The background of the slide is a faded, grayscale image of a large bridge under construction. The bridge features a prominent, curved, cantilevered section on the left side. In the foreground, the skeletal steel framework of a building or another part of the bridge is visible, with several cranes positioned around it. The scene is set over a body of water under a cloudy sky.

Bridging predictions and experiments: some case examples

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



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).

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

Automated Pipeline
HUMAN vs. MOUSE
RIKEN-BURNHAM
initiative

USA

Spain

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

Domain focused sequence analyses
Protein characterization
ACRATA, SPOC

Predicting interaction interfaces
CCR5 dimerization.

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

Automated Pipeline
HUMAN vs. MOUSE
RIKEN-BURNHAM
initiative

Analyses of Human vs. Mouse transcriptome

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

Domain focused sequence analyses
Protein characterization
ACRATA, SPOC

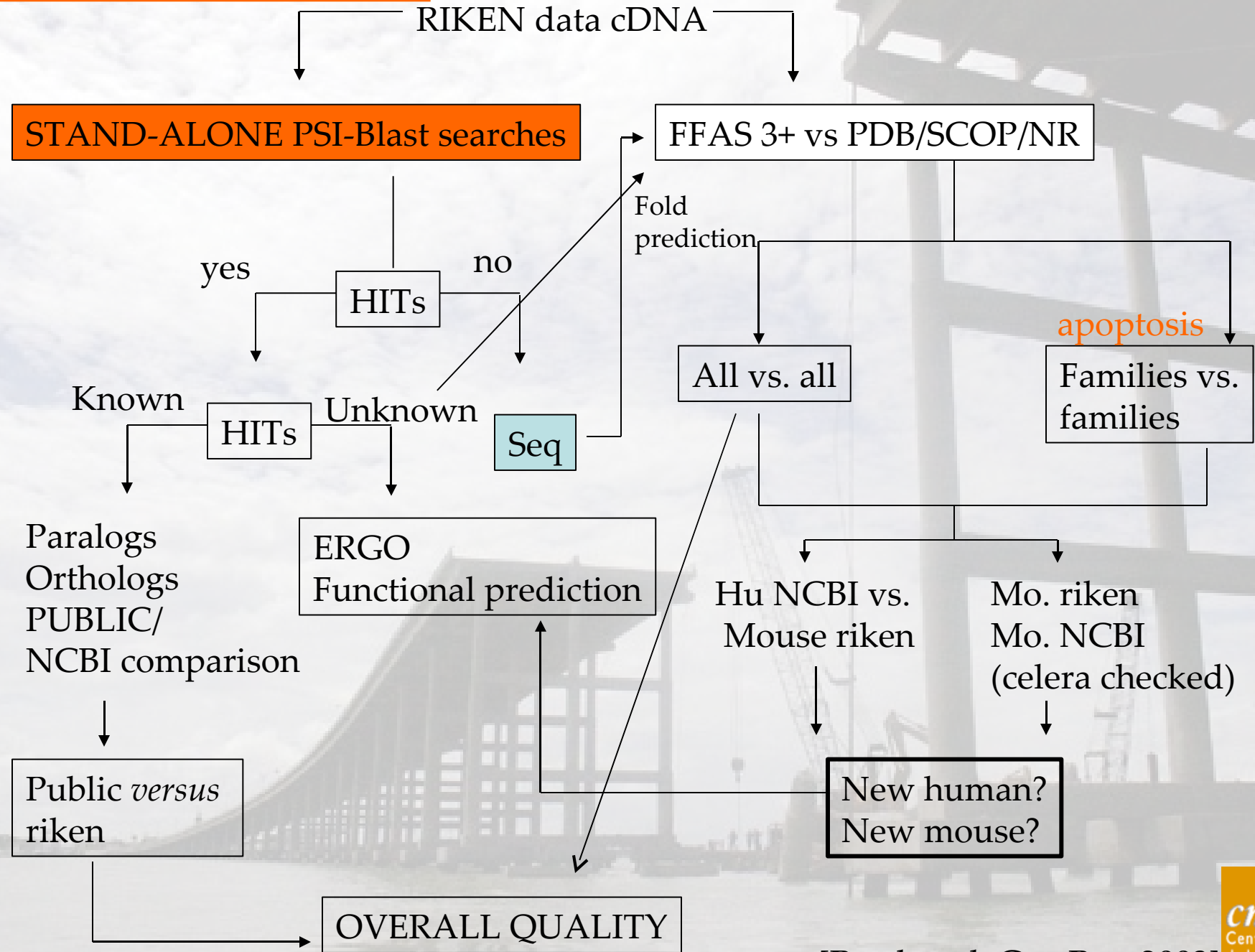
Predicting interaction interfaces
CCR5 dimerization.

GOAL:

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

[Reed et al, Gen Res, 2003]

(1) PIPELINE CREATION

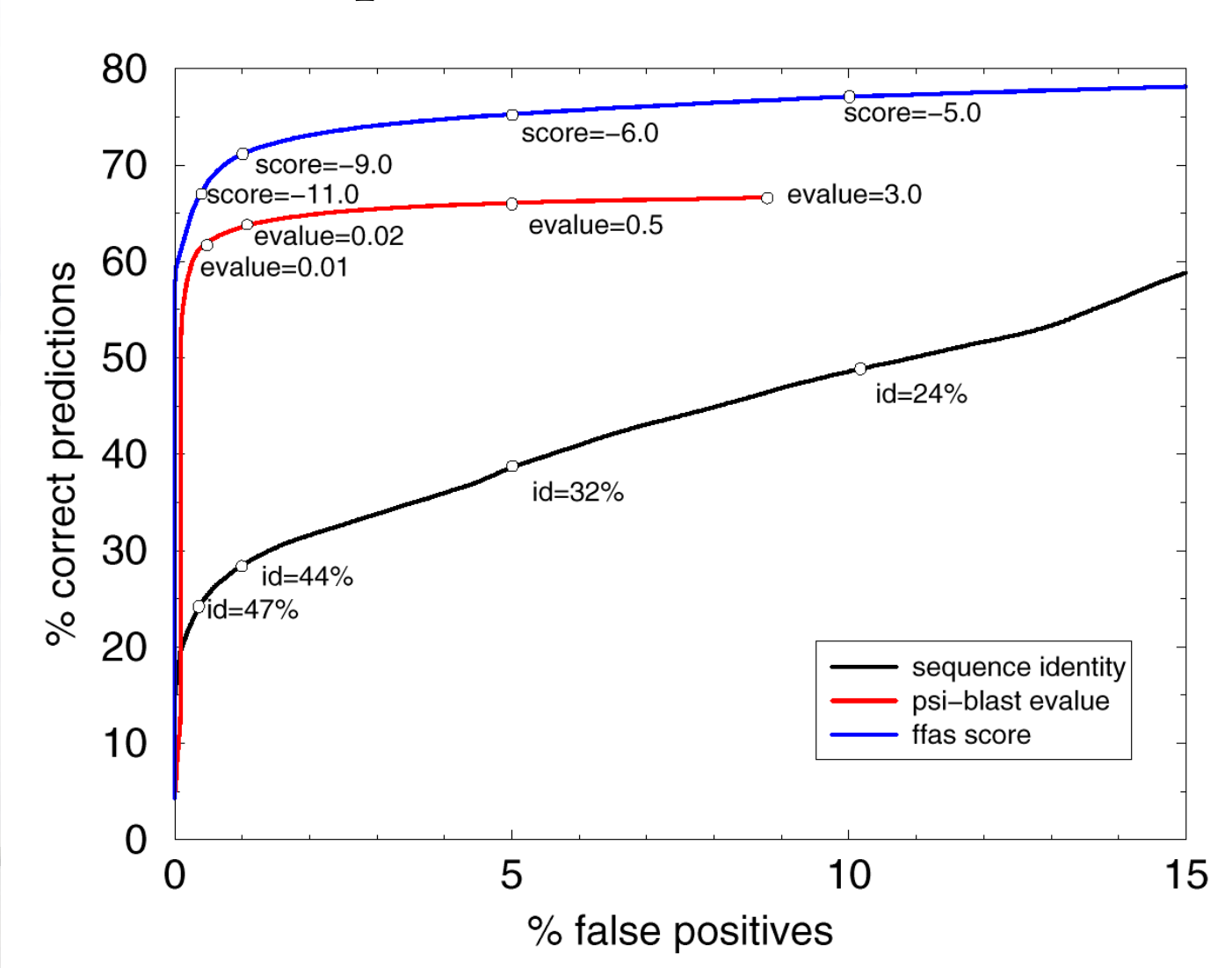


[Reed et al, Gen Res, 2003]



(1) PIPELINE CREATION

Fold prediction validation



[Reed et al, Gen Res, 2003]

Analysis of RIKEN clone collection <http://>

@ [The Burnham
Institute
Godzik Lab](#)

Mouse genome annotation development site

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 sequences 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.

(1) PIPELINE CREATION

Analysis of RIKEN FANTOM2 clone collection

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_card4 / mouse sequences \(blast\)](#)

[Set: arojas_CIITA / mouse sequences \(blast\)](#)

[Set: arojas_CLANA / mouse sequences \(blast\)](#)

[Set: arojas_cryopyrin / mouse sequences \(blast\)](#)

[Set: arojas_NAC / mouse sequences \(blast\)](#)

[Set: arojas_NAIP / mouse sequences \(blast\)](#)

[Set: arojas_NALP / mouse sequences \(blast\)](#)

[Set: arojas_NOD / mouse sequences \(blast\)](#)

[Set: cstehlik_card / mouse sequences \(blast\)](#)

[Set: gsalvesen_granzyme-B / mouse sequences \(blast\)](#)

[Set: izapata_TEF / mouse sequences \(blast\)](#)

[Set: arojas_apaf / mouse sequences \(pdb-blast\)](#)

[Set: arojas_card4 / mouse sequences \(pdb-blast\)](#)

[Set: arojas_CIITA / mouse sequences \(pdb-blast\)](#)

[Set: arojas_CLANA / mouse sequences \(pdb-blast\)](#)

[Set: arojas_cryopyrin / mouse sequences \(pdb-blast\)](#)

[Set: arojas_NAC / mouse sequences \(pdb-blast\)](#)

[Set: arojas_NAIP / mouse sequences \(pdb-blast\)](#)

[Set: arojas_NALP / mouse sequences \(pdb-blast\)](#)

[Set: arojas_NOD / mouse sequences \(pdb-blast\)](#)

[Set: cstehlik_card / mouse sequences \(pdb-blast\)](#)

[Set: gsalvesen_granzyme-B / mouse sequences \(pdb-blast\)](#)

[Set: izapata_TEF / mouse sequences \(pdb-blast\)](#)

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

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

Results: 1- 10994 of 10994

#	Query	Result vs.	Best score	%id	Best hit
1	ri A530099J19 FX00144G09 2774 seqid=45616 311..1210 Cat01	pdb	-95.200	14	lkad_A mol:protein length:360 Ccr2B
2	ri 6820446M14 FX00650M23 2778 seqid=66146 526..1374 Cat01	pdb	-89.100	43	laoi mol:protein length:228 Gp70
3	ri D330027J18 FX00192L12 3914 seqid=57248 520..1368 Cat01	pdb	-89.100	43	laoi mol:protein length:228 Gp70
4	ri D430035L01 FX00194P21 4391 seqid=51143 493..1341 Cat01	pdb	-89.100	43	laoi mol:protein length:228 Gp70
5	ri G431002I12 PH00003I01 8284 seqid=62999 2732..7797 Cat03 +5759	pdb	-87.200	17	irvi_A mol:protein length:556 Reverse Transcriptase Non-Nucleoside Binding

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

Results: 1- 249 of 249

#	Query	Result vs.	Best score	%id	Best hit
1	r1 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	r1 0610008C05 R000001B12 1232 seqid=7 77..1120 Cat01	human_genome	0.000e+00	71	gi 15990668 emb AL590548.2 AL590548 Human DNA sequence from clone RP11-299N6 on chromosome 20, complete sequence [Homo sapiens] /len=32047 (round:0)
3	r1 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 DGC Region, complete sequence /len=35739 (round:0)
4	r1 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	r1 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)

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.

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

Protein domain	Human	Mouse	Only in Mice
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	
I κ B	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 Group² and GSL Members,^{3,5} and Adam Godzik¹

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

Overall, 29 additional orthologs/paralogs not in public databases

0 novel mouse gene signature domains, compared to human

ANT2, ANT3, VDAC1, VDAC2, VDAC3, Beclin, BI-1, RTN4, etc.

Prediction of binding mode.

(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 bioinformatics tools, homology modeling to predict binding Interfaces.

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

(2) USE OF HOMOMOLOGY MODELING

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

WHERE IS THE PAAD DOMAIN?

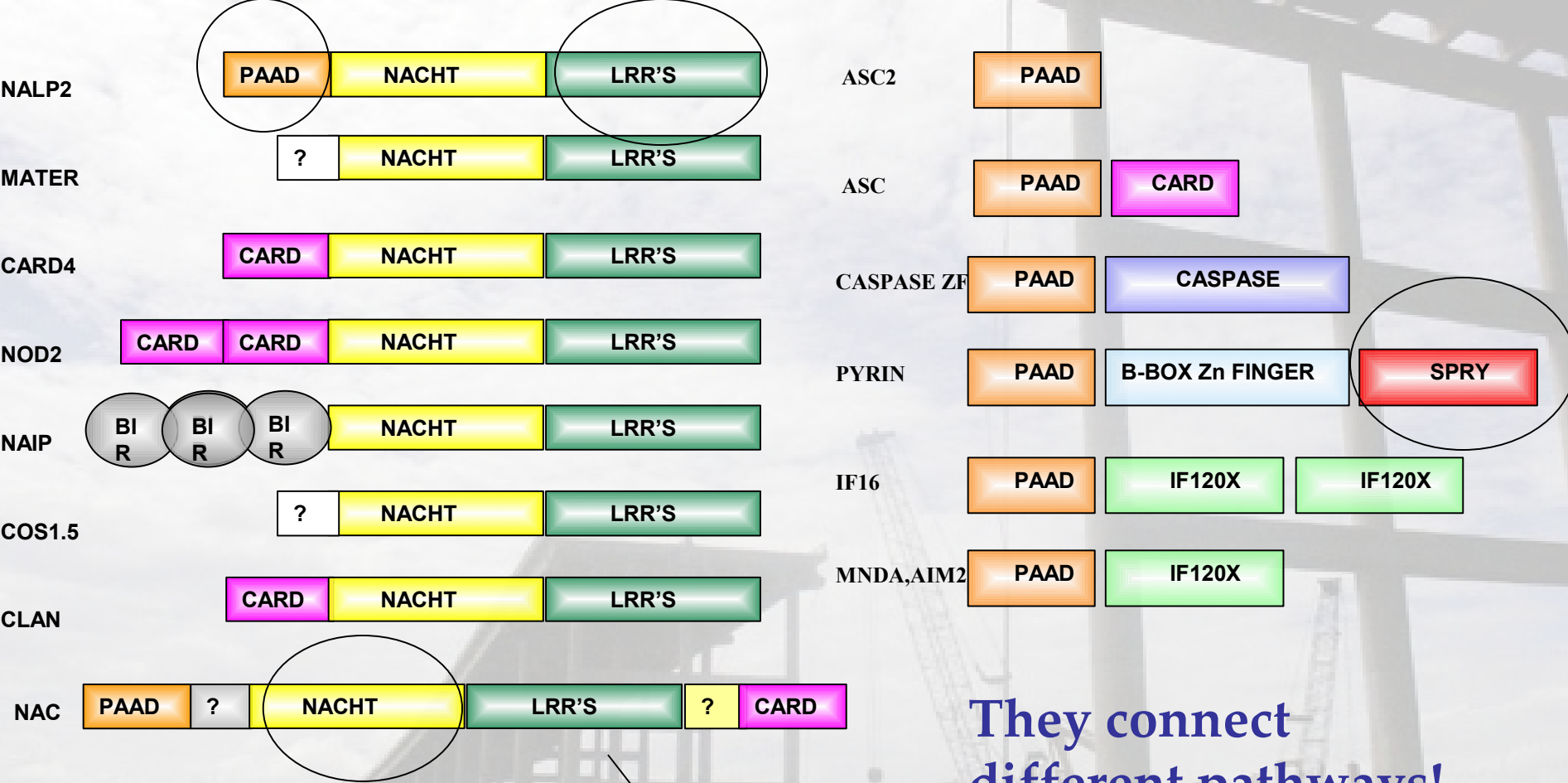
Nacht family: PAN/NALPs/DEFCAP/PYCARD,
CATERPILLER
(Tschopp et al, Nature, 2003)

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

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

(2) USE OF HOMOMOLOGY MODELING

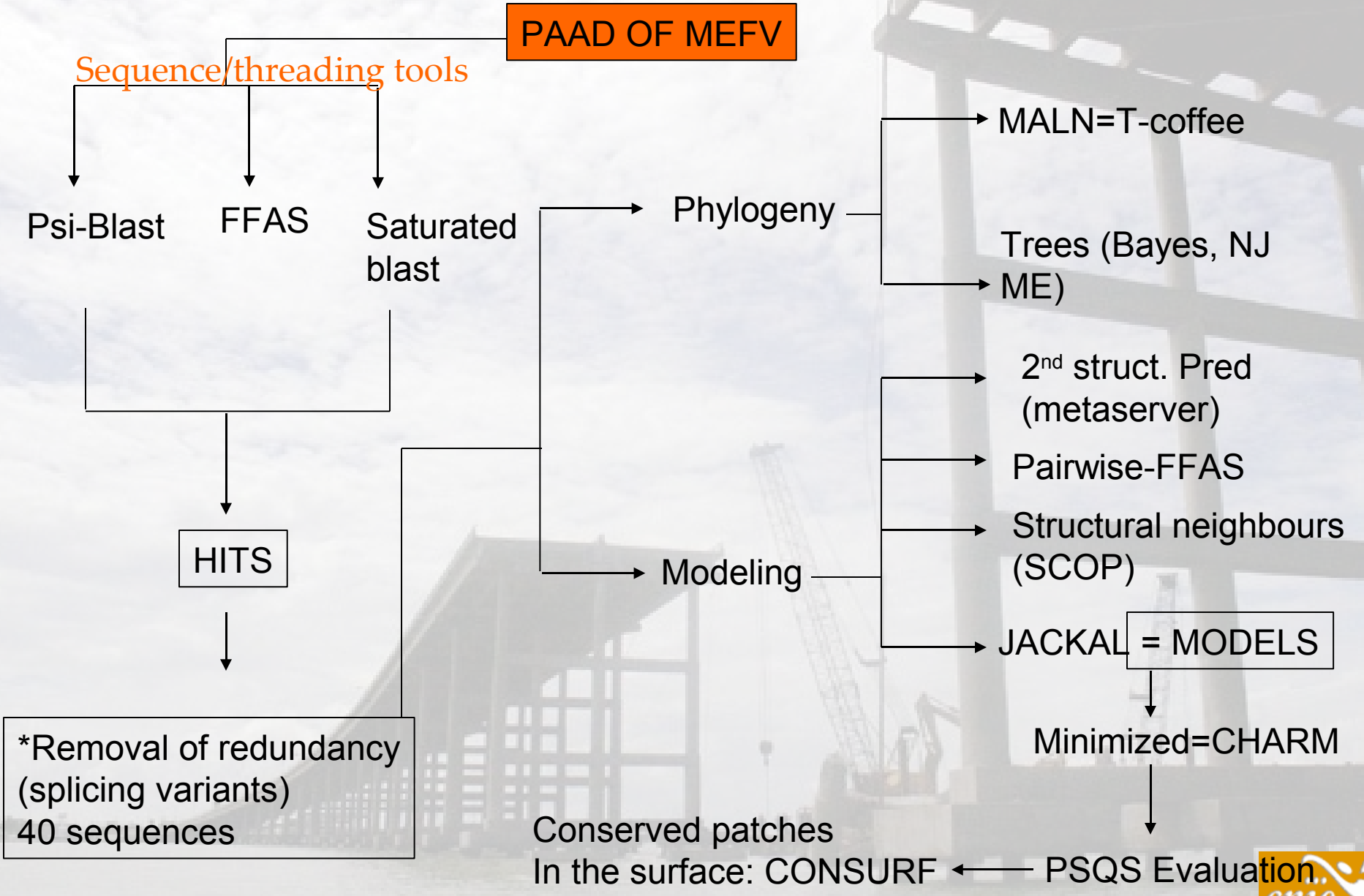
DOMAIN ARCHITECTURES



They connect different pathways!

Sensors!

(2) USE OF HOMOMOLOGY MODELING



*Removal of redundancy (splicing variants) 40 sequences

Conserved patches
In the surface: CONSURF

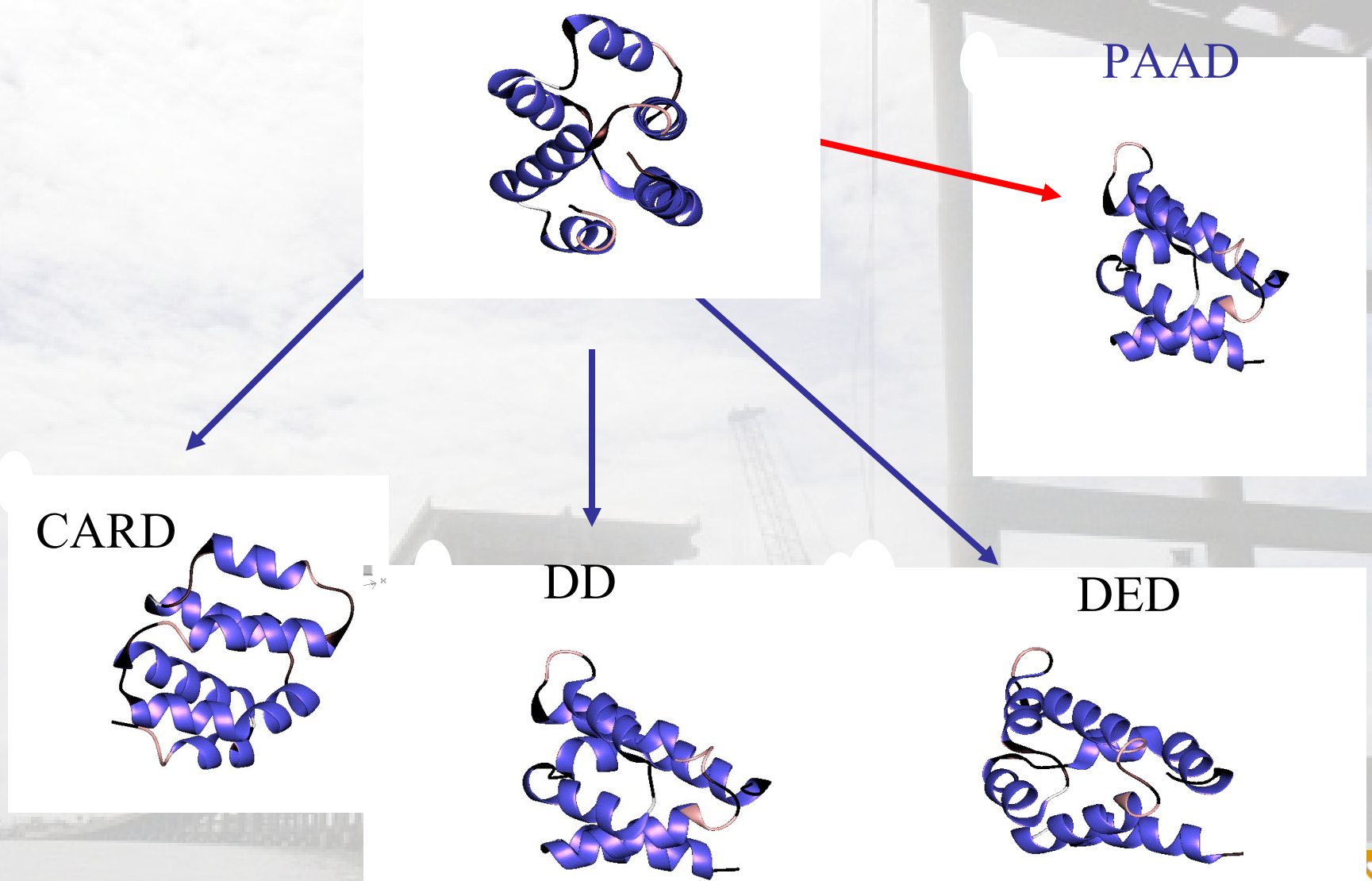
PSQS Evaluation

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



(2) USE OF HOMOMOLOGY MODELING

ANCESTORAL DOMAIN



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

Pyrin
 Sec_str
 MEFV_Mouse
 ASC_Human
 ASC-PENDING-Mouse
 PYC1_Human
 MEFV_Rat
 MEFV_Human
 AF427617_1_Human
 ASC1_zebrafish
 LOC280619_Mouse
 AF233434_1_Zebrafish
 AF327410_1_Zebrafish
 CAAB01003190_Fugu
 CAAB01007457_Fugu

PAN
 Sec_str
 PAN2_Human
 PAN3_Huamn
 PAN10_Huamn
 PAN4_Huamn
 PAN1_Huamn
 PAN7_Huamn
 PAN8_Huamn
 PAN11_Huamn
 PAN6_Huamn
 PAN5_Huamn

