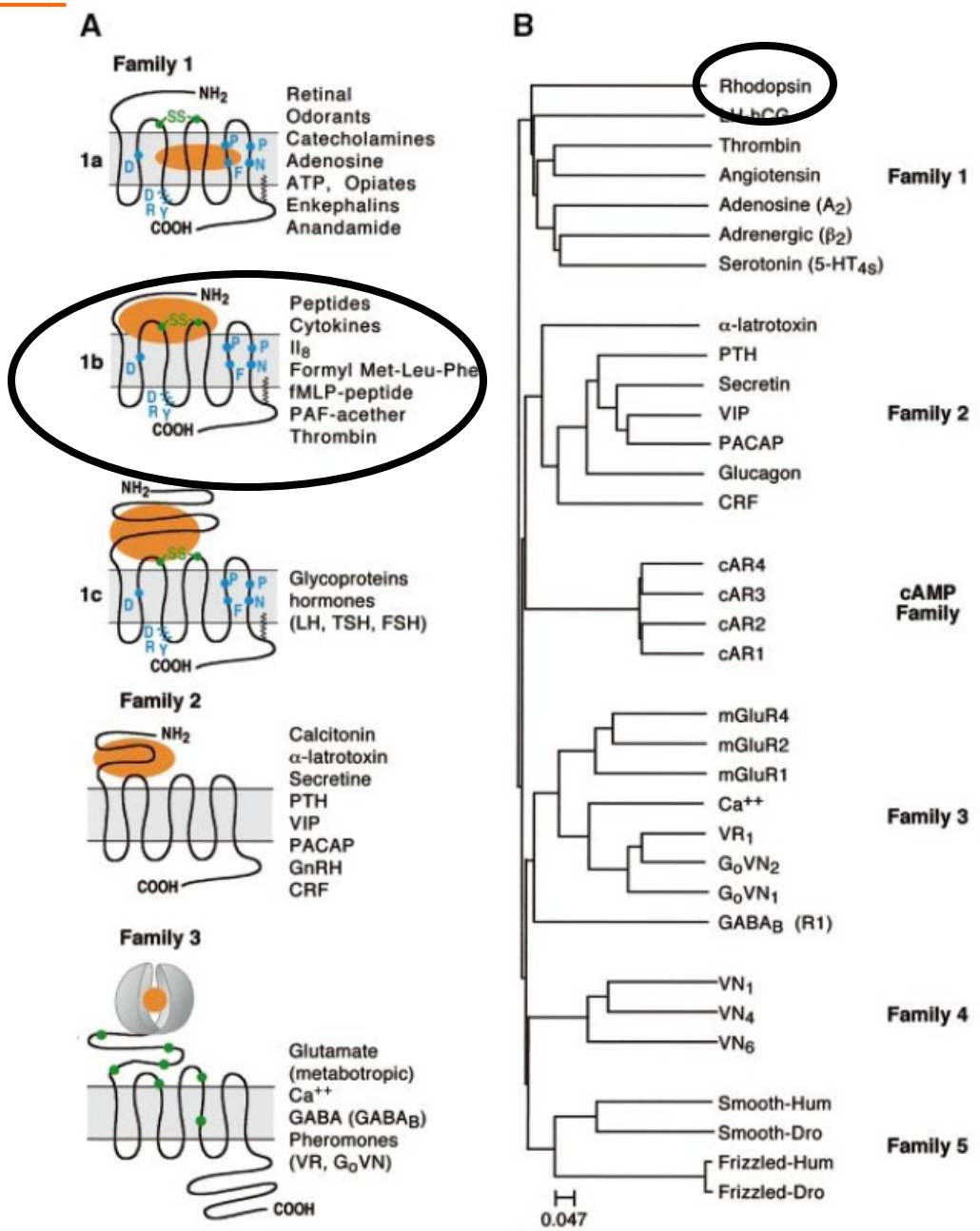
Remarks:

- The numbering system have been modified for class A GPCRs: it describes now the helices location instead of transmembrane domains. It is based on the structure determined by G. Schertler and coll (Li et al. 2004).
- The G proteins have been removed from the GPCRDB and will be available in several weeks via another database dedicated to GPCR interacting partners.

Class A: Rhodopsin like
Class B: Secretin like
Class C: Maetabotropic glutamate/pheromone
Class D: fungal pheromone
Class E: cAMP receptors (Dictyostelium)
Frizzleed/Smoothened family

cnio Centro Nacional de Investigaciones Oncológicas

(4) CCR5



J.Bockaert and J.P.Pin

How many GCRP?

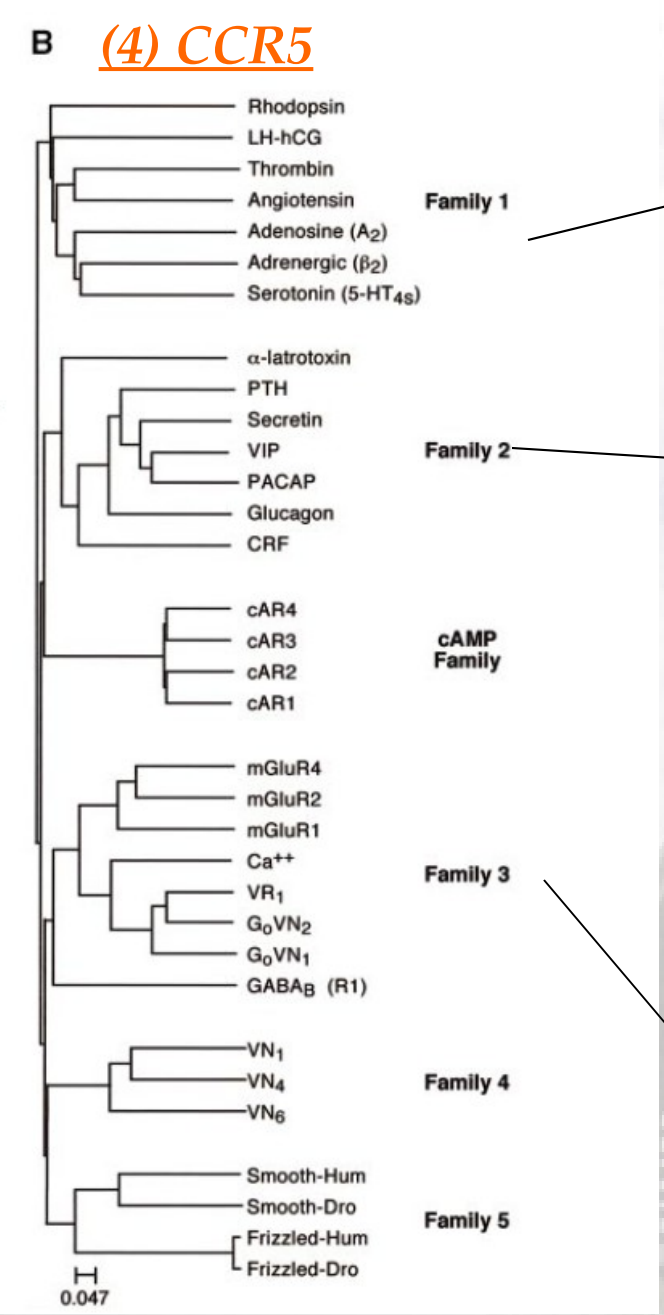
2495 entries!

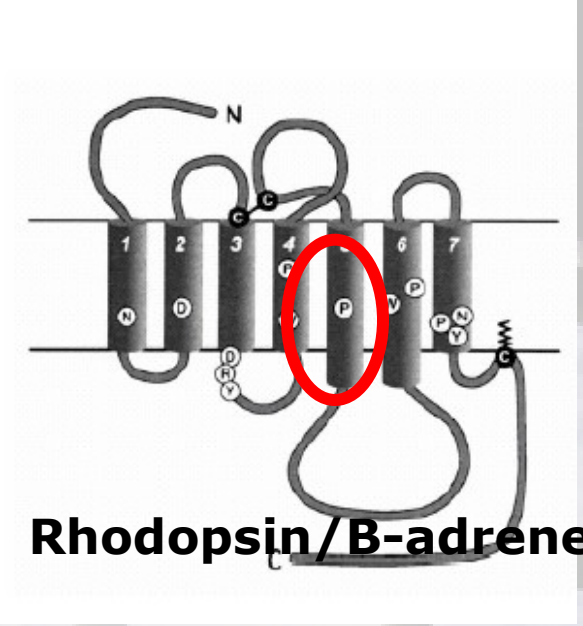
→ **CLASS A**
(16 subfamilies)

- No sequence similarity!
- The tunning to bind ligand-G prots is regulated by RNA editing and phosphori-lation

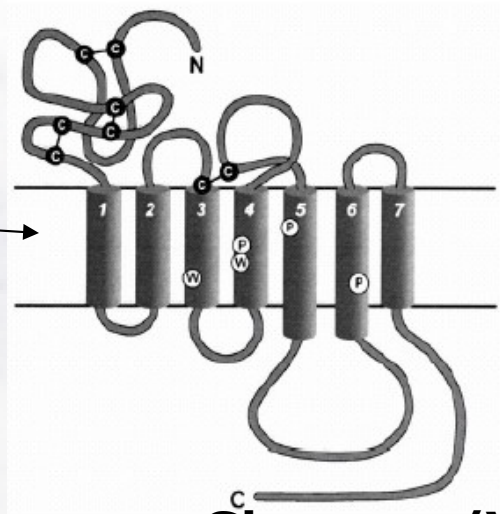
cnio
Centro Nacional de Investigaciones Oncológicas

(4) CCR5

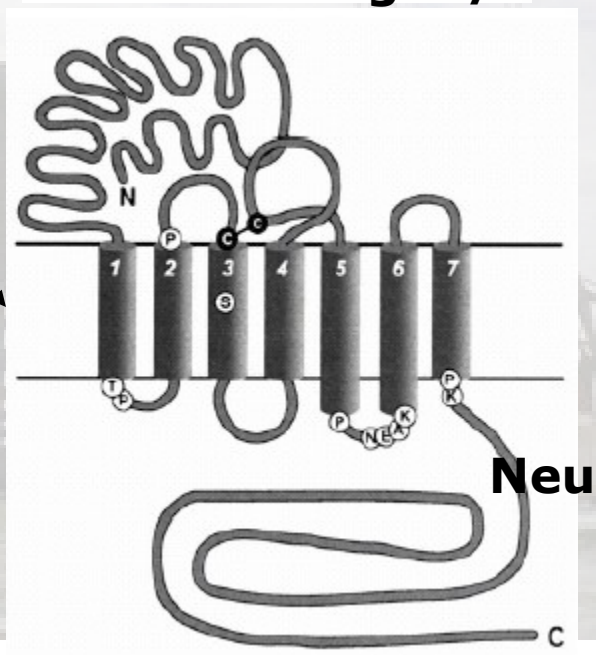




Rhodopsin/B-adrenerg



Glucagon/VIP/Calcitonin



Neurotransmiter/Ca

cnio
Centro Nacional de Investigaciones Oncológicas

CCR5

well documented (aminergic R) =505

Class A

Sub-families

- **Amine**
- Peptide
- Hormone protein
- (Rhod)opsin
- Olfactory
- Prostanoid
- Nucleotide-like
- Cannabinoid
- Platelet activating factor
- Gonadotropin-releasing hormone
- Thyrotropin-releasing hormone & Secretagogue
- Melatonin
- Viral
- Lysosphingolipid & LPA (EDG)
- Leukotriene B4 receptor
- Class A Orphan/other

Sub-families

Muscarinic
acetylcholine
Adrenoceptors
Dopamine
Histamine
Serotonin
Octopamine
Trace amine

