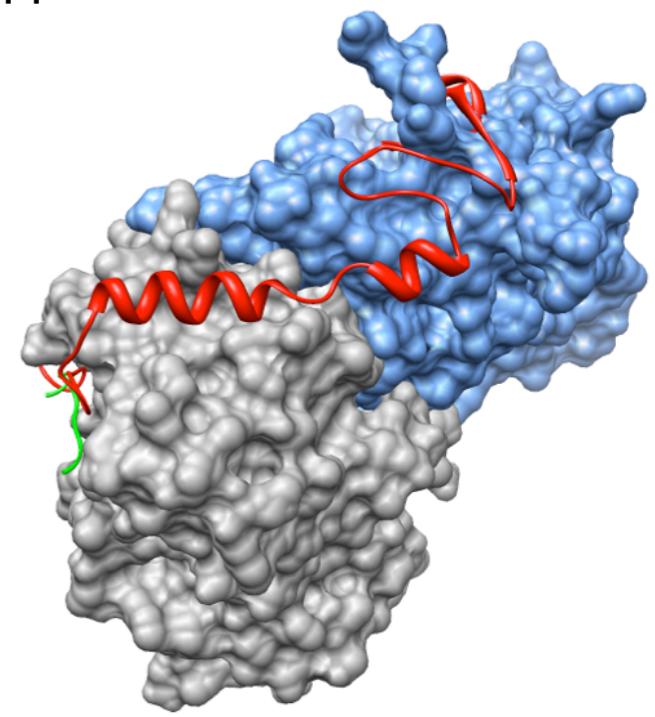
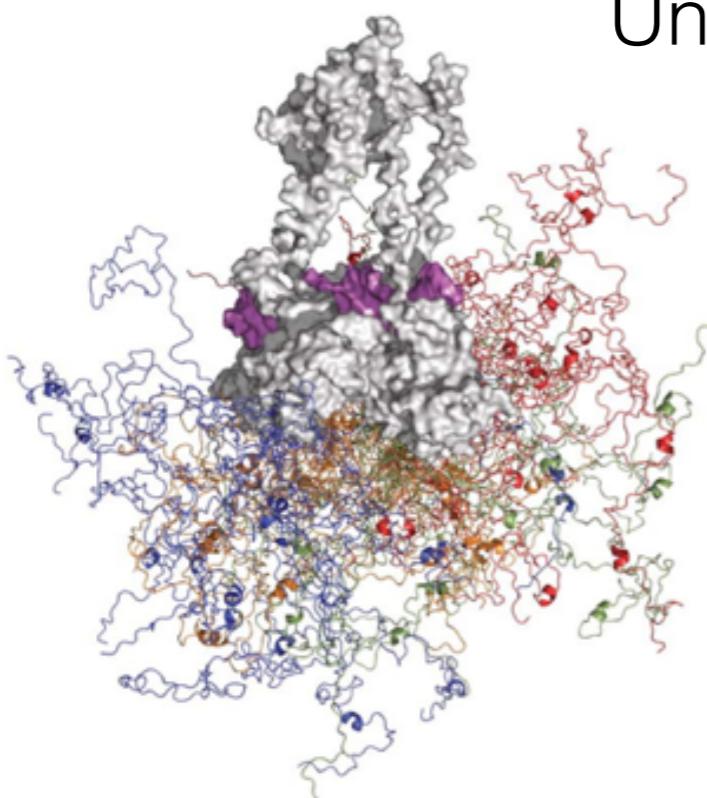


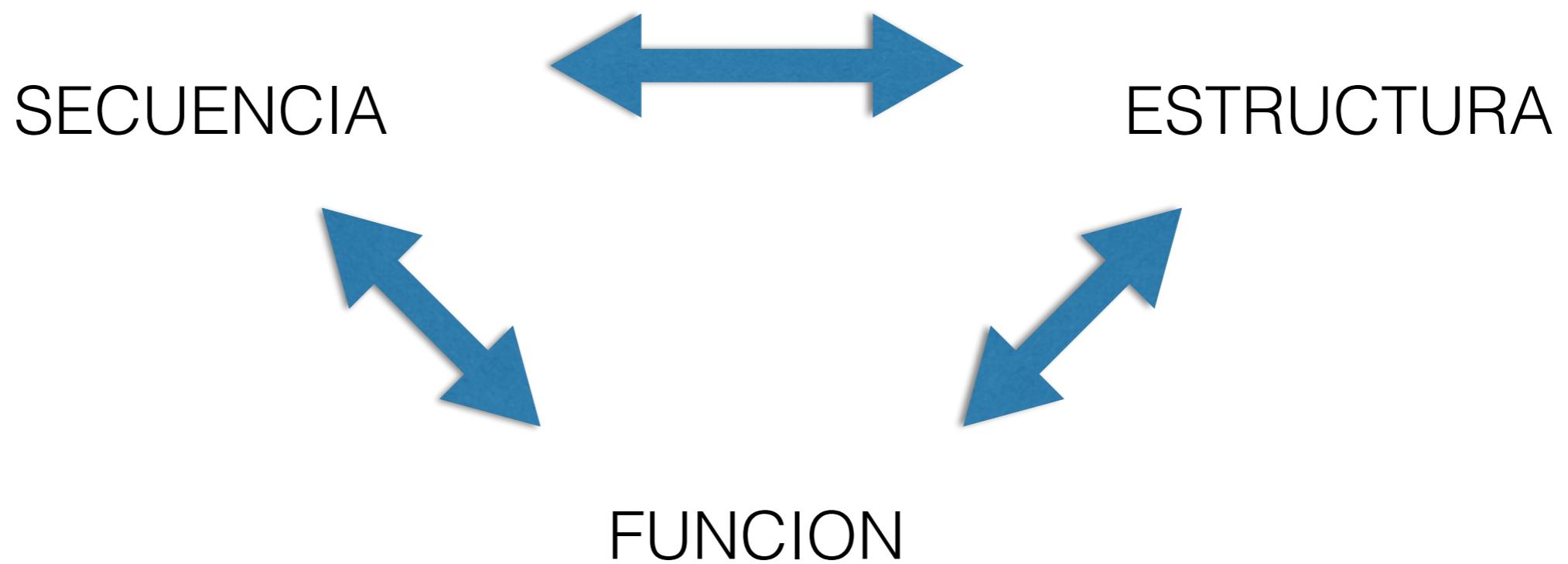
PREDICCION DE DESORDEN Y MOTIVOS FUNCIONALES EN PROTEINAS

Prof. Lucía Chemes

Instituto de Investigaciones Biotecnológicas
Universidad Nacional de San Martín



PROTEINAS



Las secuencias y las estructuras de las proteínas evolucionan y la información contenida en ambas nos dice mucho acerca su función

PROTEINAS

actualmente contamos con un gran número de secuencias y estructuras, que nos permiten elucidar regiones funcionales en las proteínas



Bases de datos de secuencias

~80.000.000



(500.000 curadas)



Bases de datos de estructuras

~160.000

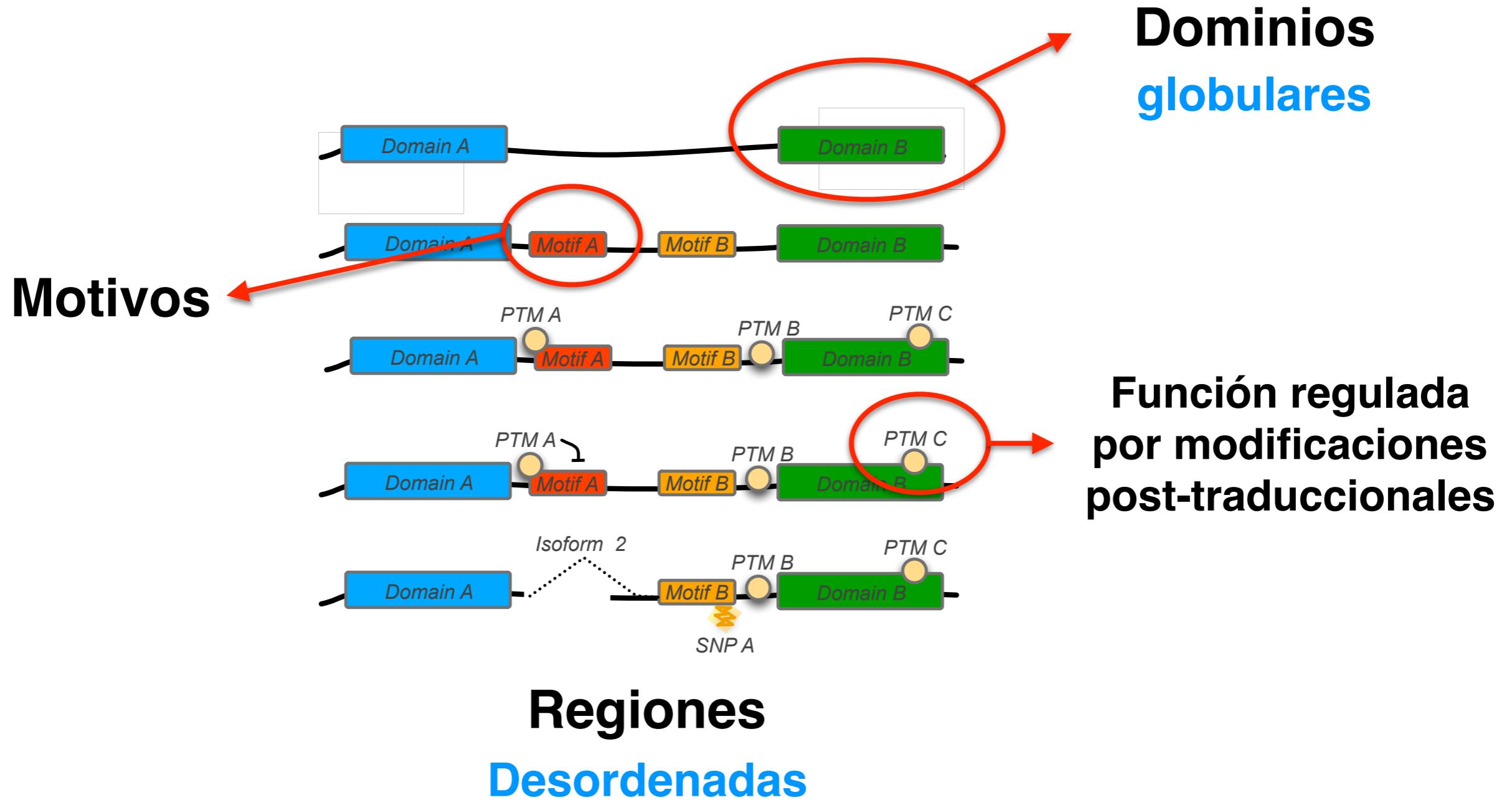


Bases de datos de regiones desordenadas

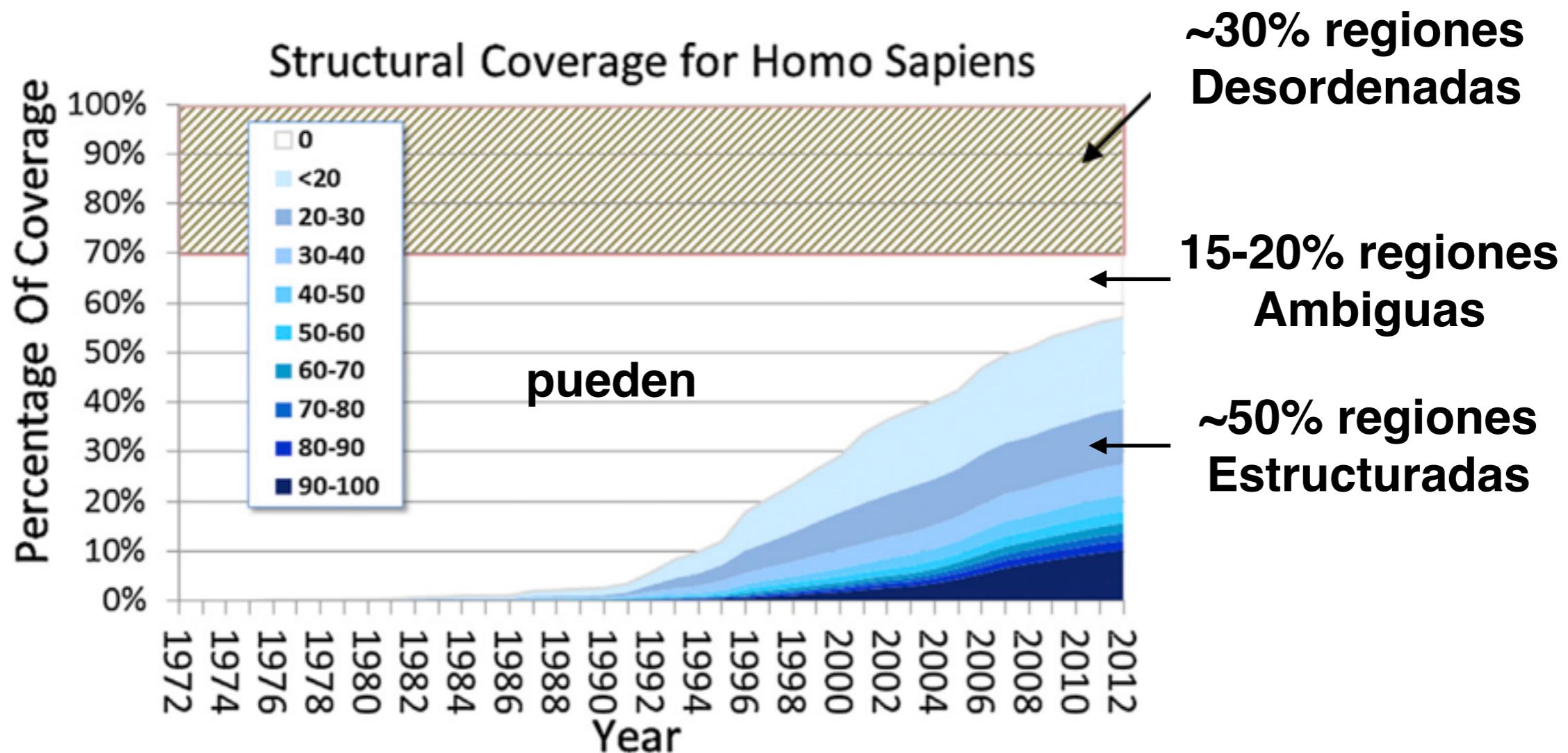
~3.000



Las proteínas tienen una organización MODULAR. Tanto las regiones globulares como las desordenadas tienen funciones importantes

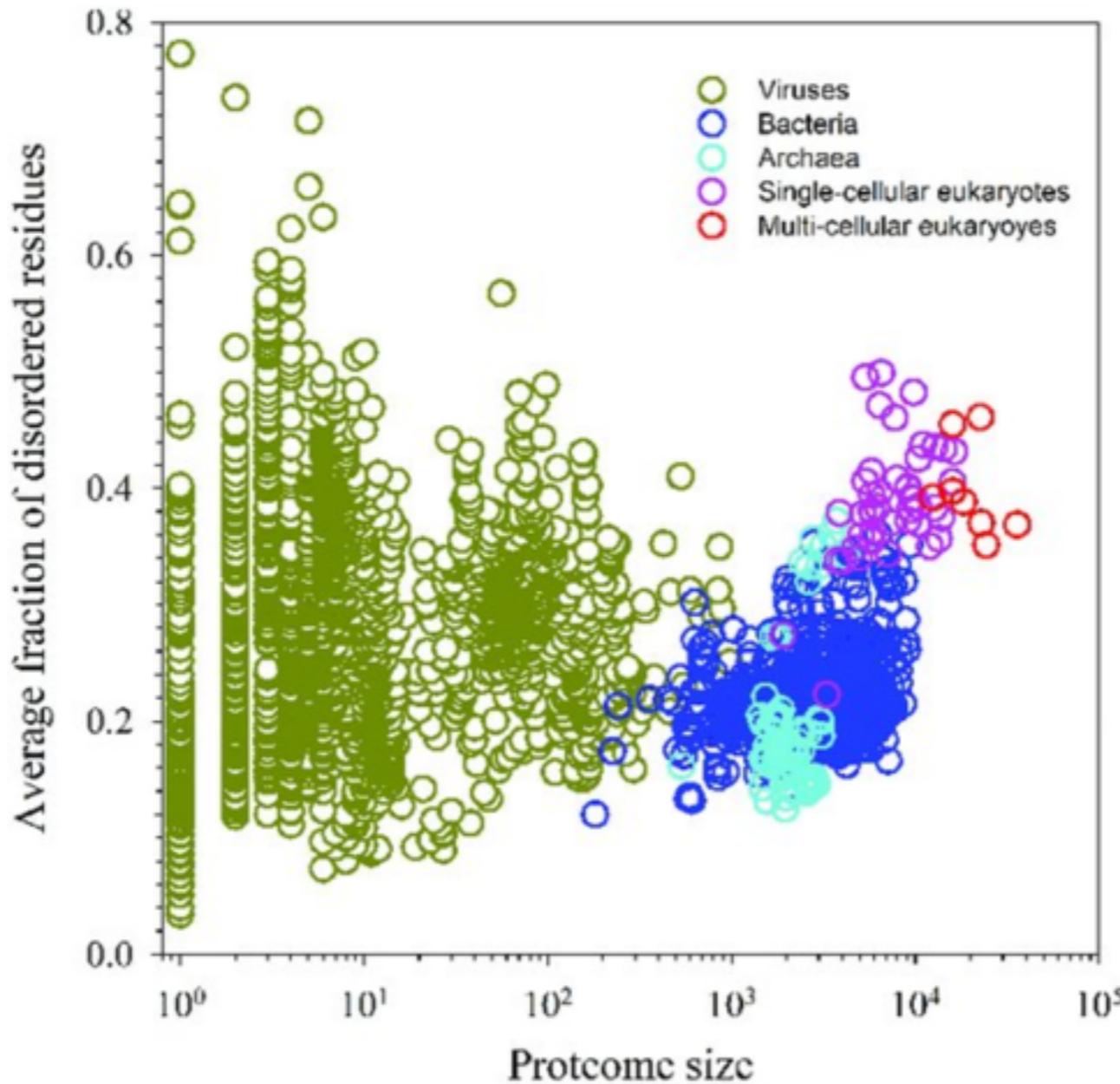


Qué fracción del proteoma humano corresponde a regiones estructuradas versus desordenadas?