AIM
 Sec_str
 AIM2_Human
 AIM2_Mouse
 AIM2_Rat

IFI
 Sec_str
 If1204_Mouse
 If1203_Mouse
 MNDA_Human
 If116_Human
 If1205_Mouse
 LOC226690_Mosue
 LOC240922_Mouse
 LOC240921_Mouse
 LOC235882_Mouse
 M74124_Mouse

Virus
 Sec_str
 18L_Yaba Like Disease
 SPV014 Swinepox
 GP013L_Rabbit Fibroma
 M013L_Myxoma

1 2 3 4 5 6
 HHHHHHHHHHHH--HHHHHHHHH---HHHHHHHHH-----HHHHHHH-----HHHHHHHHH---HHHHHHH
 DHLNLTLEELLPYDFEKFKFKLQNTSLEKQHSKIPRGHMOMA-RPVKLASLLITYYGEEYAVRLTLQILRATNQRQLAEELR
 DALLDALENLTAEELKKFKLKLKLSVPLREGYGRIPRGALLSM-DALDLTDKLVSFYLETYGAELTANVLRDMGLQEMAGQLQ
 DALLDALENLSGDELKKFKMKLLTVQLREGYGRIPRGALLQM-DALDLTDKLVSYLSEYGLELTMTVLRDMGLQELAEQLQ
 EAILKLVLENLTPEELKKFKMKLGTVPRLREGFGRIPRGALGQL-DIVDLTDKLVASYYEDYAAELVAVLRDMRMLLEEAARLQ
 DHLNLTLEELLPYELEKFKFKLHTTSLEKQHSRIPLSLVKMA-RPICKLTRLLLTYYGEEYAVRLTLQILRATNQRQLAEELH
 DHLNSTLEELVPYDFEKFKFKLQNTSVQKEHSRIPRSQIQRA-RPVKMATLLVTYYGEEYAVQLTQVLRATNQRLLAEELH
 CKLARYLELEDVDLKKFKMHLLEDYPPQKGCIPLRGQTEKA-DHVDLTLMIDFNNGEEKAWAMAVWIFAAINRRDLYEKAK
 EHLQEFEDLGADNLRKFKSKLGD---RRQEPVTKSAIEKLKDEIDLADLVGVVTSKDAVSVTVEILRAIKCNAVADDLL
 EALLWALNDLENSFKTLKFHLRDVT---QFHLARGELESL-SQVDLASKLISMYGAQEA VRVVSRL LAMNLMELVDYLN
 DHLQDALSNI GADNLRRFQSR LGD---RKQEPVRKSTIEKLKDEIDLVDLVMTFTSD-AVSVTVDIRGKICNAVAEELL
 QLLSDVLEDLVEAELKFTQRLW-IGVKPGEVPIPRGKLQNA-DRQDVVDSMVQYQSE-AGTITVQTLRKIKQNERAKRLE
 --LLKILELDEKFKFTKFWYLT-LDLLENCNPIPRAHLDQA-SRIETVDKLLRSYSEETA VKITNEALRRMNMTKASEELM
 KLLKDFLDELDDTMLREFKWYLGQHK-ERGSRPIQRSQLENT-SRTEVTDKLVQAYGAEGAVVTTVDVLYRMRNLNDLATQL-

---HHHH-----HHHHHHHH---HHHHHHHH-----HHHHHHHHHHHHHHHH-----HHHHHHHHHH---HHHHHHHHH
 FGLMNYLEELKKEEFKFKKEHLKQITLQLELQKIPWTEVVKASREELANLLIKHYEEQQAWNITLRIQKMDRDKLCMKVMR
 ELLLALEELLSQEQLKFRHKLKRDYV--PDGRSIPWGRLEPADAVDLAEQLAQFYGPEPALEVARKTKRADARDVAAQLQE
 FDLLWYLENLSDKFQSFKKYLARKIL---DFKLPQFPLIQMTKEELANVLPISYEGQYIWNMLFSIFSMRREDLCKRIIG
 NGVMLYMRNVSHEELQRFKQLLTFE--LSTGTMPITWQVETASWAEVWHLLIERFPGRRAVDVTSNIFAIMNCDKMCVVVRR
 FNLQALLEQLSQDELSKFKYLITTSLAHELQKIPHKVEVDKADGKQVLEILTTCHDSYVWEMASLQVFEKMHMRDLSERAK
 WTLQTLLEQLNDELKSKFSLWAPLEDVLRKTPWSEVEADGKLAELVNTSSENWIRNATVNIIEEMNLTELCMKMA
 FGLLYLEELNKEELNTFKLFLKE-TMEPEHGLTPWNEVKKARREDLANLMKKYYPGEKAMSVSLKIFGKMNLKDLKERAKE
 YGLQWCLYELDKKEFQTFKELLKKSSESTTCSIQFIEENANVECLALLHEYYGASLAWATSIIFENMNLRTLSEKARD
 CRLSTYLEELEAVELEKFKLYLGT-T-ELGEGKIPWSEMEKAGPLEMAQLLITHFGEPAWRLALSTFERINRKDLWERGQR
 EALLWALSDLEENDFKKLFYLRDITLSEGGPPLARGELEGLIPVDLAELI-SKYGEKEAVKVVLKGLKVMNMLELVDQLSH

HHHHHHHH--HHHHHHHHHH---HHHHH---HHHHHHHH-----HHHHHHHH---HHHHHHHH
 ILLTGLDNITDEELDRKFFFLSDTFNIIATGKLHTANRIQVATLMIQAGAVSAVMKTIIRIFQKLNVMLLAKRLQE
 MLLTGLDHITEEELKRFKYFALTFQIARSTLDVADRTLEADHLIQSAGAASAVTKAINIFQKLNVMHIANALEE
 MLLTGLDHITEEELKRFKYFALTFNIPRKTNLNIADRTLEADQLIQSAGAASAVAKAISIFQKLNVMYDIKALEE

HHHHHHH---HHHHHHHHHH---HHHHH---HHHHHHHH-----HHHHHHHH-----HHHHH
 IVLLRGLLECINKHYFSLFKSLLARLNLERDNQEQYTTIQIANMMEEKFPADSGLGKLEFCEEVPALKR-AEILKK
 IVLLKGLLENMEDIYQFRTVKSLLRKLKLTKKMQEDYDRIQADWVEDKFPKDAGLDKLIKVCEHIKDLL-LAKKLT
 ILLKGFELMDDYHFTSIRKSLLAYLGLTTKMQEEYMRIRKITDLMEKKFQGVACLDKLLELAKDMPSLK-LVNNLRK
 IVLLKGLEVINDIYHFRVKSLLSNLKLNLKMRREYDKIQIADLMEEKFRGDAGLGKLIKIFEDIPTLEDLAETLKK
 IVLLRGLLECINKHYFSLFKSLLARLNLERDNQEQYTTIQIANMMEEKFPADSGLGKLEFCEEVPALRKRAEILKK
 IVLLSGLEYMNDYMFRAKSLLNHLKLTKNMQDDYDRINLADLMEEKFPEDAGLSKLEFCEVCEIPELAAARVDILRK
 IVLLTGLMGINDHDFRNVKSLLSKLKLNR-MQDQYDVRKIADLMEDKFPKDAVDQLIKLYKQIPGLD-IANKLN
 IVLLSGLEYMNDYMFRAKSLLNHLKLTKNMQDDYDRIKIADLMEEKFPEDAGLSKLEFCEVCEIPELD-HVDILRK
 IVLLEGLENMEDIYQFRTVKSLLRKLKLTKKMQEDYDRIQADWVEDKFPKDAGLDKLIKVCEHIKDKLDLAKKLT
 LVLLEGLELECINKHQFNLFKSLMVKLNLEEDNQEYTTIQIANMMVKKFPADAGLDRLINFCERVPTLKKRAEILKK

-HHHHHH--HHHHHHHHHH---HHHHH---HHHHHHHH-----HHHHHHHH---HHHHHHHH
 SAILFSLVEDVTHYQFKILIFLTKDLNLSDEEKQILDVDFAEKLFQYTPGIKSLYFLEKALSMVNPNAKYARSNIN
 YTIISVLERLTPYQFKTLFLIQDINISNDDINVLDVLDLAIKIMNKYNNYRAITFYLKYLIRINTE-YISGTLQ
 GVIIITVLENLTDYQFKMFLYLVTDLRINPVEKEKIDRIDLAYKISELYPGHSYIEFMKQVYIPNKKHWDSLLK
 GVIIITVLENLSDYQFKMFLYLAMEPLYIERAEKEKIDRIDLAHKIISQYLGTDYIEFMKRVYIPNKKHWDSLLK

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

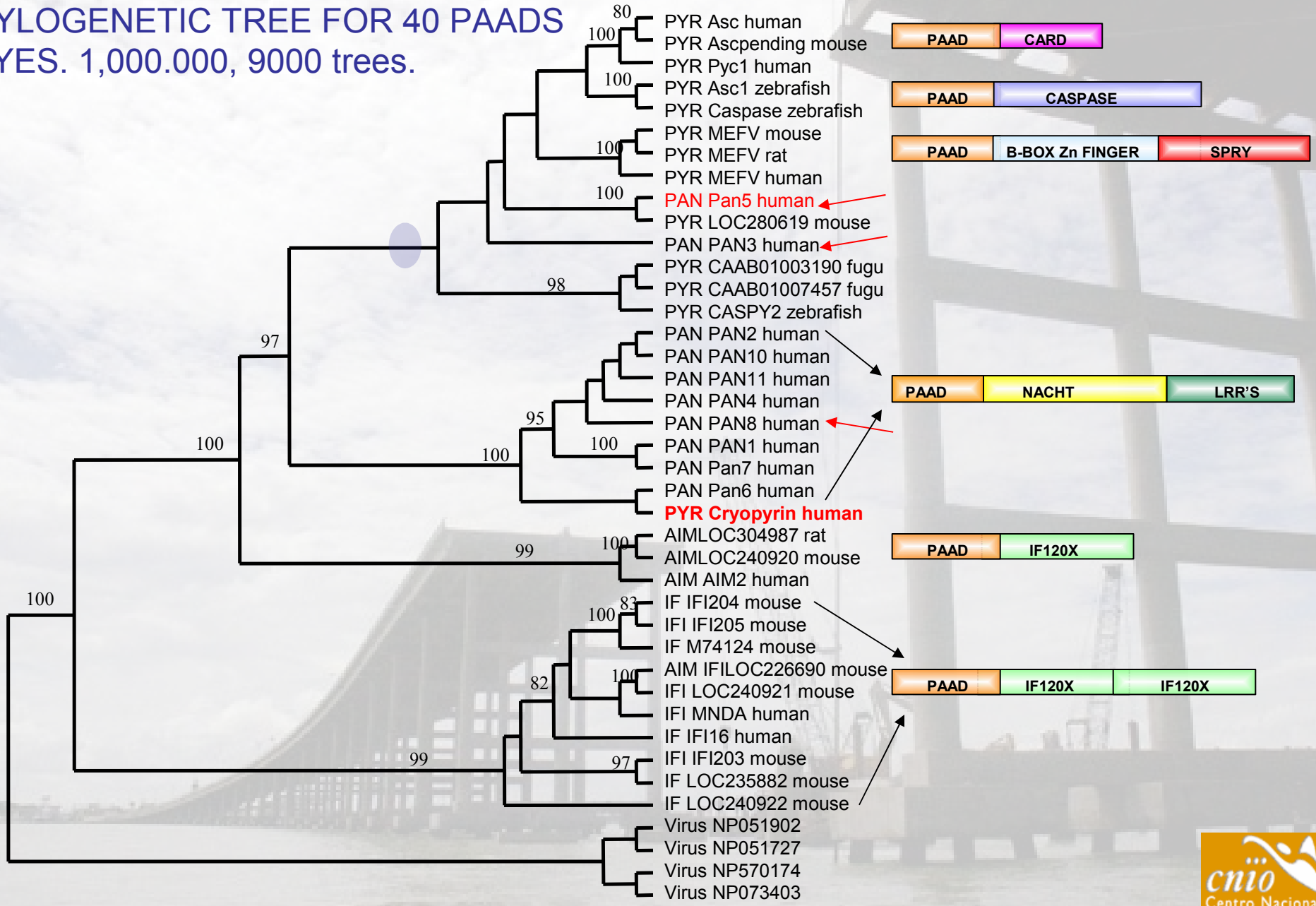
Sec. Structure Prediction

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



(2) USE OF HOMOMOLOGY MODELING

PHYLOGENETIC TREE FOR 40 PAADS
 1,000,000, 9000 trees.

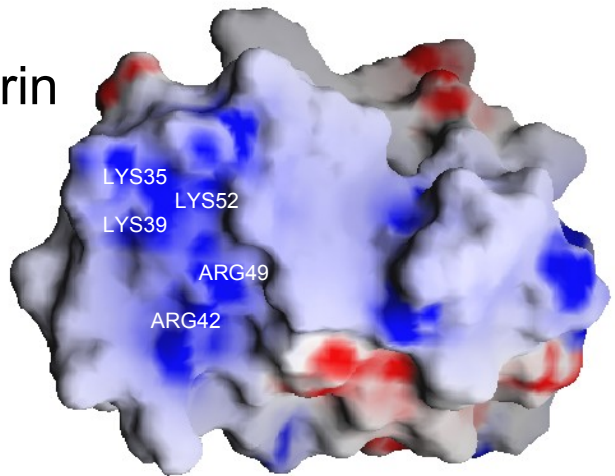


[Liu,T et al, Prot Sci, 2003]

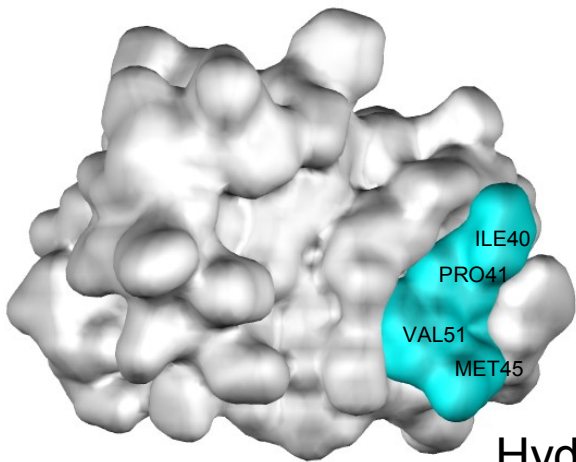
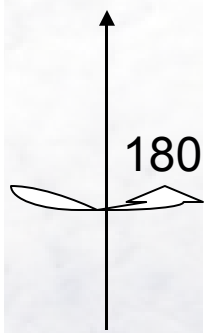


(2) USE OF HOMOMOLOGY MODELING

pyrin

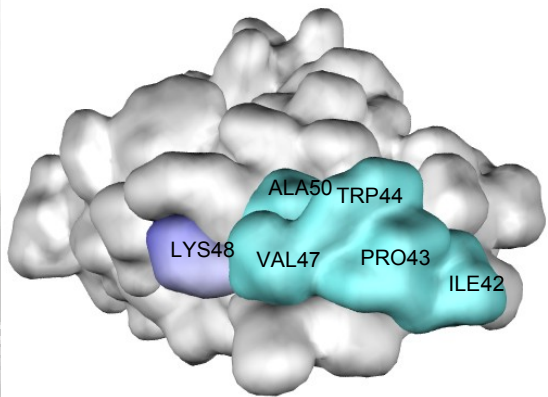


Charged patch



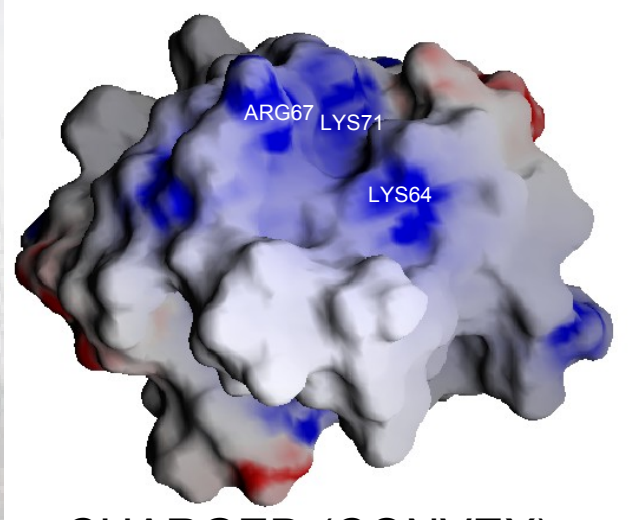
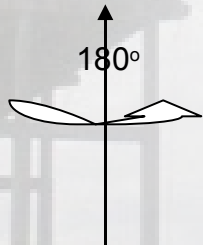
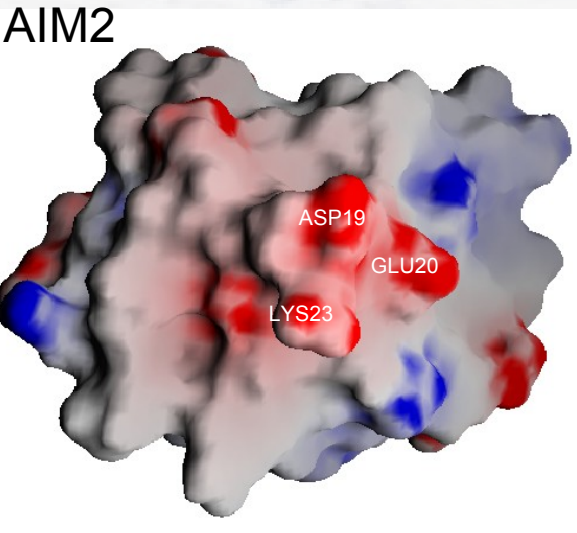
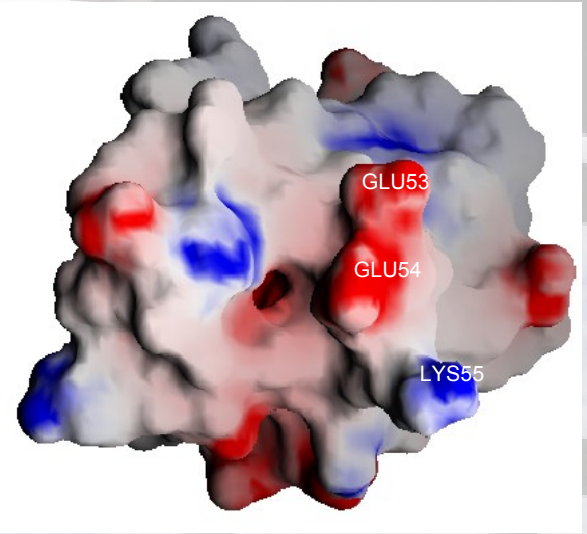
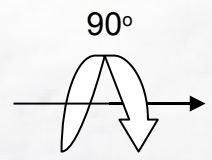
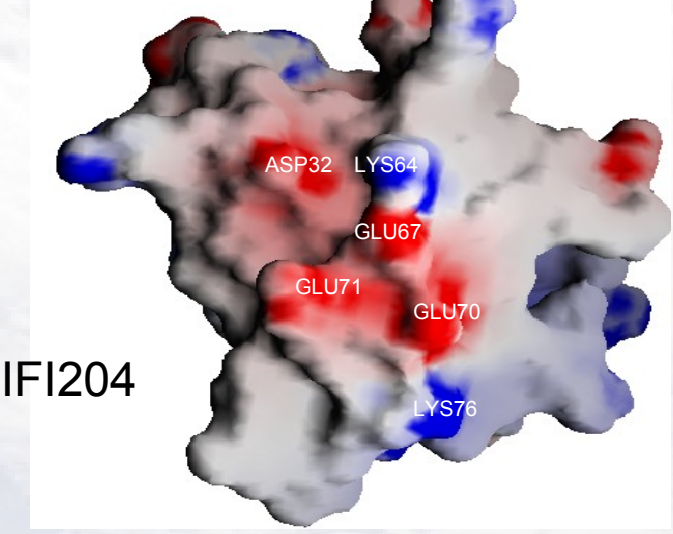
Hydrophobic patch

Pan2/NALP4



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

(2) USE OF HOMOMOLOGY MODELING



- CHARGED (CONCAVE)

+ CHARGED (CONVEX)

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

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

IFI6

PAAD

IFI120X

IFI120X

MNDA,AIM2

PAAD

IFI120X

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”**

(3) USE OF HOMOMOLOGY MODELING

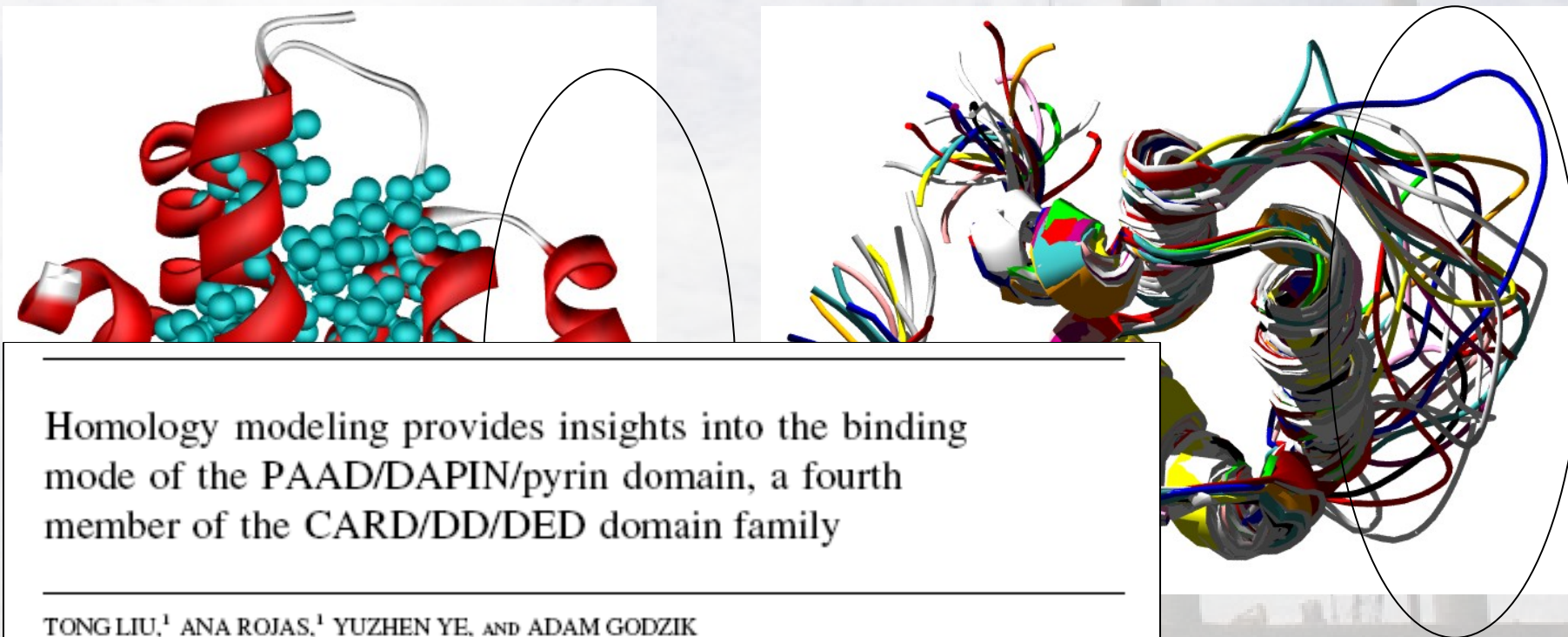
PAAD is a 6 alpha helical bundle

Helix 3 is disordered

Binding patches correctly predicted

Real structure 1PN5

Released October 2003



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]

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:

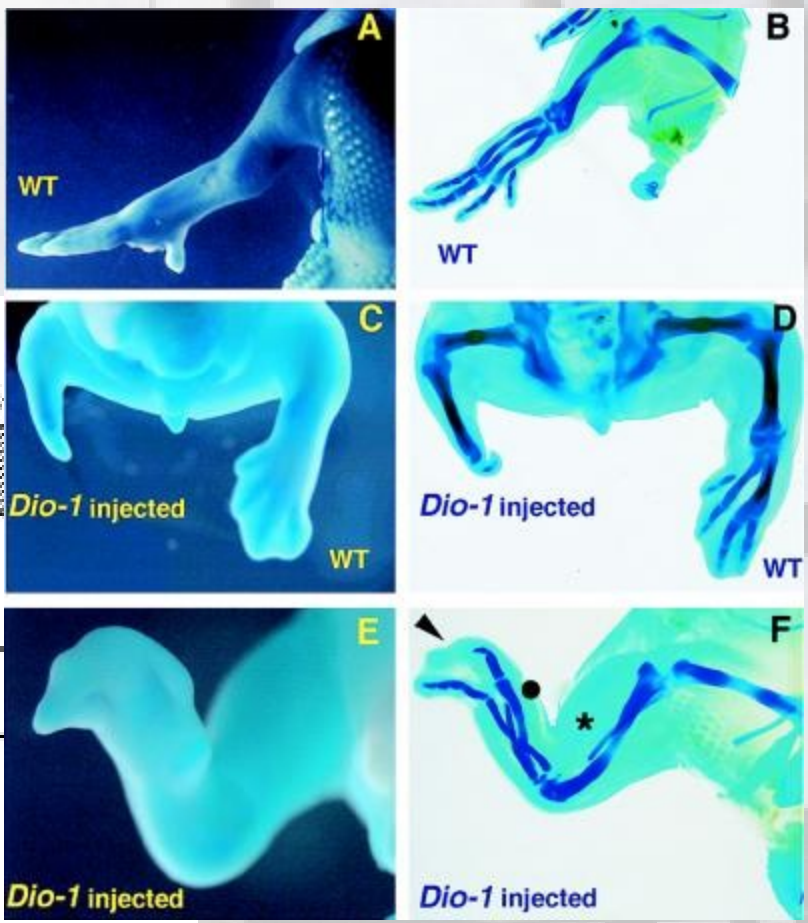
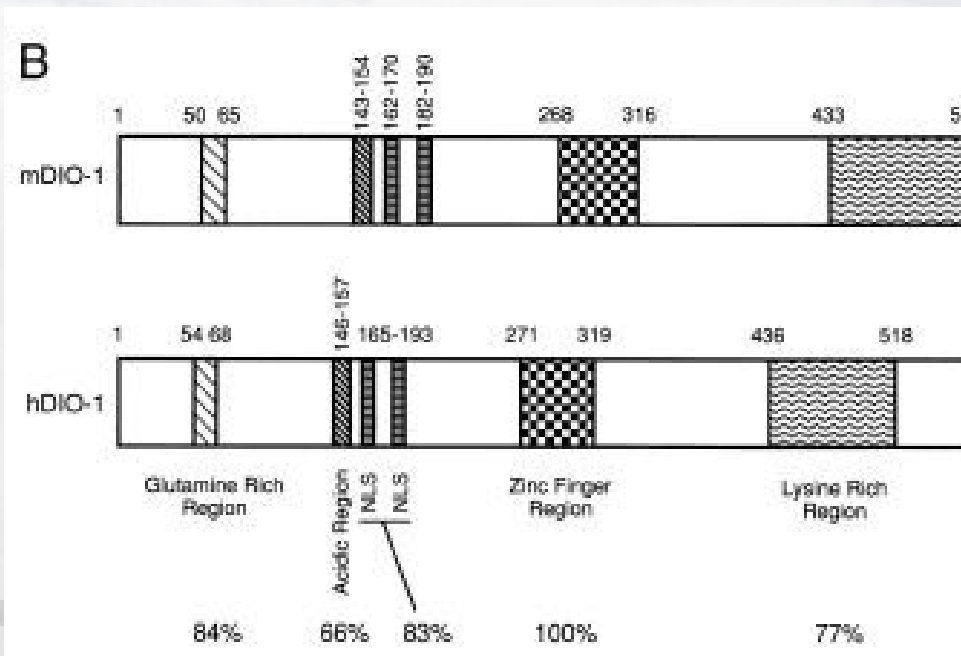
Use domain focused sequence searching to get insights into novel Function.