Sub-families

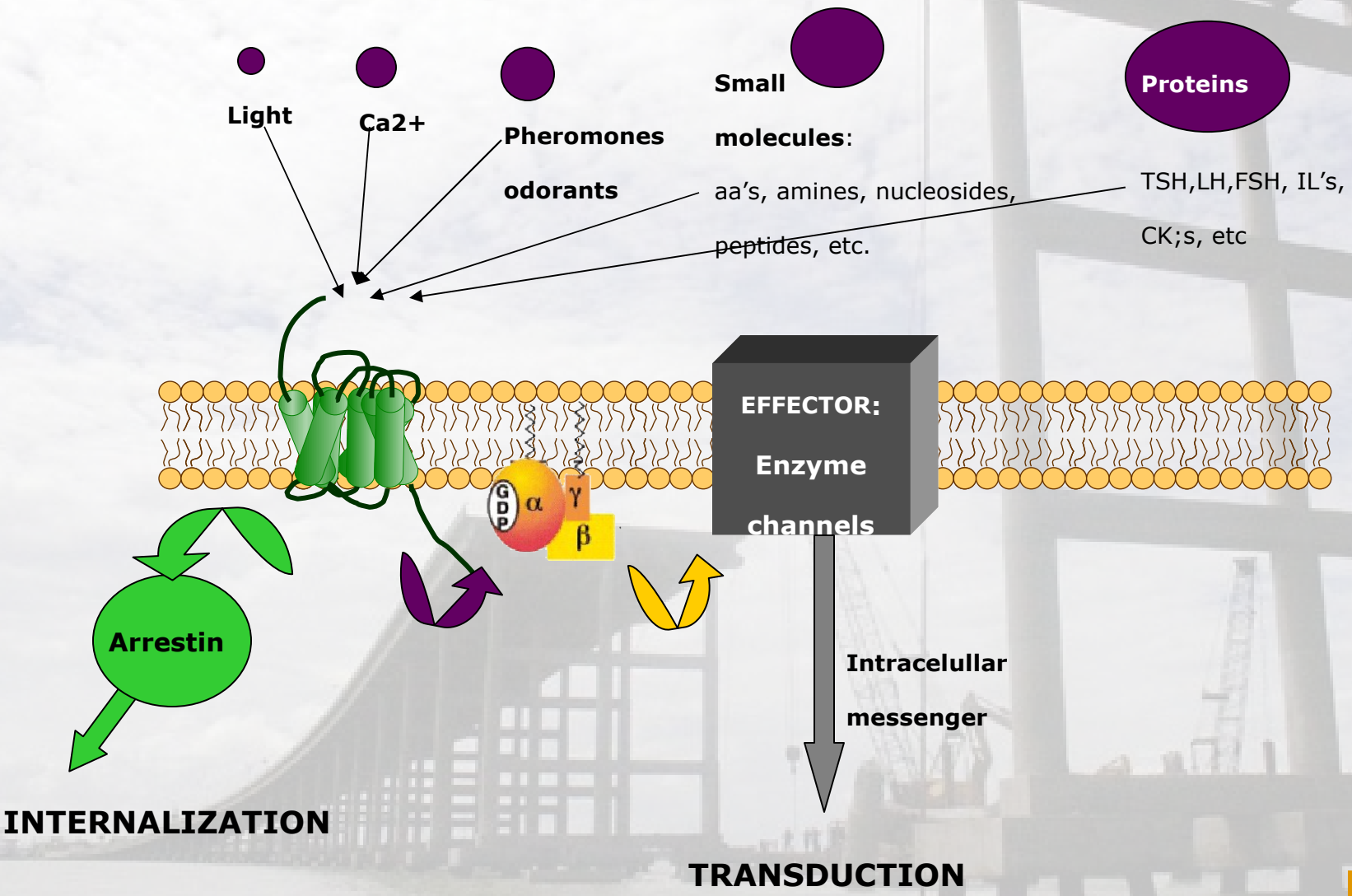
Angiotensin
Bombesin
Bradykinin
C5a anaphylatoxin
Fmet-leu-phe
APJ like
Interleukin-8
Chemokine → **357 Highly similar**
Cholecystokinin CCK
Endothelin
Melanocortin
Duffy antigen
Neuropeptide Y
Neurotensin
Opioid
Somatostatin
Tachykinin
Vasopressin-like
Galanin like
Proteinase-activated like
Orexin & neuropeptides FF,QRFP
Urotensin II
Adrenomedullin (G10D)
GPR37 / endothelin B-like
Chemokine receptor-like
Neuromedin U like
Somatostatin- and angiogenin
-like peptide
Allatostatin C / drostatin C
Melanin-concentrating hormone receptor

Prokineticin receptors
Other peptide receptors

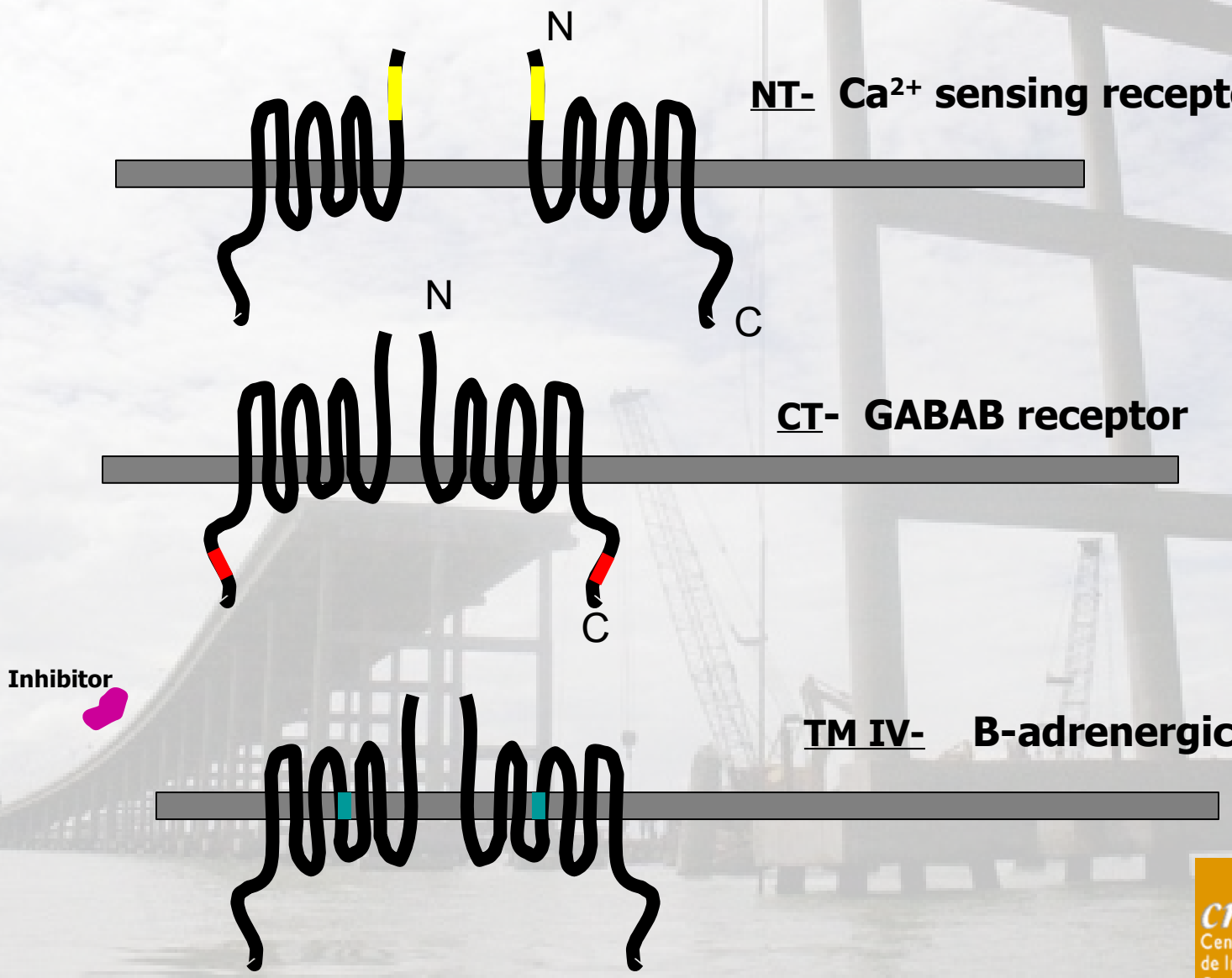
1094!

cnio
Centro Nacional de Investigaciones Oncológicas

GCRP: Ligand binding



The GCPR's dimerize



GCRP: The problem

The two main events here are:

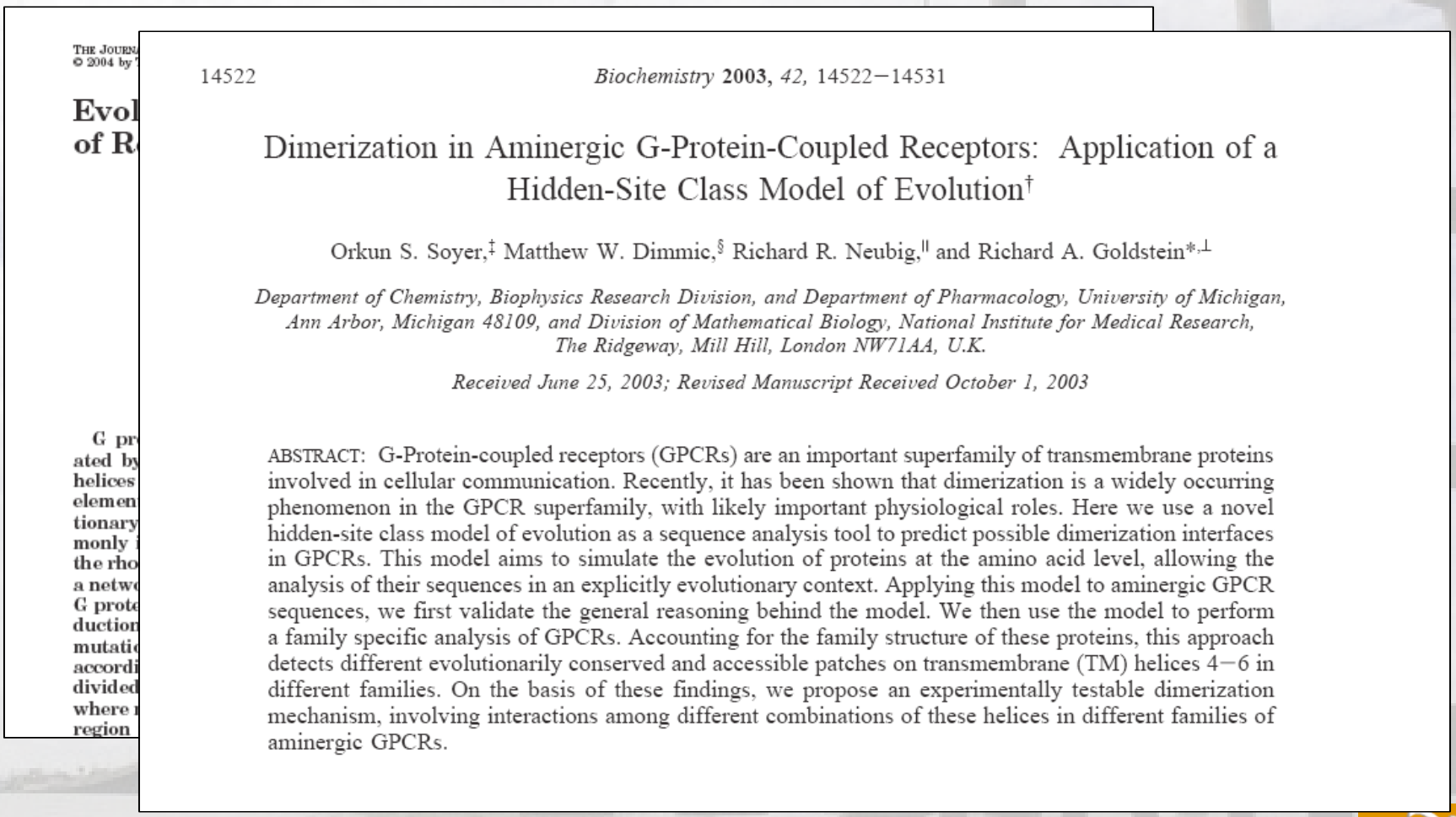
- Binding specificity.
- Dimerization/Oligomerization.

Then, we have two aims:

- ***Can we predict the signals and distinguish them at the sequence level?***
- ***Which residues are involved in dimerization?***

(4) CCR5

- ***Existing methods to detect important residues:***



14522 *Biochemistry* **2003,** *42,* 14522−14531

Dimerization in Aminergic G-Protein-Coupled Receptors: Application of a Hidden-Site Class Model of Evolution†

Orkun S. Soyer,‡ Matthew W. Dimmic,§ Richard R. Neubig,∥ and Richard A. Goldstein*,⊥

Department of Chemistry, Biophysics Research Division, and Department of Pharmacology, University of Michigan, Ann Arbor, Michigan 48109, and Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW71AA, U.K.

Received June 25, 2003; Revised Manuscript Received October 1, 2003

ABSTRACT: G-Protein-coupled receptors (GPCRs) are an important superfamily of transmembrane proteins involved in cellular communication. Recently, it has been shown that dimerization is a widely occurring phenomenon in the GPCR superfamily, with likely important physiological roles. Here we use a novel hidden-site class model of evolution as a sequence analysis tool to predict possible dimerization interfaces in GPCRs. This model aims to simulate the evolution of proteins at the amino acid level, allowing the analysis of their sequences in an explicitly evolutionary context. Applying this model to aminergic GPCR sequences, we first validate the general reasoning behind the model. We then use the model to perform a family specific analysis of GPCRs. Accounting for the family structure of these proteins, this approach detects different evolutionarily conserved and accessible patches on transmembrane (TM) helices 4−6 in different families. On the basis of these findings, we propose an experimentally testable dimerization mechanism, involving interactions among different combinations of these helices in different families of aminergic GPCRs.

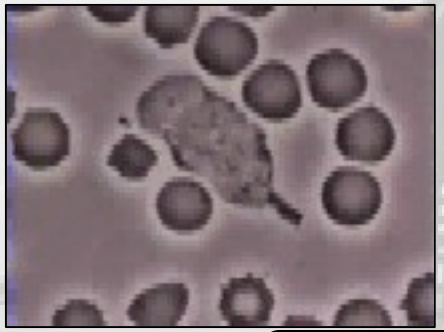
Same approach but use Probabilistic trees (MrBayes!)

Hannenhalli & Russell. **JMB** (2000). 306:61-76.