En azul se muestra la fracción de aminoácidos que pueden asignarse a estructuras conocidas de la PDB dependiendo del %identidad de secuencia

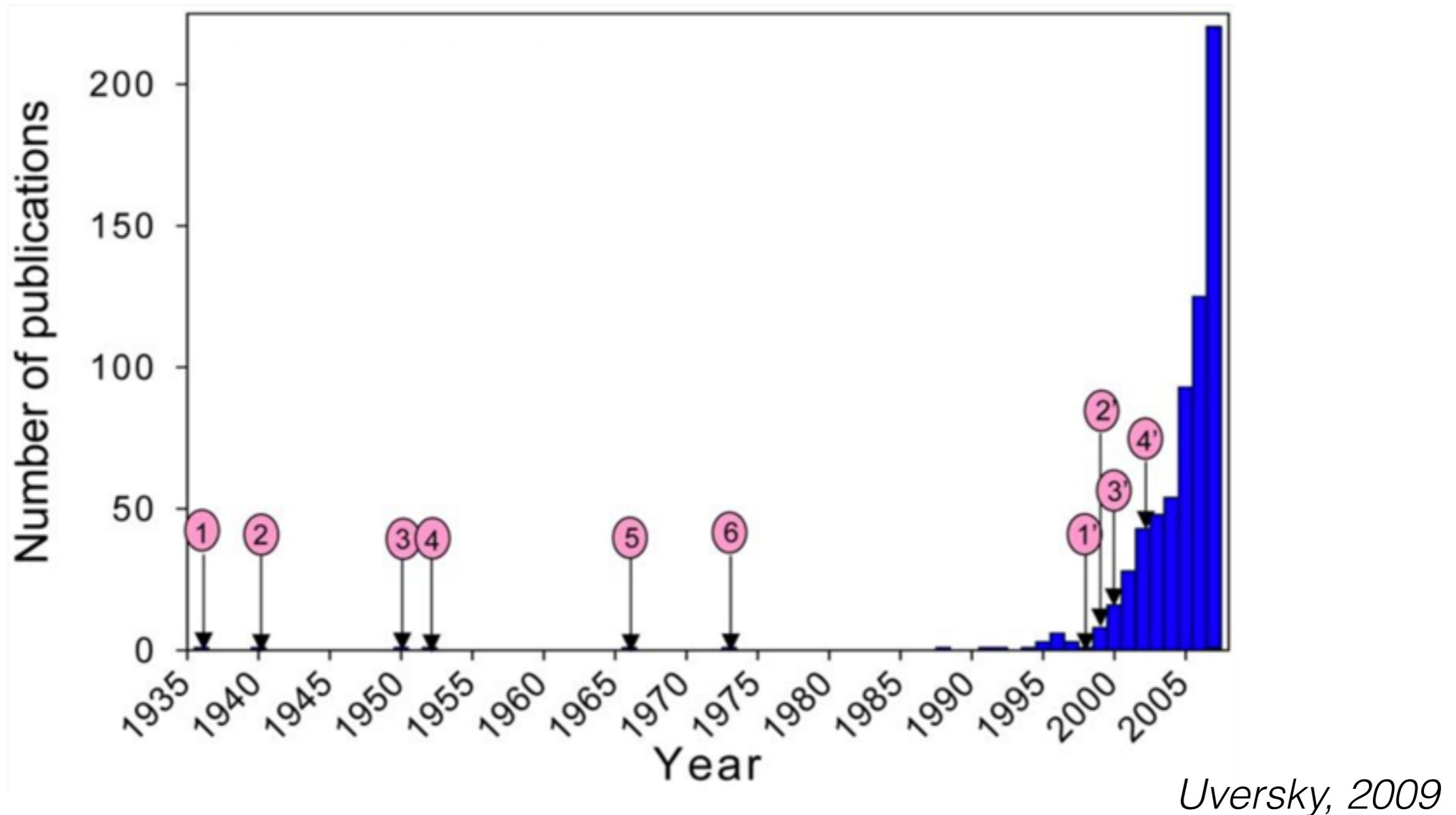
Todos los proteomes tienen la misma cantidad de desorden?



Virus 5-60%
Bacteria ~ 20%
Archaea ~ 20%
Eucariotas ~ 40%

Los organismos eucariotas, y los virus que infectan a eucariotas, tienen mayores proporciones de desorden en sus proteomas

El desorden en proteínas se empezó a estudiar hace
muy poco...



Palabras claves: intrinsically disordered, natively unfolded, intrinsically unstructured, intrinsically unfolded, intrinsically flexible



Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm

Peter E. Wright* and H. Jane Dyson*

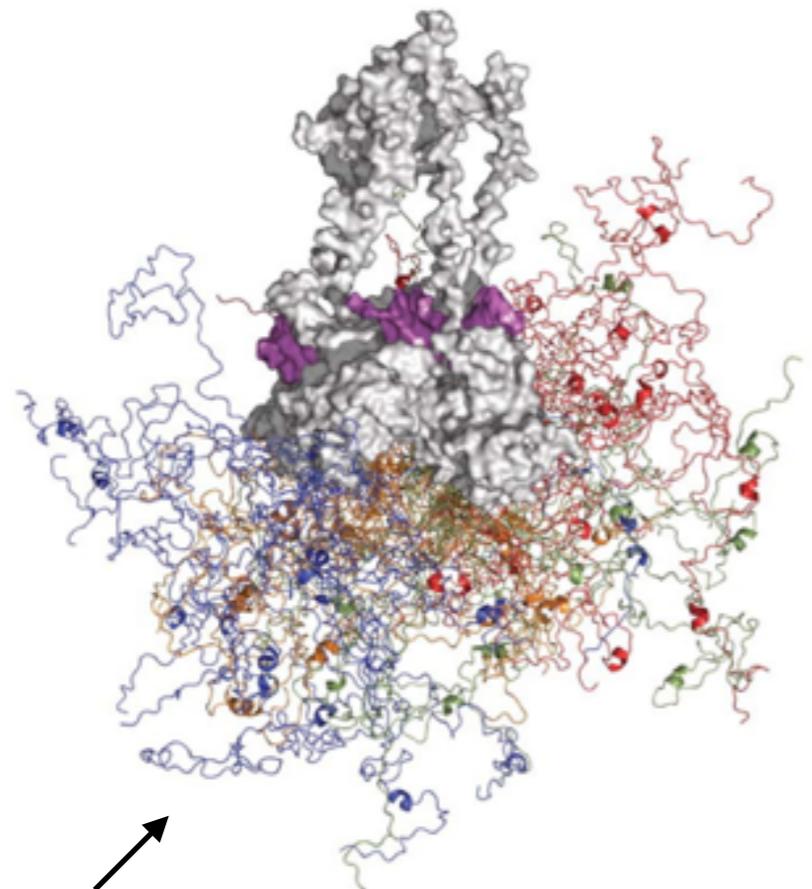
Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA 92037, USA

A major challenge in the post-genome era will be determination of the functions of the encoded protein sequences. Since it is generally assumed that the function of a protein is closely linked to its three-dimensional structure, prediction or experimental determination of the library of protein structures is a matter of high priority. However, a large proportion of gene sequences appear to code not for folded, globular proteins, but for long stretches of amino acids that are likely to be either unfolded in solution or adopt non-globular structures of unknown conformation. Characterization of the conformational propensities and function of the non-globular protein sequences represents a major challenge. The high proportion of these sequences in the genomes of all organisms studied to date argues for important, as yet unknown functions, since there could be no other reason for their persistence throughout evolution. Clearly the assumption that a folded three-dimensional structure is necessary for function needs to be re-examined. Although the functions of many pro-

Cómo definimos una región o proteína intrínsecamente desordenada?

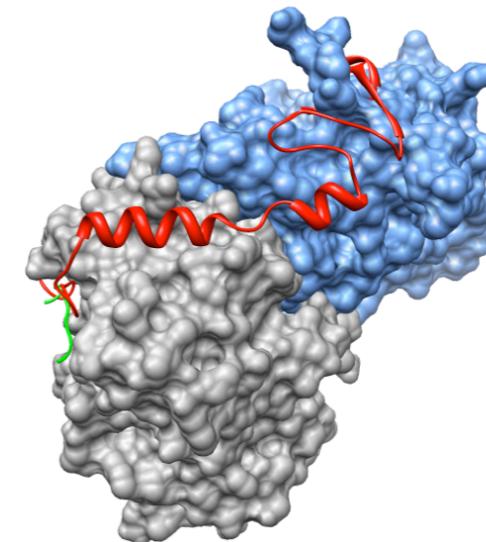
IDP: proteína intrínsecamente desordenada

IDR: región intrínsecamente desordenada (10-500 res)

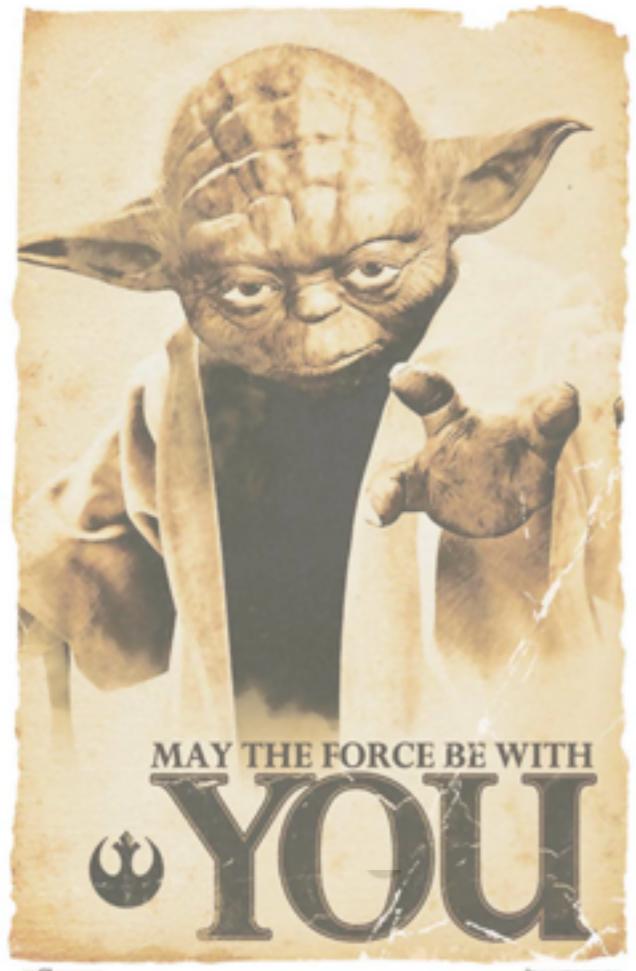


Regions que no adquieren estructura 3D definida en el estado nativo

Pueden adquirir estructura al unirse a dominios globulares o entre sí

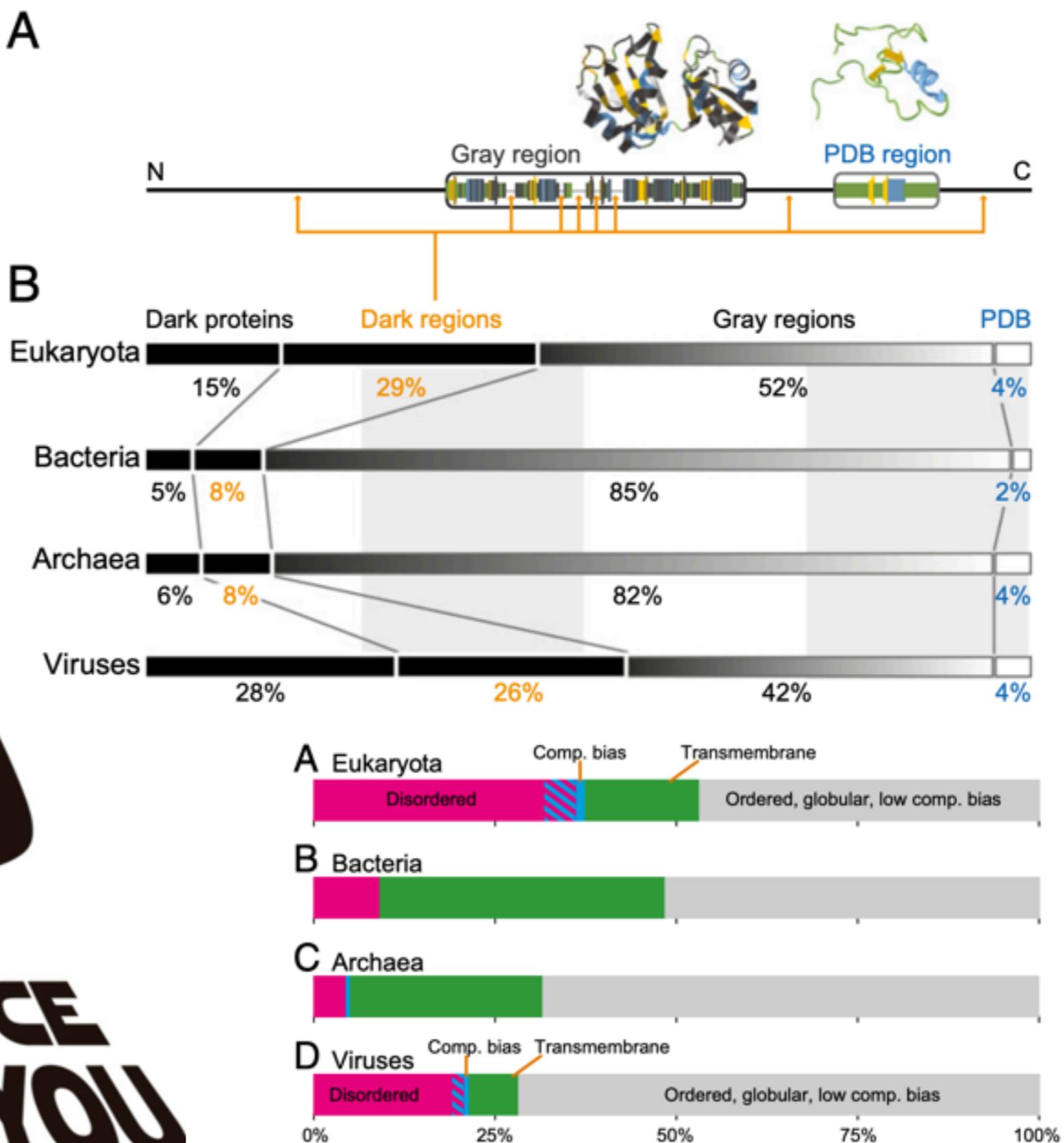


Existen como un ensamble de conformaciones



Unexpected features of the dark proteome

Nelson Perdigão^{a,b}, Julian Heinrich^c, Christian Stolte^c, Kenneth S. Sabir^{d,e}, Michael J. Buckley^c, Bruce Tabor^c, Beth Signal^d, Brian S. Gloss^d, Christopher J. Hammang^d, Burkhard Rost^f, Andrea Schafferhans^f, and Seán I. O'Donoghue^{c,d,g,1}