[Rojas, et al, FEBS J, 2005]

(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]

(3) DIDO-1 PROTEIN ANALYSES

WHAT IS KNOWN: DEATH ASSOCIATED FACTOR GENE

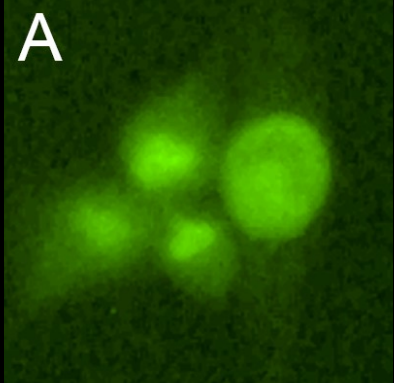
INVOLVED IN APOPTOSIS (Garcia-Domingo et al., 2003)

- **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-1 Δ NLS is a dominant negative mutant that protects cells from apoptosis.**

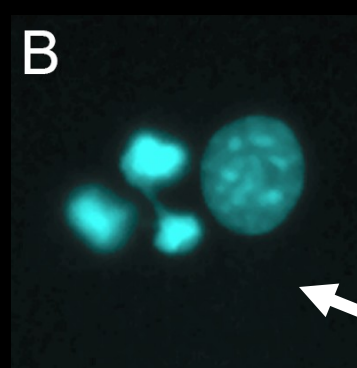
[Rojas, et al, FEBS J, 2005]

(3) DIDO-1 PROTEIN ANALYSES

ADDITIONAL EXPERIMENTAL DATA

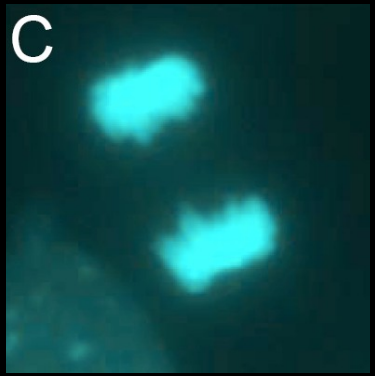


DIO-1 is present in mitotic chromosomes

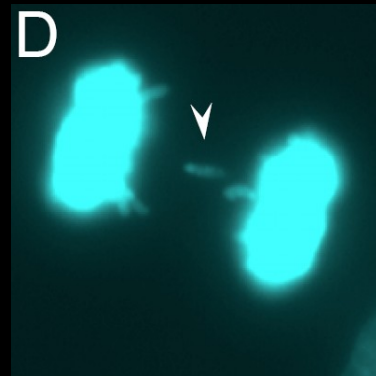


Mitosis on DIO overexpressed-cells

Asymmetric divisions!



Normal anaphase



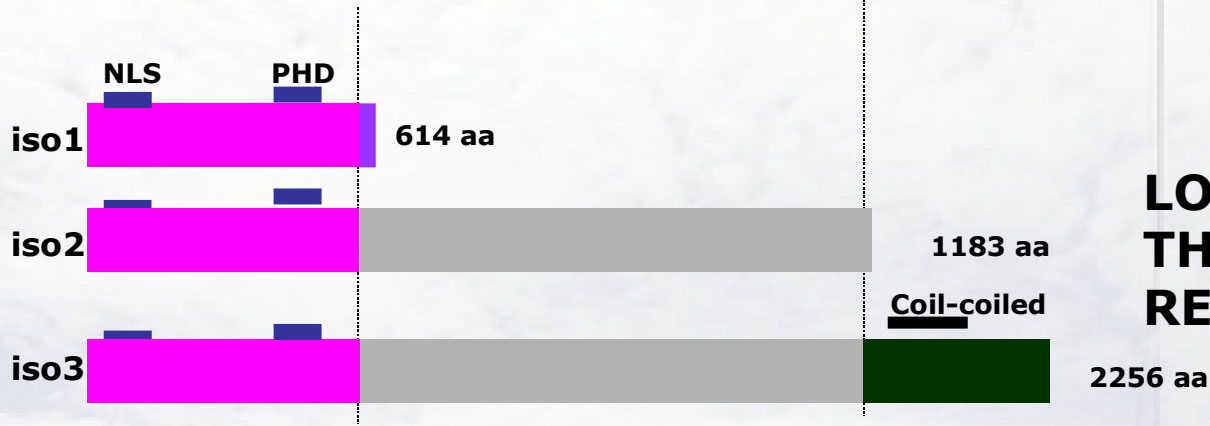
DIO-targeted cells show abnormal anaphases: lagging chromosomes

TARGETED MICE SHOW SEVERE SUB-FERTILITY!!

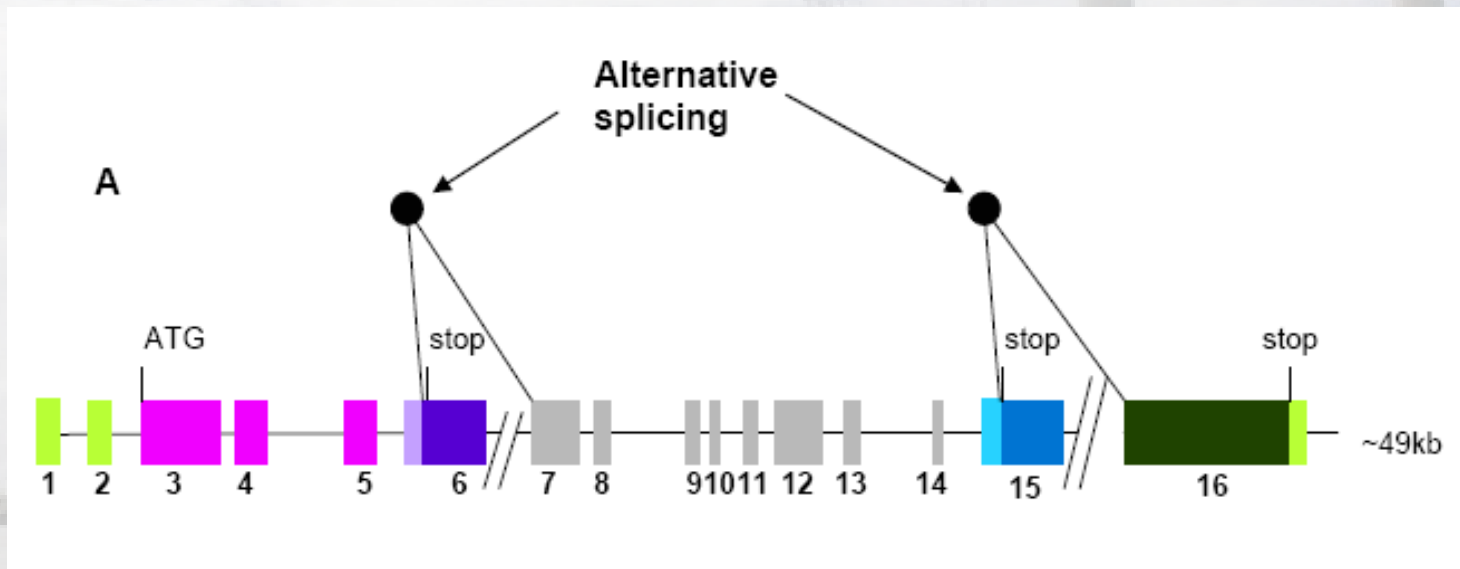
[Rojas, et al, FEBS J, 2005]

(3) DIDO-1 PROTEIN ANALYSES

The gene contains 3 splicing variants

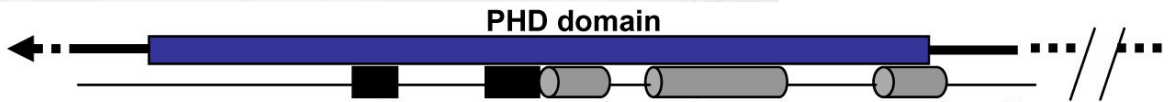
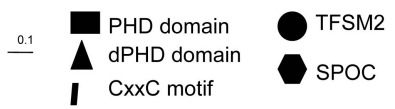
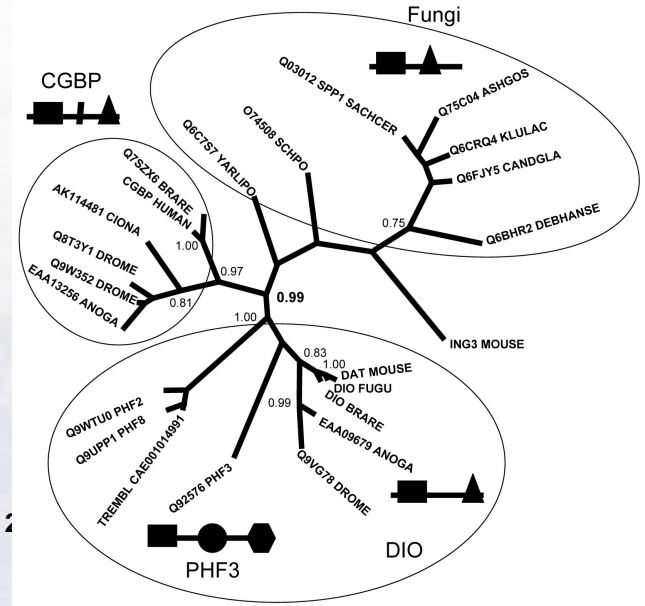
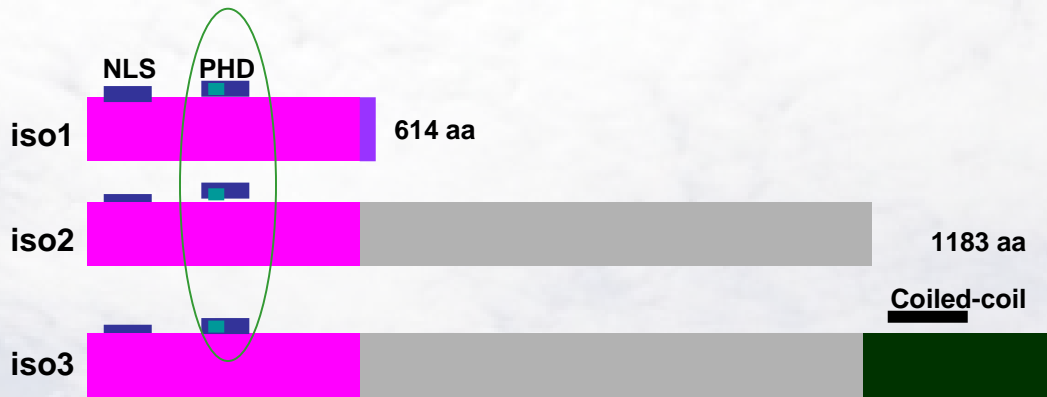


LONG PARTS OF THE PROTEIN REMAIN UNCOVERED!



[Rojas, et al, FEBS J, 2005]

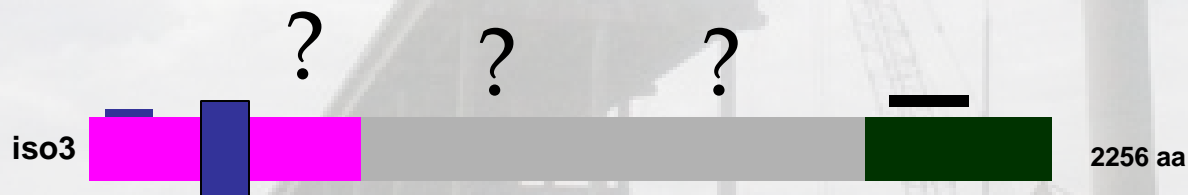
(3) DIDO-1 PROTEIN ANALYSES



Accession	Sequence	Length
DAT1_HUMAN/265-462	: DPNAIYICIRQPHNNRFMICCDRC [*] EEWFHGD ²⁰ CVGI [*] SEARGRLDERNGEDYICPNCTILQV	80
DAT1_MOUSE/262-423	: DPNAIYICIRQPHNNRFMICCDRC [*] EEWFHGD ²⁰ CVGI [*] SEARGRLDERNGEDYICPNCTILQV	80
Dio1est_Chick/274-434	: DPNAIYICIRQPHNNRFMICCDRC [*] EEWFHGD ²⁰ CVGI [*] SEARGRLDERNGEDYICPNCTILQG	80
Dio_Fugu/206-393	: DPNAIYICIRQKHNKRFMICCDRC [*] EEWFHGD ²⁰ CVGI [*] TEARGRLMERNGEDYICPNCTTKKN	80
Dio_Brare/234-402	: DPNAIYICIRQKHNKRFMICCDRC [*] EEWFHGD ²⁰ CVGI [*] TEARGRLMERNGEDYVCPNCTYQKG	80
EAA09679_Anoga/391-524	: DPDRLWICIRQPHNNRFMICCDSC [*] EDWFHGKCVNI [*] TKAMGQQMEQDGI ⁴⁰ EWTCPNCLK.KK	80
Q9VG78_Drome/907-1094	: DPNKLWICIRQPHNNRFMICCDL [*] CEDWFHGT ²⁰ CVGV [*] TKAMGTD ⁴⁰ MENKIDWKC ⁶⁰ PKCVKRQE	79
Q03012_Sacce/19-126	: TGEDVYICIRKRPDYGELMVGCDG [*] CDDWFHFT ²⁰ CLH ⁴⁰ IP ⁶⁰ QFKDLV ⁸⁰ ...SFYCPYQ...	80
O74508_Schpo/115-208	: HQRPLIYICIQKPD [*] DGSWMLG ²⁰ CDG [*] CEDWFHGT ²⁰ CVNI [*] PSYNDLTV...QYFCPKCTEE..	52
CGBP_HUMAN/23-404	: ENAP [*] IYICIRKPD [*] INCFMIG ²⁰ CDN [*] CNEWFHGD ²⁰ CIRI ⁴⁰ TEKMAKATR...EWYCRECREKDP	54
CGBP_MOUSE/23-408	: ENAP [*] IYICIRKPD [*] INCFMIG ²⁰ CDN [*] CNEWFHGD ²⁰ CIRI ⁴⁰ TEKMAKATR...EWYCRECREKDP	76
Q7SZX6_Brare/22-311	: ENAP [*] IYICIRKSD [*] INCFMIG ²⁰ CDN [*] CNEWFHGH ⁴⁰ CINV ⁶⁰ TEKMAKATR...EWYCCQCRARDP	76
GB_AK114481_Ciona/5-319	: EDAT [*] EYICIRSS [*] DIDRFMIM ²⁰ CDE [*] CEWFHGD ²⁰ CIRI ⁴⁰ SEKEARYIK...YFYCKTCQGKNP	76
EAA13256_Anoga/23-223	: QDGA [*] YICIRSS [*] DSSRFMIG ²⁰ CDA [*] CEEWYHGD ²⁰ CINV ⁴⁰ SEKEAKHK...HYYCQRCKEEDP	76
Q9W352_Drome/57-302	: QEDQ [*] AYICIRSS [*] DCSRFMIG ²⁰ CDG [*] CEEWYHGD ²⁰ CIGI ⁴⁰ TEKEAKHK...QYYCRRCKKENP	76
	5C6C M6 CD C W5Hg C6 6 e 5 C C	76

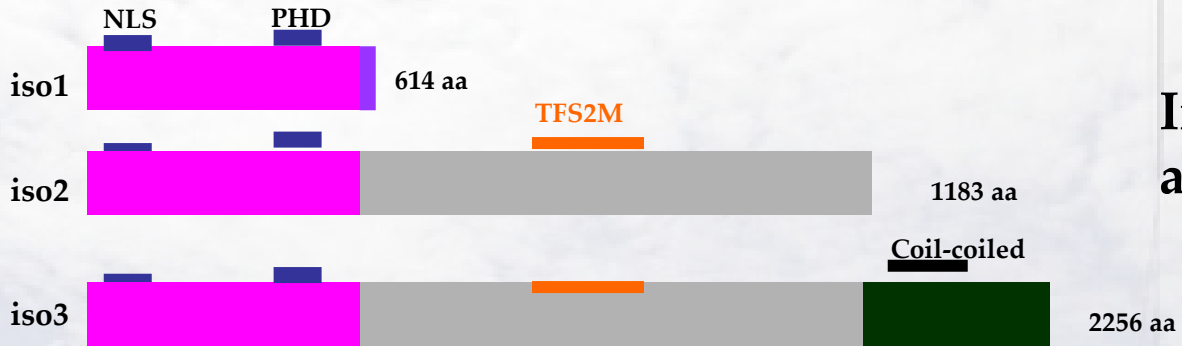
Then...

WHAT ELSE CAN BE FOUND IN THESE UNCOVERED REGIONS?

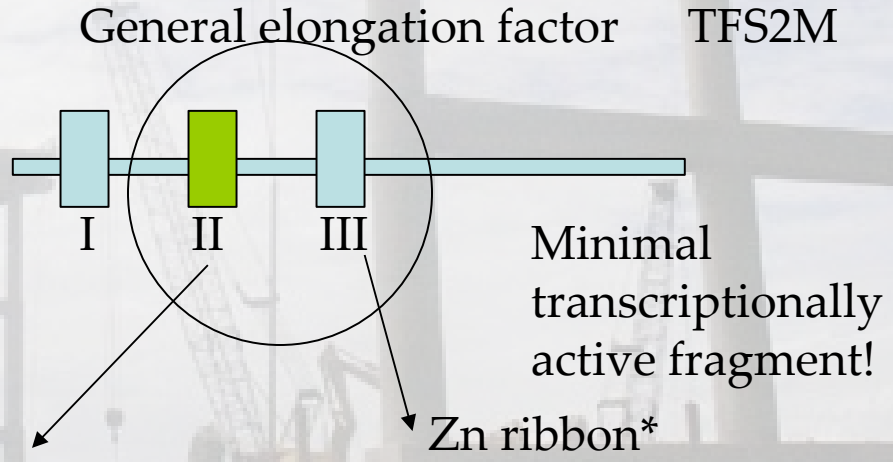
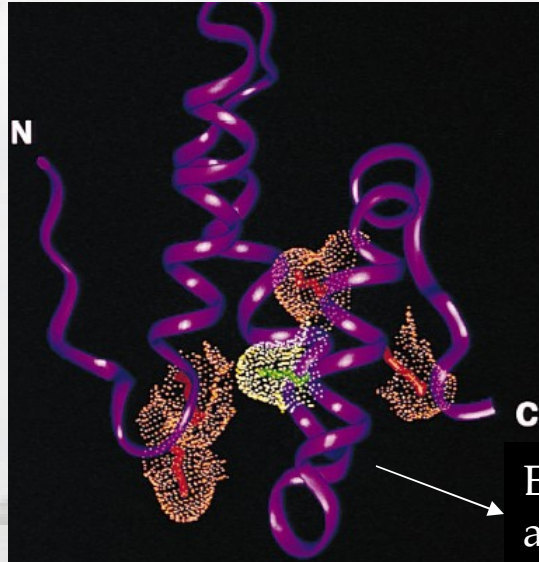


(3) DIDO-1 PROTEIN ANALYSES

METHODS AUTOMATIC SEARCHES



**Involved in DNA Binding!
a regulatory domain**



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

[Rojas, et al, FEBS J, 2005]

(3) DIDO-1 PROTEIN ANALYSES

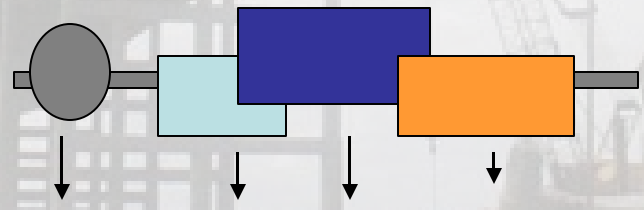
METHODS: manual **Selecting regions**



Multiple alignment
T-COFFEE,
MUSCLE, etc

TO ENRICH PROFILE!

PROFILE BUILDING



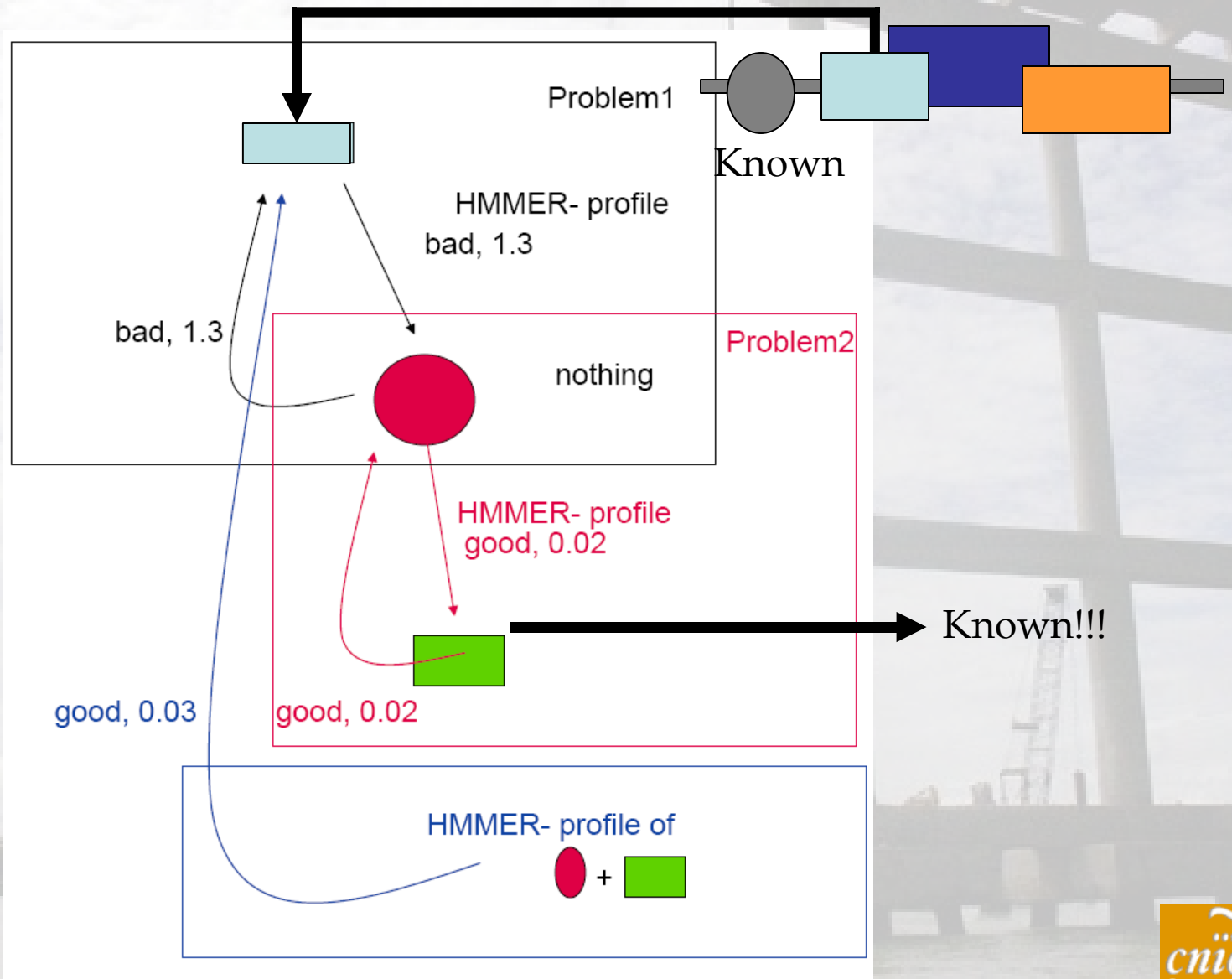
HMMER/PSI-BLAST SEARCHES in Uniprot90



(3) DIDO-1 PROTEIN ANALYSES

METHODS: HMMER strategy

Intermediate profile searching!