Chemokines: biological functions

CHEMOKINES

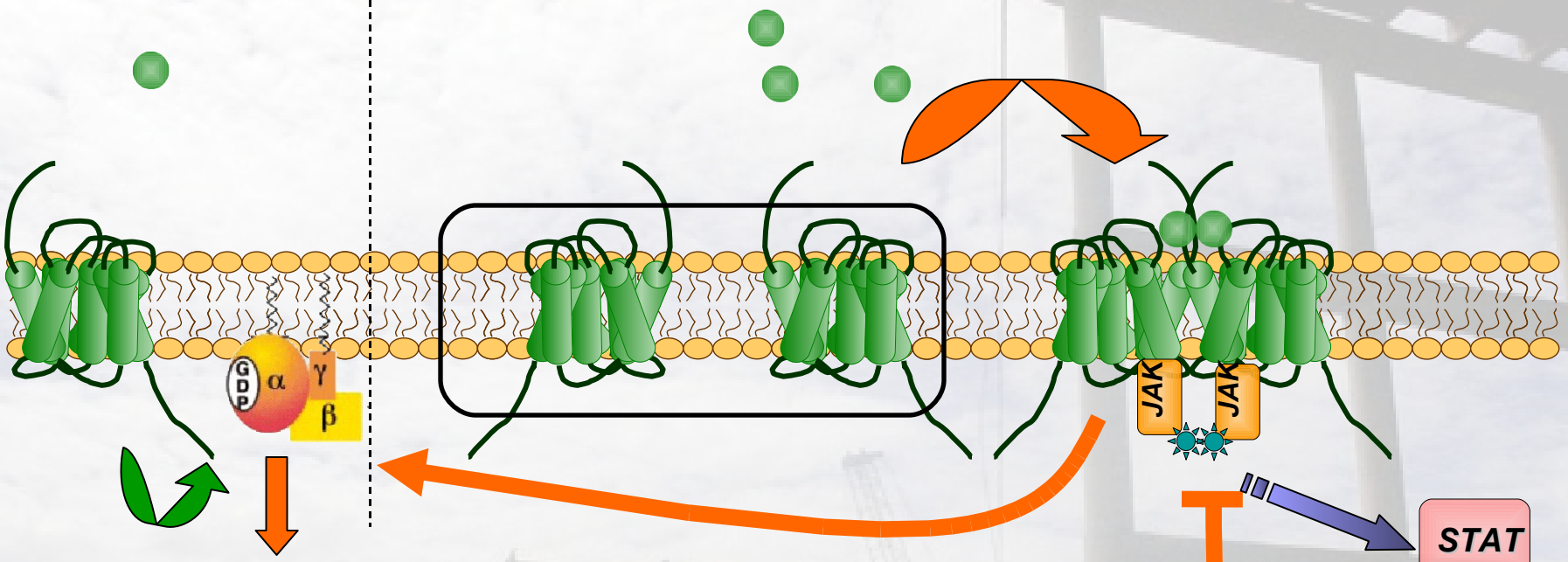
- Wound healing
- Th1/Th2 development
- Angiogenesis
- Tumor metastasis
- Cell recruitment
- Inflammation
- Organogenesis
- Lymphoid trafficking



Rogers D.
Vanderbilt University (1950s)

Rossi & Zlotnik. **Annu. Rev. Immunol.** (2000). 18:217-242.

CK'S SIGNALING



ADHESION
CHEMOTAXIS
POLARIZATION
INTERNALIZATION
GENE EXPRESSION

Thelen (2001)

Mellado et al, (2001)

cnio
Centro Nacional de Investigaciones Oncológicas

Our strategy

TEST CASE: CHEMOKINES, known to dimerize.

Steps:

1.- Alignment selection.

2.- Tree determinants searching=finding residues.

3.- Selecting regions.

4.- Mapping and rough model generation based on bovine Rhodopsin (to visually represent the results).

5.- Experimental validation

Alignment selection

TEST CASE: CHEMOKINES

(http://www.gpcr.org/7M/)

- **Clustering**: to obtain a representative alignment containing groups: CCR1-9, CXCR3-5, and IL8A-B (**total 61**).

- **Different levels** of redundancy tested (75-100%). A redundancy level of 95% selected to compensate the number of sequences and alignment bias reduction

- **Realignment** using T-COFFEE with secondary structure predictions taking into account the rhodopsin model.

Our strategy

TEST CASE: CHEMOKINES, known to dimerize.

Steps:

1.- Alignment selection.

2.- Tree determinants searching=finding residues.

3.- Selecting regions.

4.- Mapping and rough model generation based on bovine Rhodopsin (to visually represent the results).

5.- Experimental validation

Finding residues

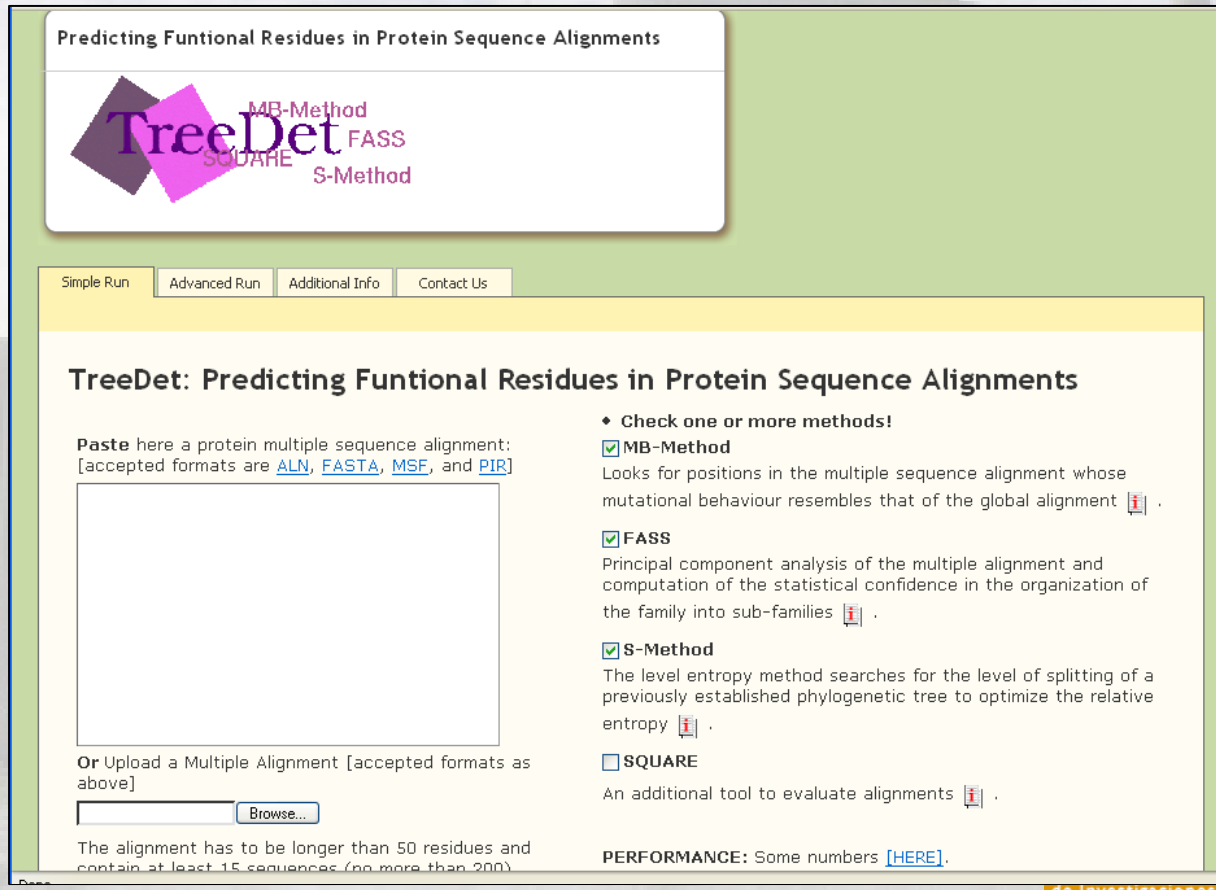
Basics: **Homodimerization** specificity is trying to avoid promiscuous dimerization between homologous sequences!

Dimerization-focused strategy: obtaining the best subfamily division (as many subfamily groups as possible).

TREE DETERMINANT SEARCHING

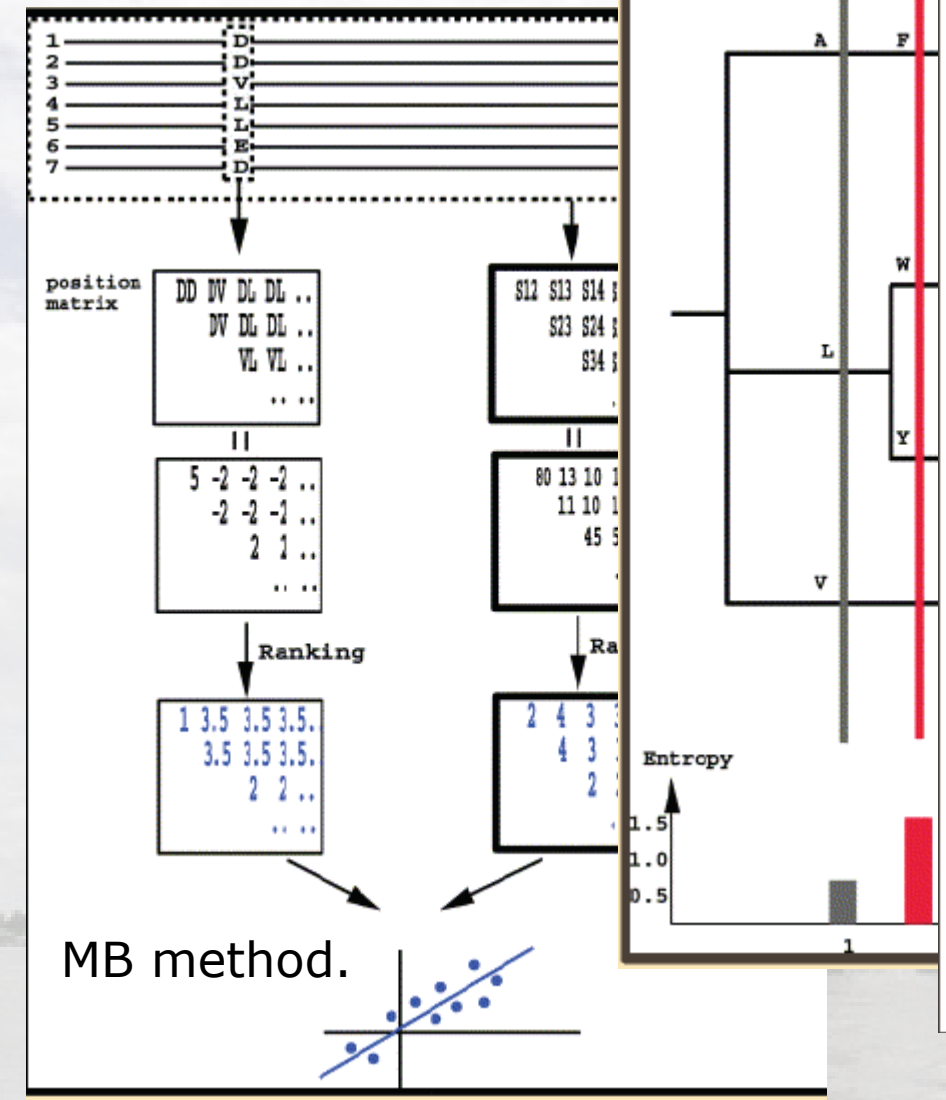
- Level entropy
- Mutational behaviour

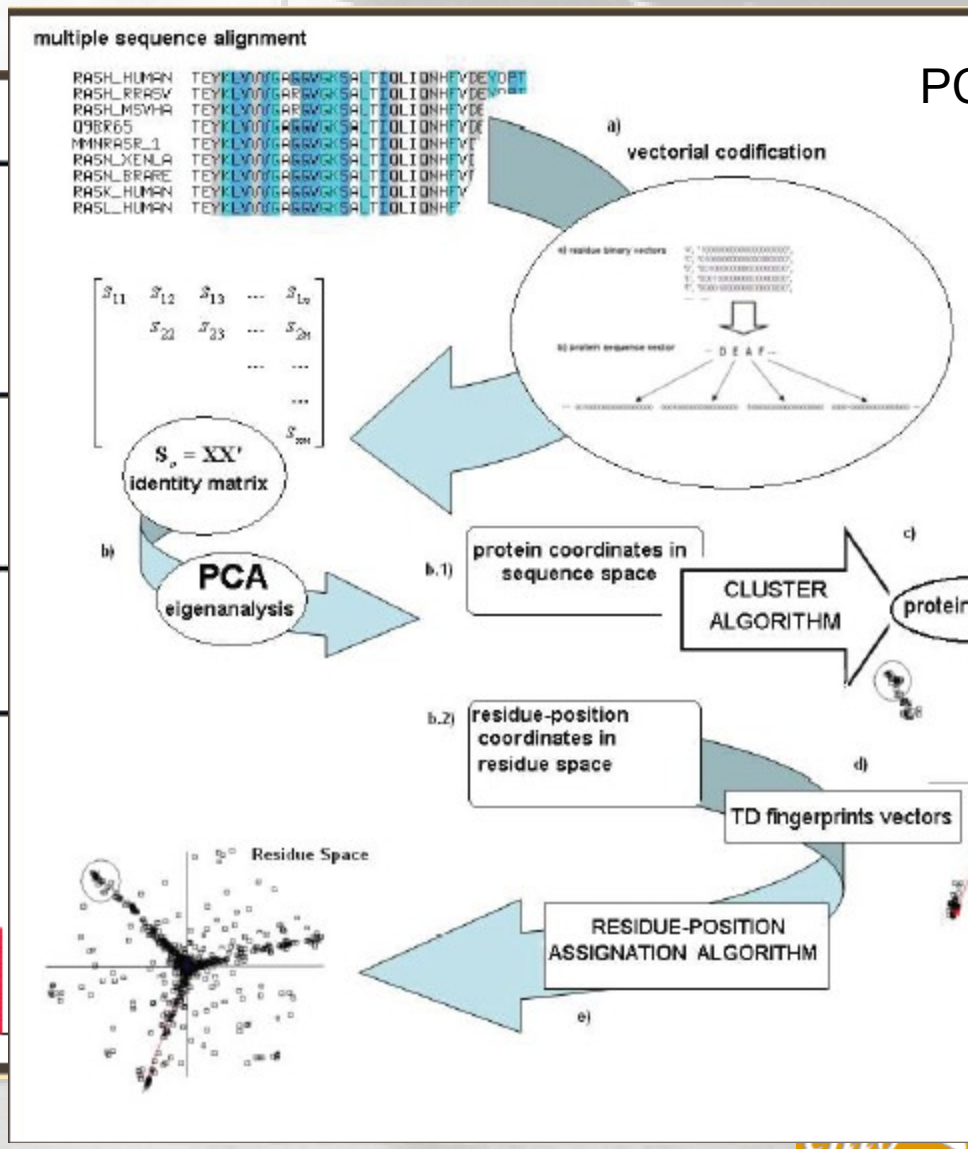
- Sequence Space Automated Method



- What the methods do?

Predict functional sites using different approaches.