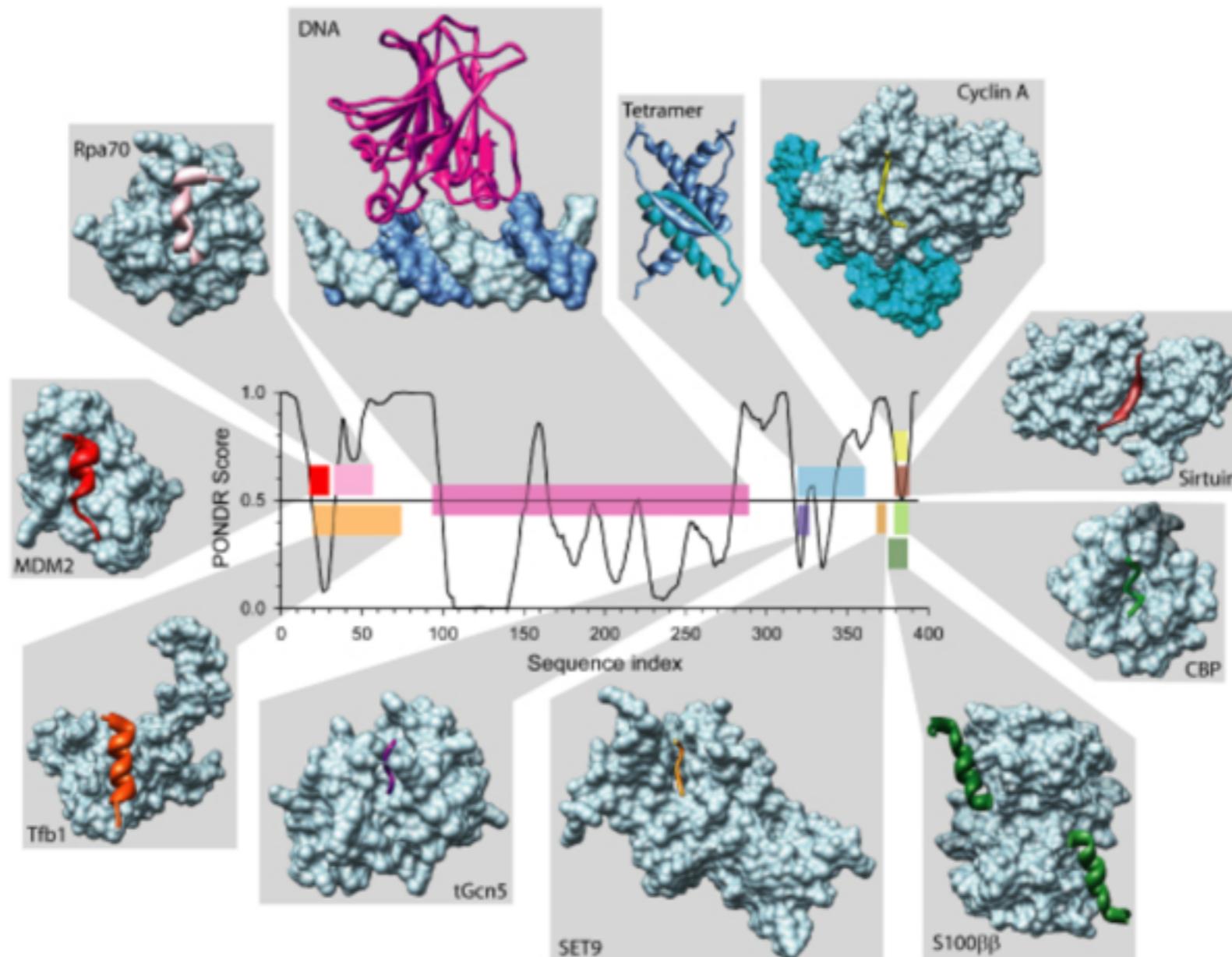


Perdigao, PNAS 2016

p53 es la proteína desordenada más estudiada

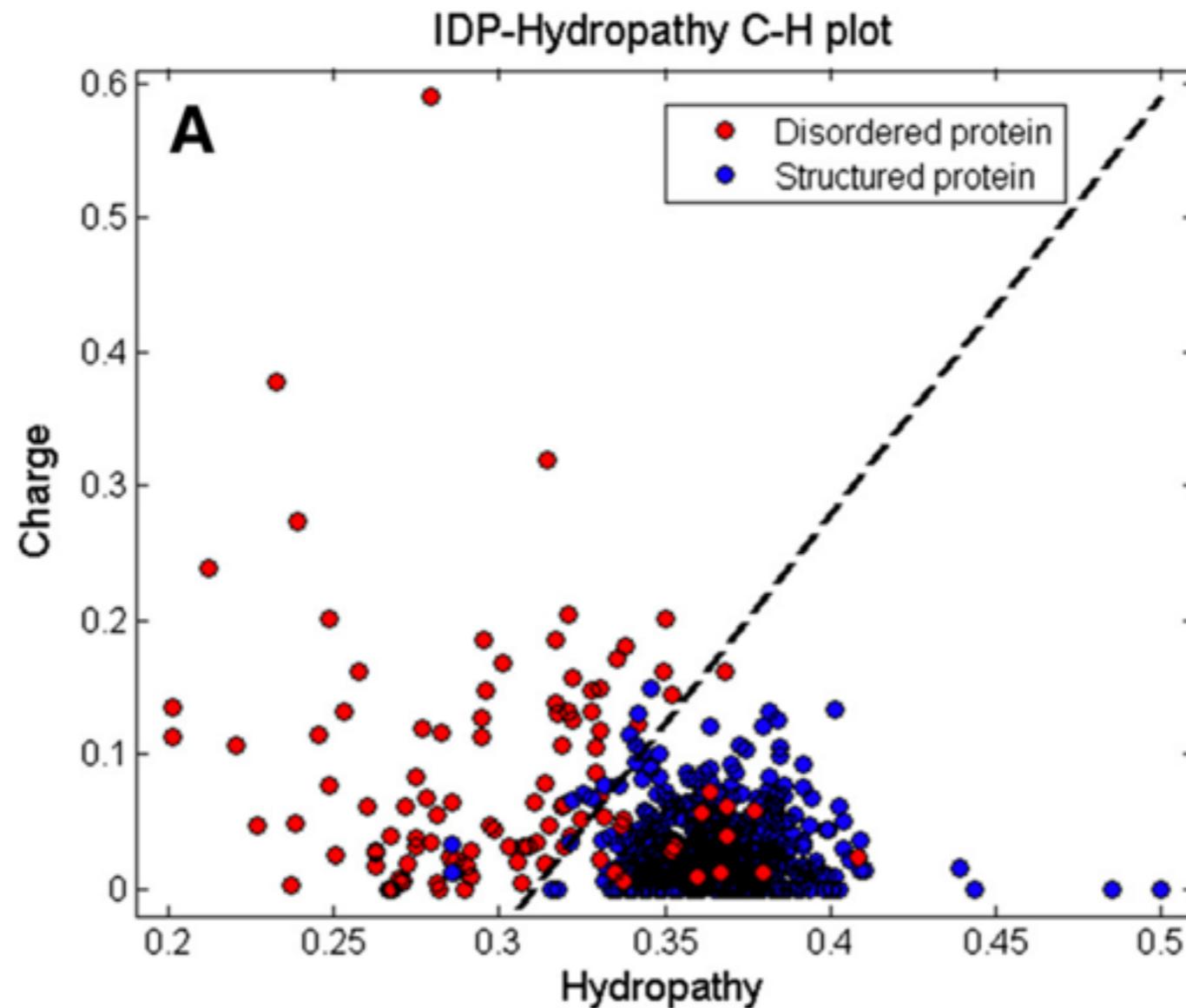
muchas IDPs son importantes en cáncer y señalización

***p53
altamente
mutada en
cáncer humano***



Regiones N-terminal y C-terminal desordenada que puede formar muchas interacciones diferentes

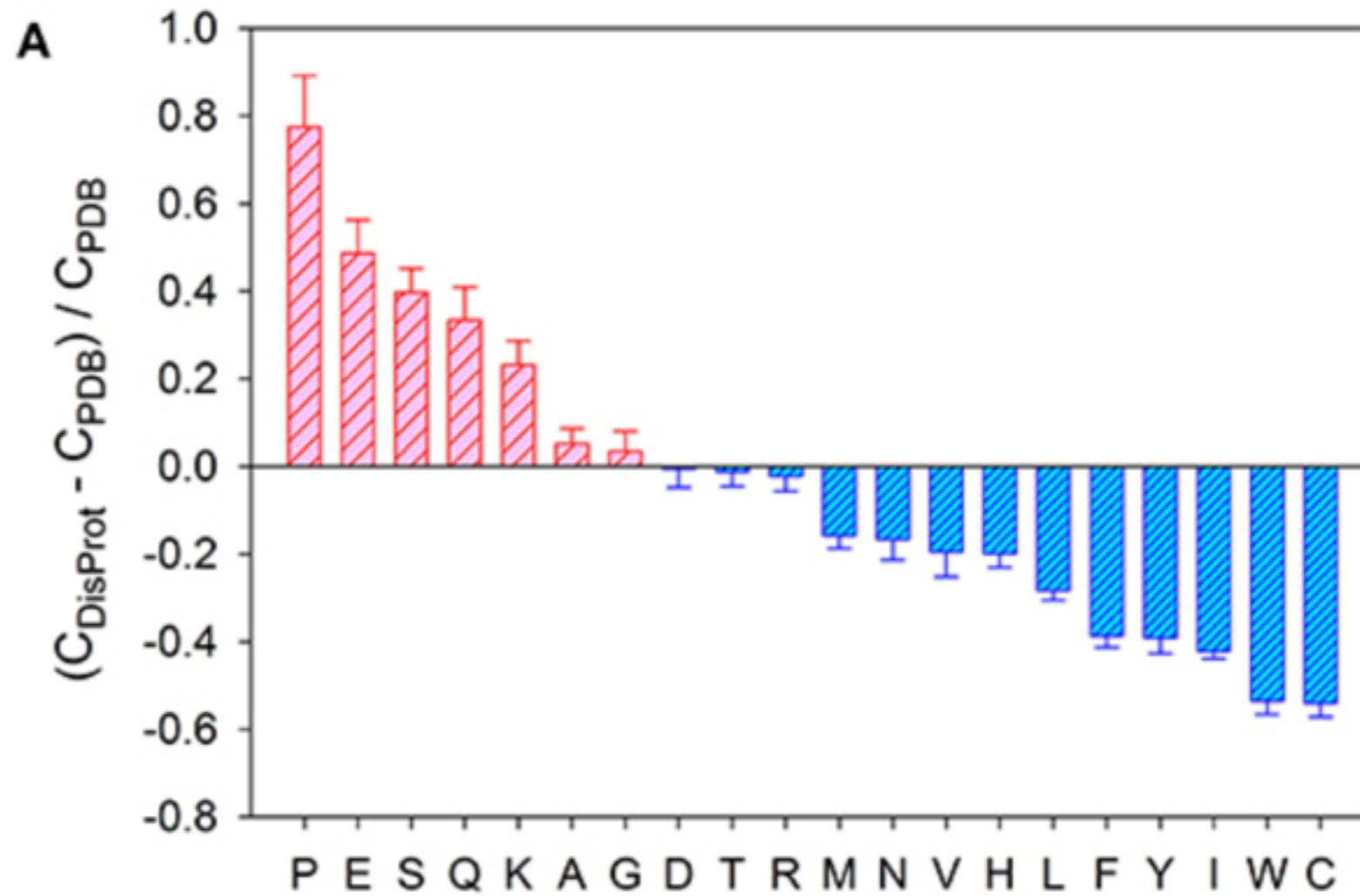
Existen propiedades de secuencia que distinguen a las regiones desordenadas?



Uversky, Proteins (2000) 41:415

Las regiones desordenadas son ricas en aminoácidos cargados y están depletadas en aminoácidos hidrofóbicos

Existen propiedades de secuencia que distinguen a las regiones desordenadas?



estas regiones tienen un *bias* (desviación) composicional

- Alta proporción de Prolina
- Alta proporción de residuos polares (Ser, Glu, Asp, Lys, Arg)
- Baja proporción de residuos hidrofóbicos (Trp, Tyr, Cys, Phe, Leu, Met, Val)

Las regiones desordenadas se caracterizan por

- Tener regiones de “Baja Complejidad” o ‘low complexity regions’ LCRs
- También definidas como regiones con un *bias* composicional

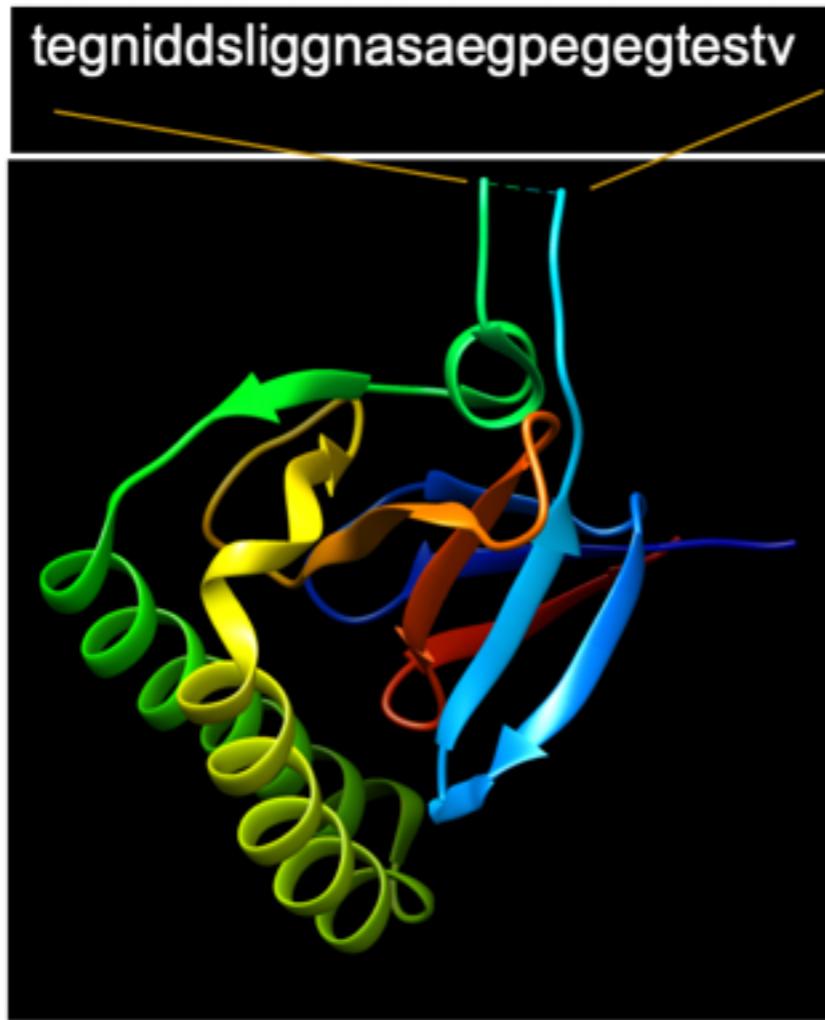
LCRs

Regiones con menor variabilidad de aminoácidos que una secuencia promedio

Repeticiones de un mismo aminoácido (polyQ)

Repeticiones de motivos cortos: dipéptidos, tripéptidos: (CAG) n

Cómo se detecta experimentalmente el desorden?



XRay

regiones faltantes
“missing residues”

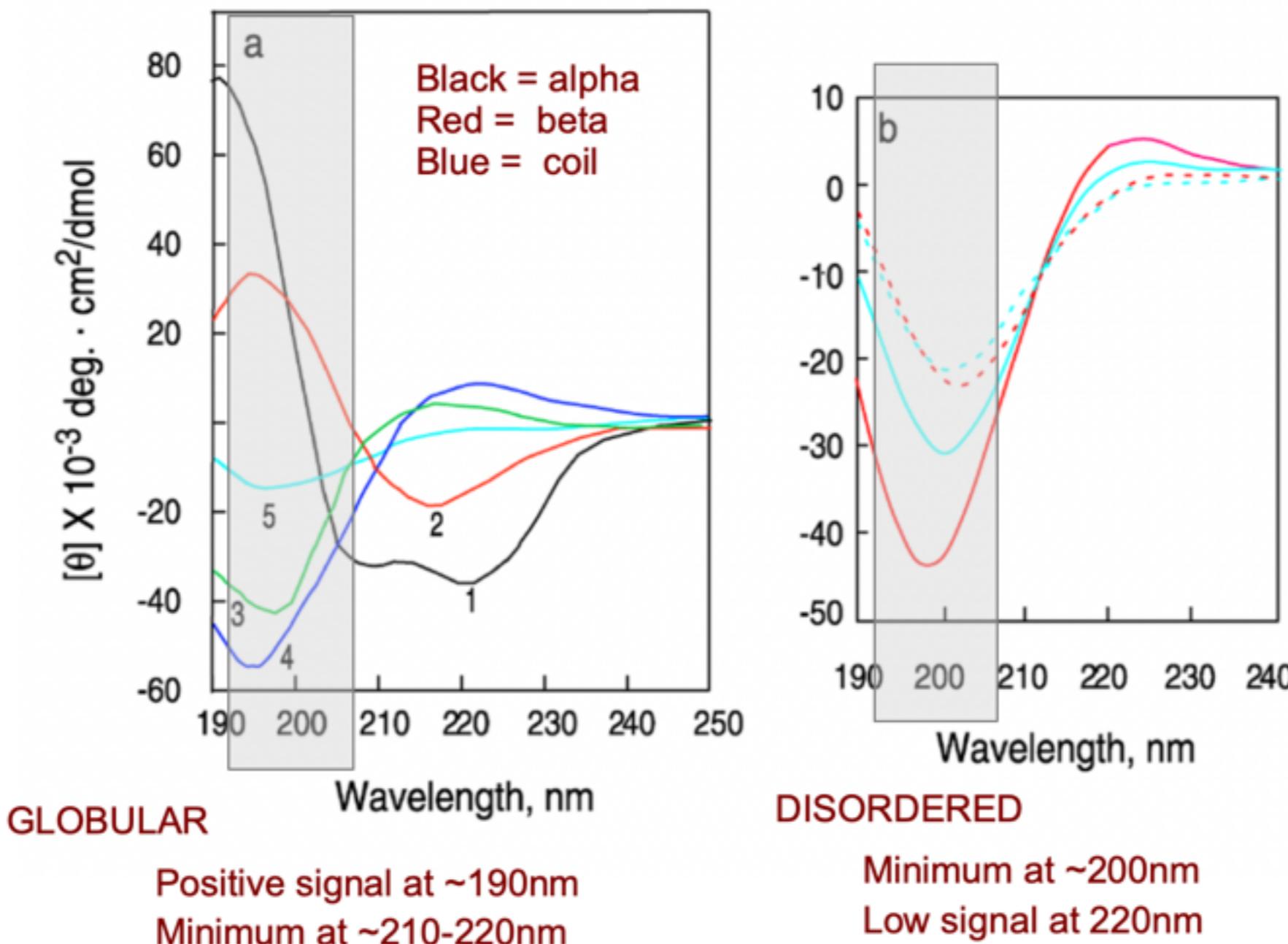


RMN

regiones flexibles
(con muchas conformaciones)