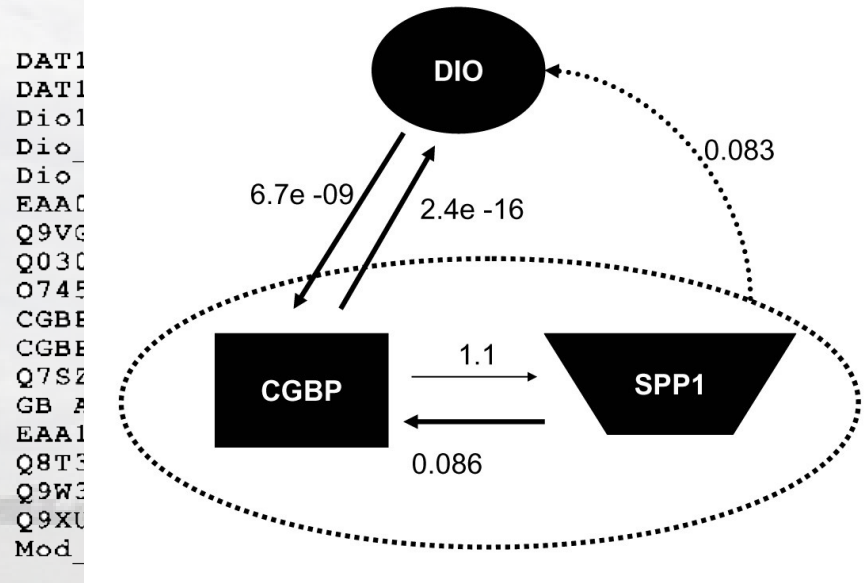
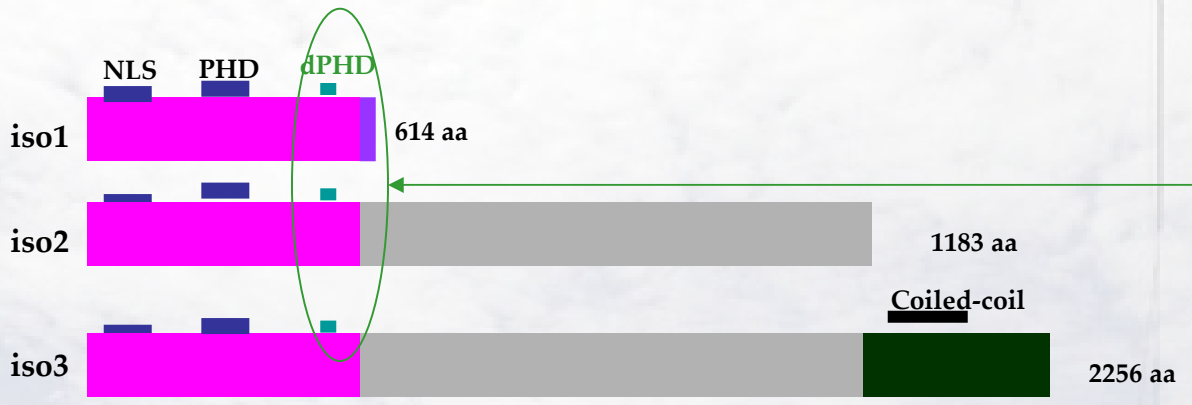


[Rojas, et al, FEBS J, 2005]



(3) DIDO-1 PROTEIN ANALYSES

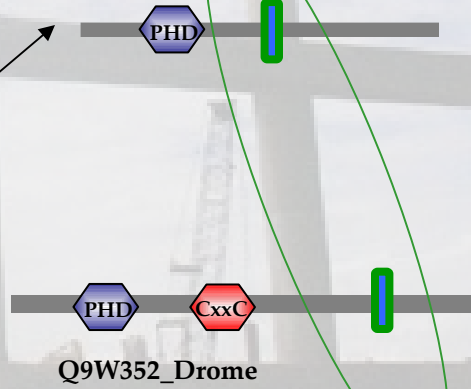
METHODS HMMER ANALYSES I



```

20      *
CSND CILKHA AATM
CSND CILKHA AATM
CSND CILKHA AATM
CGND CILRHA AAAA
CGND CILRHA AAAA
CSDECIRKHA SSTI
CGEE CIRKYA QSAL
CSEEHGREFVNDIW
CSDKHGVDFFREKV
CSDDCGMKLANRI
CSDDCGMKLANRI
CSEDCGMKLANRI
CSHECGRLLARNRL
CSDECGMKLA TSRI
CSDECGSNLA IERL
CSDKCGFNLA TKRI
CSDEC GKELAR MRL
CSDEPCGLILA KMRL
Cs c a
    
```

SPP1:
Chromatin
stability

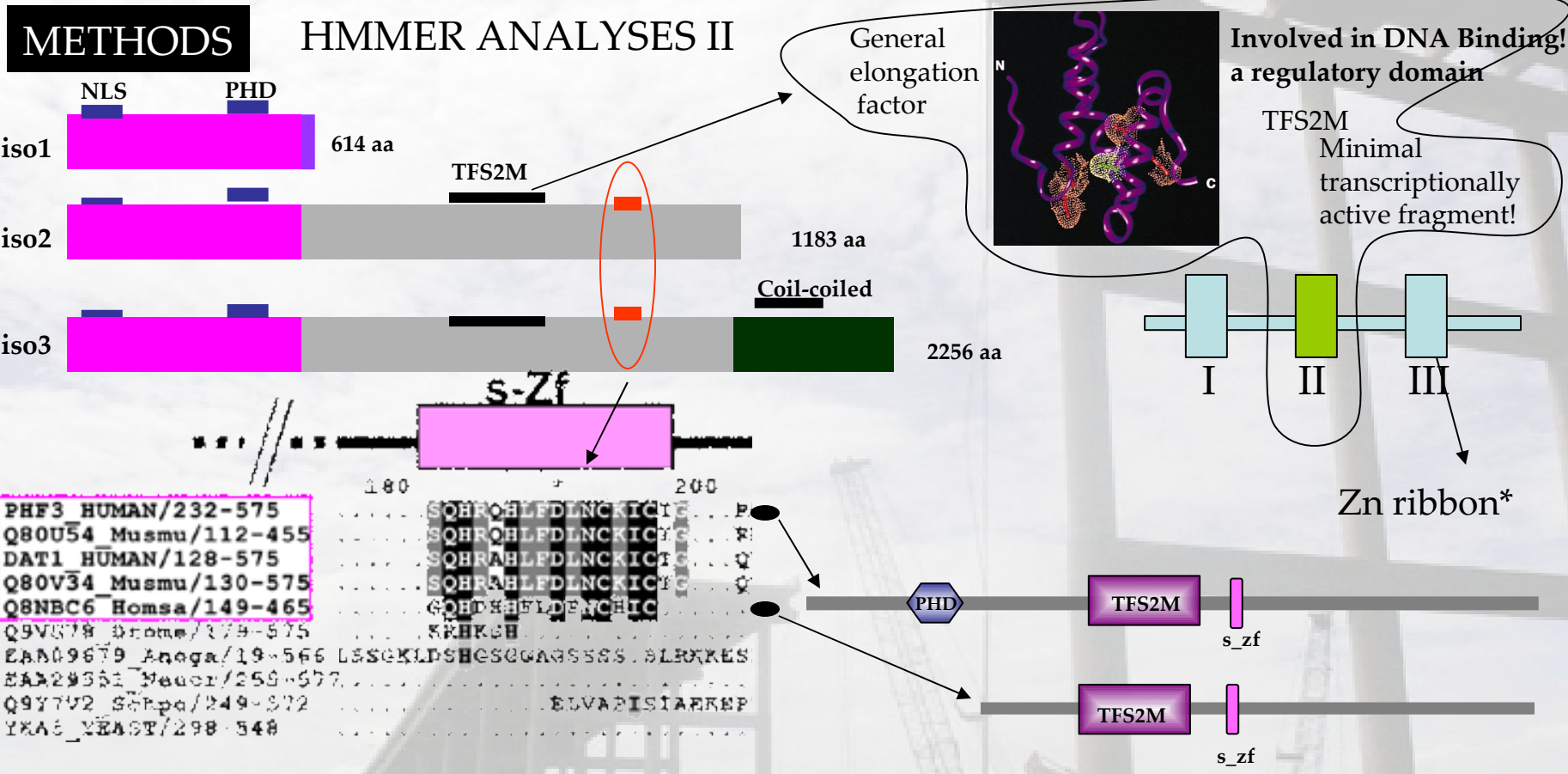


Q9W352_Drome
CGBP: DNA
binding!

[Rojas, et al, FEBS J, 2005]



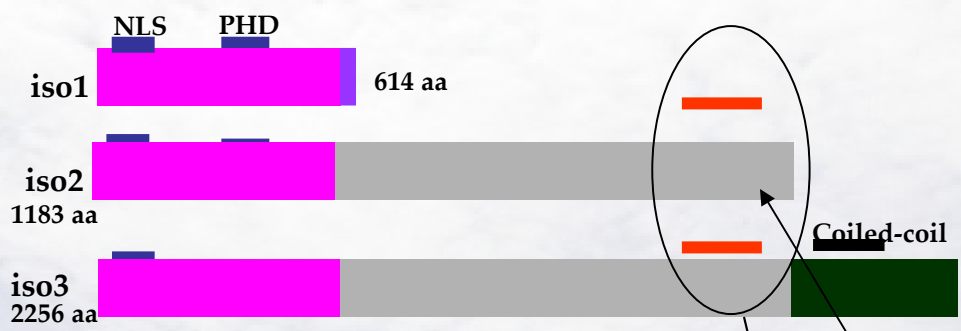
(3) DIDO-1 PROTEIN ANALYSES



[Rojas, et al, FEBS J, 2005]

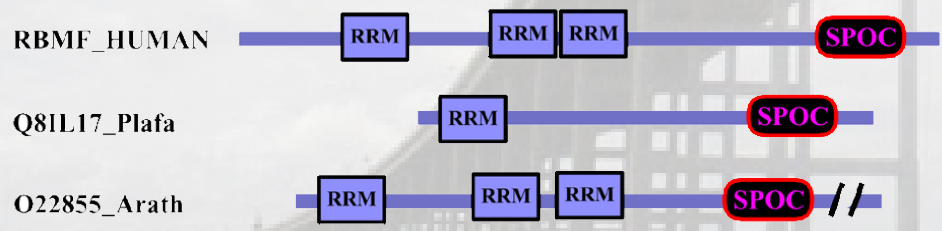
(3) DIDO-1 PROTEIN ANALYSES

METHODS HMMER ANALYSES III



SPOC: Protein-protein interaction

0.083 0.05



(3) DIDO-1 PROTEIN ANALYSES

METHODS

HMMER

BMC Bioinformatics



Research article

Open Access

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

Luis Sánchez-Pulido*¹, Ana M Rojas¹, Karel H van Wely², Carlos Martínez-A² and Alfonso Valencia¹

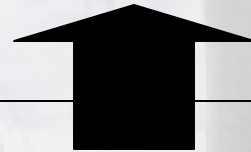
Address: ¹Protein Design Group, Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, E-28049 Madrid, Spain and ²Department of Immunology and Oncology, Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, E-28049 Madrid, Spain

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

* Corresponding author



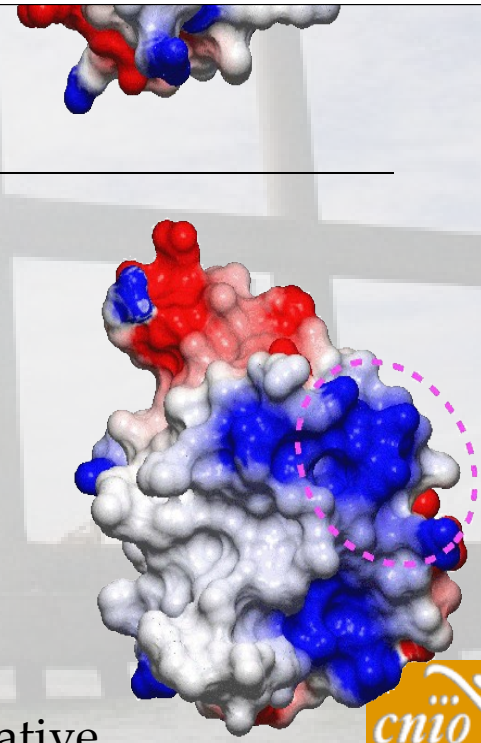
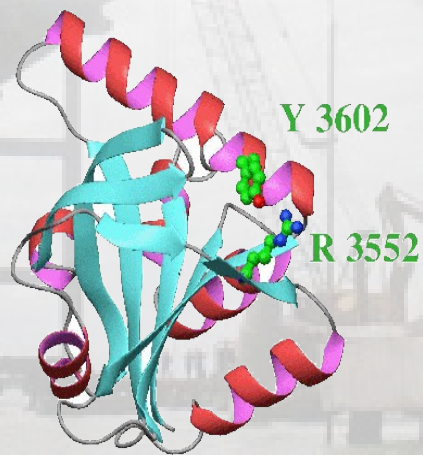
Model



SPOC: Protein-protein interaction



RBMF_HUMAN



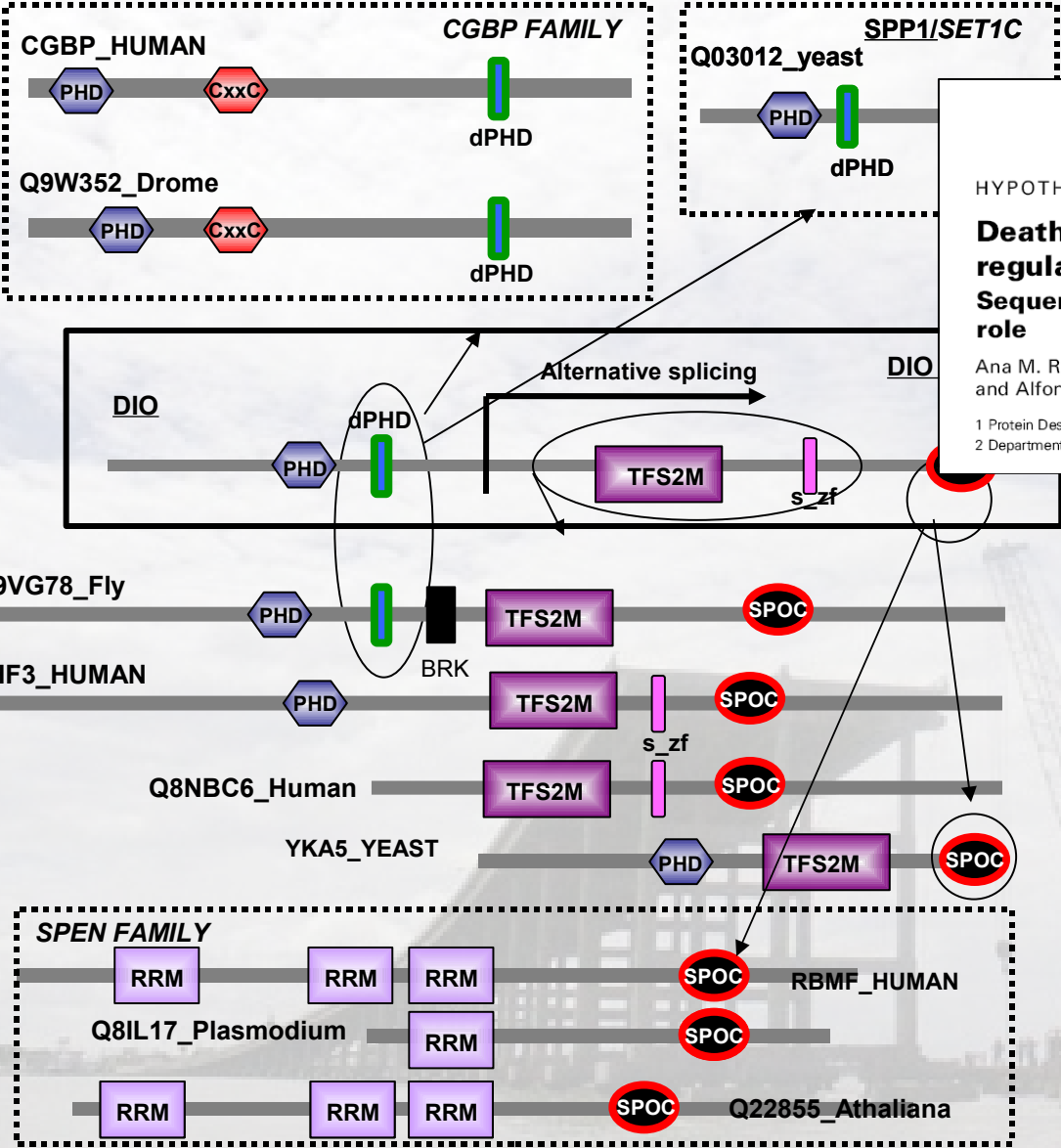
Comparative modeling

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



(3) DIDO-1 PROTEIN ANALYSES

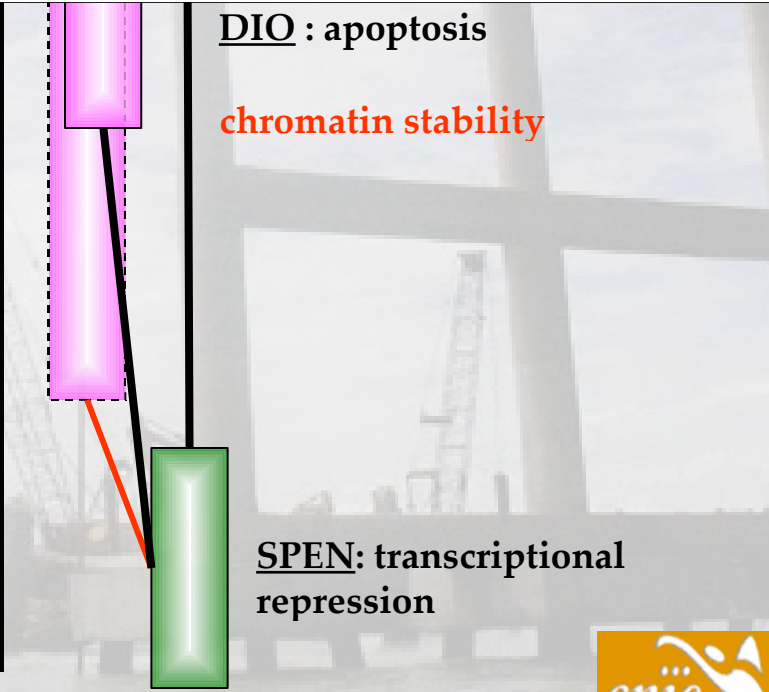
OVERVIEW



HYPOTHESIS
Death inducer obliterator protein 1 in the context of DN regulation
Sequence analyses of distant homologues point to a novel functional role

Ana M. Rojas^{1,2}, Luis Sanchez-Pulido¹, Agnes Fütterer², Karel H. M. van Wely², Carlos Martinez and Alfonso Valencia¹

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



(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.

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

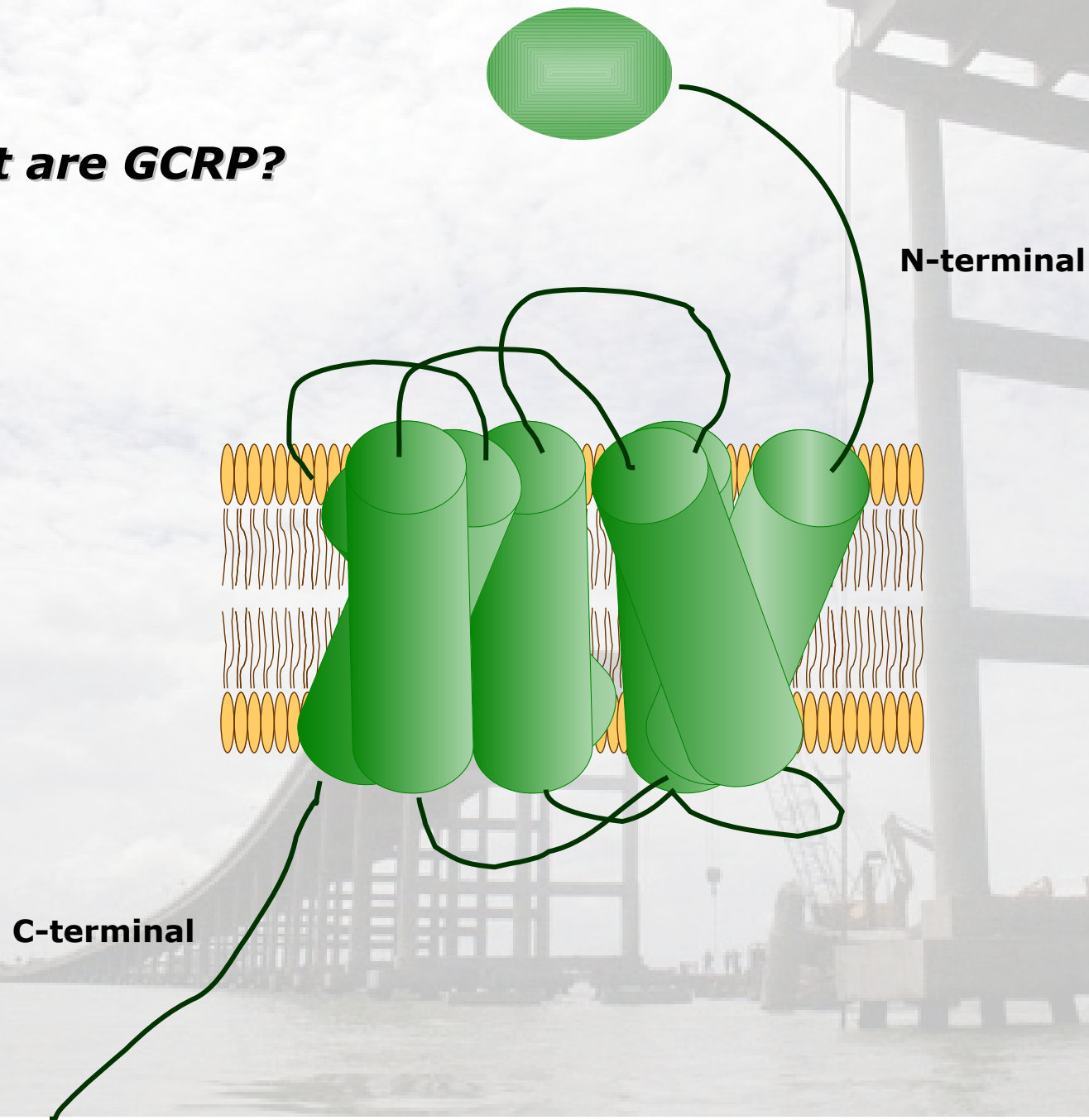
(4) Predicting interaction interfaces
CCR5 dimerization.

GOAL:

Use bioinformatics tools to predict residues involved in binding.

(4) CCR5: dimerization

What are GCRP?



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.

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



Copyright (C) 2005, [GPCRDB](http://www.gpcrdb.org).