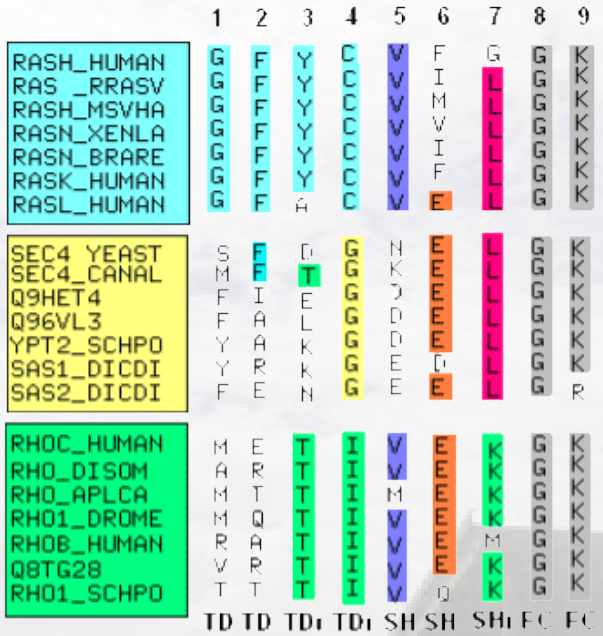


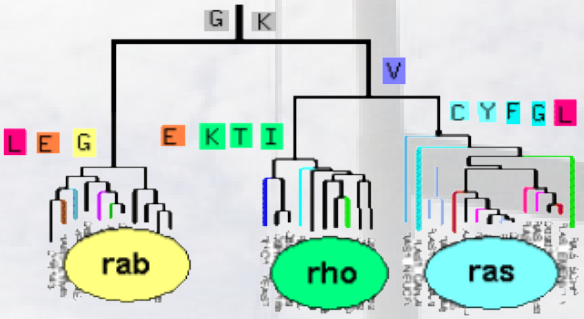
PC

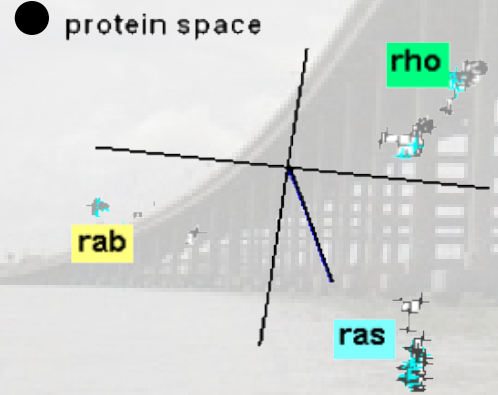
Sequence Space: overview

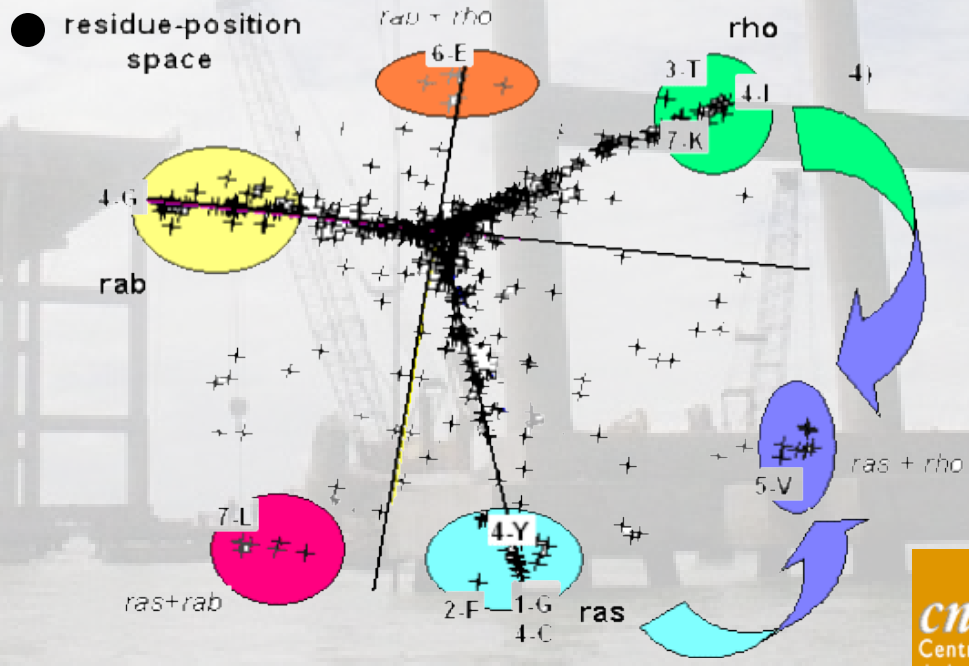
Casari, G. et al. **Nat. Struct. Biol** (1995). 2:171-178.

An example:

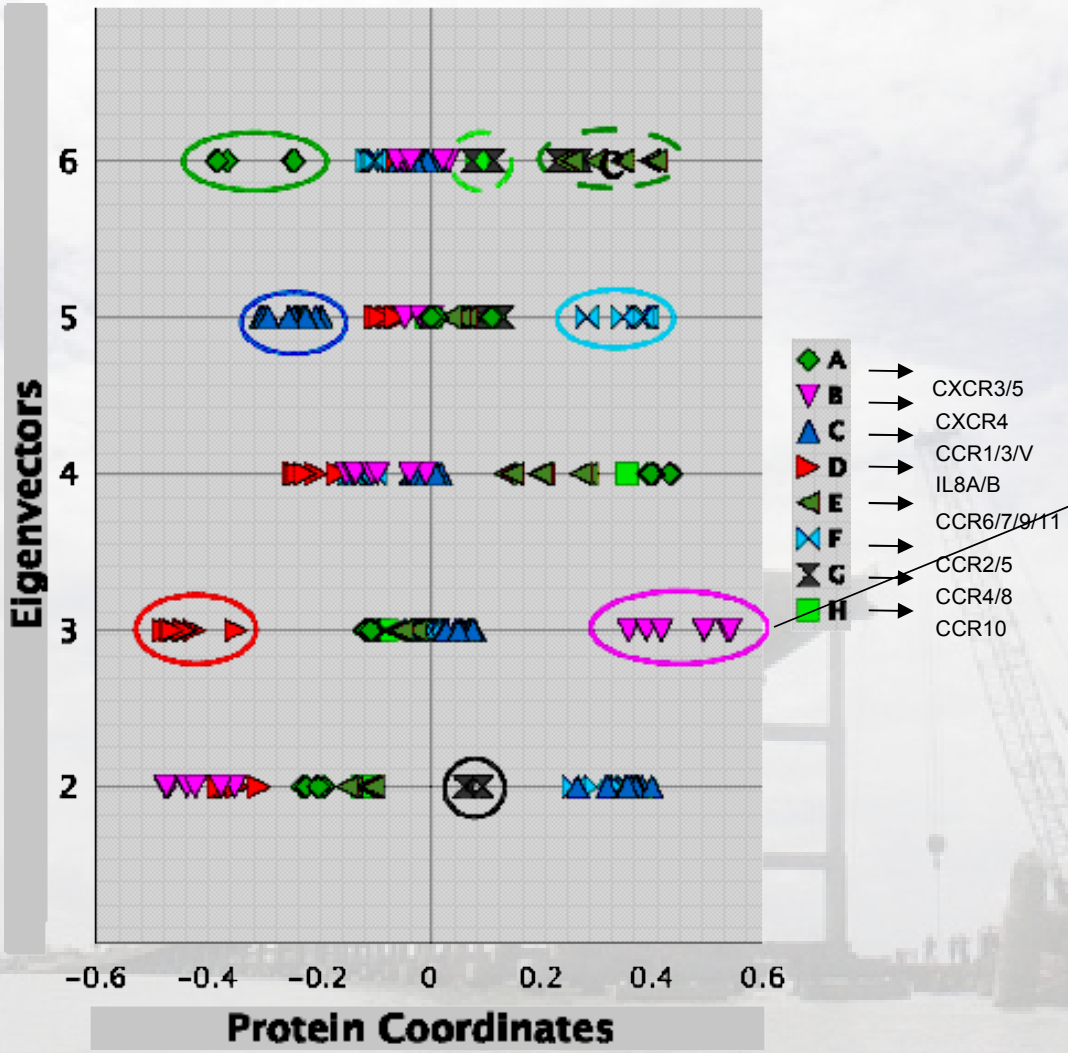




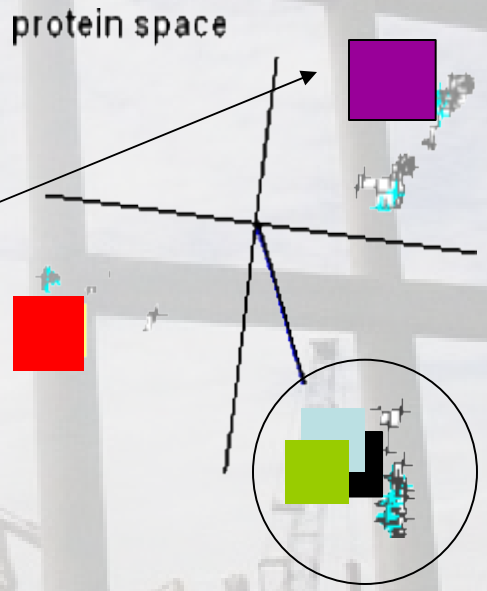




Sequence Space: Clustering results



CXCR3/5
CXCR4
CCR1/3/V
IL8A/B
CCR6/7/9/11
CCR2/5
CCR4/8
CCR10



cnio
Centro Nacional de Investigaciones Oncológicas

Our strategy

TEST CASE: CHEMOKINES, known to dimerize.

Steps:

1.- Alignment selection.

2.- Tree determinants searching=finding residues.

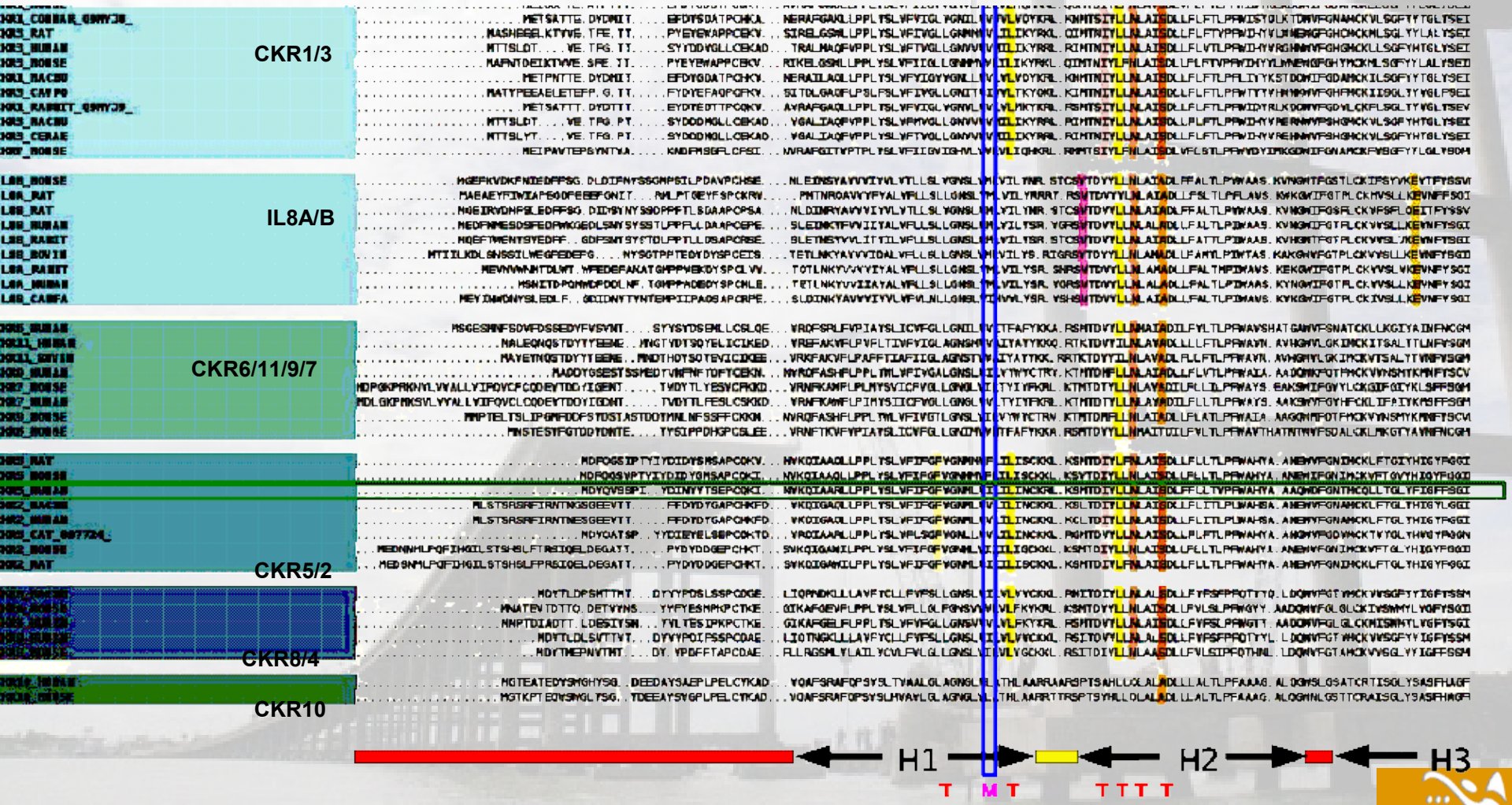
3.- Selecting regions.

4.- Mapping and rough model generation based on bovine Rhodopsin (to visually represent the results).

5.- Experimental validation

Tree-determinants: Clustering results

Residues obtained by Sequence-Space family division.



Our strategy

TEST CASE: CHEMOKINES, known to dimerize.

Steps:

1.- Alignment selection.