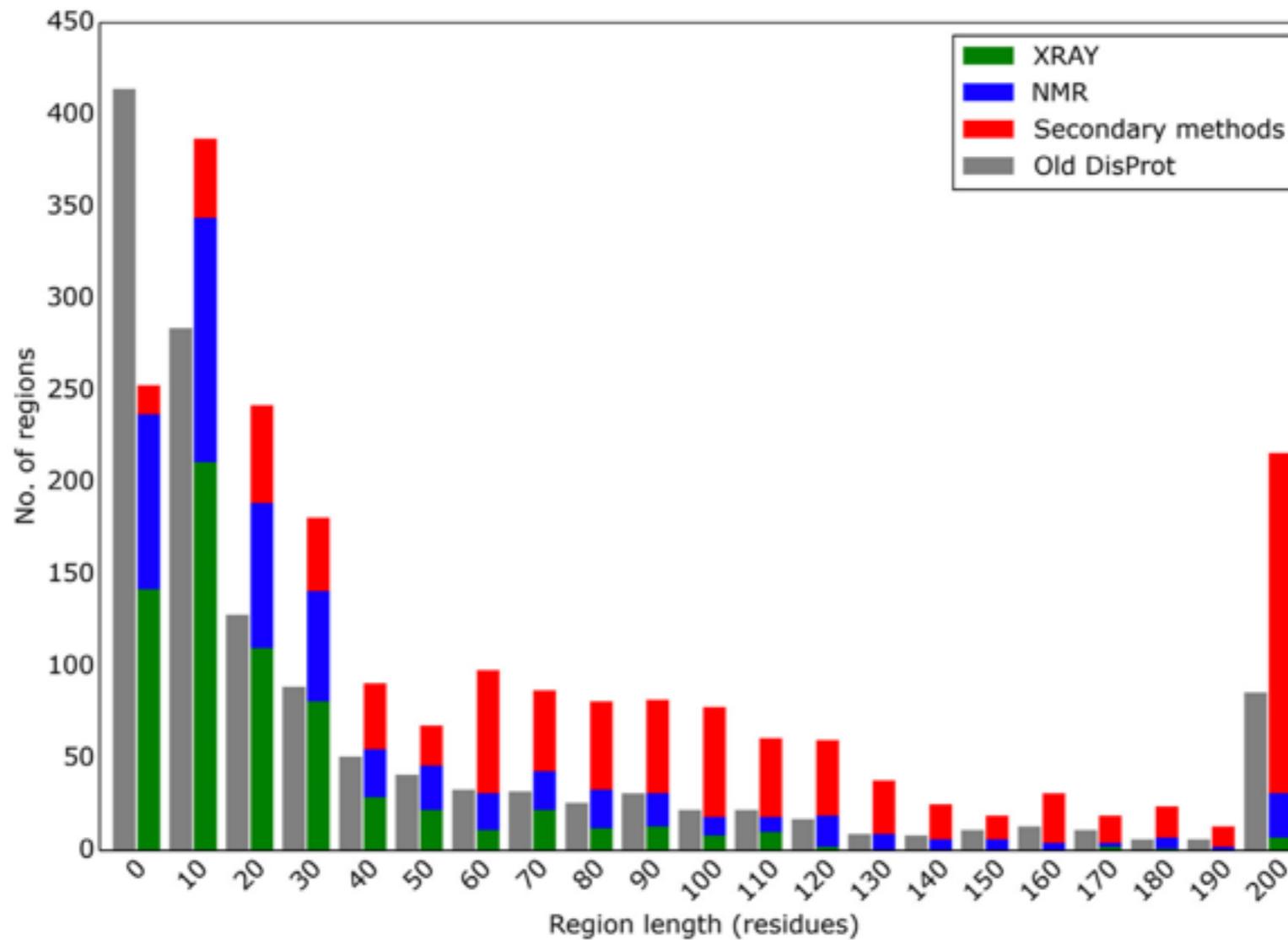
a veces es imposible cristalizar

Falta de estructura secundaria regular en solución



Métodos secundarios
Dicroísmo Circular

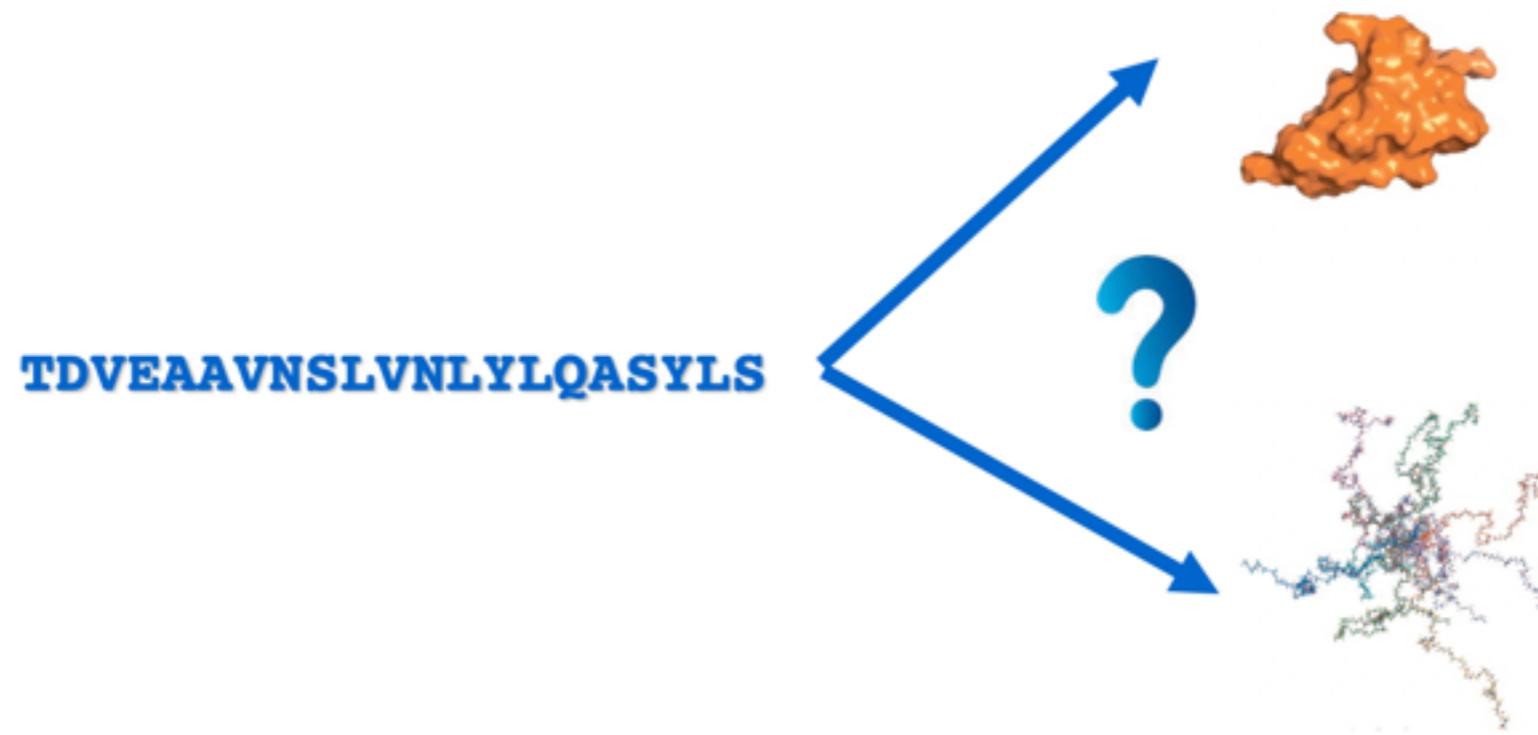
Métodos utilizados para detectar desorden



Métodos secundarios: Espectrometría de masa, SAXS, Dicroísmo Circular

Estos sets de datos nos sirven para entrenar predictores

Predicción de regiones desordenadas



a partir de una secuencia aminoacídica:
cómo predecimos si es globular o desordenada?

Existen diferentes métodos para predecir desorden

Exactitud ~80%

pueden detectar diferentes “tipos” de desorden (regiones largas vs cortas)

- Métodos basados en propiedades biofísicas o composición aminoacídica

GlobPlot, IUPred, FoldIndex, UCON, TOP-IDP

- Métodos basados en Machine Learning

PONDR, VSL2, Disopred, DisEMBL, RONN

Sets de entrenamiento:

Estructuras globulares de la PDB

Residuos “missing” o “faltantes” de la PDB, RMN

Bases de datos experimentales de desorden: DISPROT

Problemas:

Los sets de entrenamiento son pequeños

Los sets de entrenamiento no son balanceados

Métodos basados en Machine Learning

DISOPRED PONDR

Trained in missing residues
from X-ray structures

Trained with NMR data

SVM with linear kernel

Assign label: D or O

.....AMDDMLSPDIEQWFTED.....

Raw profile from PSI-BLAST Log File

Position-based scoring matrix used																							
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V				
-3	-4	-4	-4	-3	-4	-4	-4	-2	-1	-1	-4	-1	8	-5	-3	-3	0	2	-2				
0	-1	-1	3	-4	3	4	1	-1	-4	-4	0	-3	-4	-2	-1	-2	-4	-3	-3				
0	-1	2	1	-3	4	0	-1	-2	-4	-3	1	-2	-4	-2	2	0	-4	-3	-3				
-2	-3	-4	-5	-2	-3	-4	-6	-4	0	6	0	0	-1	-4	-3	-2	-4	-2	0				
0	-3	-1	-2	-3	0	-2	4	-3	-3	0	-2	-2	-4	-3	3	1	-4	-4	-3				
0	2	0	4	1	2	1	-2	-4	-4	0	-3	-4	-3	1	-2	-5	-4	-4					
-1	5	3	-2	-4	-1	-1	1	-2	-1	-4	1	-3	-4	-3	1	-2	-5	-4	-4				
-2	-3	-4	-5	-3	-3	-4	-5	-4	3	4	-1	1	2	-4	-3	-2	-3	-1	0				
-2	3	2	-2	-4	2	1	-3	-2	-3	-3	1	1	-4	-3	2	1	-4	-3	-1				
0	2	3	1	-4	0	0	0	-2	-4	-4	1	-3	-4	-3	2	0	-5	-4	-4				
5	-3	-3	-3	-2	-3	-3	-2	-3	1	-2	-3	-2	1	-3	0	1	-4	-2	0				
-1	-4	-5	-5	-3	-4	-4	-5	-4	3	3	-4	2	3	-5	-3	-2	5	-1	2				
0	3	3	0	-4	3	0	1	-2	-4	-4	1	-3	-4	-3	1	-1	-4	-3	-4				
-1	0	1	0	-4	1	-1	-1	-2	-4	-3	5	-2	0	-3	0	-2	-4	0	-3				
-2	-3	-1	-5	-3	-3	-4	-5	-4	3	4	0	4	2	-4	-3	-2	-3	-2	0				
0	3	0	-2	-3	-1	0	0	-2	0	0	1	0	-1	-3	2	0	-4	-3	0				
-1	1	3	-2	-4	0	-2	4	-2	-4	-4	0	-3	0	-3	0	0	-3	0	-4				



F(inp)

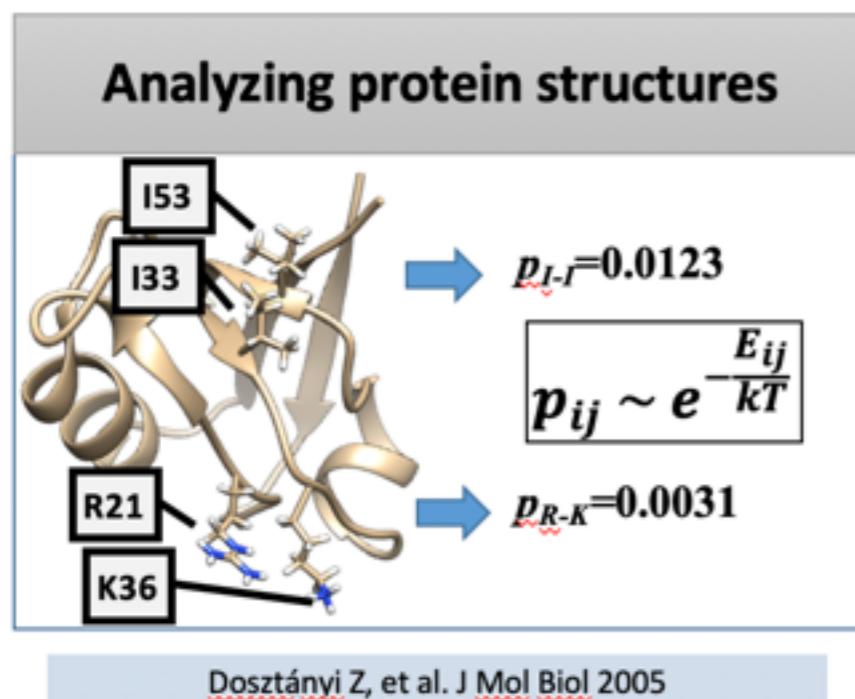


D O

Métodos basados en composición aminoacídica: IUPred

Las proteínas globulares forman interacciones estabilizadoras
Las proteínas IDPs no lo hacen

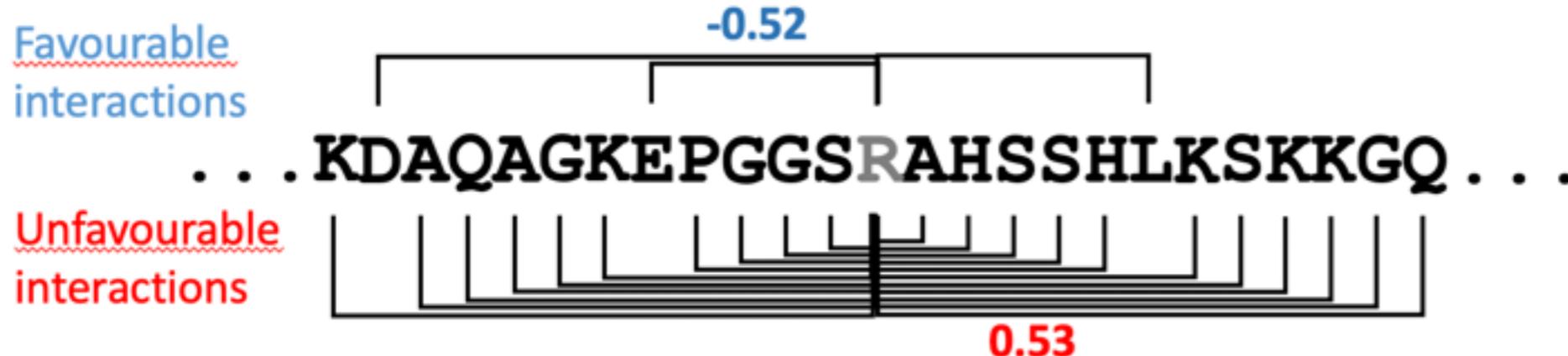
Construimos una función de energía que representa esta propiedad se basa en un potencial estadístico



Statistical force field

$$E = \begin{matrix} & R & E & I & \dots \\ R & 0.17 & -0.52 & 0.14 & \dots \\ E & -0.52 & 0.60 & 0.37 & \dots \\ I & 0.14 & 0.37 & -1.16 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{matrix}$$