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

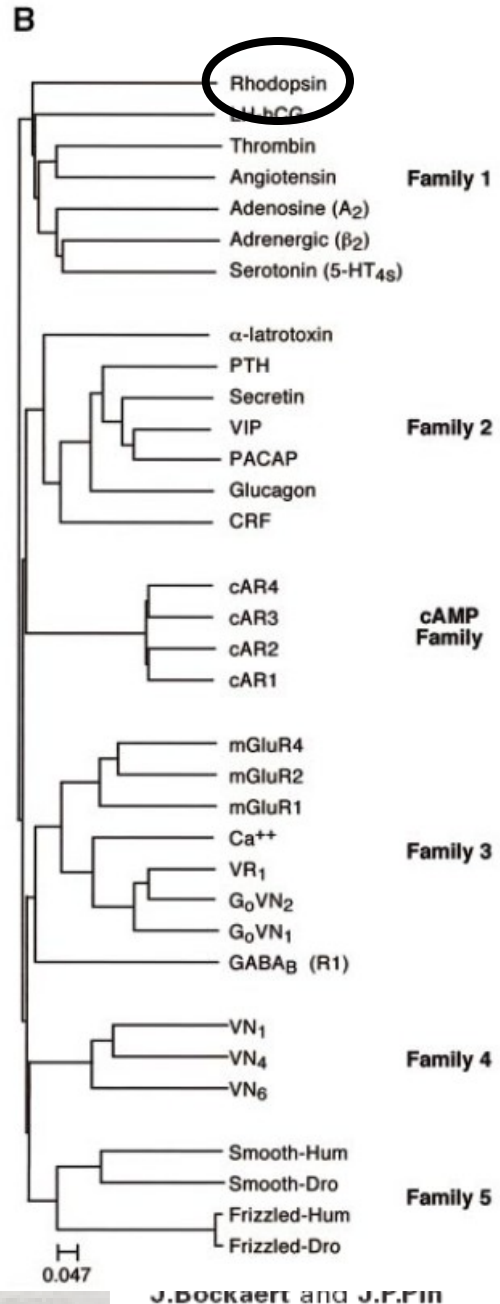
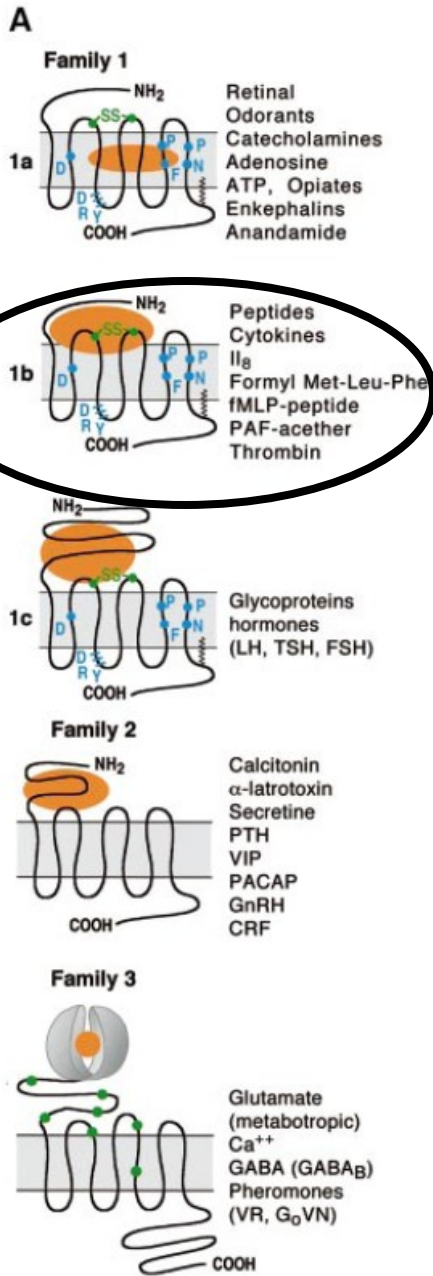
Class C: Maetabotropic glutamate/pheromone

Class D: fungal pheromone

Class E: cAMP receptors (Dictyostelium)

Frizzleed/Smoothened family

(4) CCR5



J. DOCKAERT and J.F. PIN

How many GCRP?

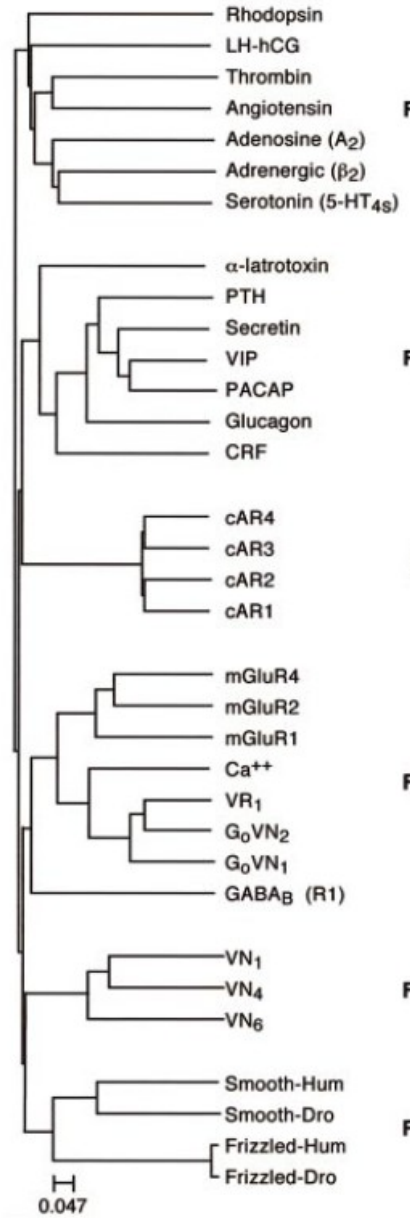
2495 entries!

CLASS A
(16 subfamilies)

- No sequence similarity!

- The tuning to bind ligand-G prots is regulated by RNA editing and phosphorylation

B (4) CCR5



Family 1

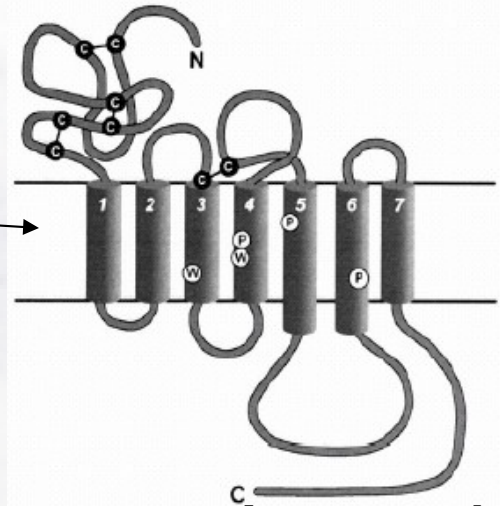
Family 2

cAMP Family

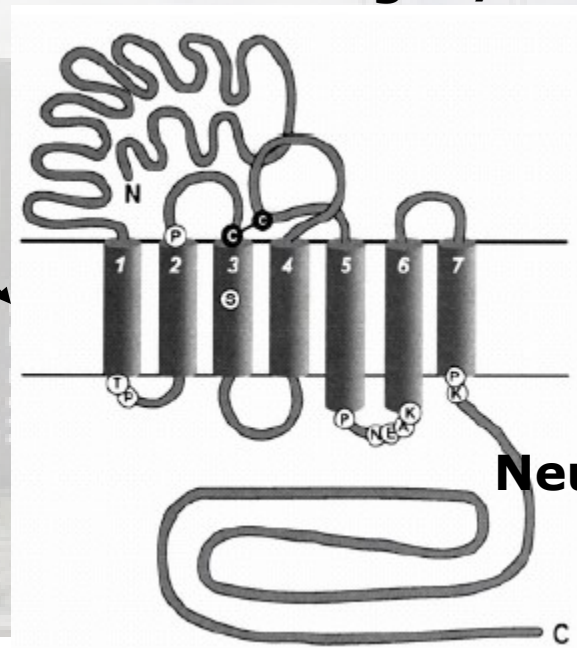
Family 3

Family 4

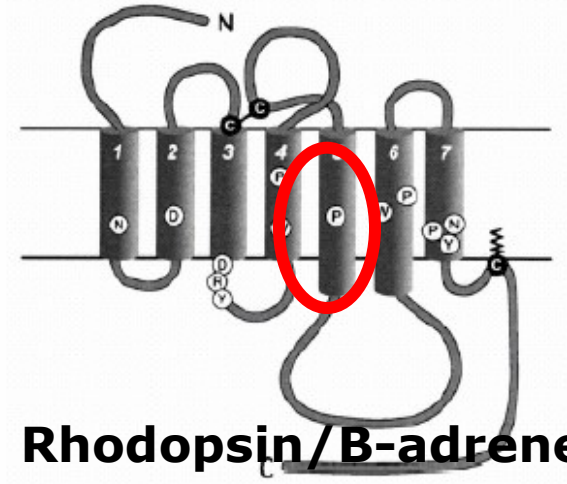
Family 5



Glucagon/VIP/Calcitonin



Neurotransmitter/Ca



Rhodopsin/B-adrenergic

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

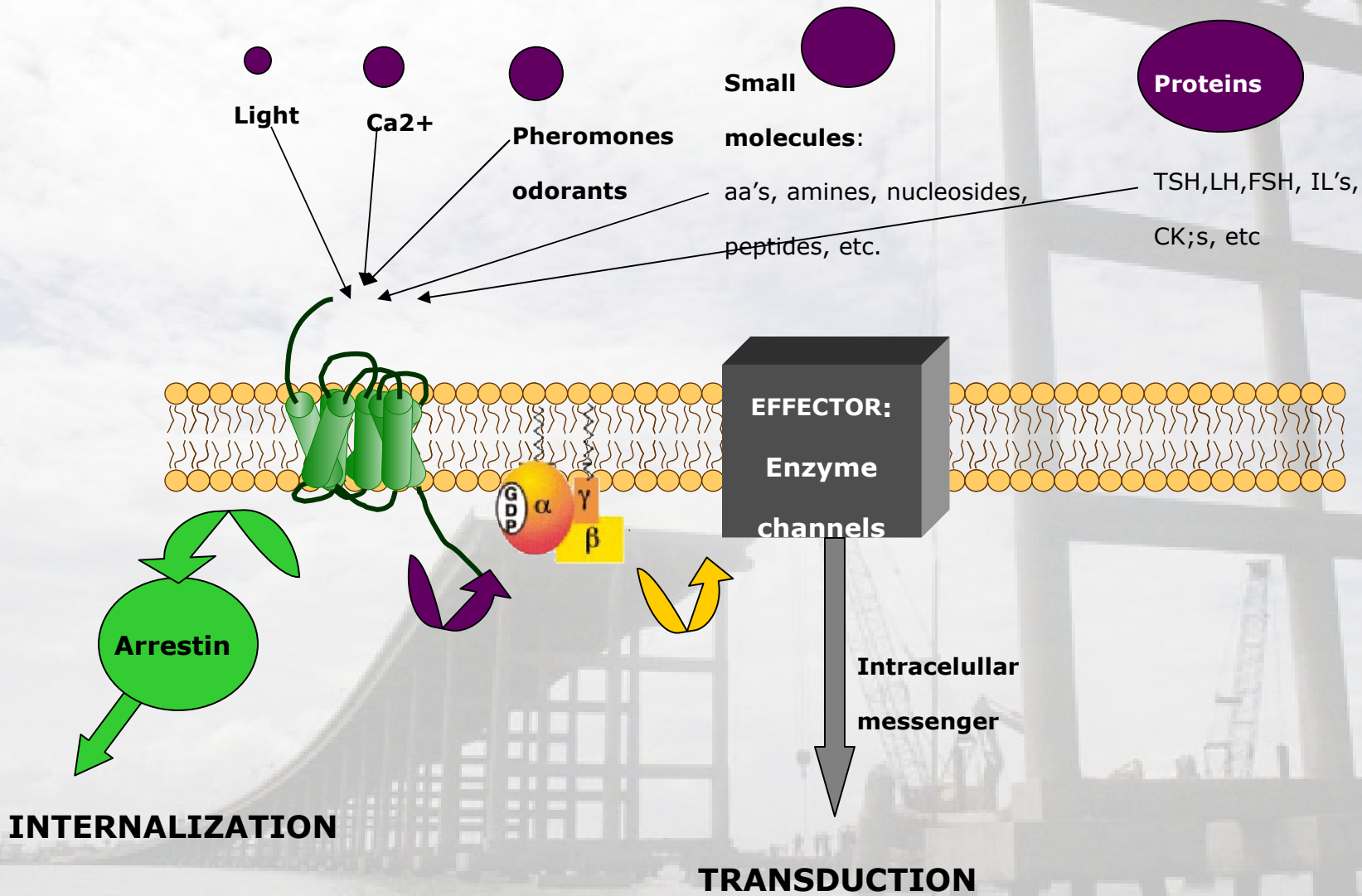
- **Sub-families**
- Muscarinic
 - acetylcholine
 - Adrenoceptors
 - Dopamine
 - Histamine
 - Serotonin
 - Octopamine
 - Trace amine

- Sub-families**
- Angiotensin
 - Bombesin
 - Bradykinin
 - C5a anaphylatoxin
 - Fmet-leu-phe
 - APJ like
 - Interleukin-8
 - Chemokine**
 - 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

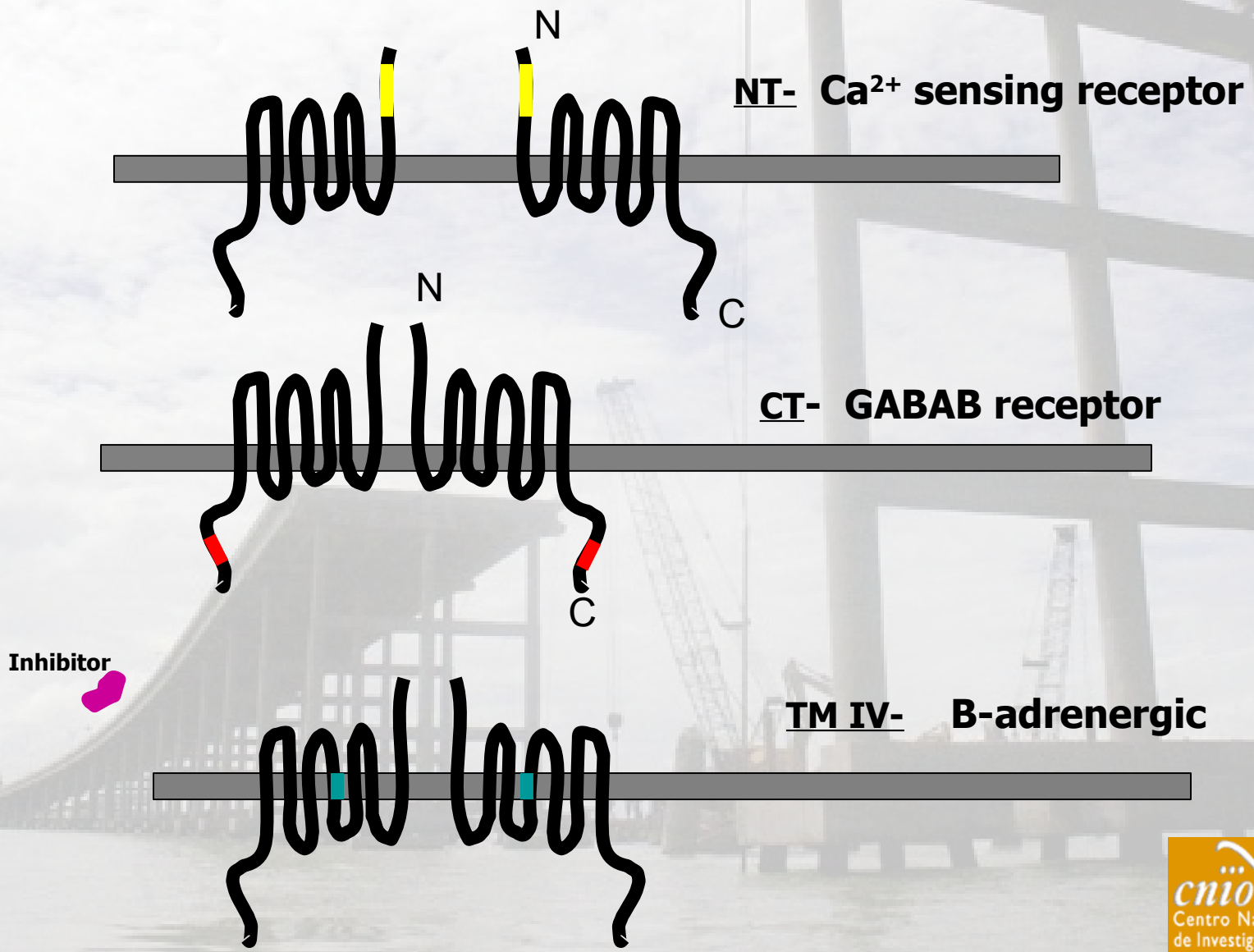
1094!

357 Highly similar

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?***

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

THE JOURNAL
© 2004 by T

Evol
of R

G pr
ated by
helices
elemen
tionary
monly
the rho
a netw
G prote
duction
mutati
accordi
divided
where
region

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^{*,-1}

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.

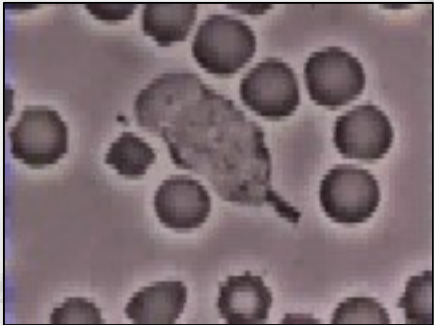
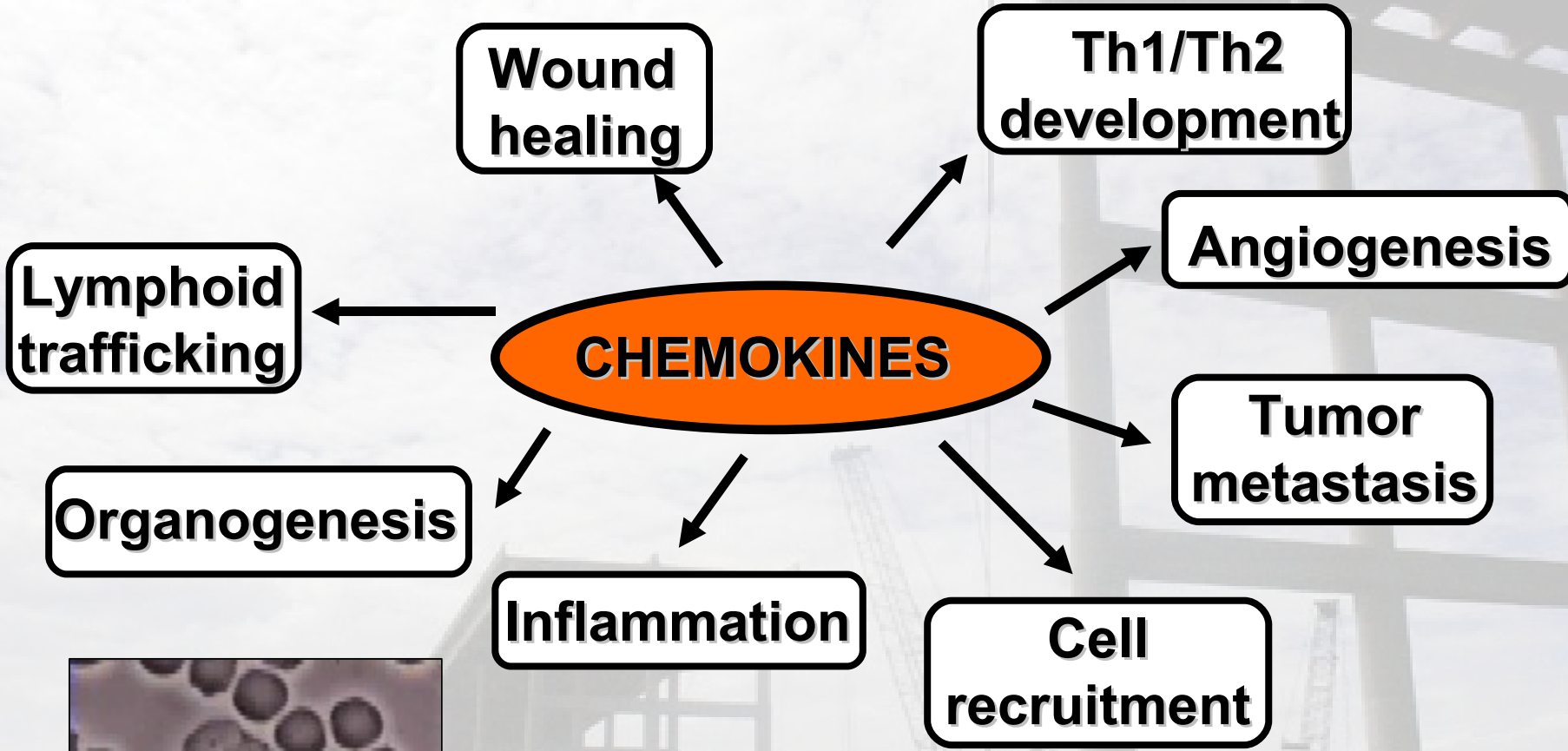
Received June 25, 2003; Revised Manuscript Received October 1, 2003

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



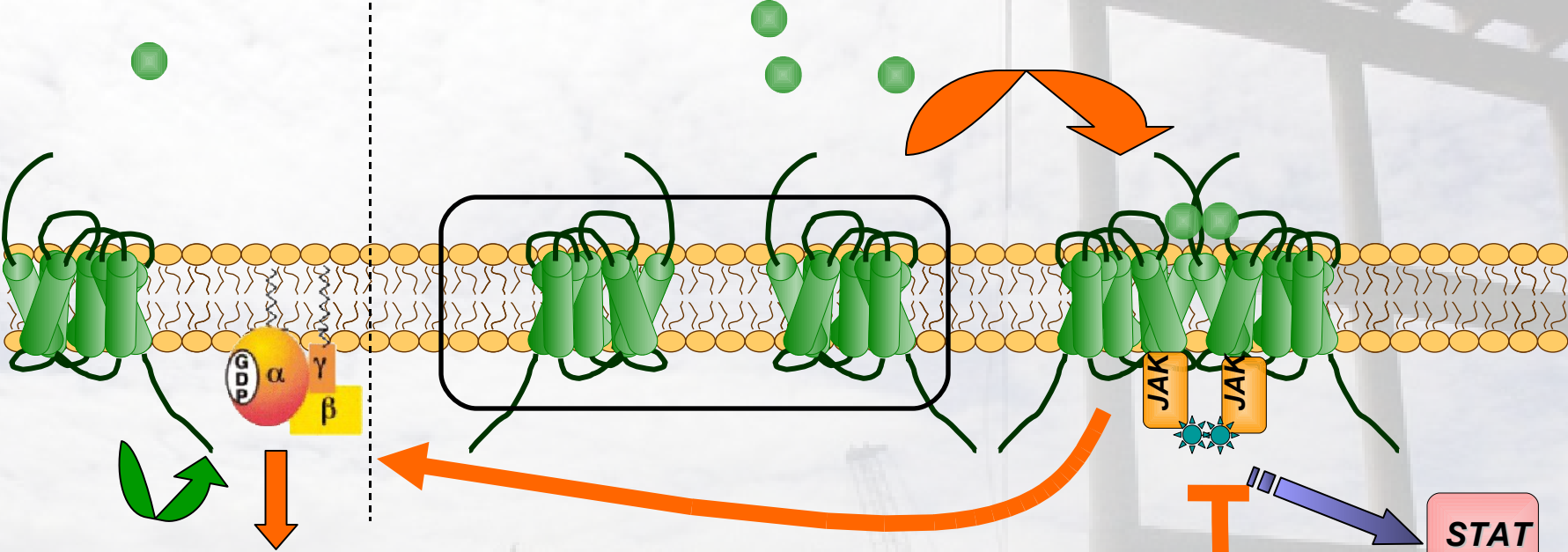
Rogers D.
Vanderbilt University (1950s)

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

CK'S SIGNALING

Gi dependent

Gi independent



**ADHESION
CHEMOTAXIS
POLARIZATION
INTERNALIZATION
GENE EXPRESSION**

Mellado et al, (2001)

Thelen (2001)

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**

(4) CCR5 **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

Predicting Functional Residues in Protein Sequence Alignments

TreeDet
SQUARE MB-Method FASS S-Method

Simple Run Advanced Run Additional Info Contact Us

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]

The alignment has to be longer than 50 residues and contain at least 15 sequences (no more than 200).