2.- Tree determinants searching=finding residues.

3.- Selecting regions.

4.- Mapping and rough model generation based on bovine Rhodopsin (to visually represent the results).

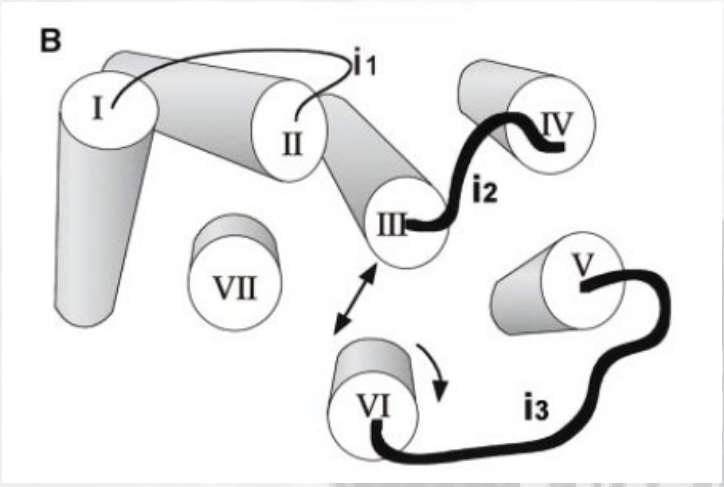
5.- Experimental validation

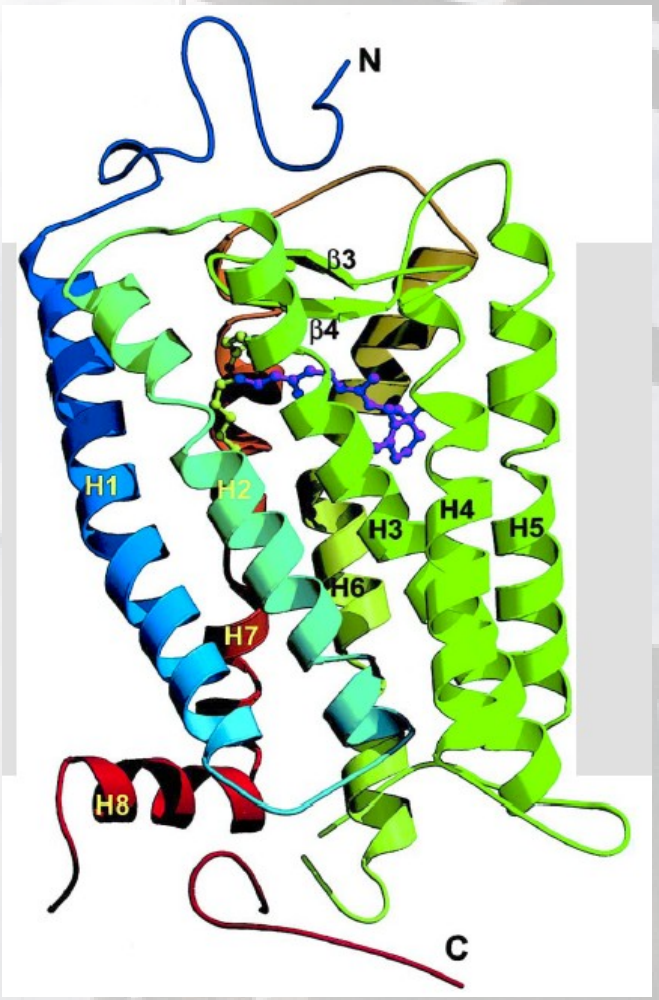
Problem: there is no Structure!!!!

1F88 (2000) 2.8A

1GZM (2002) 2.65A

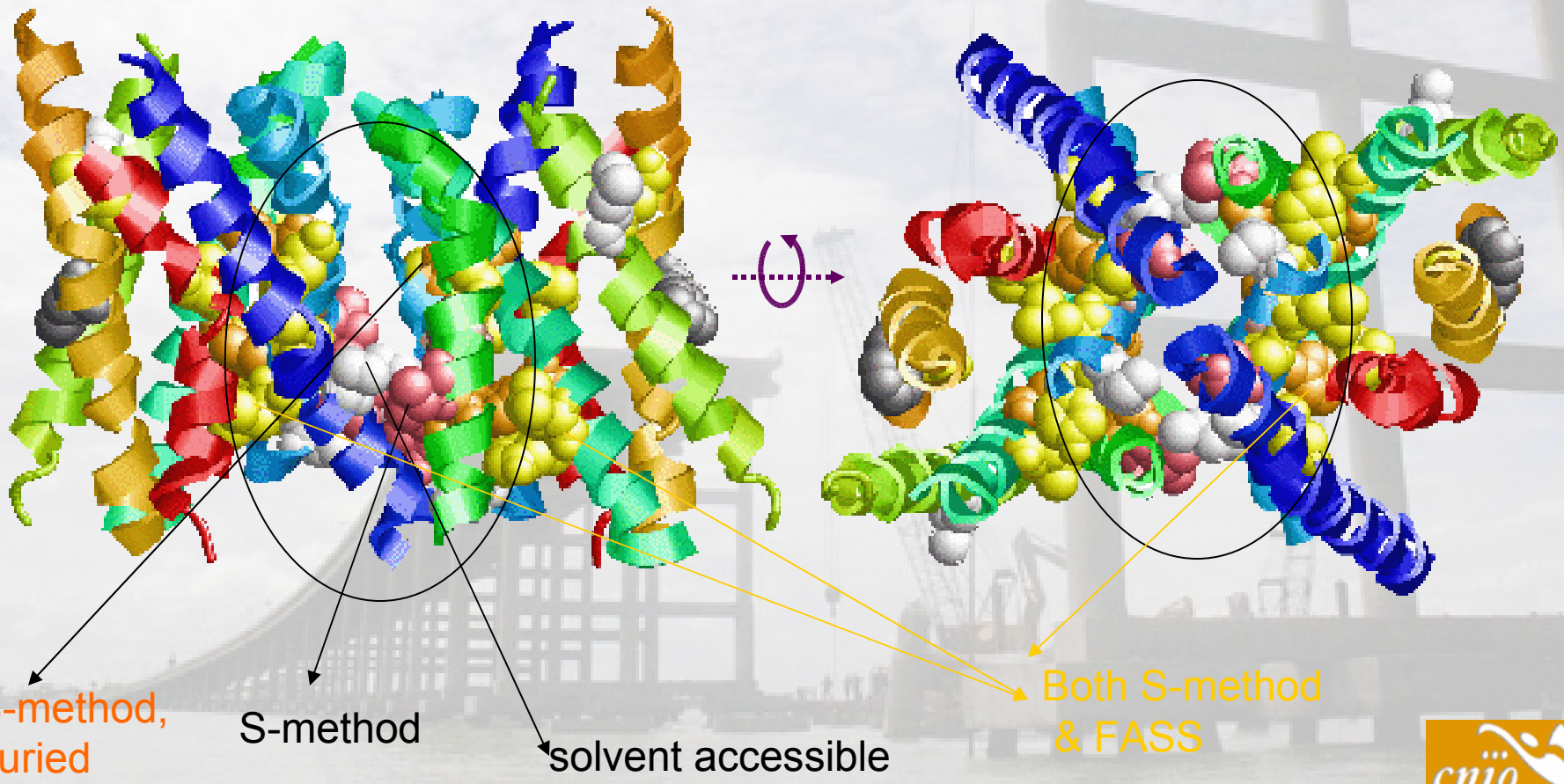
No similarity but there is a “central core” or “bundle”



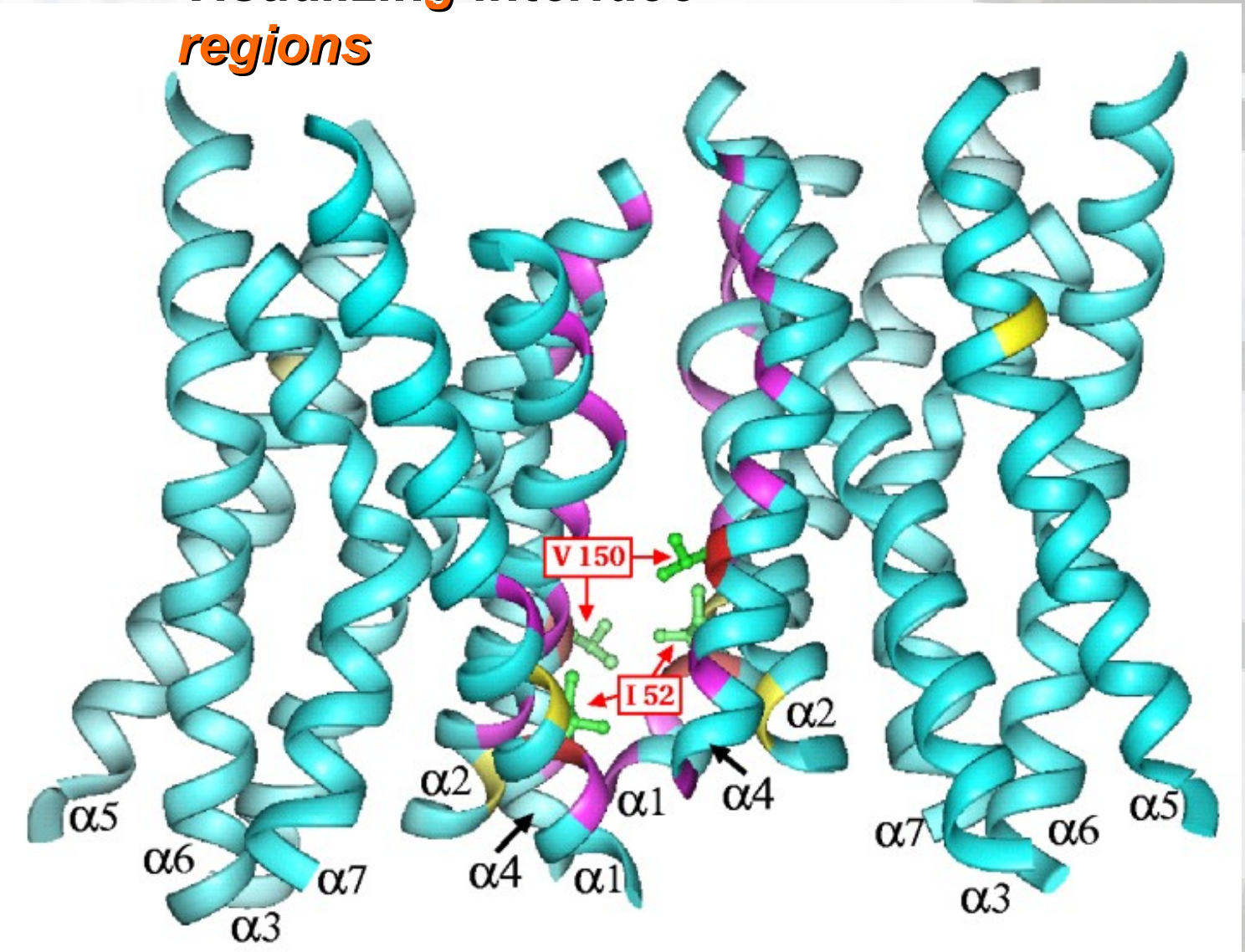


Visualizing interface regions

Region selection and then, residue selection (not necessarily the TD's)



(4) CCR5 Visualizing interface regions



Our strategy

TEST CASE: CHEMOKINES, known to dimerize.

Steps:

1.- Alignment selection.

2.- Tree determinants searching=finding residues.

3.- Selecting regions.

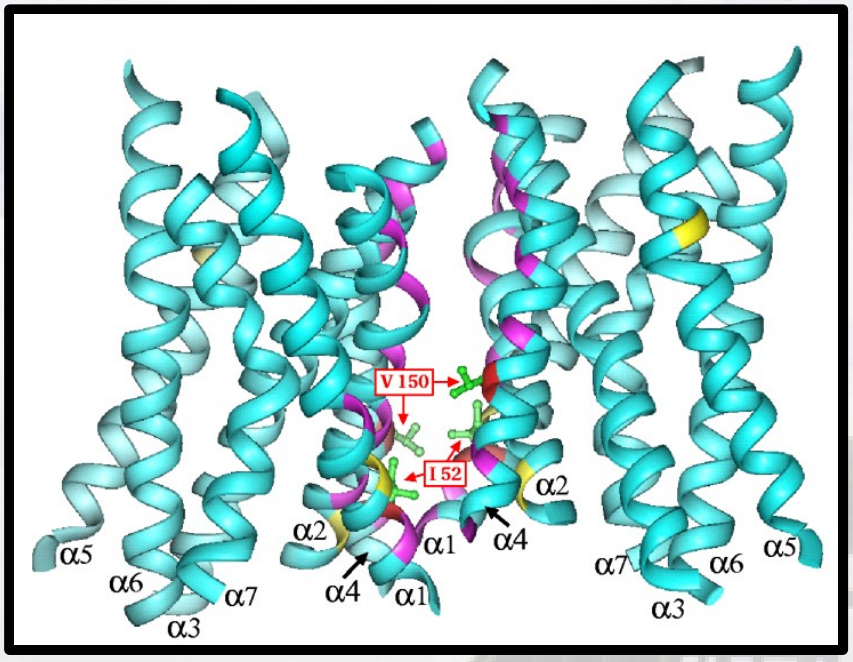
4.- Mapping and rough model generation based on bovine Rhodopsin (to visually represent the results).