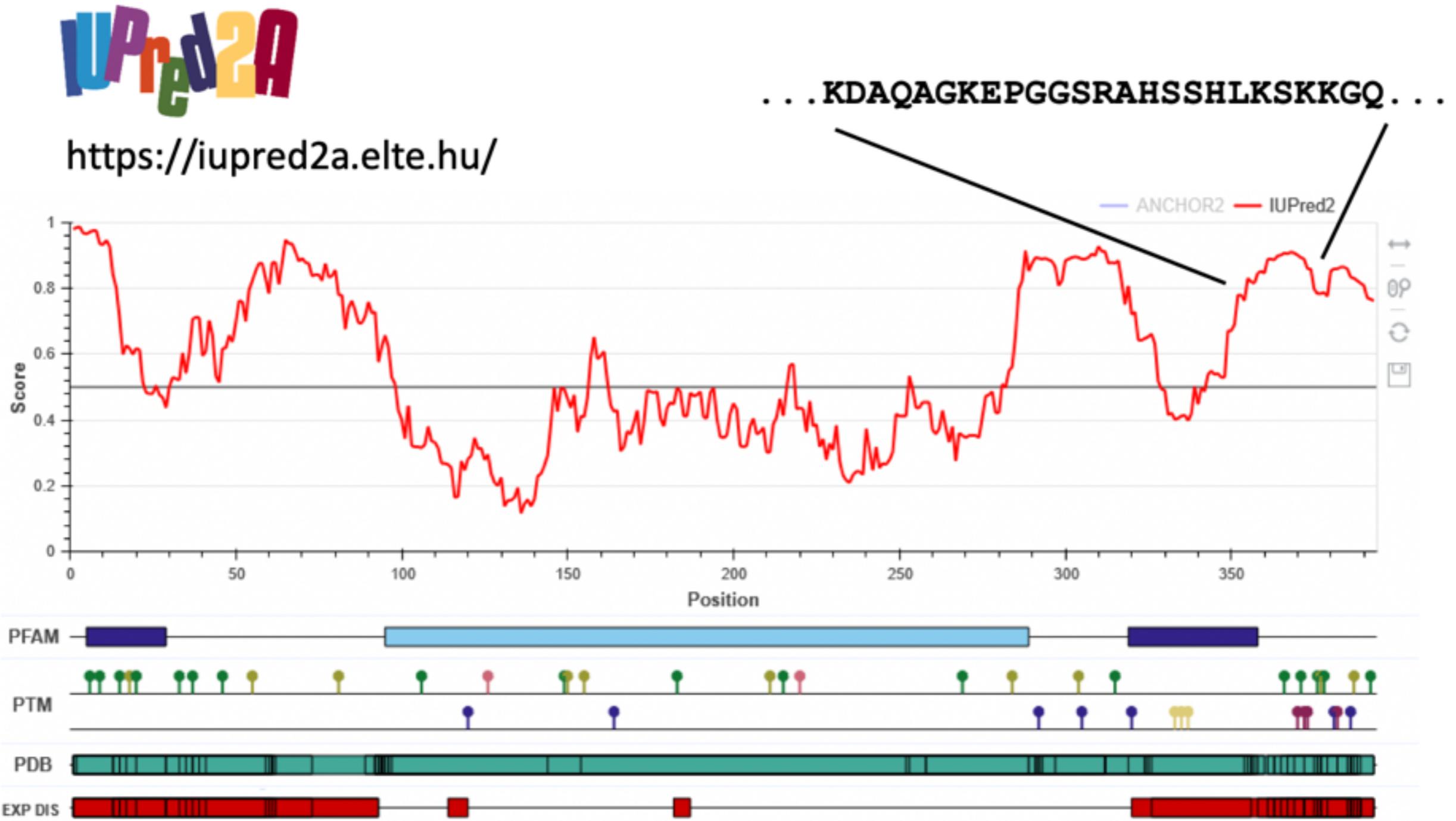
Boltzman distribution



se suman todas las contribuciones en una ventana X

Ejemplo de un output de IUPred

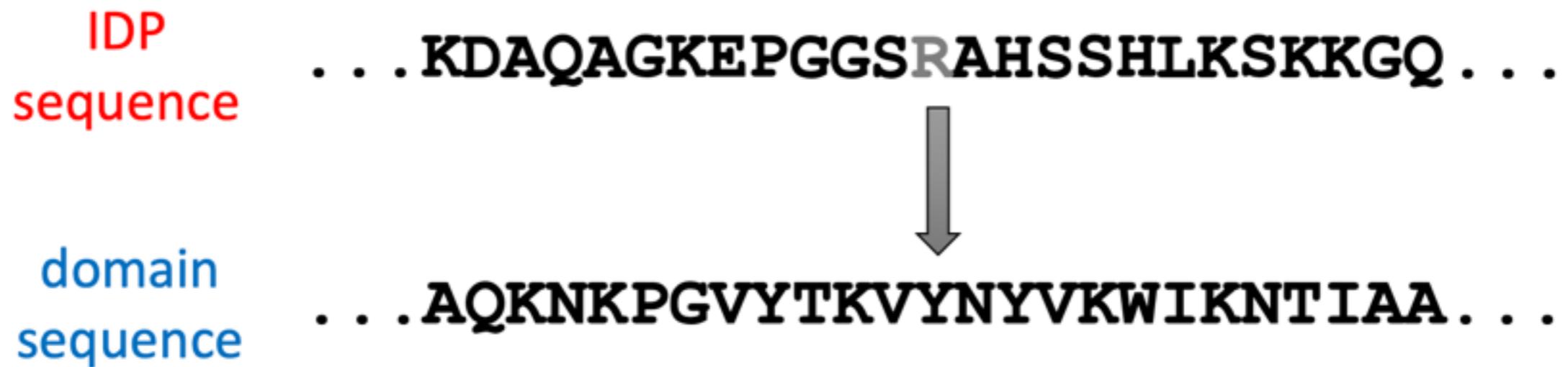
Score > 0.4-0.5 = región desordenada



Métodos basados en composición aminoacídica: ANCHOR

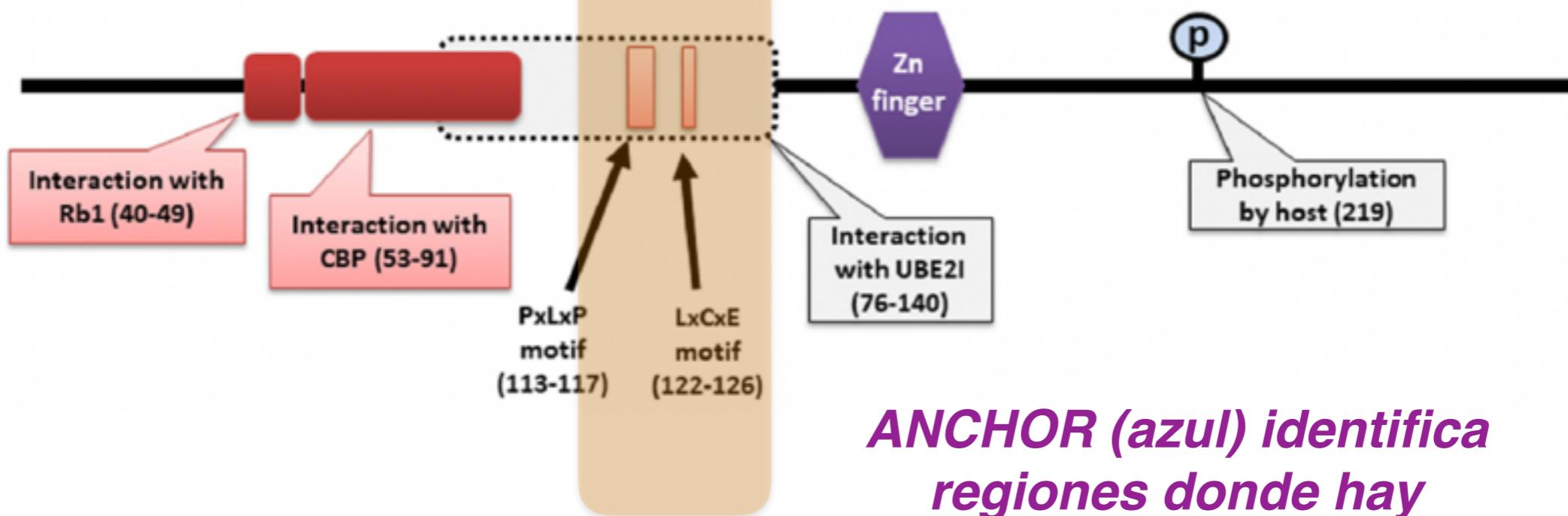
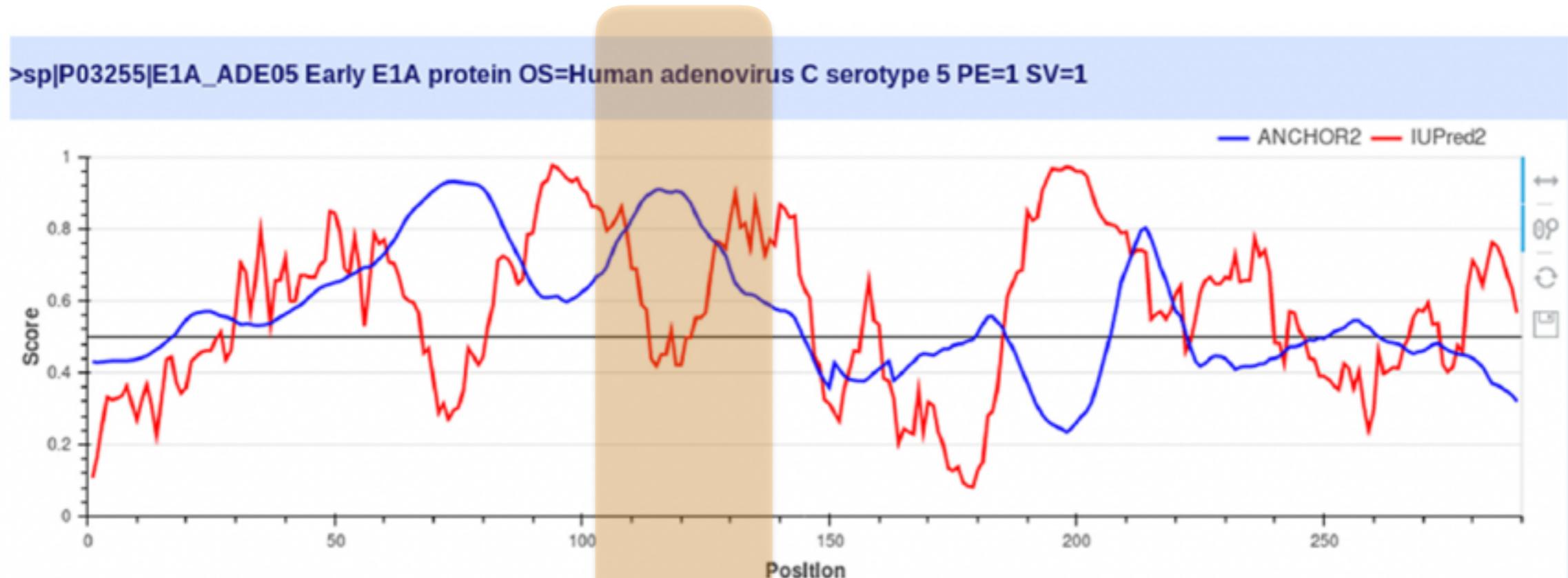
Los motivos de interacción dentro de regiones desordenadas
Deberían tener propiedades similares a un dominio globular

Representa un cambio de entorno del motivo



*Cuál es la ganancia energética de mover un residuo de un entorno desordenado a uno ordenado?
(simula lo que pasa cuando ocurre la interacción)*

Ejemplo de un output de IUPred + ANCHOR



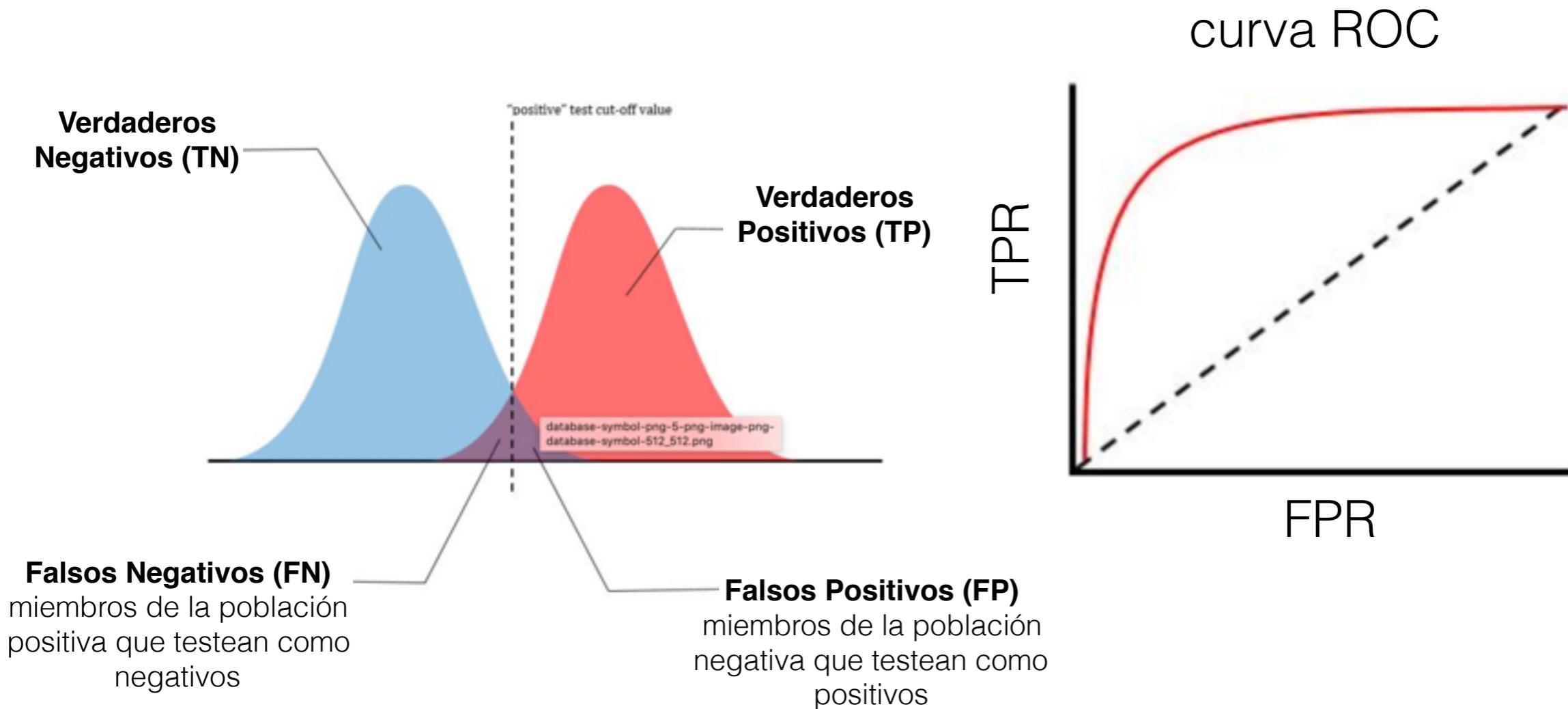
***ANCHOR (azul) identifica
regiones donde hay
motivos de unión***

Cómo medimos cuan “ bueno” es un método predictivo?

Evaluamos cuan bien puede separar los verdaderos positivos (TP) de los verdaderos negativos (TN)

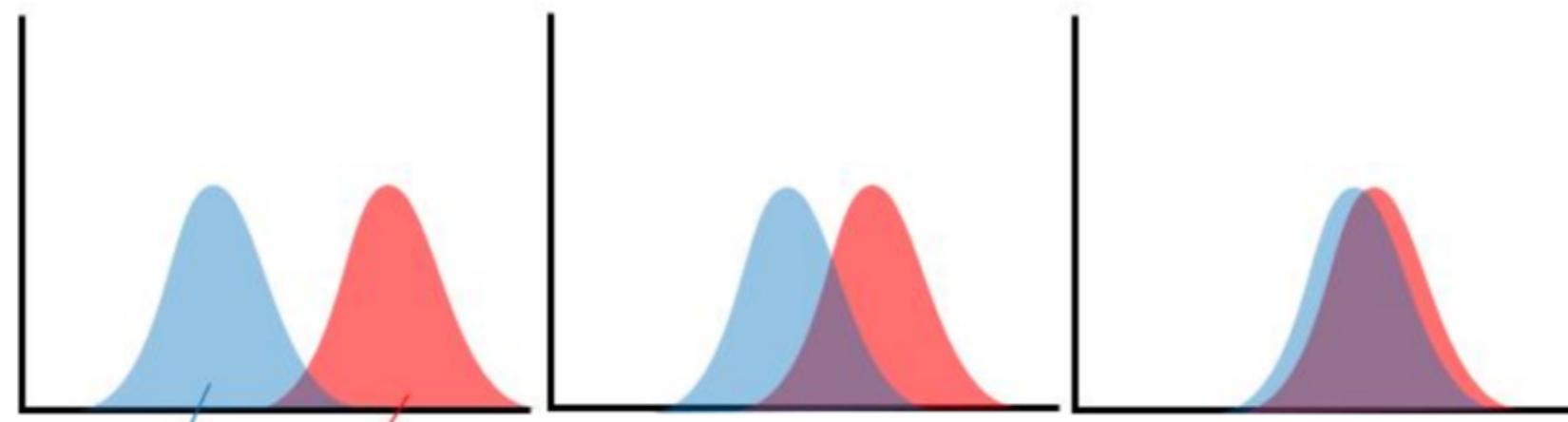
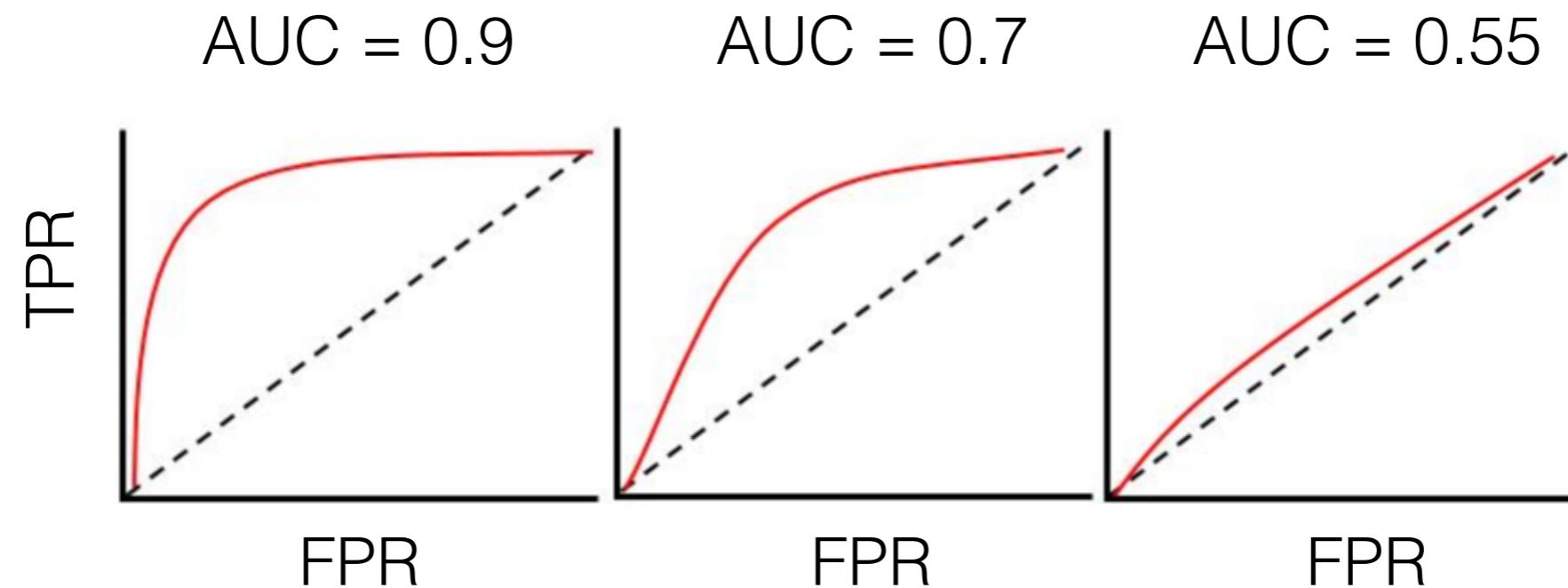
TP = regiones IDPs