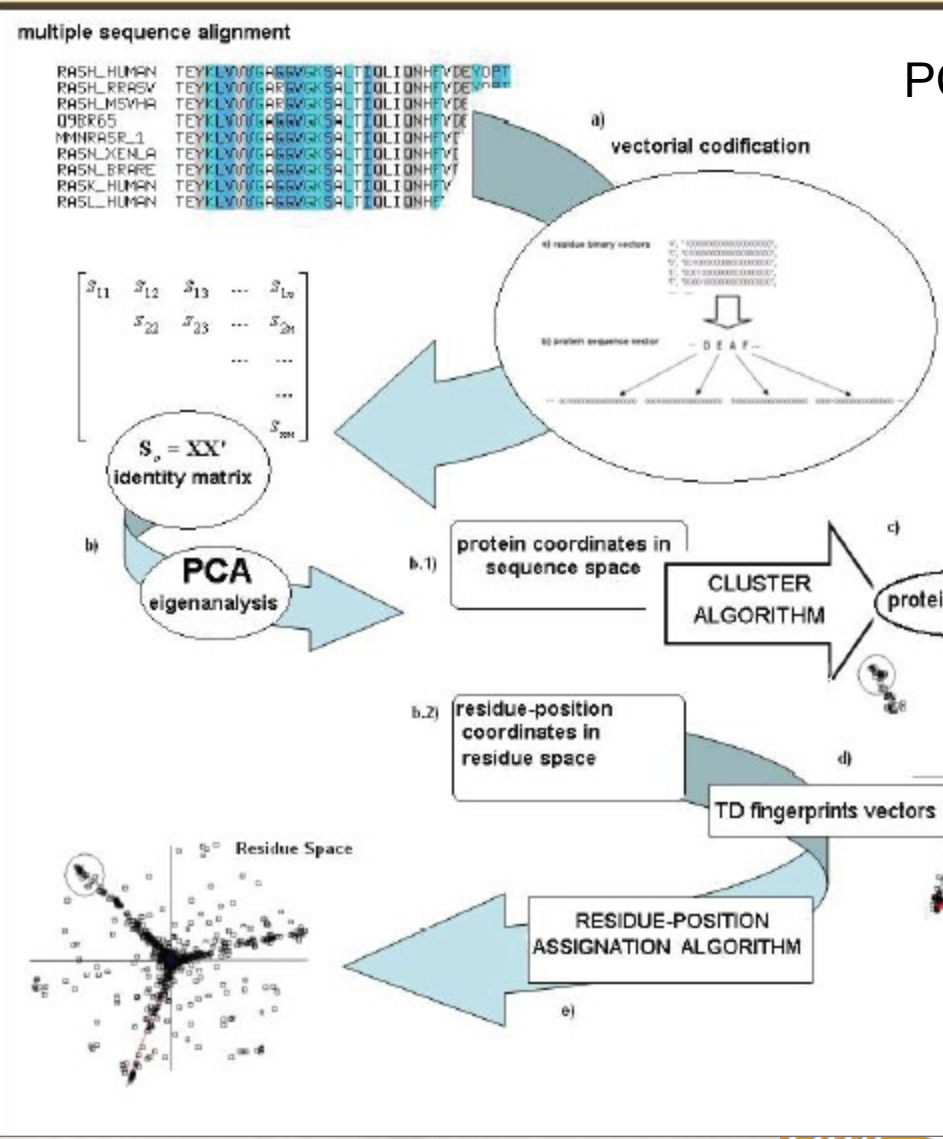
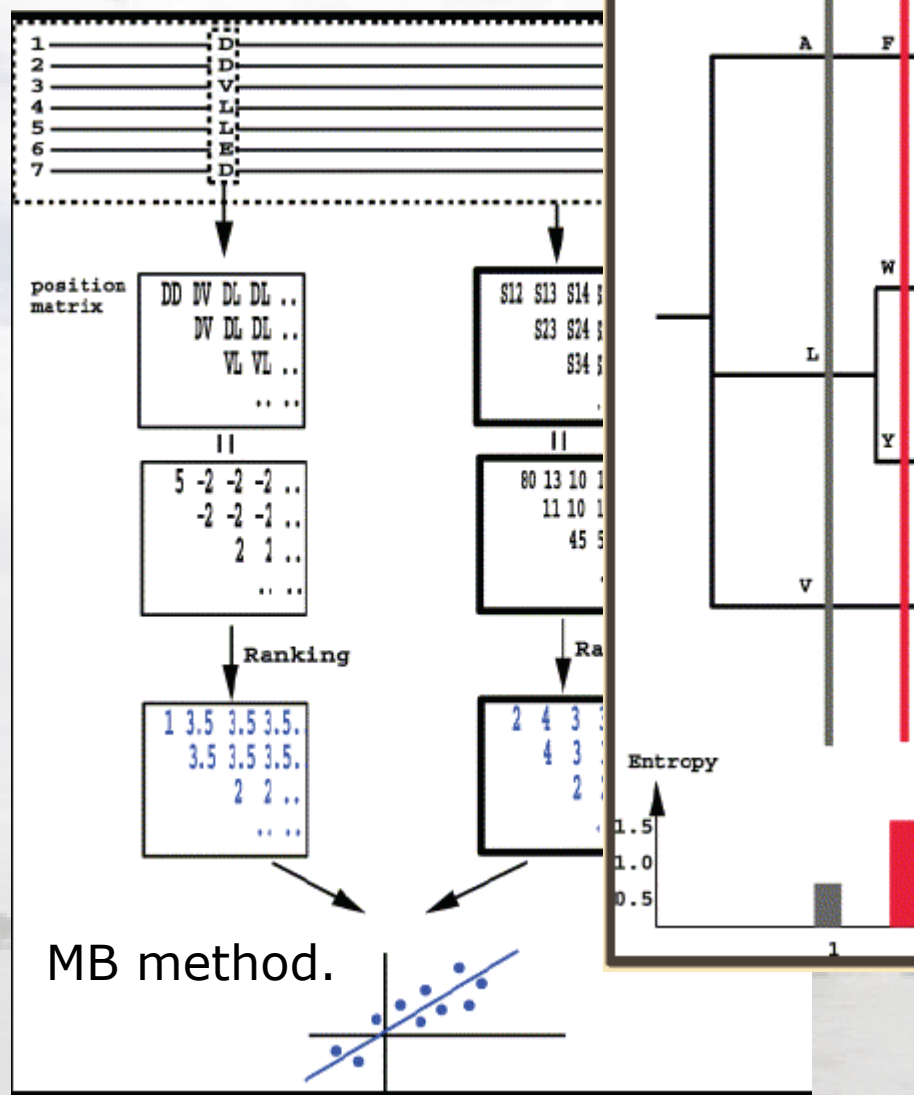
Check one or more methods!

- MB-Method**
Looks for positions in the multiple sequence alignment whose mutational behaviour resembles that of the global alignment [\[i\]](#).
- FASS**
Principal component analysis of the multiple alignment and computation of the statistical confidence in the organization of the family into sub-families [\[i\]](#).
- S-Method**
The level entropy method searches for the level of splitting of a previously established phylogenetic tree to optimize the relative entropy [\[i\]](#).
- SQUARE**
An additional tool to evaluate alignments [\[i\]](#).

PERFORMANCE: Some numbers [\[HERE\]](#).

(4) CCR5

• What the methods do?
Predict functional sites
 using different approaches.



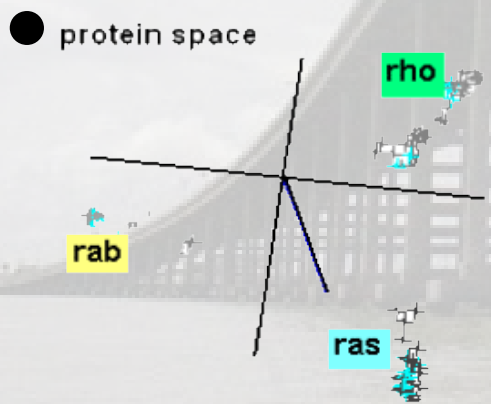
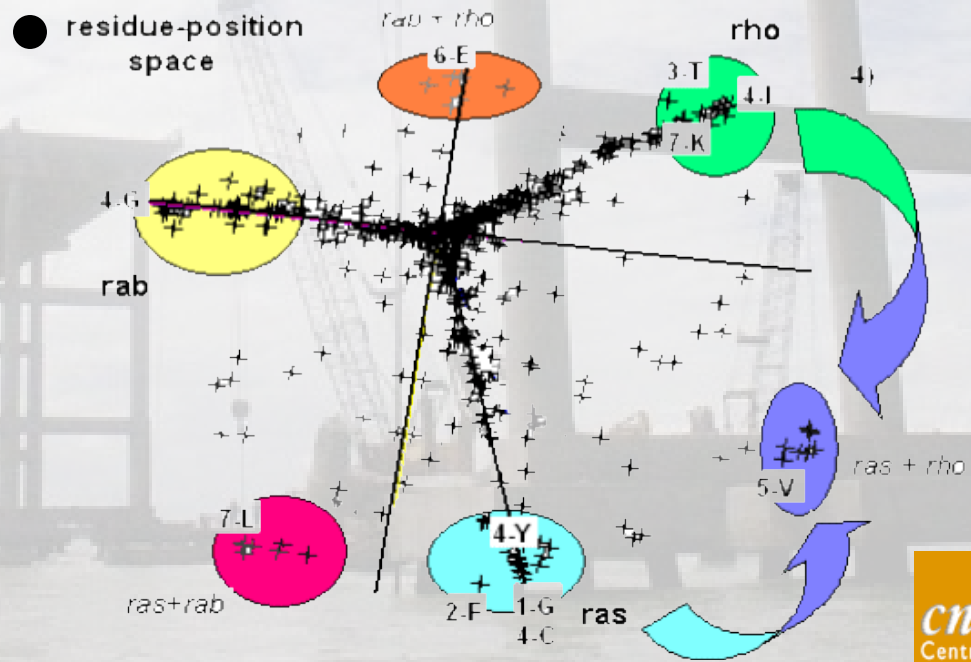
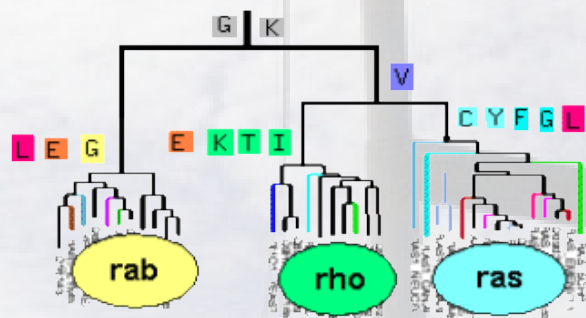
(4) CCR5

Sequence Space: overview

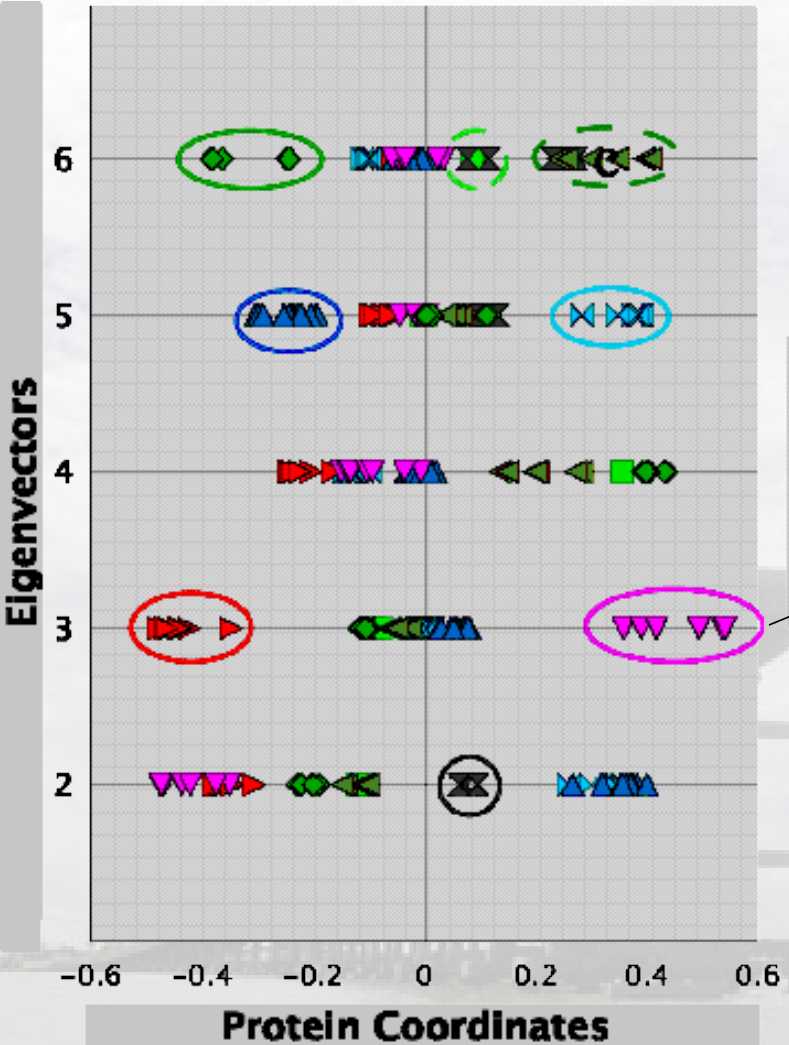
Casari, G. et al. Nat. Struct. Biol (1995). 2:171-178.

An example:

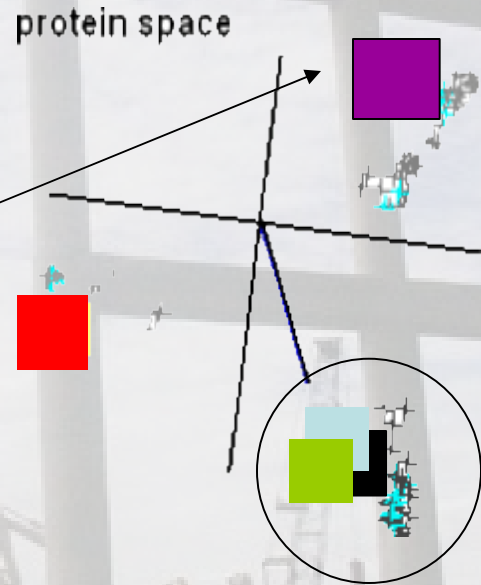
	1	2	3	4	5	6	7	8	9
RASH_HUMAN	G	F	Y	C	V	F	G	G	K
RAS_RRASV	G	F	Y	C	V	I	L	G	K
RASH_MSVHA	G	F	Y	C	V	M	L	G	K
RASN_XENLA	G	F	Y	C	V	V	L	G	K
RASN_BRARE	G	F	Y	C	V	I	L	G	K
RASK_HUMAN	G	F	Y	C	V	F	L	G	K
RASL_HUMAN	G	F	Y	C	V	F	L	G	K
SEC4_YEAST	S	F	D	G	N	F	L	G	K
SEC4_CANAL	S	F	T	G	K	F	L	G	K
Q9HET4	M	I	E	G	V	F	L	G	K
Q96VL3	F	A	L	G	V	F	L	G	K
YPT2_SCHPO	F	A	K	G	D	F	L	G	K
SAS1_DICDI	Y	R	K	G	E	F	L	G	K
SAS2_DICDI	F	E	N	G	E	F	L	G	K
RHOC_HUMAN	M	E	T	I	V	E	K	G	K
RHO_DISOM	A	R	T	I	V	E	K	G	K
RHO_APLCA	M	T	T	I	M	E	K	G	K
RHO1_DROME	M	Q	T	I	V	E	K	G	K
RHOB_HUMAN	R	A	T	I	V	E	M	G	K
Q8TG28	V	R	T	I	V	E	K	G	K
RHO1_SCHPO	T	T	T	I	V	E	K	G	K



Sequence Space: Clustering results



- ◆ A → CXCR3/5
- ▼ B → CXCR4
- ▲ C → CCR1/3/V
- ▶ D → IL8A/B
- ◀ E → CCR6/7/9/11
- ⊗ F → CCR2/5
- G → CCR4/8
- H → CCR10



Our strategy

TEST CASE: CHEMOKINES, known to dimerize.

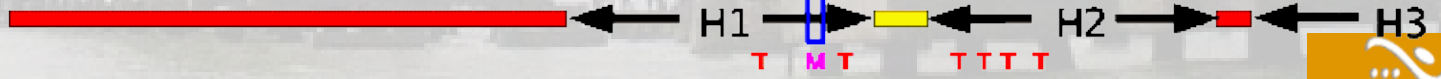
Steps:

- 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**

(4) CCR5 Tree-determinants: Clustering results

Residues obtained by Sequence-Space family division.

<p>CR3_CCR5A_00939... CR3_RAT... CR3_HUMAN... CR3_MOUSE... CR3_RACCO... CR3_CAT... CR3_RABBIT_00939... CR3_RACCO... CR3_CENAE... CR3_MOUSE...</p>	CKR1/3	<pre> MEYSATTE DYEDEIT EFDYDQATPQKAV NERAFGAQLPPL YSLVFIQGLVGNLVL VLVQYKRL KKHITSITLNLAIADLFLFTLPFWIDYKTDQVFGNANCKVLGSGFYTGLTSEI NASHSELKYVE TFE TT PVEYKAPCKEV SIRELSM LPPYSLVFTVGLGNMNL VILCKYKRL QMHNITVLLNLAIADLFLFTYFPWDIVYLNHENGFGHGWCKLNSGLYYLALYSEI HTLSLOT VE TRG PT SYDODHLLCEKAD TRALMAQVPP YSLVFTVGLGNMNL VILCKYKRL RIMHNITVLLNLAIADLFLFTVLPWDIVYLNHENGFGHGWCKLNSGFYTLGYSEI NAFNDDELKAVG SRE IT PVEYKAPCKEV RIKELGSM LPPYSLVFIQGLGNMNL VILCKYKRL QMHNITVLLNLAIADLFLFTVLPWDIVYLNHENGFGHGWCKLNSGFYTLGYSEI METPNTTE DYEDEIT EFDYDQATPQKAV NERAILAQLPPL YSLVFIQYQGLNL VLVQYKRL KKHITSITLNLAIADLFLFTLPFWIDYKSTDDHFDGANCKVLGSGFYTLGYSEI MATYEEAELETFF G TT FYDYFAQDFKRY SIDLGAQFLP YSLVFIQGLGNLIT VLVQYKRL KIMHNITVLLNLAIADLFLFTLPFWIDYVYHNNHNVFGHGWCKLISGLYTGLTSEI MEYSATTT DYEDEIT EYDDEITPQKAV ANRFAQLPPL YSLVFIQGLVGNLVL VLVQYKRL RSHITSITLNLAIADLFLFTLPFWIDYKTDQVFGSALCKFLSGLYTGLTSEI HTLSLOT VE TRG PT SYDODHLLCEKAD YSALIAQVPP YSLVFTVGLGNMNL VILCKYKRL QMHNITVLLNLAIADLFLFTLPFWIDIVYHNNHNVFGHGWCKLNSGFYTGLTSEI HTLSLYT VE TRG PT SYDODHLLCEKAD YSALIAQVPP YSLVFTVGLGNMNL VILCKYKRL RIMHNITVLLNLAIADLFLFTLPFWIDIVYHNNHNVFGHGWCKLNSGFYTGLTSEI MELPAVTEPSYNTVA KADFMSGLCFSCI WVARAGITVPT YSLVFITVGLGNMNL VILCKYKRL RSHITSITVLLNLAIADLFLFTLPFWIDYKTDQVFGHGWCKLVSGFYTGLTSEI </pre>
<p>L8A_MOUSE... L8A_RAT... L8A_RAT... L8A_HUMAN... L8A_RABET... L8A_BOVIN... L8A_RABIT... L8A_HUMAN... L8A_CANFA...</p>	IL8A/B	<pre> HGEFKVDKFNDEDFPSG DLDPNYSGNPPIPLDVAVPHSE NLEDNSYAVNYIVLVTLTSLVGNLNL VLVLYNR STCSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV NABAIEYFIWZAPDFEBFGNIT RLPLPTGEYSPCKRV PHTNROAVVYVYVALYVLLSLGNSLNL VLVLYNR STCSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HGEIKVDHFLSDFPSG DIDNINYSNPPTLLSQAAPPSA NLDIMRYAVVYVYVALYVLLSLGNSLNL VLVLYNR STCSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HEDFMNESDGFDFPQDGLSMYSSTLPPFLDAAPCPSE SLEDKCPVWIITVLLVLLSLGNSLNL VLVLYNR YQFSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HGFTEWESYVEDFF GDSMYSITLPLPFLDAPCPSE BLETSYVYVITVLLVLLSLGNSLNL VLVLYNR STCSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HTIILKDL SNSGLWGFDEPFG NYSOTPHTEGDYDYPCEIS TETLNKYAVVYVIALVLLSLGNSLNL VLVLYNR STCSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HEVNWVHIDLMT WFEDEFAKATGHPNBDYSPCLVY HEVNWVHIDLMT WFEDEFAKATGHPNBDYSPCLVY TETLNKYAVVYVIALVLLSLGNSLNL VLVLYNR YQFSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HSHKTDPOHWCPDOLF TGHPPHBDYSPCLVY TETLNKYAVVYVIALVLLSLGNSLNL VLVLYNR YQFSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV HEYIDWNYSLDLF GGDINNYTNTDPHPPAOSAPCPSE SLDINNYAVVYVYVYVLLVLLSLGNSLNL VLVLYNR YQFSYTDVYLLNLAIADLFLFPWAAS KVNQHTFGSTLCKIPSYKVEYTFYSSV </pre>
<p>CR6_MOUSE... CR6_HUMAN... CR6_BOVIN... CR6_MOUSE... CR6_MOUSE... CR6_MOUSE... CR6_MOUSE...</p>	CKR6/11/9/7	<pre> HSGSNNFSDFWSSDSEDFYVSVMT SYSYSDEMLLCSLQE YRQSRFLVPIATSLICVGLLGNLIT VTFAYYKRL RSHITDITVLLNLAIADLFLFPWAASHATGANVFSNATCKLKGIIAIFNFKGM NALEQNQSDTYTTEEME MNGTVDTSYELCLIKED YRFAKVFLVPLTIVFYIQLAGNSHNY IYATYKRL RTKIDITVLLNLAIADLFLFPWAASHATGANVFSNATCKLKGIIAIFNFKGM NAYETNGSDTYTEEME MNDTHQTSOYELCLIKED YRFAKVFLVPLTIVFYIQLAGNSHNY IYATYKRL RTKIDITVLLNLAIADLFLFPWAASHATGANVFSNATCKLKGIIAIFNFKGM NAGDYGSESTSHEDTUNHFDTFTCKML NWRQFASHPLPPL YMLVFIQGLGNLNL VLVLYNR KTHITDITVLLNLAIADLFLFPWAASHA AADQVFGSTPCKVYSHNKHPFVSCV HDQKPKRNLYVYALLYIPWCPDDEYTDYIGENT TVDYTLYESYCPKID YRNFKAMPFLPIMYSICVGLGNLNL VLVLYNR KTHITDITVLLNLAIADLFLFPWAASHA BAKSHIFGFLGCKVDFGKYLSPFSGM HDLQKPKRNLVYALLYIPWVCLDDEYTDYIGENT TVDYTLFESLCKKID YRNFKAMPFLPIMYSICVGLGNLNL VLVLYNR KTHITDITVLLNLAIADLFLFPWAASHA AAKSHIFGFLGCKVDFGKYLSPFSGM NNPTELTLIPGDFDFSDTASTDOTHNLNFSDFFCKM NWRQFASHPLPPL YMLVFIQGLGNLNL VLVLYNR KTHITDITVLLNLAIADLFLFPWAASHA AAGQVFGSTPCKVYSHNKHPFVSCV HNSTEYFTGDDYONTE TYSDPHSPGSLLEE YRNFKAMPFLPIMYSICVGLGNLNL VLVLYNR RSHITDITVLLNLAIADLFLFPWAASHATNTWYFSALCKLKGIIAIFNFKGM </pre>
<p>CR8_RAT... CR8_MOUSE... CR8_MOUSE... CR8_MOUSE... CR8_MOUSE... CR8_MOUSE... CR8_MOUSE... CR8_MOUSE... CR8_MOUSE...</p>	CKR5/2	<pre> HDQGSIP TTYIYDDYSRSAPCKY HWKQIAAQLPPL YSLVFIQYVGNMNL VILCKYKRL KSHITDITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI HDQGSVP TTYIYDDYGSAPCKY HWKQIAAQLPPL YSLVFIQYVGNMNL VILCKYKRL KSHITDITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI HDYQVSP TTYDNYTSEPCCKY HWKQIAAQLPPL YSLVFIQYVGNMNL VILCKYKRL KSHITDITVLLNLAIADLFLFPWAASHA AAGHDFGNANCKVLTGTYHIGYFGGI RLSTSRSRFRINTNSGEEVIT FFDYDTGAPCKKFD WKOOGAQLPPL YSLVFIQYVGNMNL VILCKYKRL KSLTDTITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI RLSTSRSRFRINTNSGEEVIT FFDYDTGAPCKKFD WKOOGAQLPPL YSLVFIQYVGNMNL VILCKYKRL KSLTDTITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI HDYQATSP TYDIEYLSGPCCKTD WKOOGAQLPPL YSLVFIQYVGNMNL VILCKYKRL RSHITDITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI MEDNNHLPQFZKQILSTSHLFTSGLDEGATT PYDYGDEPCKT SWKIQAGMLPPL YSLVFIQYVGNMNL VILCKYKRL KSHITDITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI MEDNNHLPQFZKQILSTSHLFTSGLDEGATT PYDYGDEPCKT SWKIQAGMLPPL YSLVFIQYVGNMNL VILCKYKRL KSHITDITVLLNLAIADLFLFPWAASHA ANHFVFGNANCKVLTGTYHIGYFGGI </pre>
<p>CR9_MOUSE... CR9_MOUSE... CR9_MOUSE... CR9_MOUSE... CR9_MOUSE... CR9_MOUSE...</p>	CKR8/4	<pre> HDYTLDPKSHHTHT DNYYPDLSPPDGE LTPQNKLLLVAVTCLLPYSLGNSLNL VLVYKRL RSHITDITVLLNLAIADLFLFPWAPPTIYLDQVFGSTWCKVWSGFYIGFYSSM MNAETIDTDT DETYVNS YFYFSNHPCKTKE QTKAFGEVFLPPL YSLVFLGLPQNSHNL VLVYKRL RSHITDITVLLNLAIADLFLFPWAPPTIYLDQVFGSTWCKVWSGFYIGFYSSM NHPDIDADTT LDESITSN YVLTESDKPCKTKE GLKAFGEVFLPPL YSLVFLGLPQNSHNL VLVYKRL RSHITDITVLLNLAIADLFLFPWAPPTIYLDQVFGSTWCKVWSGFYIGFYSSM HDYTLDSLVTVIT DNYYPDLSPPDGAEL IOTNGKLLLVAVTCLLPYSLGNSLNL VLVYKRL RSHITDITVLLNLAIADLFLFPWAPPTIYLDQVFGSTWCKVWSGFYIGFYSSM HDYMEPNTVHT DY VDFEFTAPCDAE RLLGSMYVLAIVYCLVFLGLGNSLNL VLVYKRL RSHITDITVLLNLAIADLFLFPWAPPTIYLDQVFGSTWCKVWSGFYIGFYSSM </pre>
<p>CR10_MOUSE... CR10_MOUSE...</p>	CKR10	<pre> HGTATEDYSMHSYSG DEEDAYSSEPPELCYKAD YQAFSRAFQPSYSLYAAQLGAGNLNL ITHLAAARAAFPSTSAHLQALADLFLFPWAASHA ALDQMSLQSATCRTISGLYSASFLHGF HGTKPTEDYSNGLTSG TDEEAYSVGLPELCYKAD YQAFSRAFQPSYSLYAAQLGAGNLNL ITHLAAARRTTSPSTSAHLQALADLFLFPWAASHA ALDQMSLQSATCRTISGLYSASFLHGF </pre>