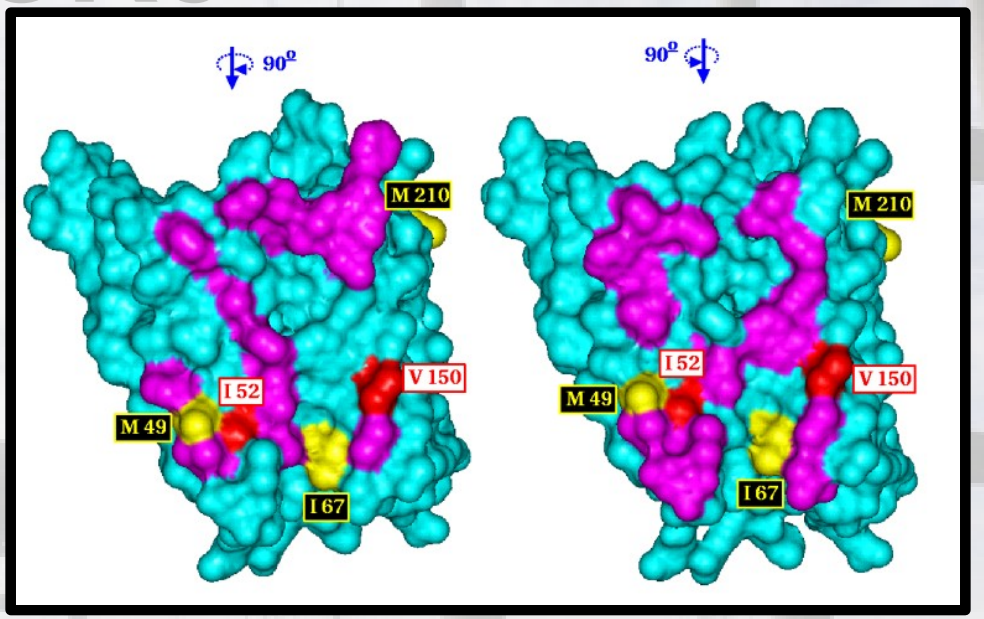
5.- Experimental validation.

Experimental validation

Chemokine receptor dimerization model

CCR5



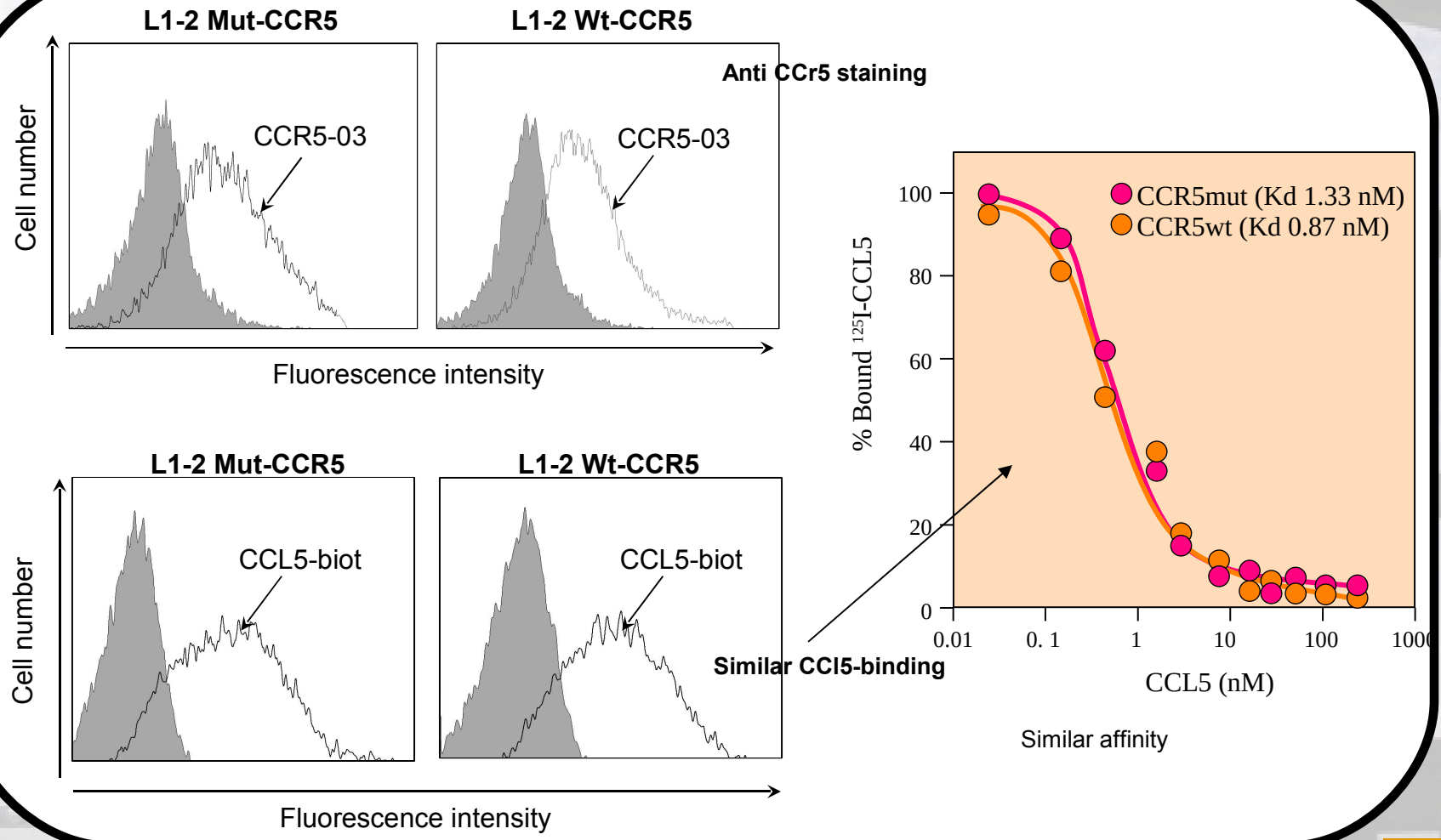


TM I : "GXXXLXXL"

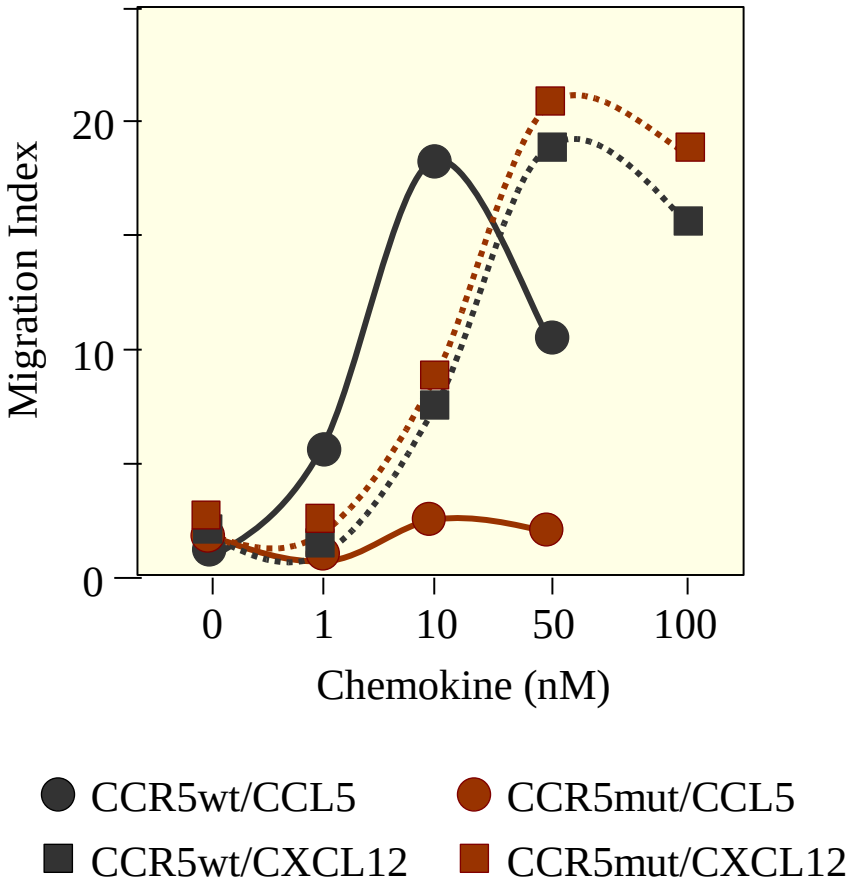
Hebert et al. **J. Biol. Chem.** (1996). 271:16384-16392.

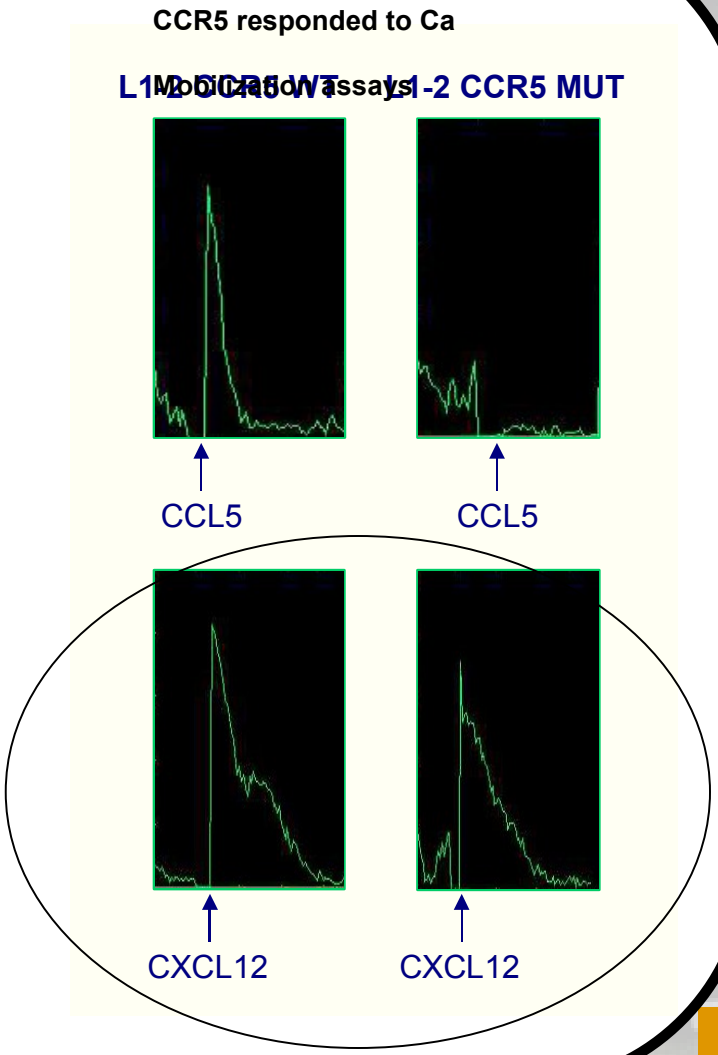
cnio Centro Nacional de Investigaciones Oncológicas

CCR5I52V/V150A and CCR5 show similar membrane expression and ligand binding



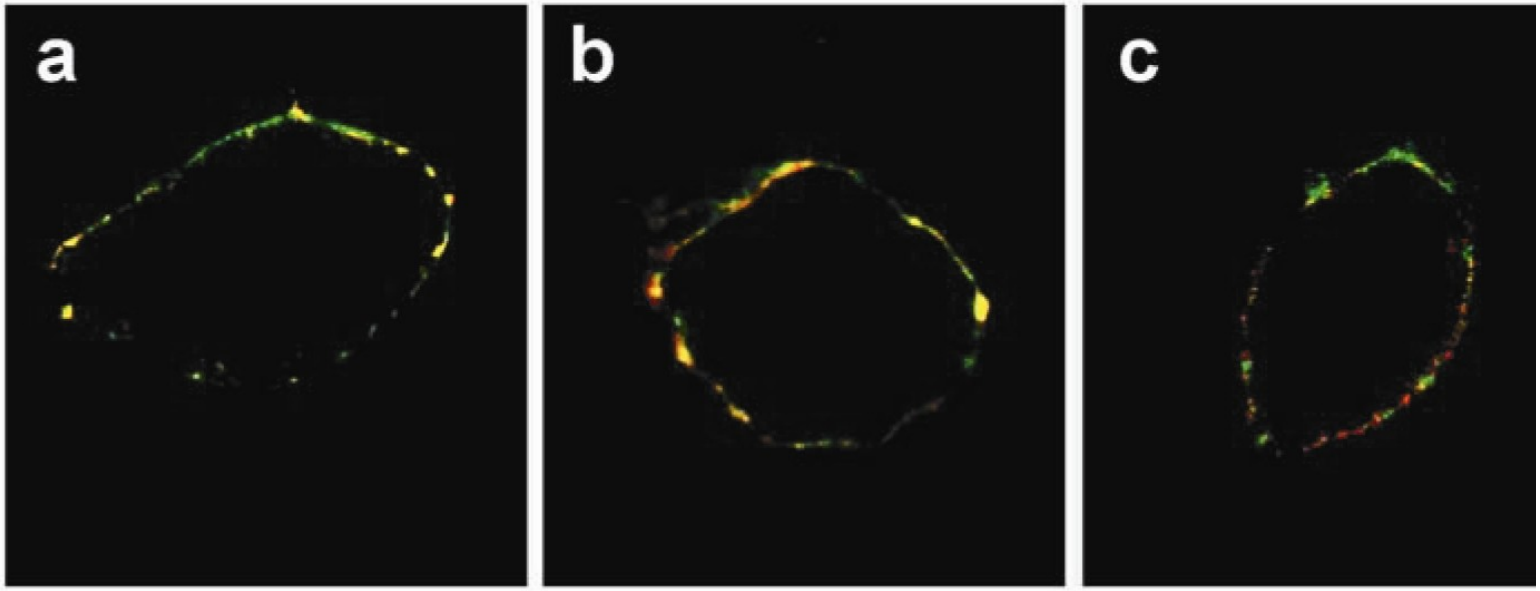
CCR5I52V/V150A is a non-functional receptor