TN = regiones globulares



Análisis de una curva ROC

AUC = 1
predictor
perfecto

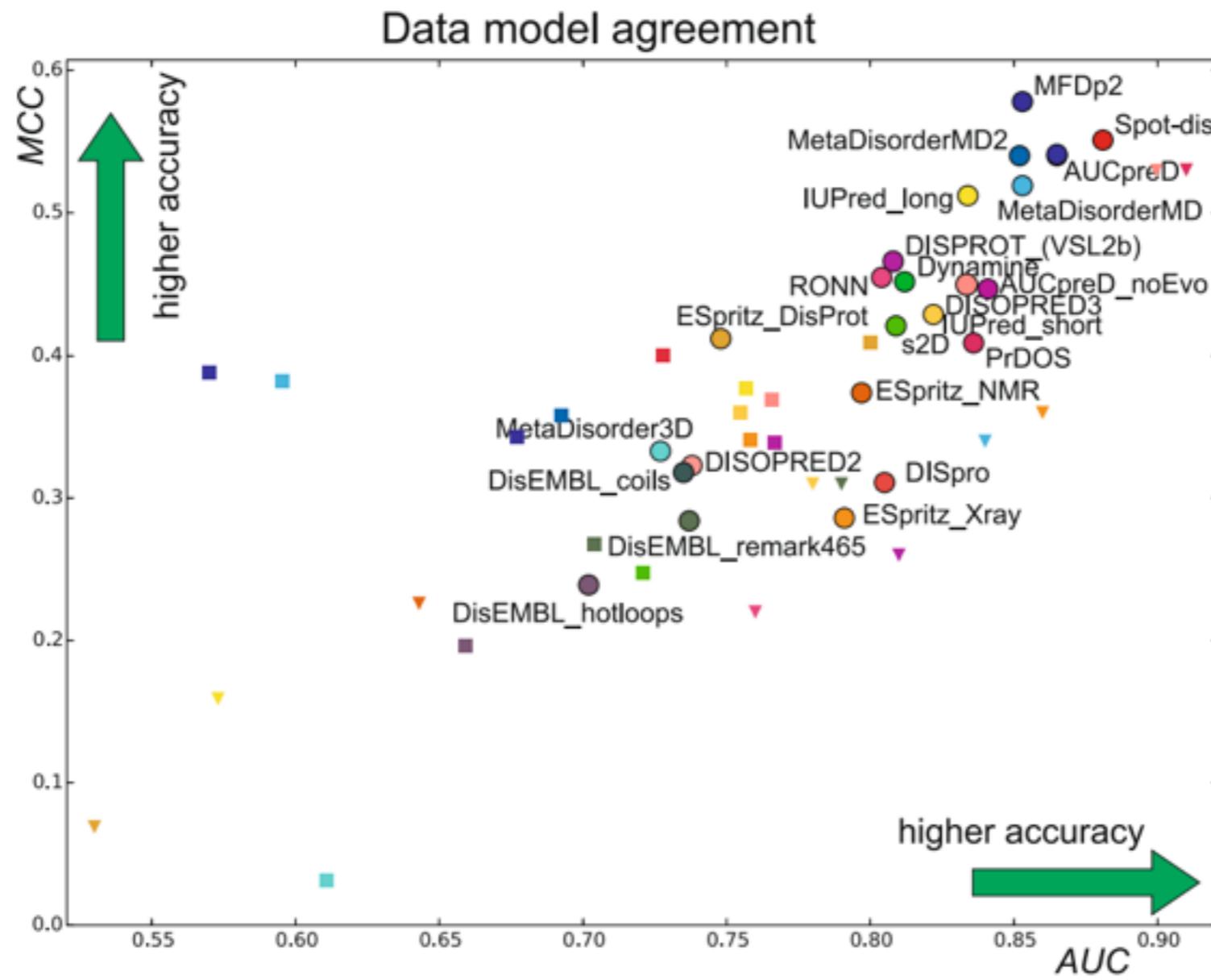


azul: true negatives

rojo: true positives

AUC = 0.5
predictor
aleatorio

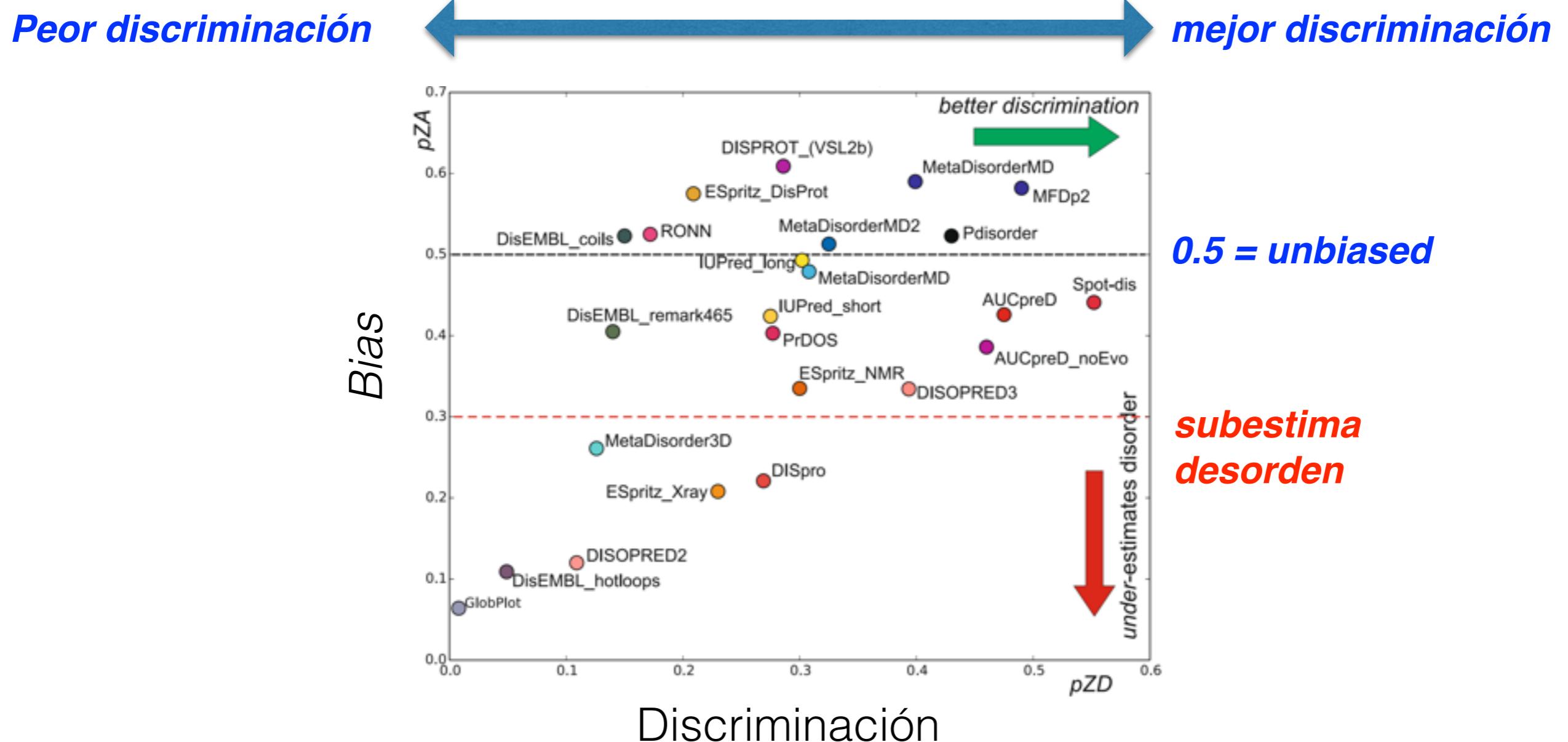
Performance de métodos de predicción de desorden



Ojo: todo depende del set de datos usando para testear

Ej. si el método es entrenado con el mismo tipo de datos que se testea va a dar mejor (entrenado con RMN, testado con TP, TN de RMN versus XRay)

Performance de métodos de predicción de desorden



los métodos varían en su capacidad de discriminar orden de desorden y en su tendencia a sobre o sub estimar el desorden

Meta-Predictores: MobiDB (lo vemos en TP)



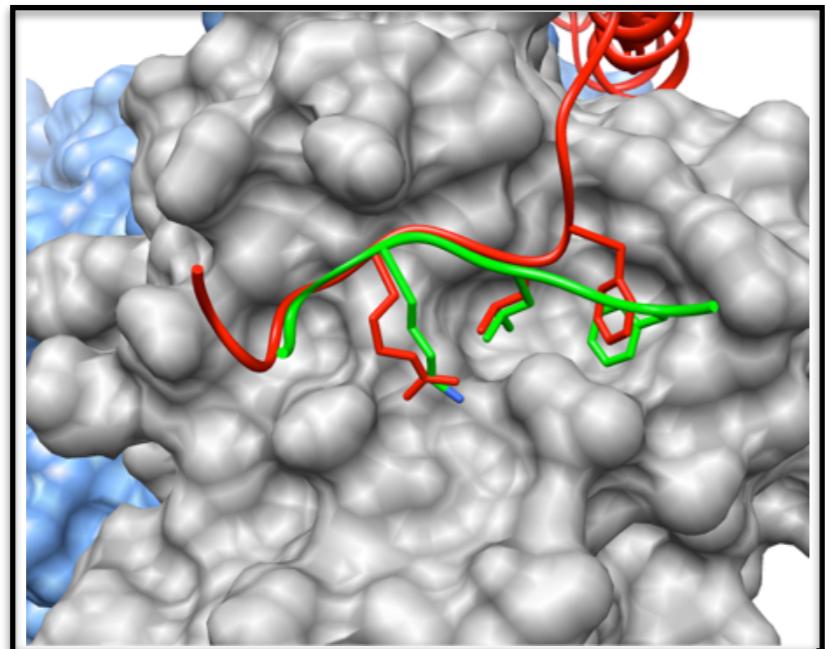
Integran información de una variedad de fuentes y predictores

Integrated resources



MobiDB lite - Mobi 2 - 82D - RCI - Anchor - FESS - RING - DisEMBL - ESpritz - GlobPlot - IUPred - Jronn - VSL2b - Pfilt - SEG

MOTIVOS FUNCIONALES EN REGIONES DESORDENADAS



motivos de unión

pueden ser representados por expresiones regulares o por PSSMs

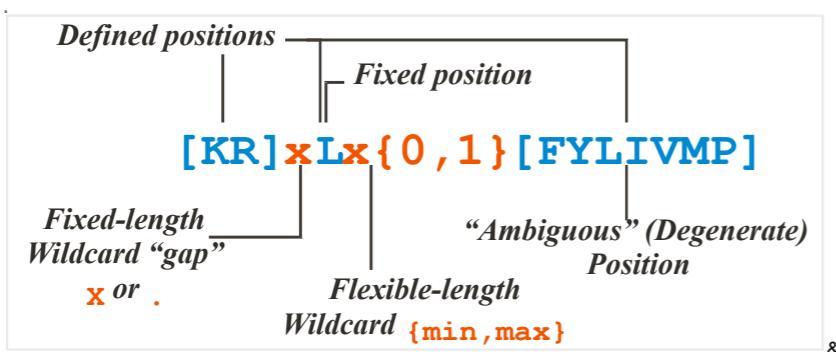


Figure 8.3. Anatomy of a SLiM.

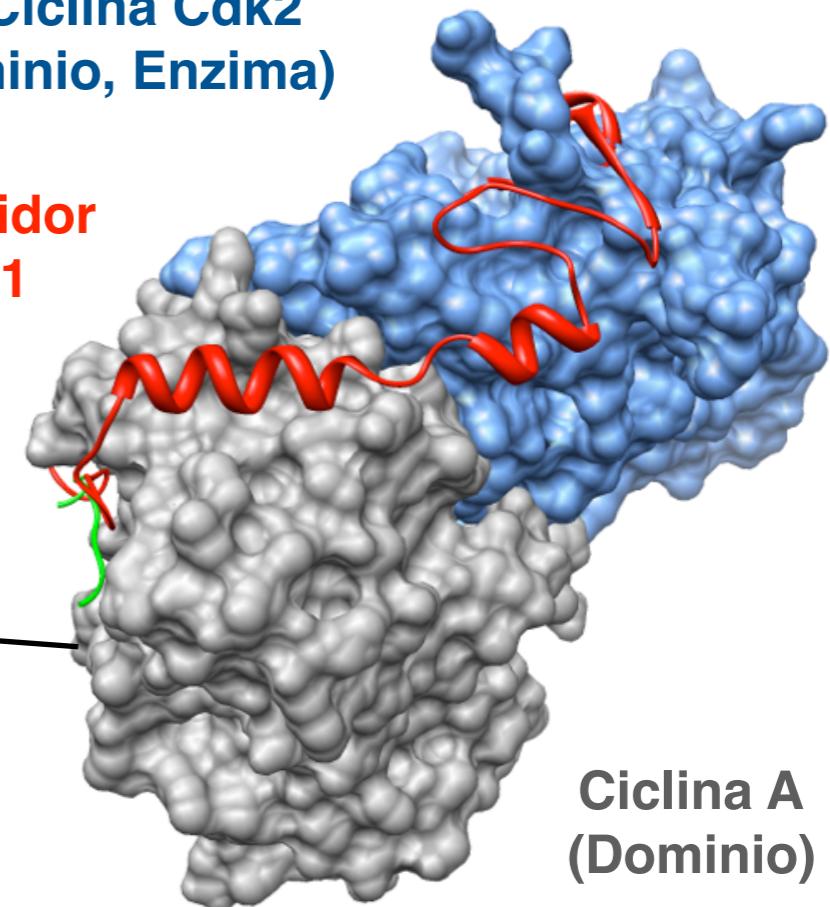
Ejemplo LOGO PSSMs

Ejemplos...

- Señal “KDEL” de retención en RER
- Motivos de unión a Integrina “RGD”
- Motivos de unión a Ciclinas “RxL”

**Quinasa dependiente
de Ciclina Cdk2
(Dominio, Enzima)**

**Inhibidor
p21**



Características de los motivos lineales o SLiMs

“**S**hort **L**inear **M**otifs”

- Se encuentran en regiones intrínsecamente desordenadas (IDP/IDR)
- Son secuencias cortas (3-10 residuos)
- Median interacciones proteína-proteína transientes (afinidad intermedia)
- Muy importantes en señalización molecular
- Utilizados ampliamente por los patógenos durante la infección

- Los SLiMs son módulos intrínsecamente desordenados que funcionan de forma independiente de su estructura terciaria

Cómo describimos a un motivo lineal o SLiM “Short Linear Motif”

Linear Motifs can be described by “Regular Expressions”

[KR]: different amino acids allowed at this position

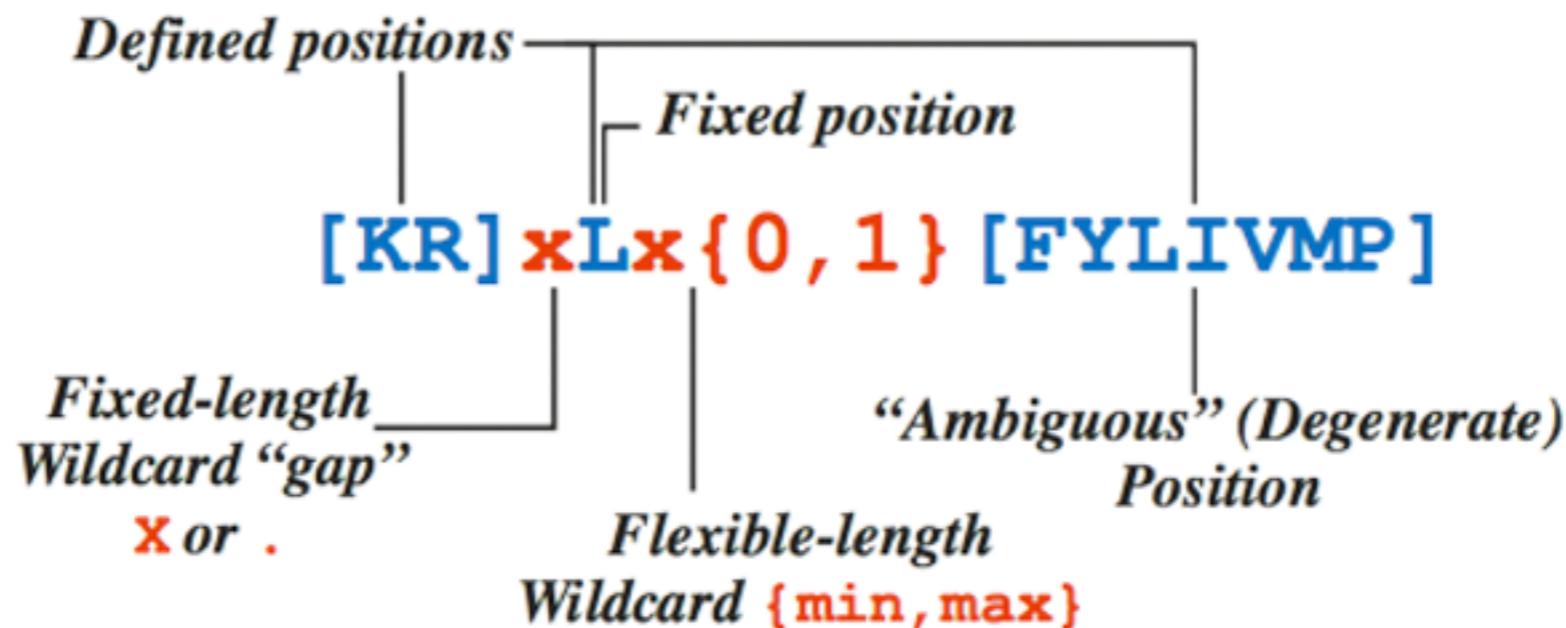
x or . : Wildcard

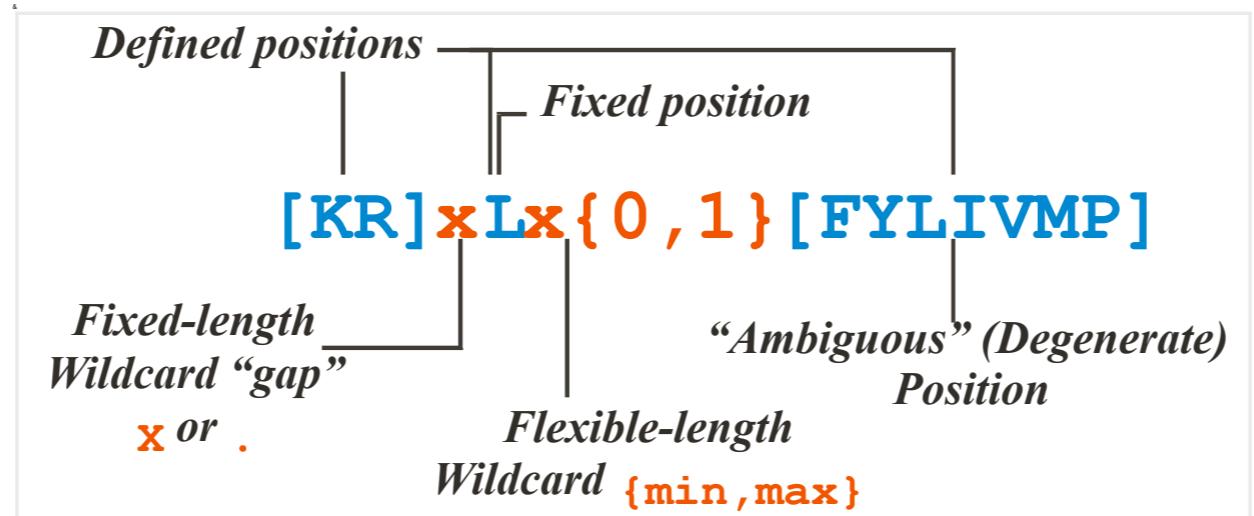
L: single amino acid “L”=Leucine

x{0,1} : variable length

[FYLIVMP]: different hydrophobic amino acids allowed

Example: DOC_CYCLIN_1:





La simbología comúnmente utilizada en expresiones regulares es:

Símbolo	Definición
.	Cualquier aminoácido es permitido
[XY]	Solo los aminoácidos X e Y son permitidos
[^XY]	Los aminoácidos X e Y están prohibidos
{min,max}	Número mínimo y máximo de veces que se puede repetir una posición
^X	El aminoácido X se encuentra en el extremo N-terminal
X\$	El aminoácido X se encuentra en el extremo C-terminal
(AB) (CD)	Se encuentran, o bien, los aminoácidos AB, o bien, los aminoácidos CD

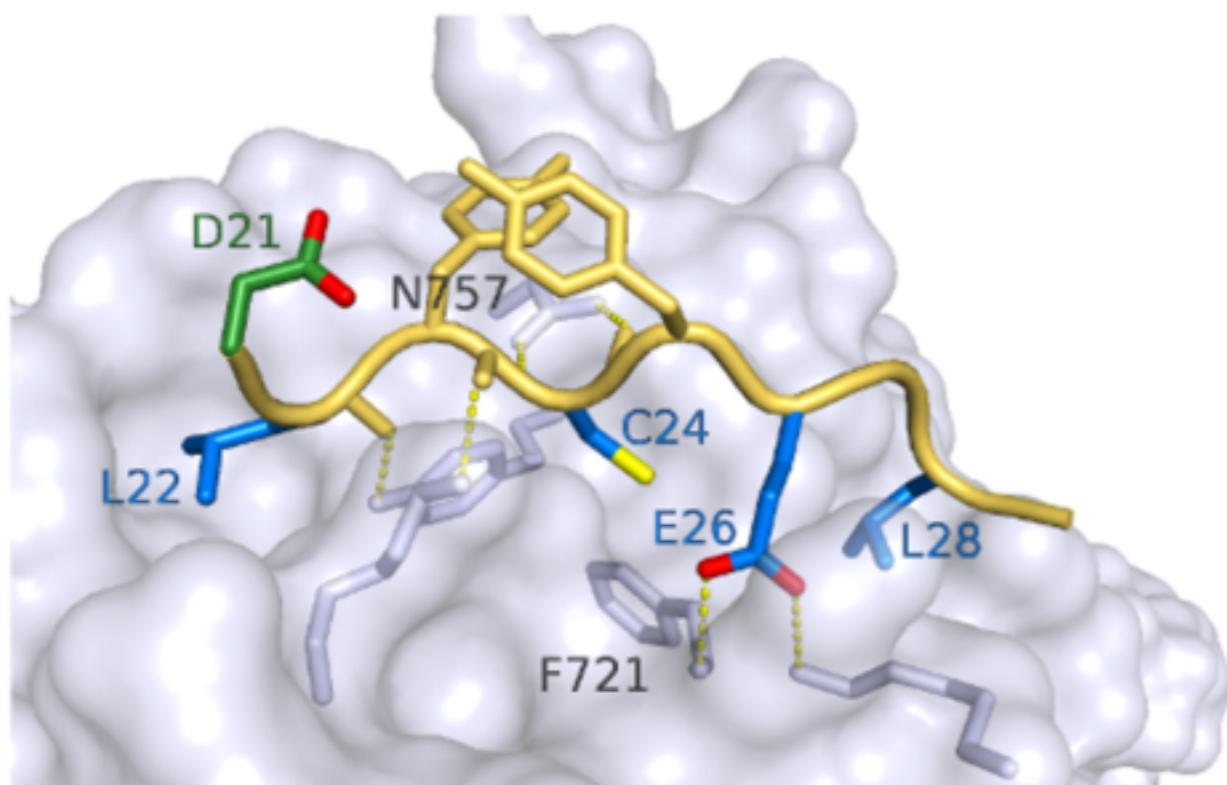
Las expresiones regulares que describen a un SLiM
pueden ser muy simples o muy complicadas

Example: CLV_PCSK_KEX2_1
RegEx: [KR] R .