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**

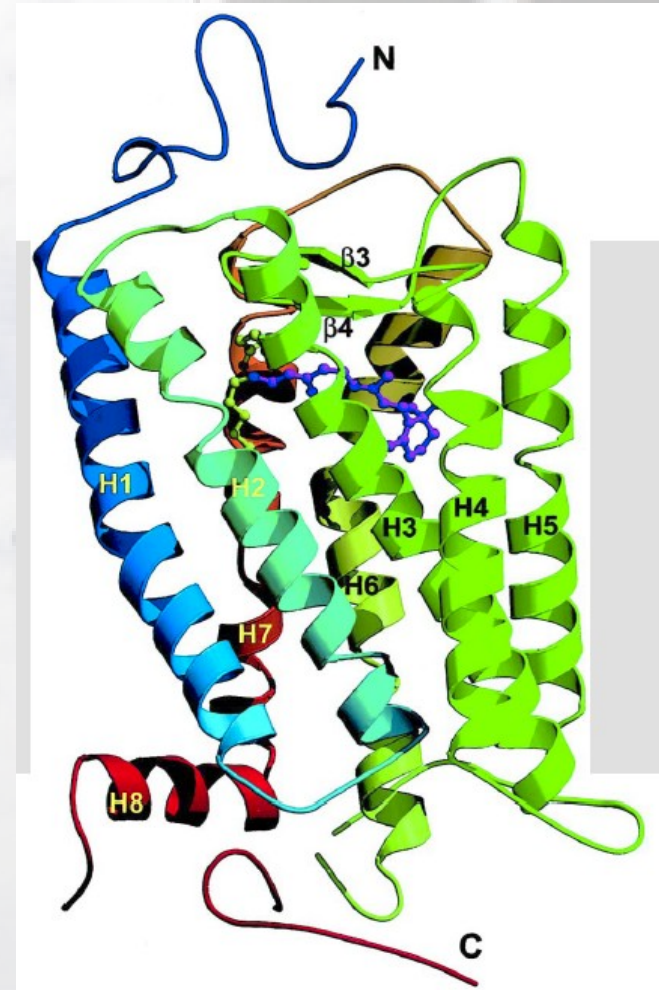
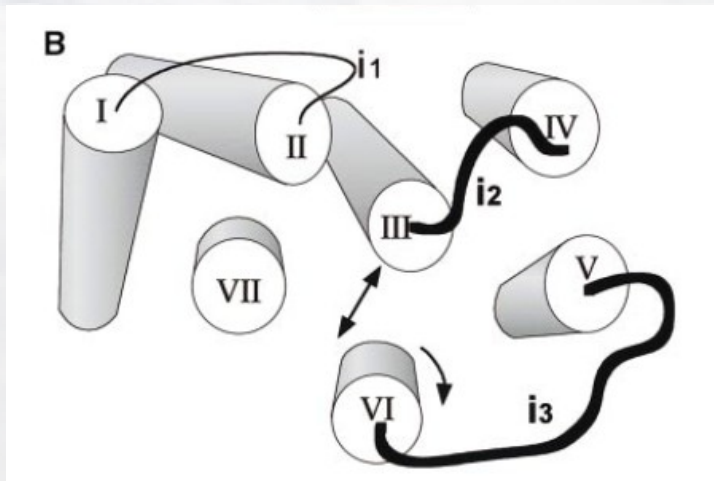
(4) CCR5

Problem: there is no Structure!!!!

1F88 (2000) 2.8A

1GZM (2002) 2.65A

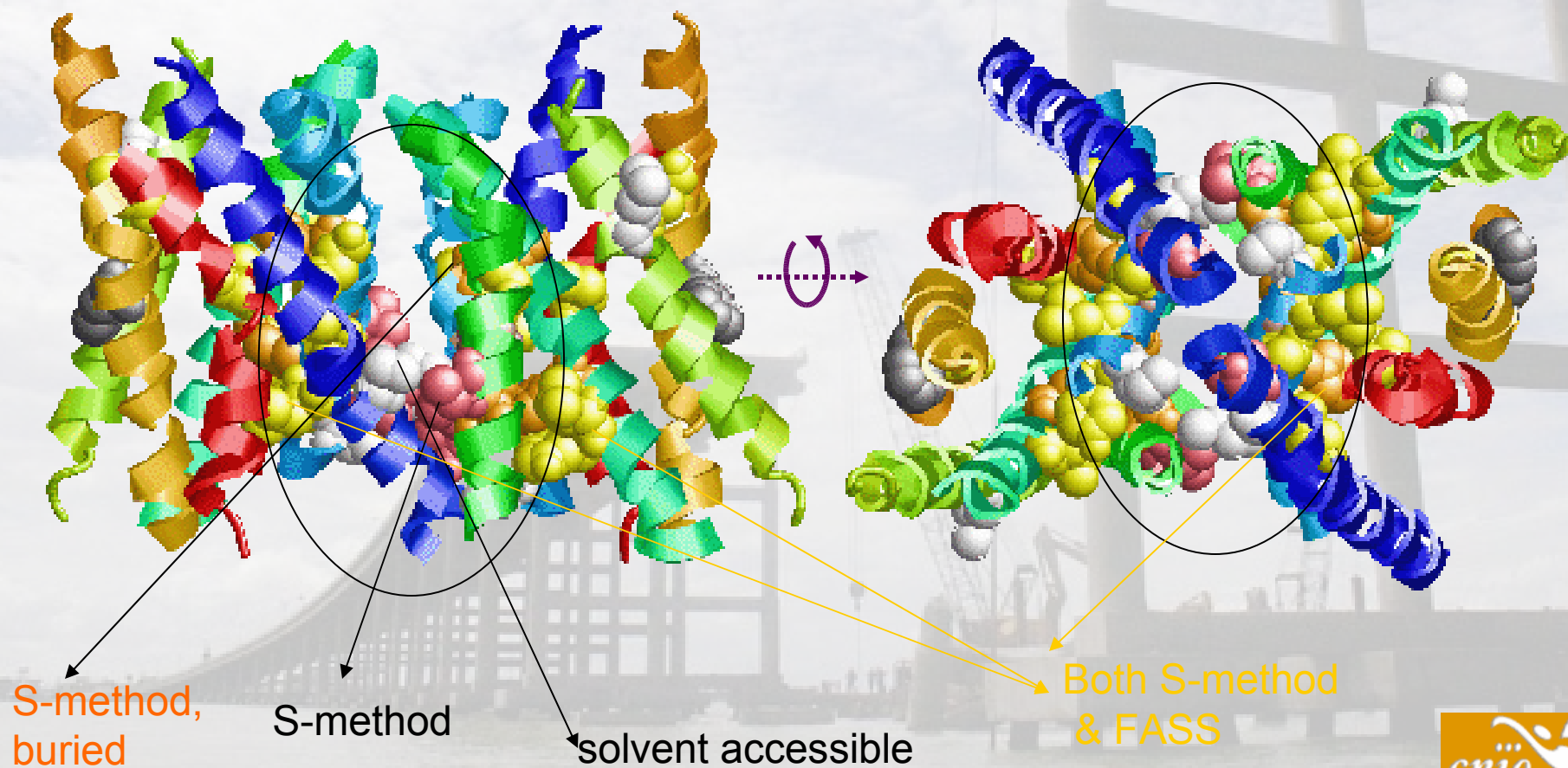
No similarity but there is a “central core”
or “bundle”



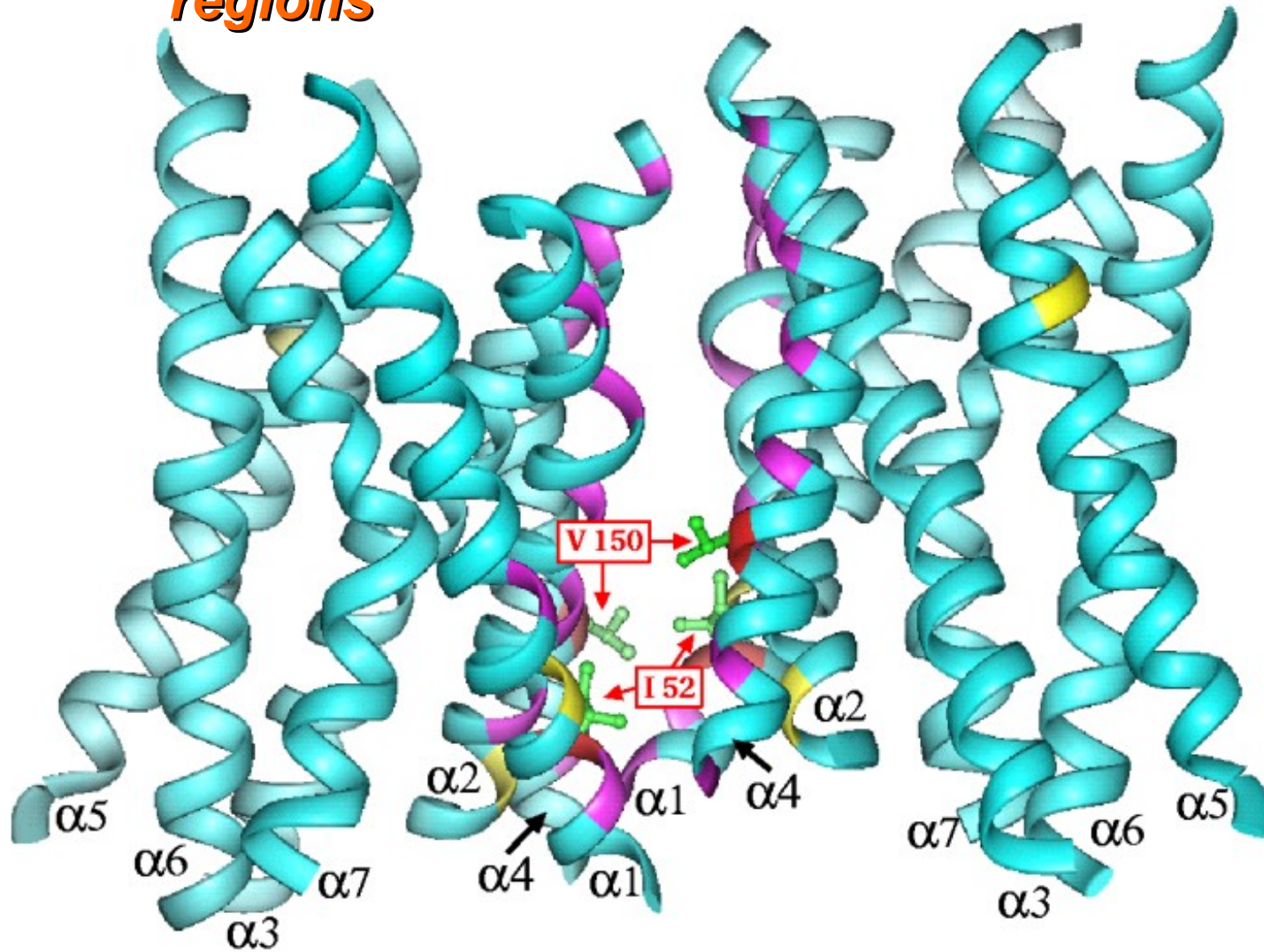
(4) CCR5

Visualizing interface regions

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



(4) CCR5 Visualizing interface regions



Our strategy

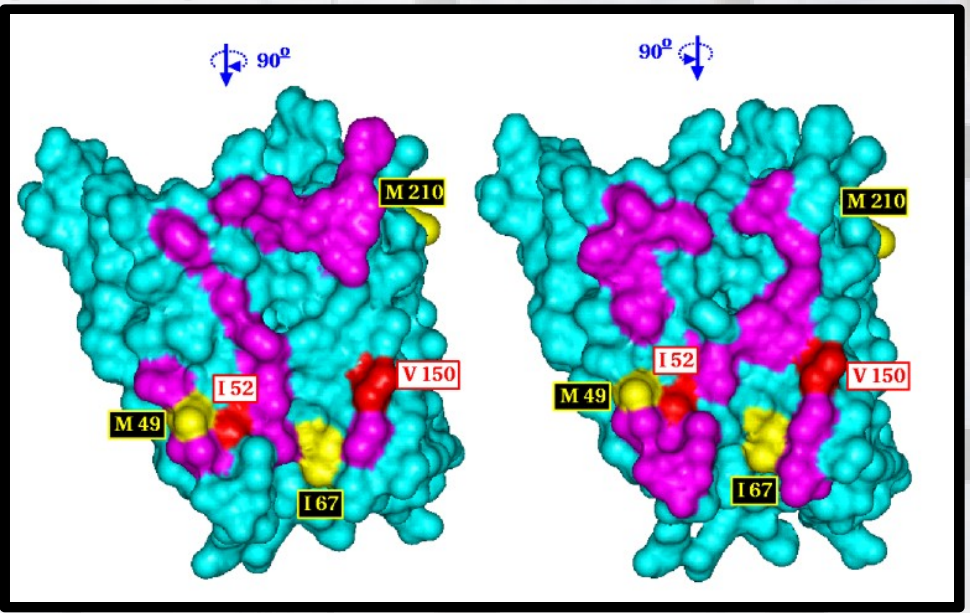
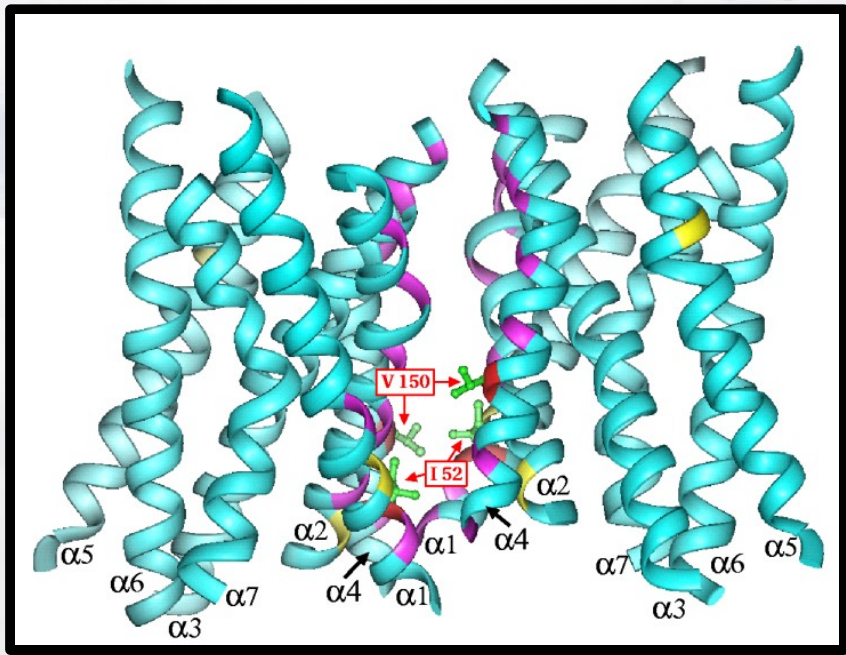
TEST CASE: CHEMOKINES, known to dimerize.

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.**

Chemokine receptor dimerization model

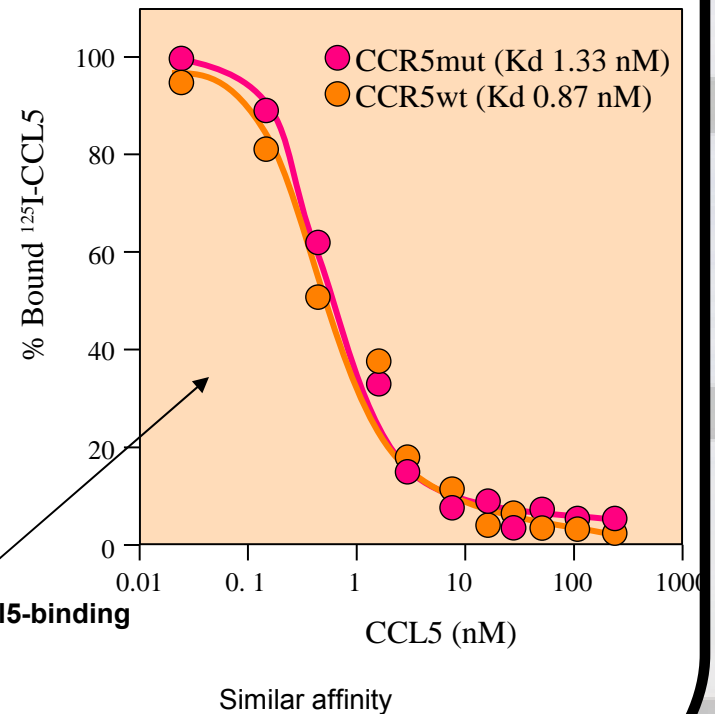
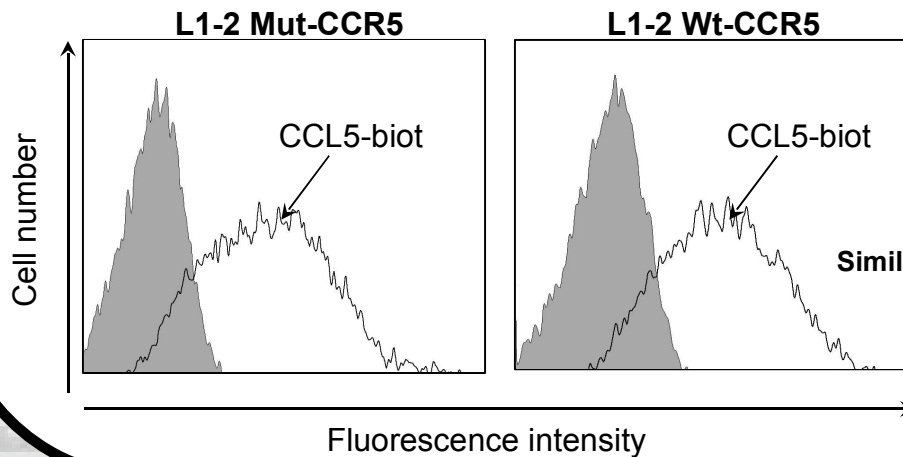
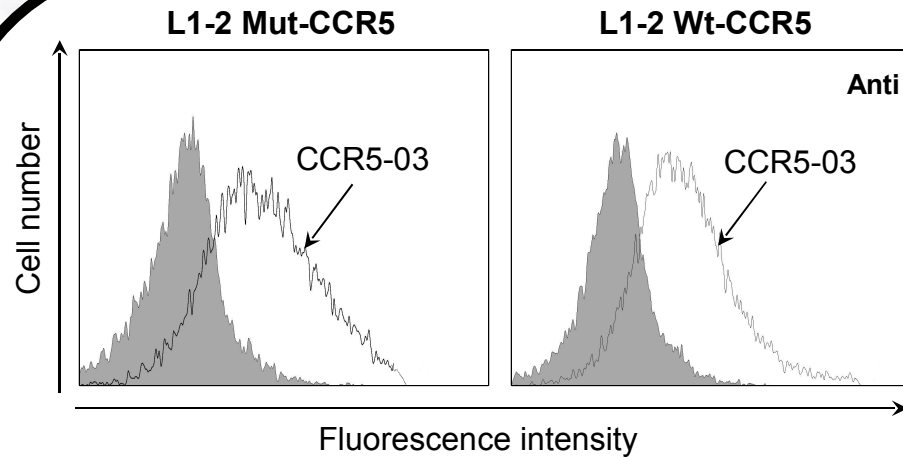
CCR5



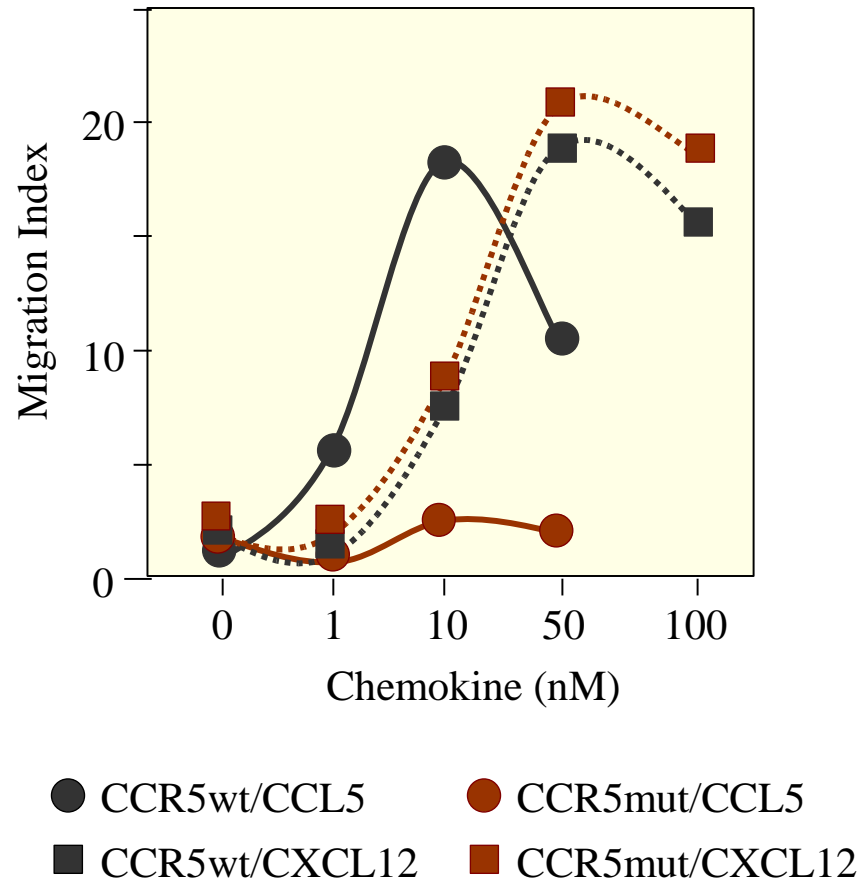
TM I : "GXXLXXL"

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

CCR5^{I52V/V150A} and CCR5 show similar membrane expression and ligand binding

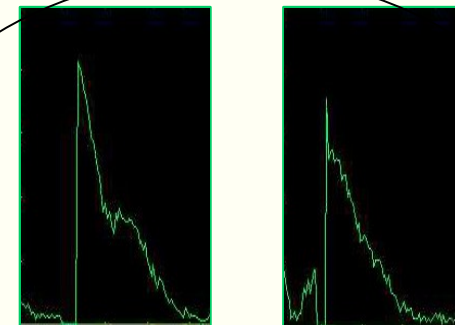
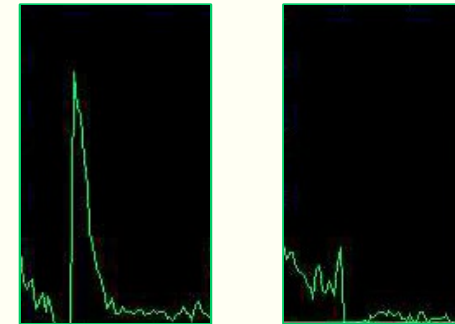