cnio
Centro Nacional de Investigaciones Oncológicas

CCR5I52V/V150A is a non-functional receptor

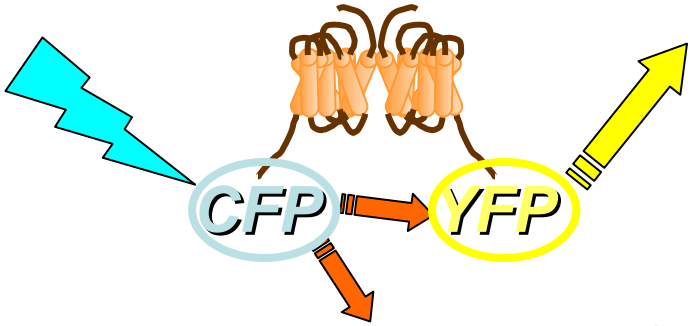


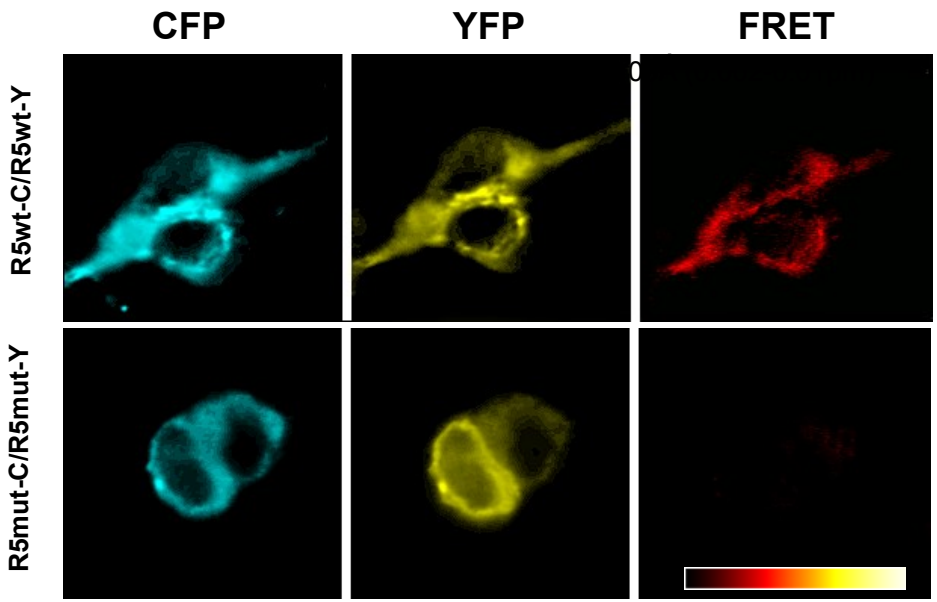
Green: CTx → Marker for rafts
Red: CCR5wt/CCR5mut/TfR
Yellow: overlay

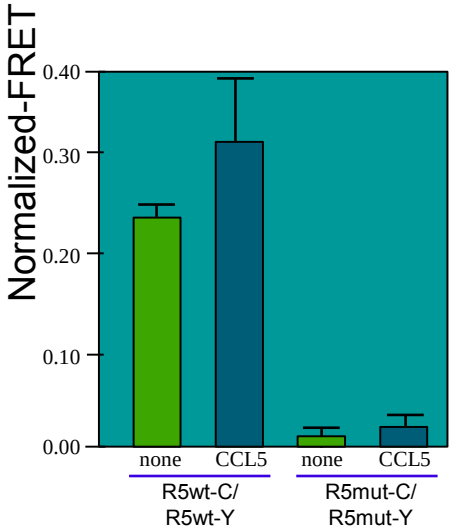
No differences in the membrane distribution

CCR5I52V/V150A does not dimerize: FRET

Fluorescence Resonance Energy Transfer





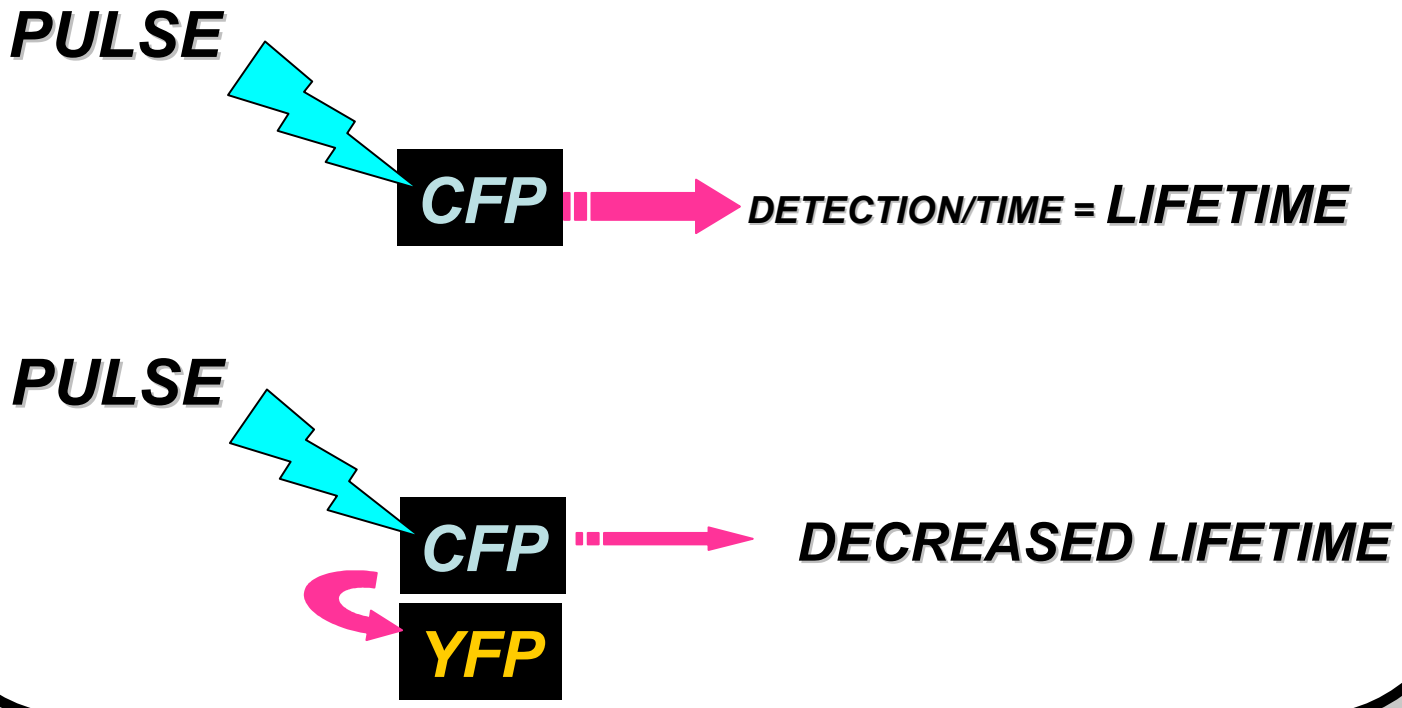


within 20-100Å (0.002-0.01µm): prot-prot interactions

cnio
Centro Nacional de Investigaciones Oncológicas

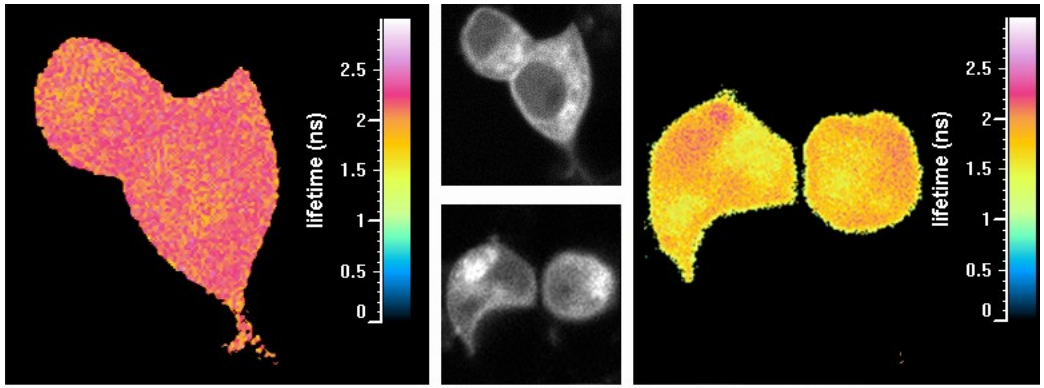
CCR5I^{52}V/V^{150}A does not dimerize: FLIM

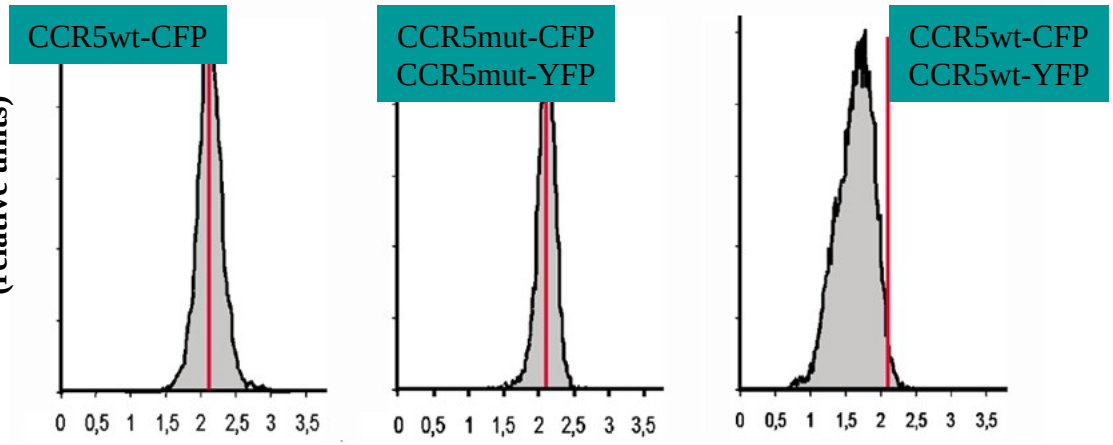
Fluorescence Lifetime IMaging



cnio
Centro Nacional de Investigaciones Oncológicas

CCR5I52V/V150A does not dimerize: FLIM



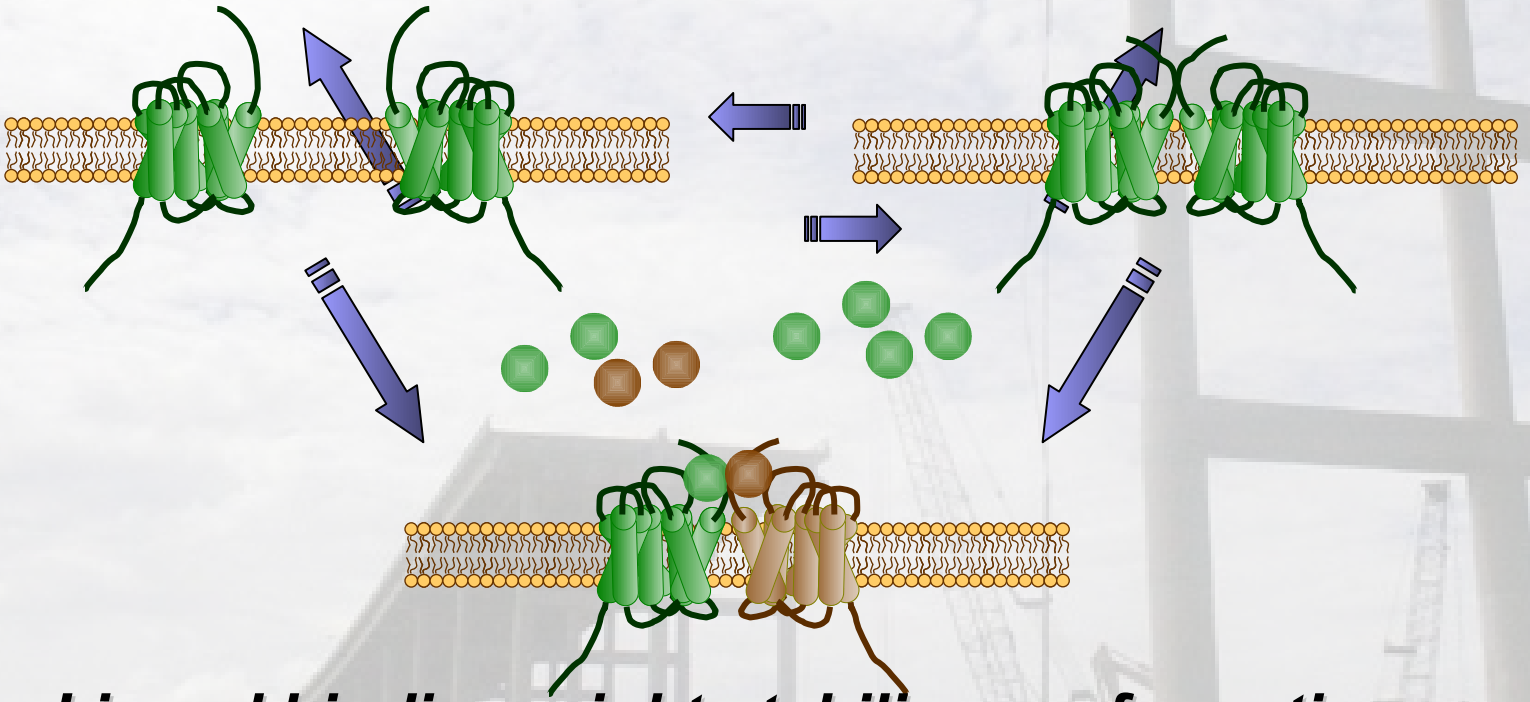


cnio
Centro Nacional de Investigaciones Oncológicas

CCR5 dimerization conclusions

CCR5 pre-exist as homodimers in absence of exogenous ligand

Chemokine receptors are in an equilibrium between several conformations: monomers, homodimers and heterodimers