Example: DOC_MAPK_HePTP_8
RegEx:

```
([LIV][^P][^P][RK]....[LIVMP].[LIV].  
[LIVMF]) | ([LIV][^P][^P][RK][RK]G.{4,7}  
[LIVMP].[LIV].[LIVMF])
```

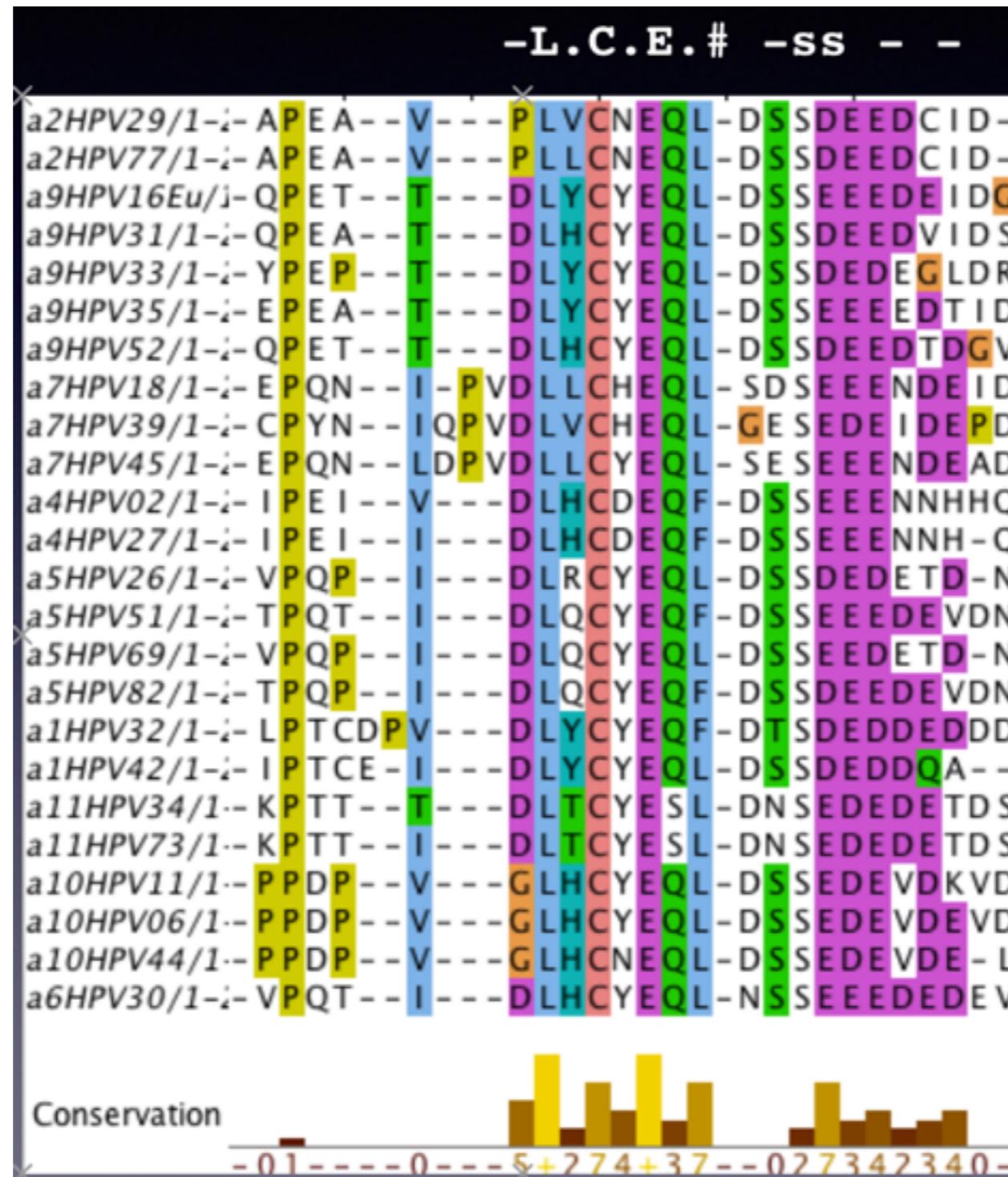
El patrón de la expresión regular representa la arquitectura del sitio de unión



	Host	K _d /μM
HDAC	DKR I A C E E -FS	20
	<i>Virus</i>	
E7	TTD L Y C YE Q -LN	0.005
E1A	DLT L T C HEAGFP	0.140
LT	EEN L FC S EE-MP	0.400
	* * *	
	Virus > Host	

Qué tipo de información evolutiva encontramos?

Alineamiento múltiple de secuencia (MSA)



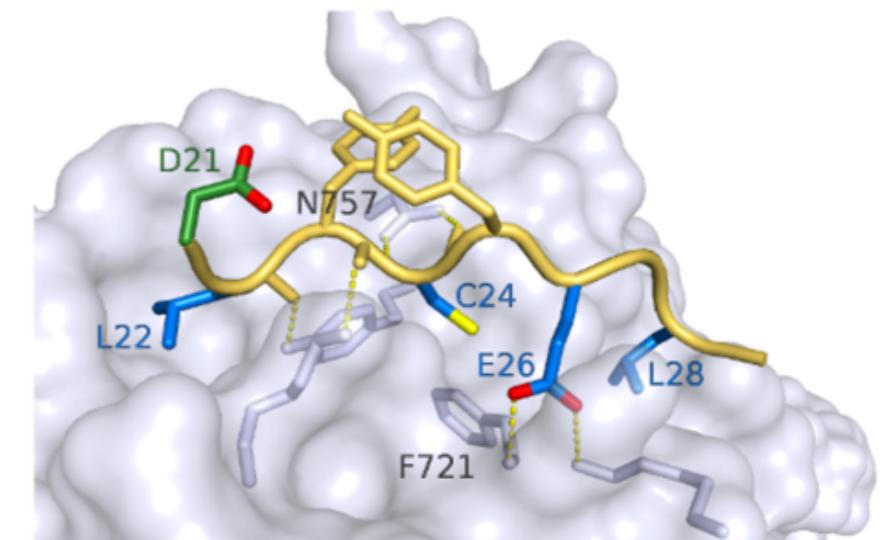
Tomamos múltiples secuencias del motivo LxCxE de varias familias de proteínas

Alineamiento múltiple de secuencia (MSA)

Entropía de Shannon aplicada a secuencias (Schneider, 1986)

$$R(l) = \log_2 20 - \left(- \sum_b f(b, l) \log_2 f(b, l) \right) - e(n)$$

$f(b, l)$ = frecuencia de aminoácido b en posición l



Mín $R(l)=0$, Máx $R(l)=\sim 4.32$

Los motivos lineales también pueden ser caracterizados por una PSSM

TYPES

CLV Proteolytic Cleavage Sites

DEG Destruction Sites (**Degrons**)

DOC Docking Sites

LIG Ligand Binding Sites

MOD Post-Translational Modification Sites

TRG Subcellular targeting sites

MODIFICATION SITES

Description:

Modification Motifs mediate **specific binding** to the **active site** of a modifying **enzyme** to allow subsequent catalytic **post-translational** modification of the target site.

Example:

Name MOD_CDK_SPxK_1

RegEx . . . ([ST]) P . [KR]

Kinase domain

P
CDK site

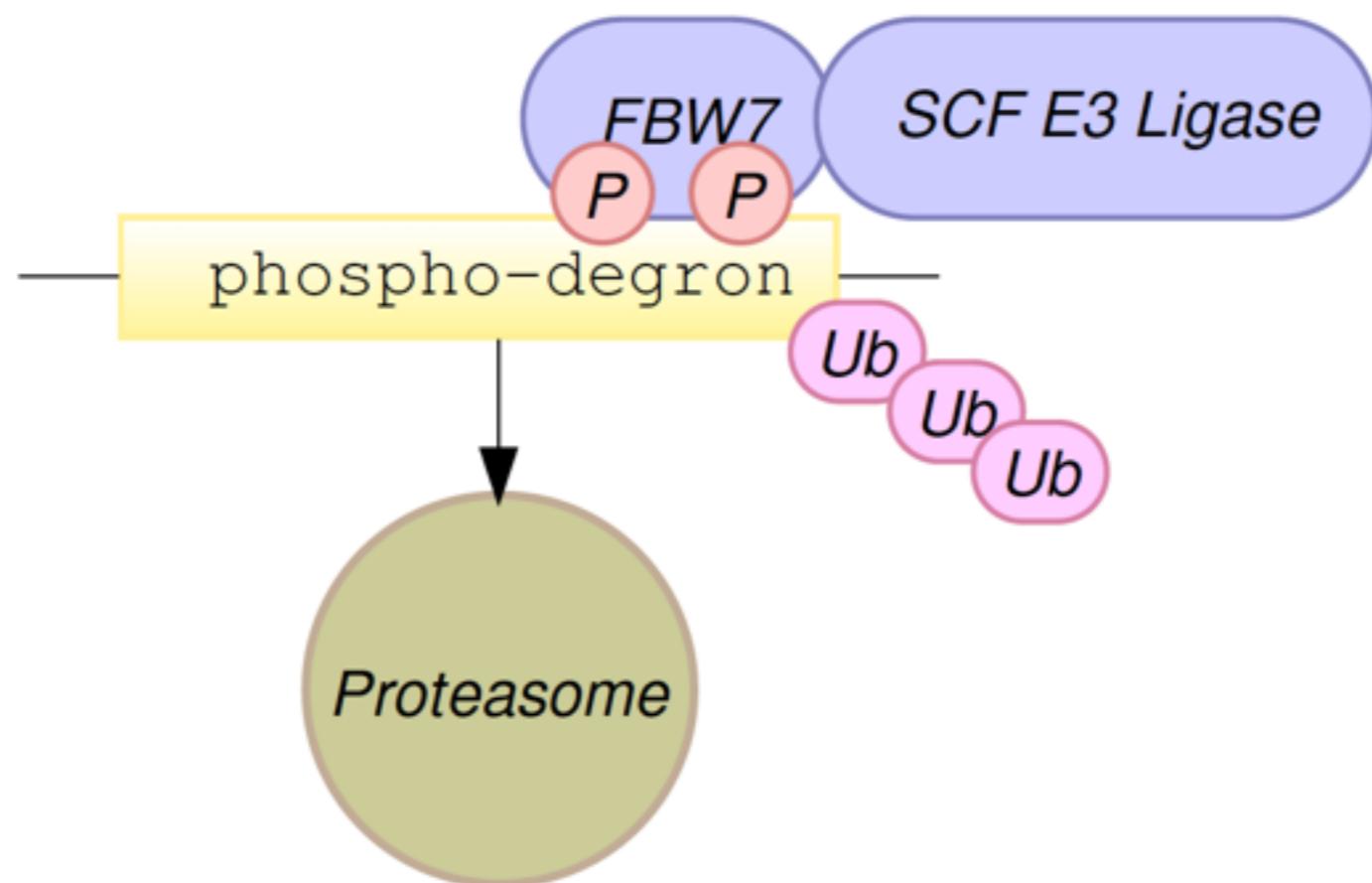
DEGRADATION MOTIFS

Description:

Degradation motifs (**Degrons**) recognized by **E3 Ubiquitin Ligase** complexes priming proteins for **degradation**, regulating protein half-life.

Example:

Name	DEG_SCF_TRCP1_1
RegEx	D (S) G . . ([ST])



TARGETING/ANCHORING MOTIFS

Description:

Targeting motifs allow a protein to bind to the **transport machinery** that **relocalizes** it to a particular sub-cellular location.

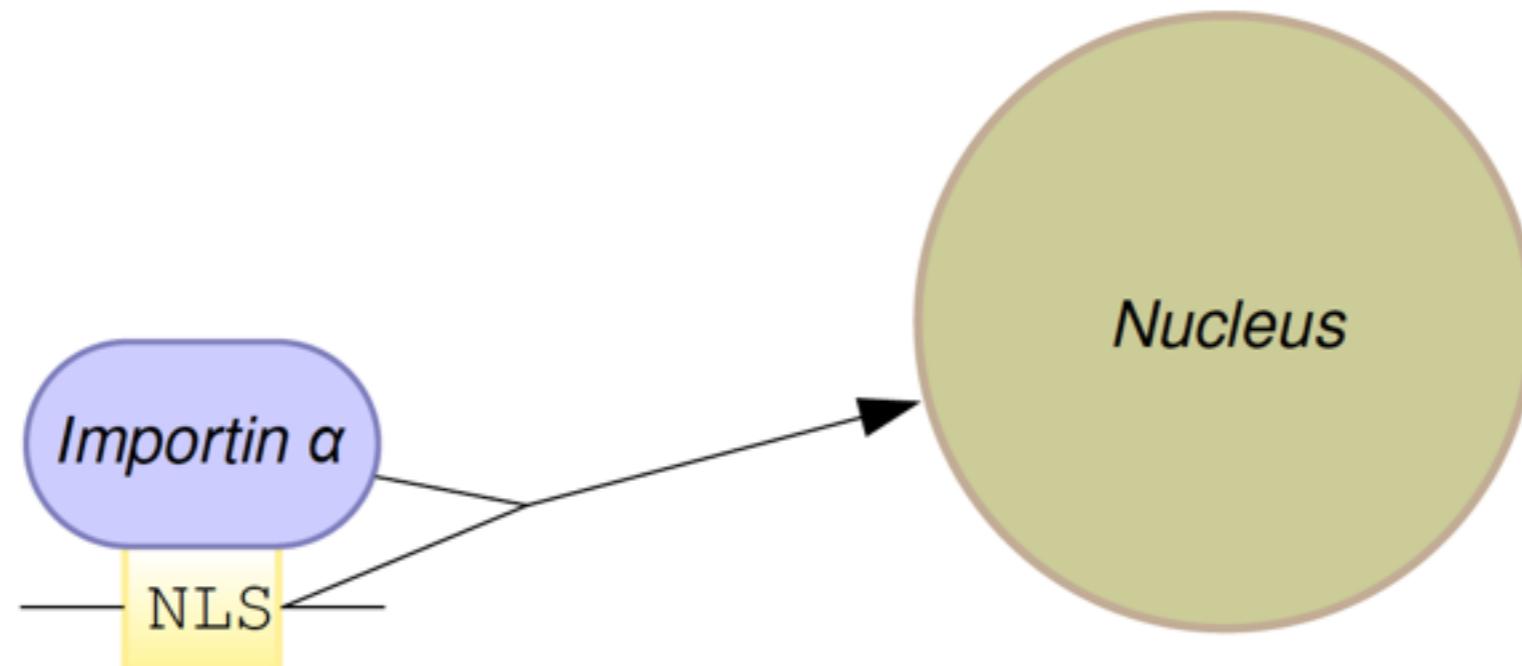
Anchoring motifs are recognized by biomolecules specific to a **sub-cellular** location and thereby **retain** the **motif**-containing protein at that location.