CCR5^{52V/V}150A is a non-functional receptor

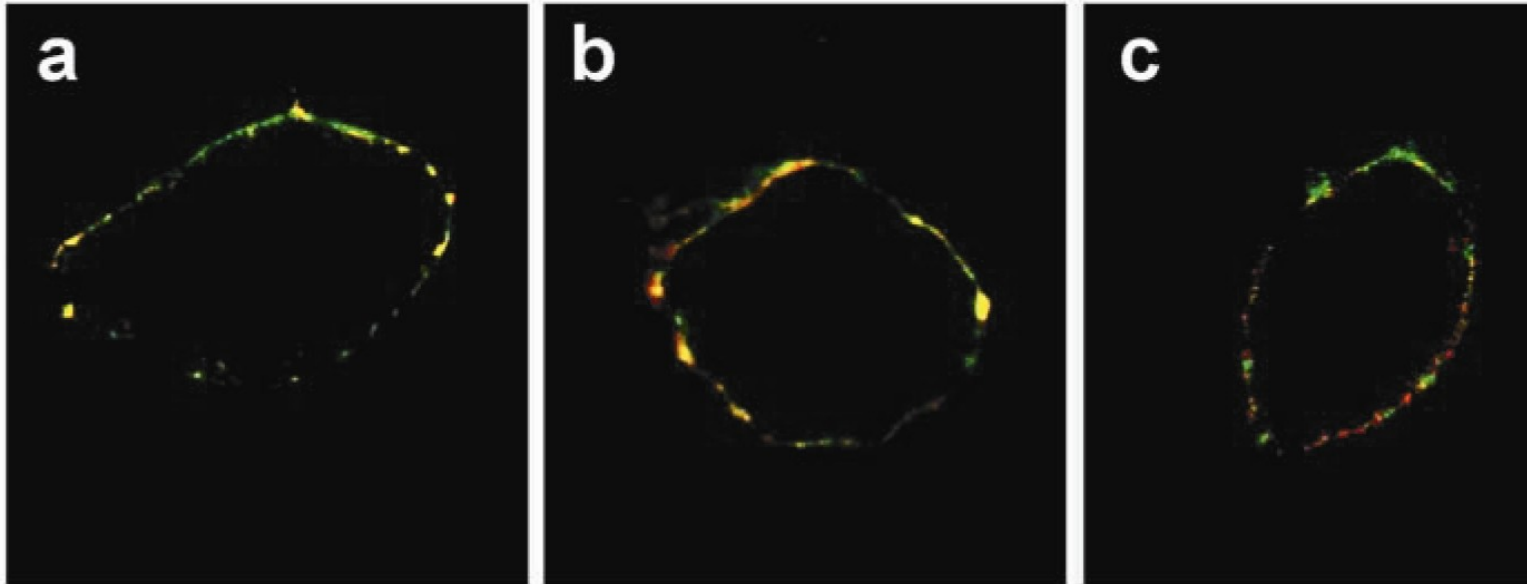


CCR5 responded to Ca

L1M20120107 assays 1-2 CCR5 MUT



CCR5I⁵²V/V¹⁵⁰A is a non-functional receptor

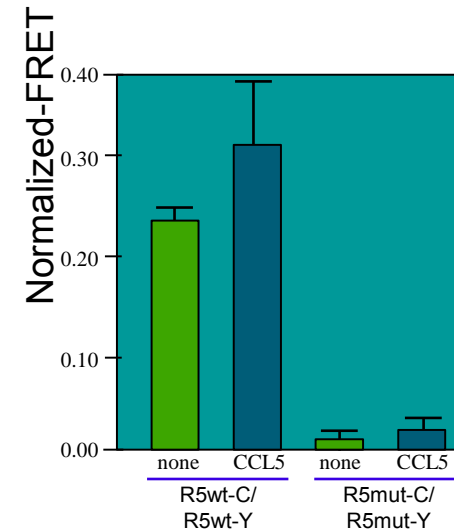
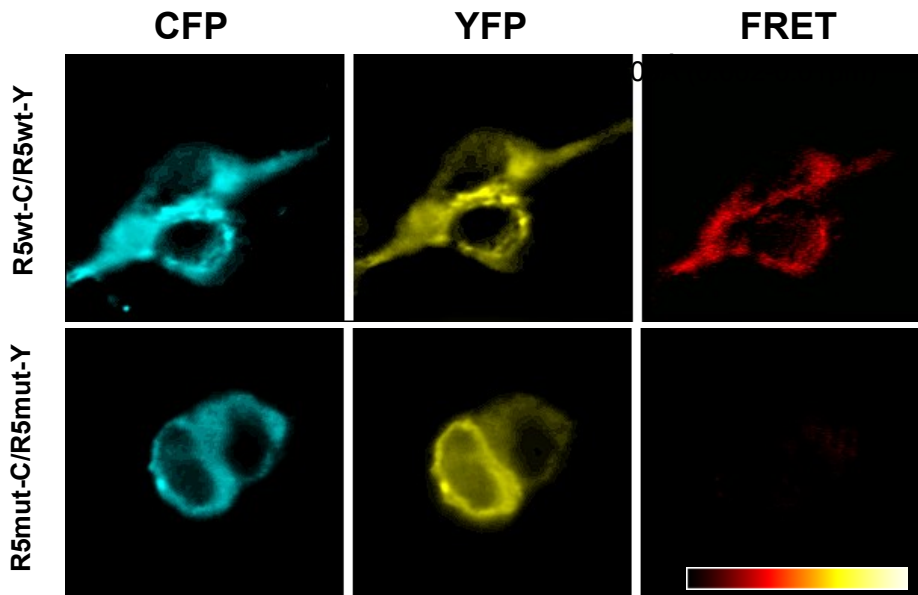
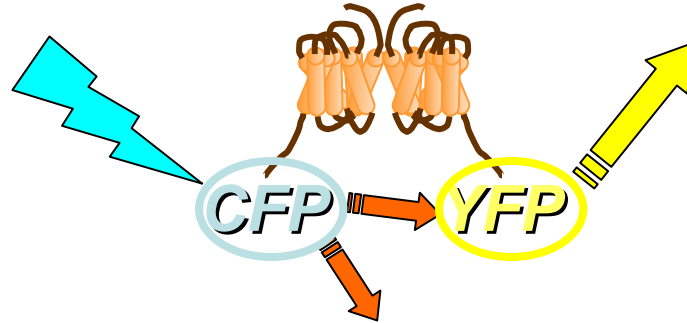


Green: CTx → Marker for rafts
Red: CCR5wt/CCR5mut/TfR
Yellow: overlay

No differences in the membrane distribution

CCR5^{I52V/V150A} does not dimerize: FRET

Fluorescence Resonance Energy Transfer

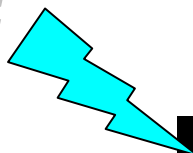


within 20-100Å (0.002-0.01µm): prot-prot interactions

CCR5^{I52}V/V¹⁵⁰A does not dimerize: FLIM

Fluorescence Lifetime IMaging

PULSE

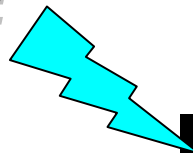


CFP



DETECTION/TIME = LIFETIME

PULSE



CFP

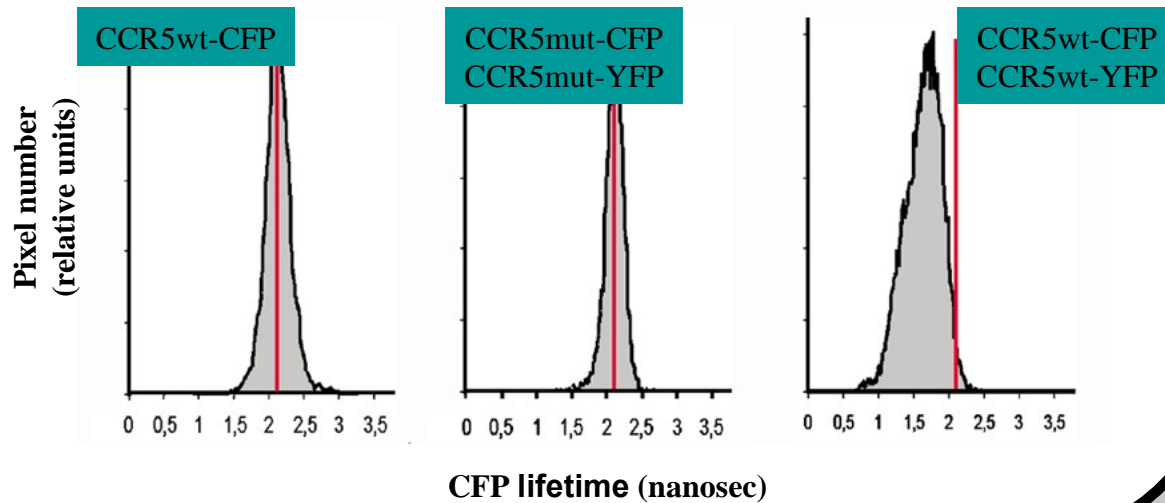
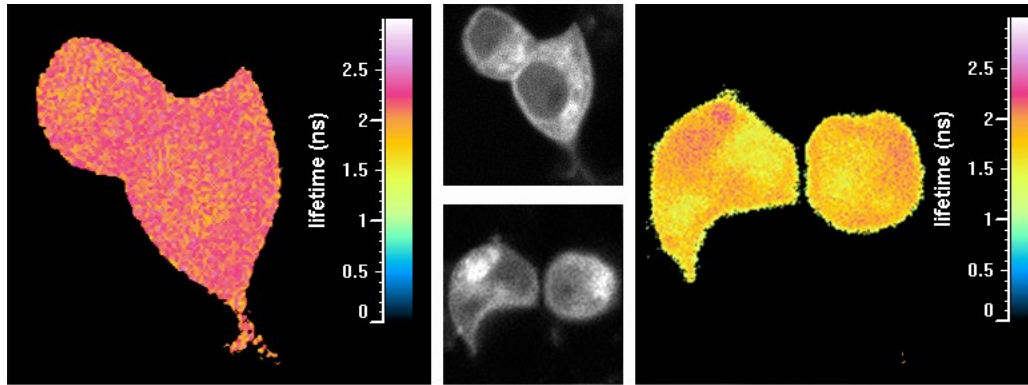


YFP



DECREASED LIFETIME

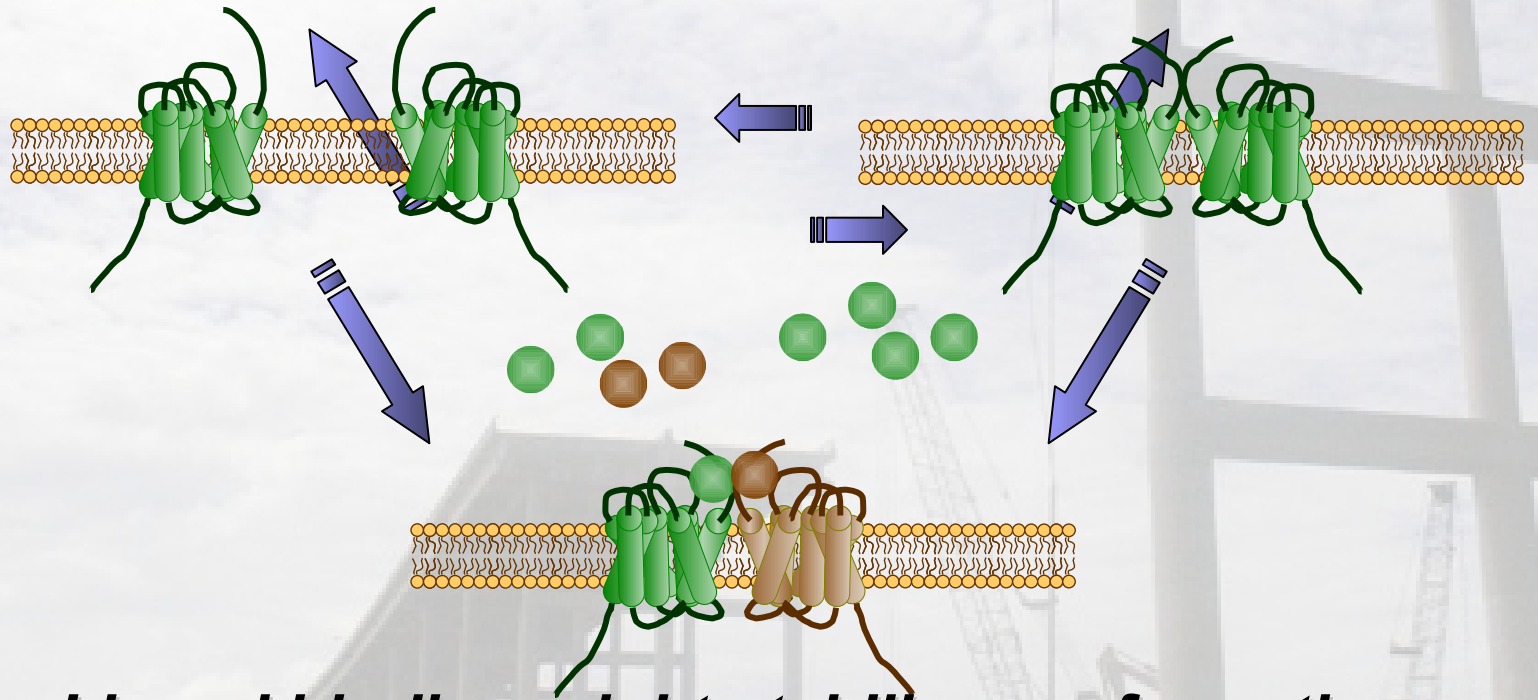
CCR5^{I52V/V150A} does not dimerize: FLIM



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



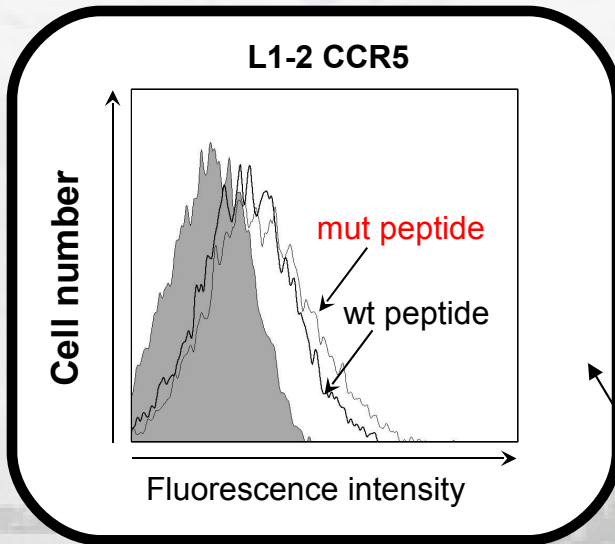
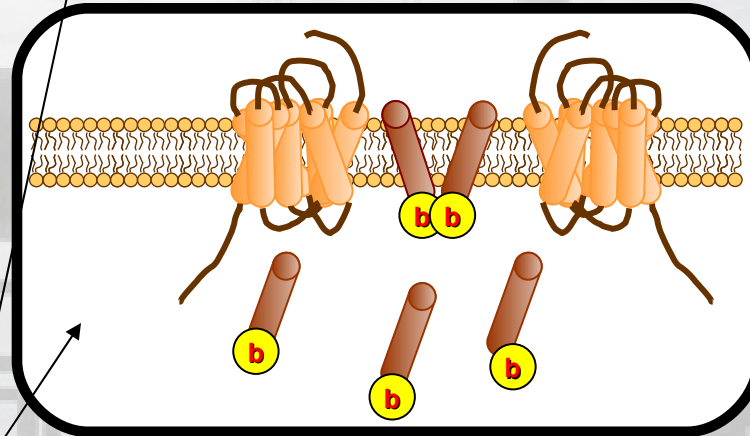
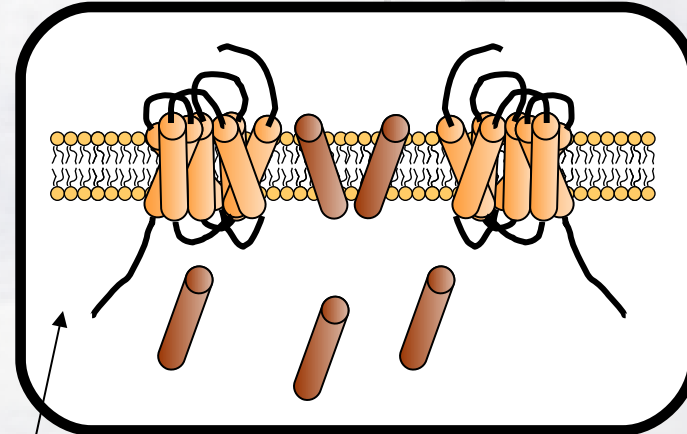
CCR5-based synthetic peptides containing I⁵² and V¹⁵⁰ block CCR5 function

wt peptide

MLVILIL + VTSVITW

mutant peptide

MLVVLIL + VTS^AITW



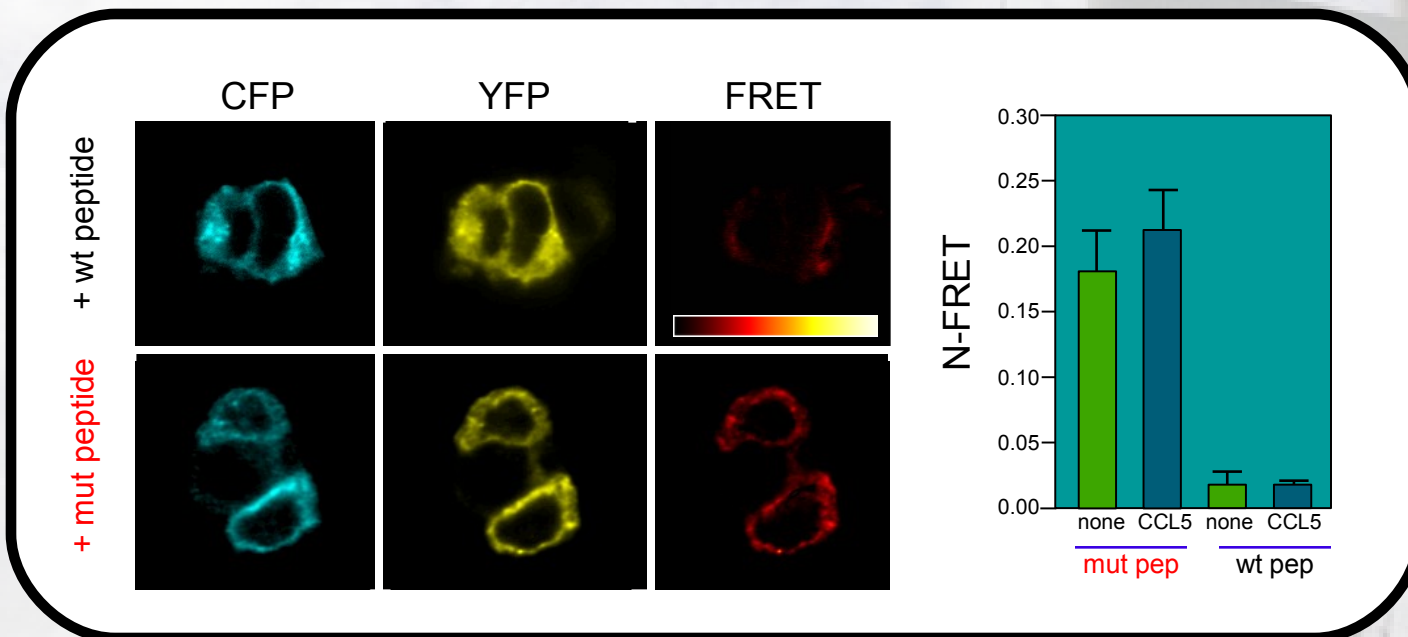
BOTH INCLUDED IN THE MB!

BIOTIN STAINING.

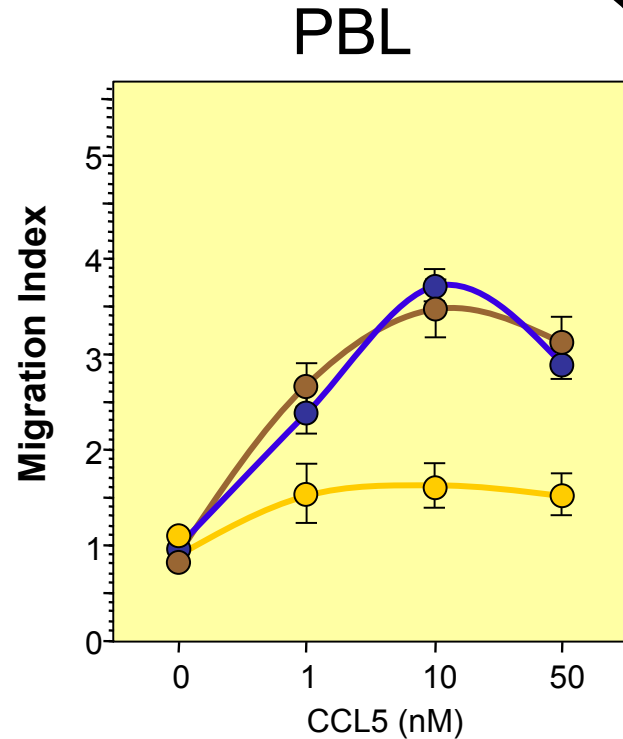
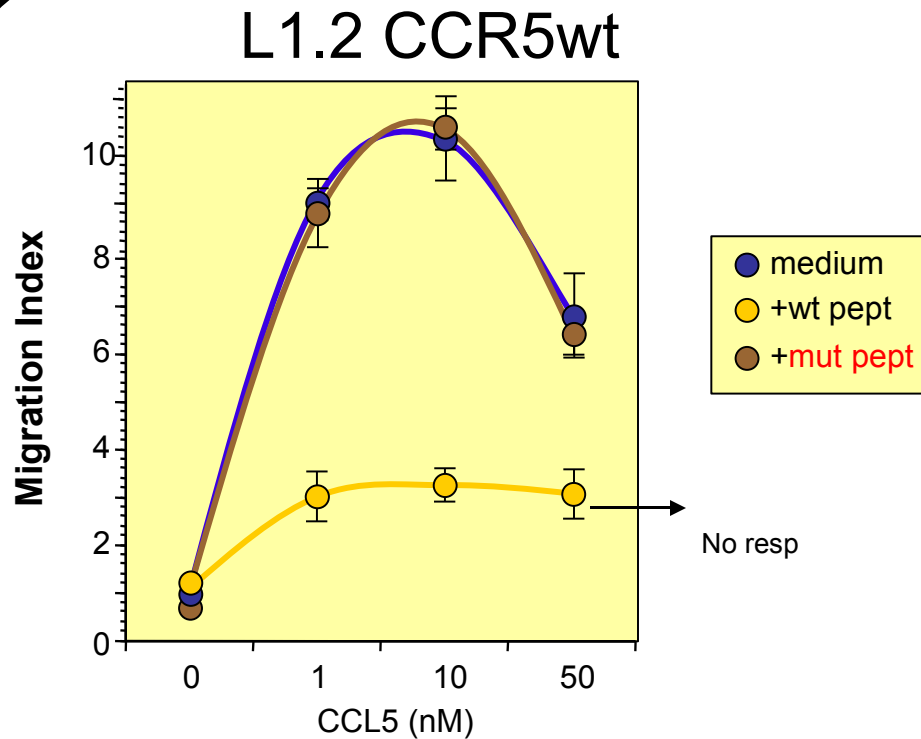
CCR5-based synthetic peptides containing I⁵² and V¹⁵⁰ block CCR5 function, dimerization

wt peptide
MLVILIL + VTSVITW

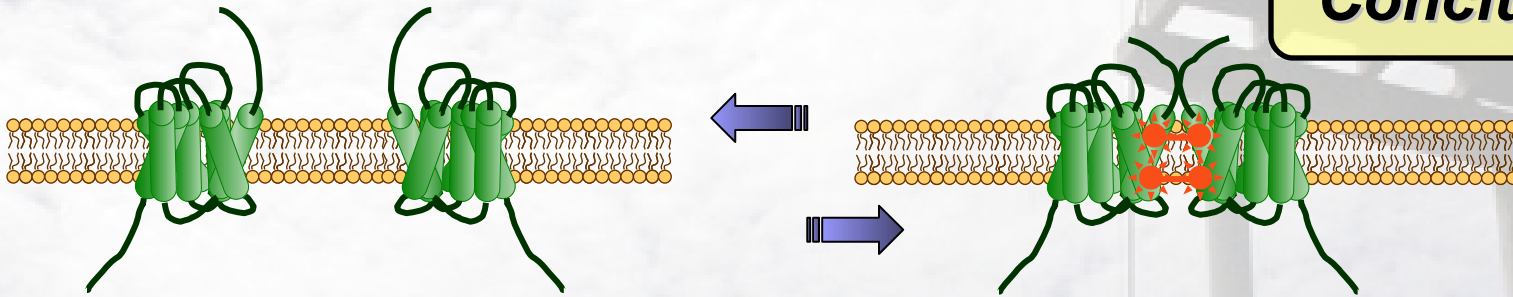
mutant peptide
MLVVLIL + VTS^AITW



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

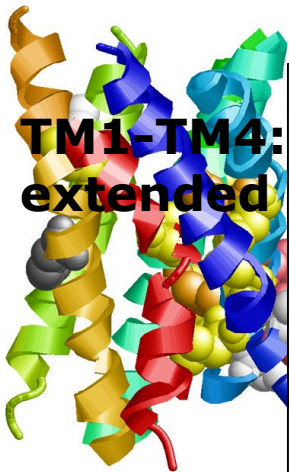


PBL: peripheral blood lymphocytes.



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

A



B

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

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

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 Juan¹, Mario Mellado², José Miguel Rodríguez-Frade², Patricia Hernanz-Falcón², Antonio Serrano², Antonio del Sol³, Alfonso Valencia¹, Carlos Martínez-A² and Ana María Rojas^{1,2*}.

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

Predicting Functional Residues in Protein Sequence Alignments



Simple Run

Advanced Run

More Info

Contact Us

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 [\[i\]](#).

FASS

Principal component analysis of the multiple alignment and computation of the statistical confidence in the organization of the family into sub-families [\[i\]](#).

S-Method

The level entropy method searches for the level of splitting of a previously established phylogenetic tree to optimize the relative entropy [\[i\]](#).

SQUARE

An additional tool to evaluate alignments [\[i\]](#).



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!



Centro Nacional
de Investigaciones Oncológicas

Alfonso Valencia
Rest of the group