Ligand binding might stabilize conformations

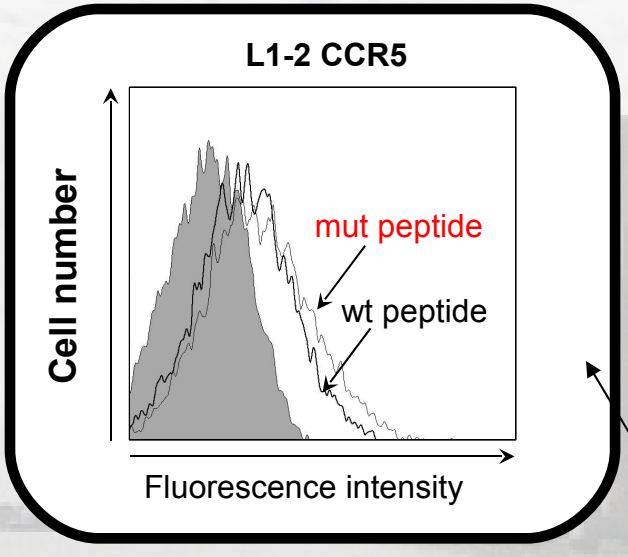
CCR5-based synthetic peptides containing I^{52} and V^{150} block CCR5 function

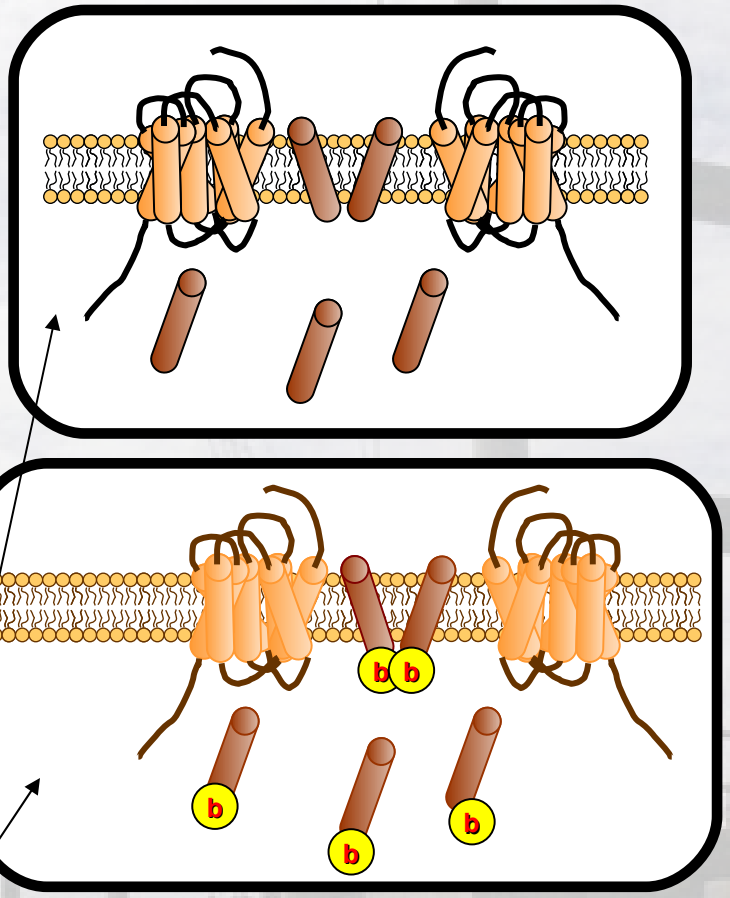
wt peptide

MLVILIL + VTSVITW

mutant peptide

MLVVLIL + VTSAITW





BOTH INCLUDED IN THE MB!

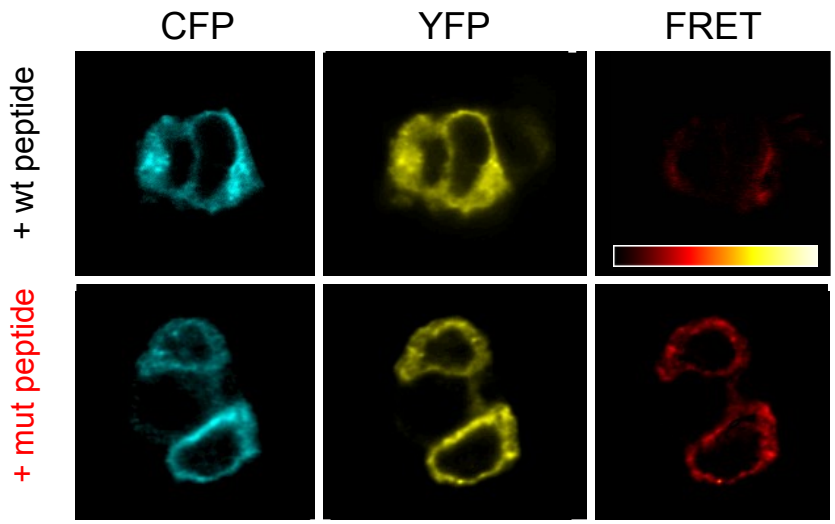
BIOTIN STAINING.

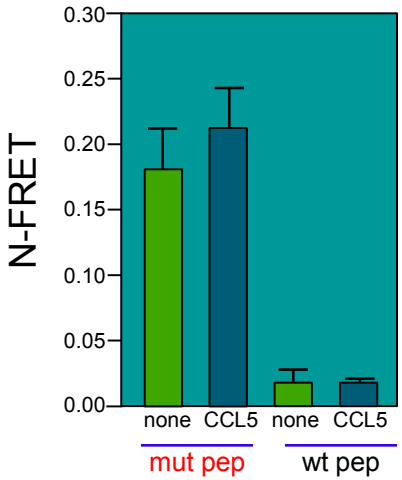
cnio Centro Nacional de Investigaciones Oncológicas

CCR5-based synthetic peptides containing I^{52} and V^{150} block CCR5 function, dimerization

wt peptide
MLVILIL + VTSVITW

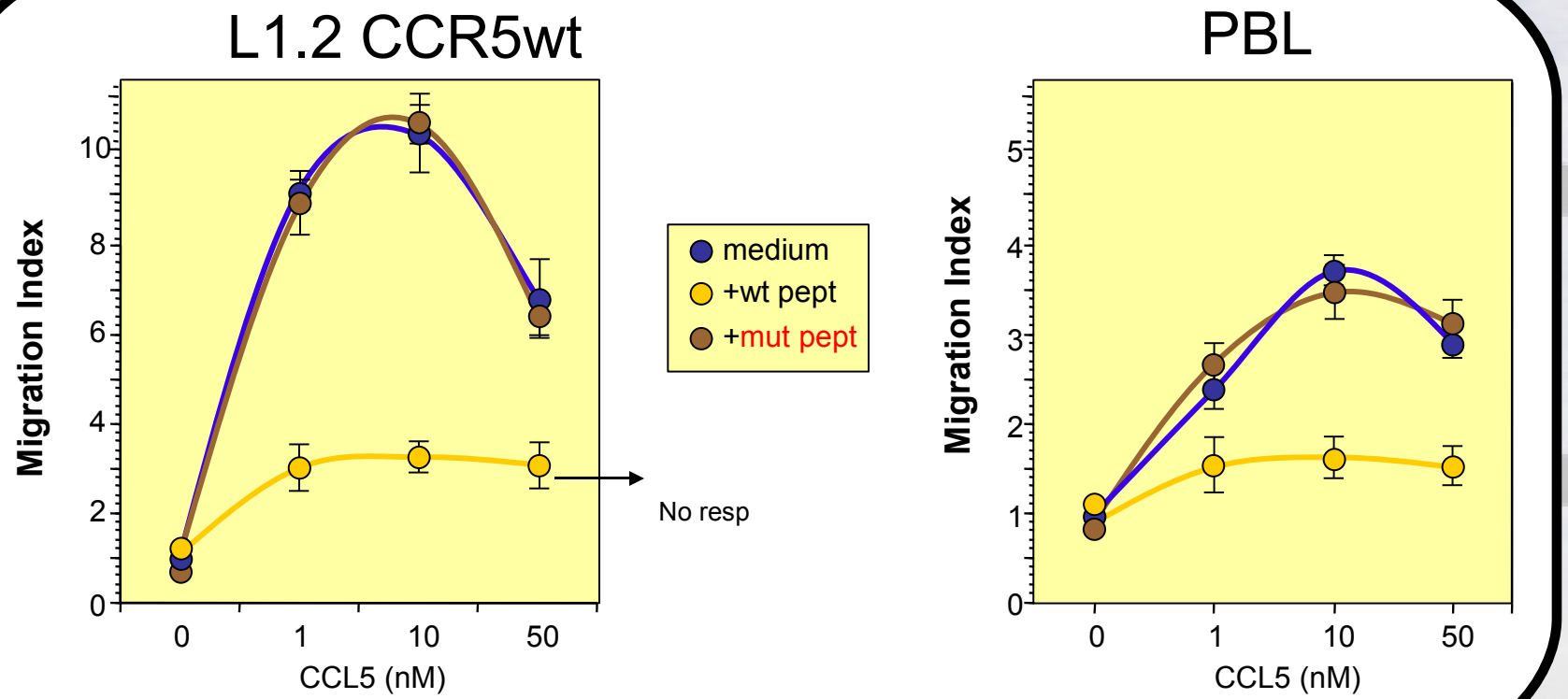
mutant peptide
MLVVLIL + VTSAITW





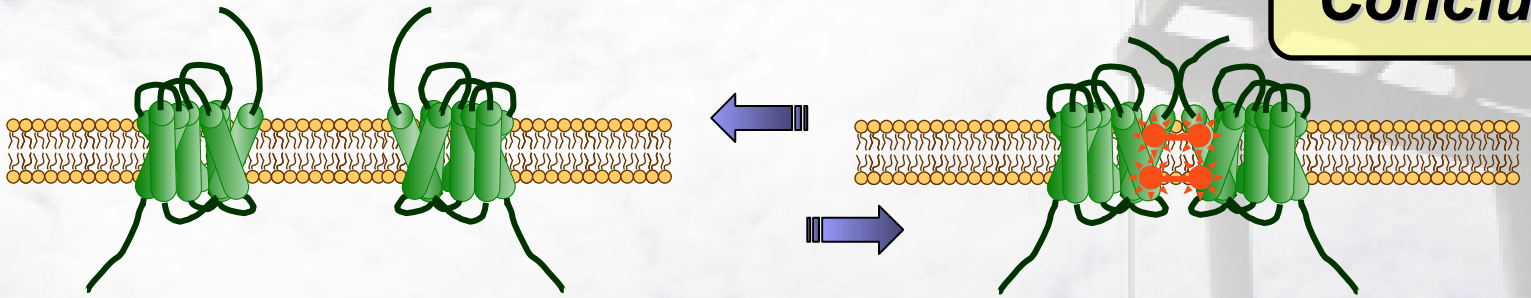
cnio
Centro Nacional de Investigaciones Oncológicas

CCR5-based synthetic peptides containing I52 and V150 block CCR5 function, migration assays



PBL: peripheral blood lymphocytes.

Conclusions



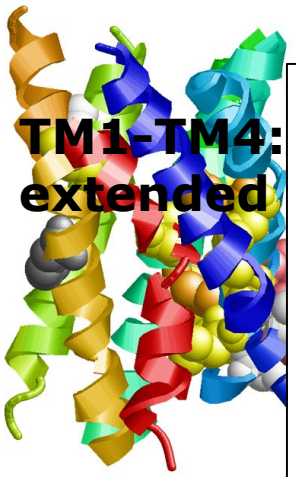
I52 in TM I and V150 in TM IV play a critical role in CCR5 dimerization and function

Hernanz-Falcon P. et al. **Nat. Immunol.** (2004). 5:216-223.

Hernanz-Falcon P. et al. **Nat. Immunol.** (2005). 6:535-536.

A

B



TM1-TM4: extended

BIOINFORMATICS

Vol. 00 no. 0 2005, pages 1-
doi:10.1093/bioinformatics/bti2

Review: collaboration between experimentalists and computational biologists.

A framework for computational and experimental methods: Identifying dimerization residues in CCR chemokine receptors.

David de Juan[1], Mario Mellado[2], José Miguel Rodríguez-Frade[2], Patricia Hernanz-Falcón[2], Antonio Serrano[2], Antonio del Sol[3], Alfonso Valencia[1], Carlos Martínez-A[2] and Ana María Rojas[1,2*].

[1]PDG, CNB-CSIC, Madrid, Spain, [2]Department of Oncology and Inmunology, CNB-CSIC, Madrid, Spain and [3] Bioinformatics Research Project. Fujirebio Inc.Tokyo, Japan.

cnio Centro Nacional de Investigaciones Oncológicas

Predicting Functional Residues in Protein Sequence Alignments

TreeDet: Predicting Functional Residues in Protein Sequence Alignments

Paste here a protein multiple sequence alignment:
[accepted formats are ALN, FASTA, MSF, and PIR]

Or Upload a Multiple Alignment [accepted formats as above]

- **Check one or more methods!**

☑ **MB-Method**
Looks for positions in the multiple sequence alignment whose mutational behaviour resembles that of the global alignment .

☑ **FASS**
Principal component analysis of the multiple alignment and computation of the statistical confidence in the organization of the family into sub-families .

☑ **S-Method**
The level entropy method searches for the level of splitting of a previously established phylogenetic tree to optimize the relative entropy .

☐ **SQUARE**
An additional tool to evaluate alignments .

ACKNOWLEDGEMENTS

BURNHAM INSTITUTE *for* MEDICAL RESEARCH

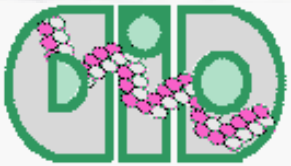
From Research, the Power to Cure

Dr. Adam Godzik

Dr. John C. Reed.

Luis Sanchez-Pulido

Sito Pazos



Mario Mellado

Karel van Wely

Thank you!

Alfonso Valencia

Rest of the group

cnio
Centro Nacional
de Investigaciones
Oncológicas