Example:

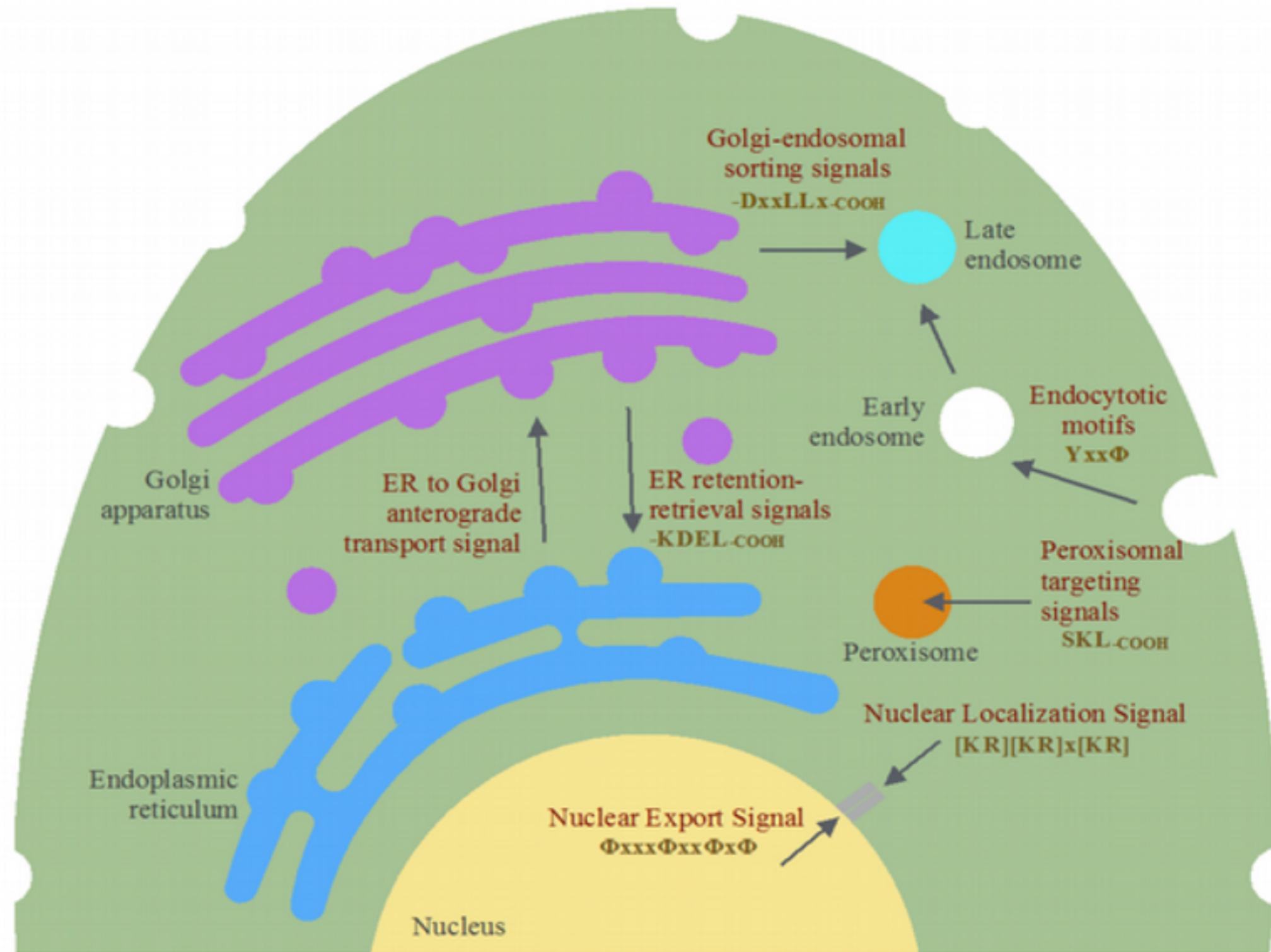
Name TRG_NLS_MonoCore_2

RegEx

[^DE] (K[RK] | RK) [KRP] [KR]
[^DE]



IMPORTANCE OF MOTIFS



Métodos para predecir SLIMs

- Expresión Regular

[L]x[C]x[E]{0,3}[LMIV]

- Shannon/PSSM



- Nuevos métodos? Elementos “fijos” (Core) + elementos “móviles/variables” ?

Core residues

⊖ { ⊖ ⊙ ⊚ } [L].C.E { ⊖ } ⊕ { ⊖ ⊙ ⊚ } ⊖ ⊖ ⊖

variable spacing

variable spacing

hydrophobic
negative/polar

Ventajas y Problemas de la predicción de SLiMs

Las definiciones son sencillas (RE, PSSM) y permiten buscar en los proteomes en forma rápida

Por su simplicidad (tamaño pequeño) y carácter degenerado (expresiones regulares), las predicciones retornarán muchos falsos positivos

Example: CLV_PCSK_KEX2_1
RegEx: [KR] R .

cuántas variantes hay?

Bases de datos de SLiMs

anotación experimental

THE ELM SERVER



**The Eukaryotic Linear Motif resource for
*Functional Sites in Proteins***

The **ELM** resource

is a **collection** of over 250 thoroughly **annotated** motif classes with over 3000 annotated instances.

It is also a **prediction tool** to detect these motifs in protein sequences employing different filters to distinguish between **functional** and **non-functional** motif instances.

elm.eu.org

TALKING ABOUT SLIMs

TYPES	CLASSES	INSTANCES
CLV Proteolytic Cleavage Sites	CLV_C14_Caspase3-7	Q7KZF4 (812-817)
DEG Destruction Sites (Degrons)	DEG_SCF_TRCP1_1	O97143 (292-297)
DOC Docking Sites	DOC_CYCLIN_1	P04637 (381-385)
LIG Ligand Binding Sites	LIG_RB_LxCxE	Q13547 (411-428)
MOD Post-Translational Modification Sites	MOD_CDK_SPxK_1	Q99741 (71-77)
TRG Subcellular targeting sites	TRG_NLS_MonoCore_2	Q969H0 (10-15)

Bases de datos de SLiMs

anotación experimental

THE ELM DATABASE



The Eukaryotic Linear Motif resource for
Functional Sites in Proteins

search ELM Database

ELM Home ELM Prediction ELM DB ELM Candidates ELM Information ELM downloads

Help

«LIG_PTB_Phospho_1»

»LIG_Rb_pABgroove_1»

LIG_Rb_LxCxE_1

Accession: ELM**E000007**

Functional site class: Rb pocket B binding ligand

Functional site description: The LxCxE motif is found in multiple host and viral interactors of the retinoblastoma protein family (Rb, p107 and p130).

ELM Description: The LxCxE motif mediates binding to a highly conserved shallow groove in the B domain of Rb. The central Cysteine is highly conserved in all instances, however the Leucine and Glutamic Acid positions tolerate substitution of physicochemically similar residues allowing a less stringent definition of [LI]xCx[DE]. The staggered arrangement, evenly spaced and one residue apart, of the conserved residues cover one side of an extended, beta-strand-like conformation and bind the groove orthogonally, not by beta augmentation like many similar staggered motifs. The Leucine and Cysteine positions bind a hydrophobic region of the groove with tight complementarity. The Glutamic Acid forms hydrogen bonds with two backbone amide groups of an alpha helix forming one side of the binding groove. The interaction is further stabilized by additional hydrogen bonds to the peptide backbone adding rigidity. Phosphorylation of Rb at Thr821 and Thr826 inhibits LxCxE binding.

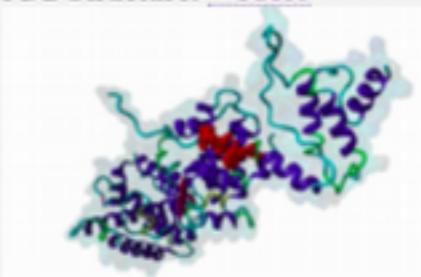
Pattern: ([DEST] | ^) . {0,4} [LI] . C.E. {1,4} [FLMIVAWPHY] . {0,8} ([DEST] | \$)

Pattern Probability: 0.0005417

Present in taxons: Metazoa Viridiplantae

Interaction Domain: RB_B (PF01857) Retinoblastoma-associated protein B domain
(Stoichiometry: 1 : 1)

PDB Structure: 1GH6



THE ELM DATABASE



LIG_Rb_LxCxE_1

■ Instance

Accession	Acc. Gene-, Name	Start	End	Subsequence	Logic	PDB	Organism	Length
ELMI003141	Q13547 HDAC1 HDAC1_HUMAN	411	428	SDKRIACEEEFSDSEEEGEG	TP	--	Homo sapiens (Human)	482

■ Instance evidence

Evidence class	PSI-MI	Method	BioSource	PubMed	Logic	Reliability	Notes
experimental	MI:0096	pull down	in vivo/in vitro		support	certain	InteractionDetection
experimental	MI:0519	glutathione s tranferase tag	in vitro	Vortmeyer,1999	support	certain	
experimental	MI:0065	isothermal titration calorimetry	in vitro	Singh,2005	support	certain	InteractionDetection
experimental	MI:0405	competition binding	in vitro	Vortmeyer,1999	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Vortmeyer,1999	support	certain	InteractionDetection
experimental	MI:0074	mutation analysis	in vivo/in vitro	Ferreira,1998	support	certain	FeatureDetection
experimental	MI:0405	competition binding	in vitro	Ferreira,1998	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Ferreira,1998	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Luo,1998	support	certain	InteractionDetection
experimental	MI:0074	mutation analysis	in vivo/in vitro	Magnaghi-Jaulin,1998	support	certain	FeatureDetection
experimental	MI:0405	competition binding	in vitro	Magnaghi-Jaulin,1998	support	certain	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Magnaghi-Jaulin,1998	support	certain	InteractionDetection
experimental	MI:0405	competition binding	in vitro	Brehm,1998	support	certain	InteractionDetection
experimental	MI:0018	two hybrid	in vivo	Brehm,1998	contradict	likely	InteractionDetection
experimental	MI:0019	coimmunoprecipitation	in vivo/in vitro	Brehm,1998	support	certain	InteractionDetection

Functional site prediction

Protein sequence

Enter Uniprot identifier or accession number: (auto-completion)
e.g. [EPN1_HUMAN](#), [P04637](#), [TAU_HUMAN](#), [\[RANDOM\]](#)

[EPN1_HUMAN](#)

EPN1_HUMAN Q9Y6I3 (Homo sapiens)

```
>EPN1_HUMAN
MSTSSLRQMKNIVHNYSSEAEIKVREATSNDPWGPSSSLMSEIADLTYNVVAFSEIMSMIWKR
LNDHGKNWRHUVYKAMTLM
EYLIKTSERVSQQCKENMYAVQTLKDFQYVDRDGKDQGVNVREKAKQLVALLRDE
DRLREERAHALKTKEKLAQTATASSAAVG
SGPPPEAEQANPQSSGEEELQLQLALAMSKEEADQPPSCGPEDDAQLQLALSLSREEHDKEERIRRGDDLR
LQMAIEESKRETGGKEESSILMDLADVFATAPAPAPTTDPNGGPAPMAAAVPTAAPTSDP
NGGPPVPPA
DPNGGPA
PTPASGDPWRPAAPAGPSVD
PWN
GT
PAPAAGEGEGPTPDPNGSSDGGVPVSGPSASDPWTPAPAFSDP
NGGSPAKPSTNGTTAAGGFDT
EPDEFSDFDR
LRTALP
TSGSSAGELELLAGEVPARSPGAFDM
SGVRGSLAEAVGSPPPAATPTPTPPTRKTPESFLGPNAALVLDLSLVS
RPGPTPPGA
KASNPFLPGGGPATGPSVTNP
FQPAPPATLTNQLRLSPVPPVPGAPPTYISPLGGPGLPPMMPPGPPAPNTNP
FLL
```

- Cell compartment (one or several):

not specified
extracellular
nucleus
cytosol
peroxisome
glycosome
glyoxisome
golgi apparatus
endoplasmic reticulum
lysosome
endosome
plasma membrane
mitochondrion

- Taxonomic Context

Type in species name (auto-completion):
[Homo sapiens](#)

- Motif Probability Cutoff:

100

[Submit](#) [Reset Form](#)

- ELM database update
A new ELM class has been annotated: [LIG_14-3-ChREBP_3](#)
- ELM database update
A new ELM class has been annotated: [LIG_IRF3_LxIS_1](#)
- ELM database update
New instances added for [LIG_Yh1_VBS_1](#)
- ELM database update
New instances have been added for [DEG_SCF_TRCP1_1](#) and [LIG_LRP6_Inhibitor_1](#)
- ELM database update
We have added new instances for: [LIG_APCC_ABBA_1](#), [LIG_APCC_ABBAyCdc20_2](#) as well as [DOC_MAPK_HePTP_8](#), [DOC_MAPK_MEF2A_6](#) and [DOC_MAPK_DCC_7](#)
- ELM Database Update
We have updated several MOD_CDK motifs and added new instances:
MOD_CDK_1 is now: [MOD_CDK_SPxK_1](#), and [MOD_CDK_SPK_2](#) [MOD_CDK_SPxxK_3](#) have been added.
- ELM database update
Several new ELM classes and instances have been added:

THE ELM PREDICTION TOOL

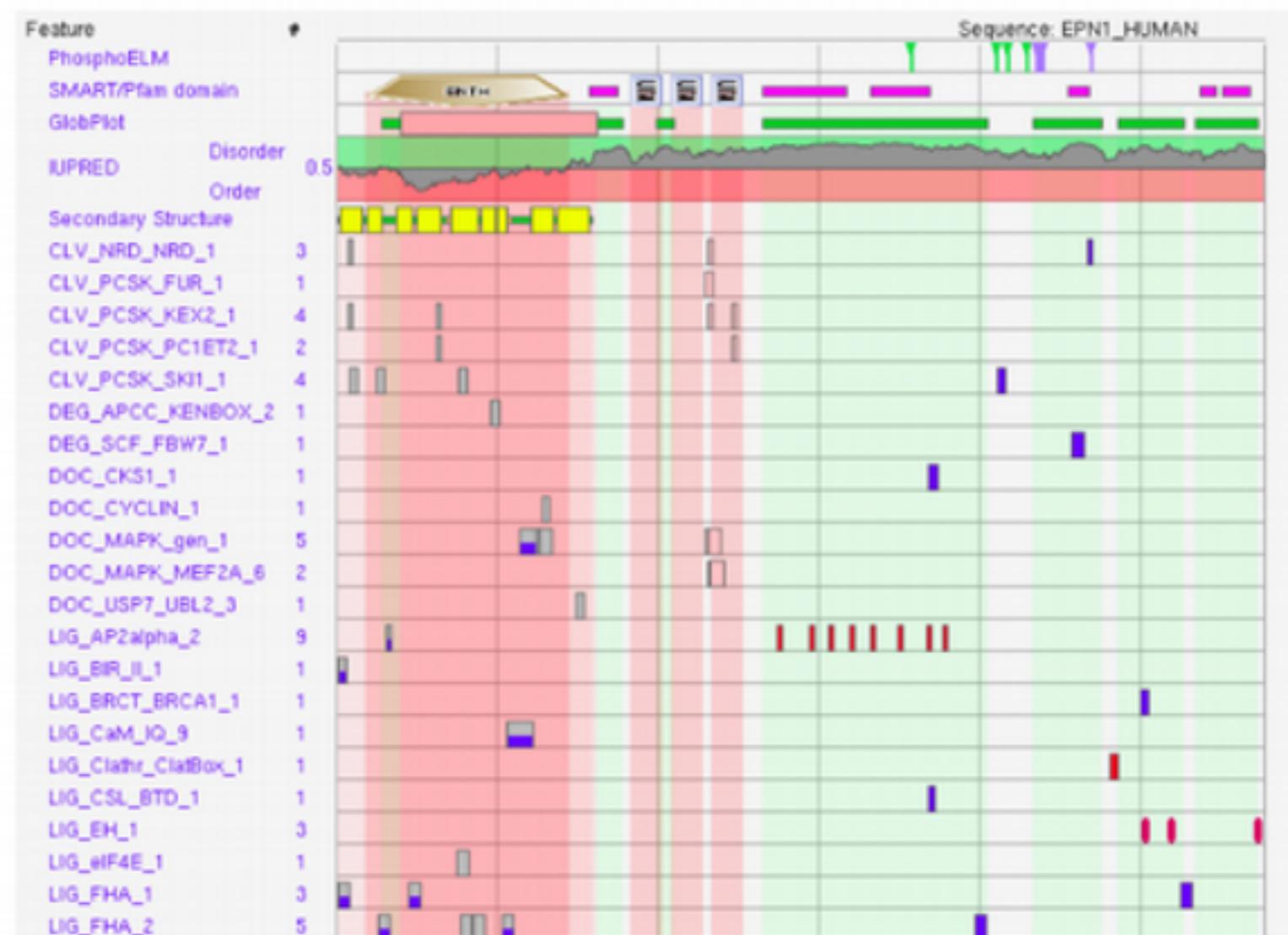


■ Summary for sequence 'EPN1_HUMAN'.

KEY

DOMAINS:	Smart/Pfam domain	Signal peptide (pred.)	Low-complexity region	Coiled-coil (pred.)	TM helix (pred.)
GLOBPLOT:	GlobDom	Disorder			
2D STRUCT:	Strand	Helix	Loop	3/10 Helix	
MOTIFS:	Favourable Context	Sparse/Smart filtered	Neutral	Annotated:	TP FP TN U < < Assigned by homology
CONSCORE:	low Conservation	medium Conservation	high Conservation		
PhosphoELM:	phosphorylated Serine	phosphorylated Threonine	phosphorylated Tyrosine		

(Mouseover the matches for more details)



THE ELM PREDICTION TOOL



KEY

DOMAINS:	Smart/Pfam domain	Signal peptide (pred.)	Low-complexity region	Coiled-coil (pred.)	TM helix (pred.)
GLOBPLOT:	GlobDom	Disorder			Help
2D STRUCT:	Strand	Helix	Loop	3/10 Helix	
MOTIFS:	Favourable Context	Sparse/Smart filtered	Neutral	Annotated:	TP FP TN U < Assigned by homology
CONSCORE:	low Conservation	medium Conservation	high Conservation		
Phospho.ELM:	Tphosphorylated Serine	Tphosphorylated Threonine	Tphosphorylated Tyrosine		

TP (True Positive): Experimental evidence supporting functionality.

FP (False Positive): Experimental evidence hinting at a function, but after careful inspection ELM annotators believe this instance to be non-functional.

TN (True Negative): Experimental evidence supporting non-functionality.

U (Unknown): Not enough evidence to determine whether this instance is functional or not.

SLIMSearch4

[Search](#)[About](#)[Help](#)

About

SLIMSearch is a short, linear motif (SLIM) discovery tool. The tool searches a proteome using a motif consensus to discover putative novel motif instances. Motif attributes known to be strong discriminators of motif functionality, such as accessibility and conservation, are calculated for each instance. Instances are also annotated with experimental, proteomic and genomic data. The tool also allows instance filtering based on keyword, interaction, localisation information. In depth ontology enrichment and conservation analysis tools are also available. Find a tutorial to understand the SLIMSearch output [here](#).

Reference:

"SLIMSearch - a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions"

Izabella Krystkowiak and Norman E. Davey

Nucleic Acids Res. 2017 Apr 6. (Epub ahead of print).

Search details

Format: WRPW motif example [?](#)

Choose species: Homo sapiens

[Homo sapiens](#)

Saccharomyces cerevisiae

Xenopus tropicalis

Schizosaccharomyces pombe

Drosophila melanogaster

Arabidopsis thaliana

Caenorhabditis elegans

+ more

Choose options

Disorder score cut-off [?](#) : 0.40



Flank length [?](#) : 5



[Submit](#)

Enter job id

[Retrieve job](#)

Jobs are stored for two weeks.

SLIMs SEARCH

View: Flank conservation | [Taxonomic range](#)

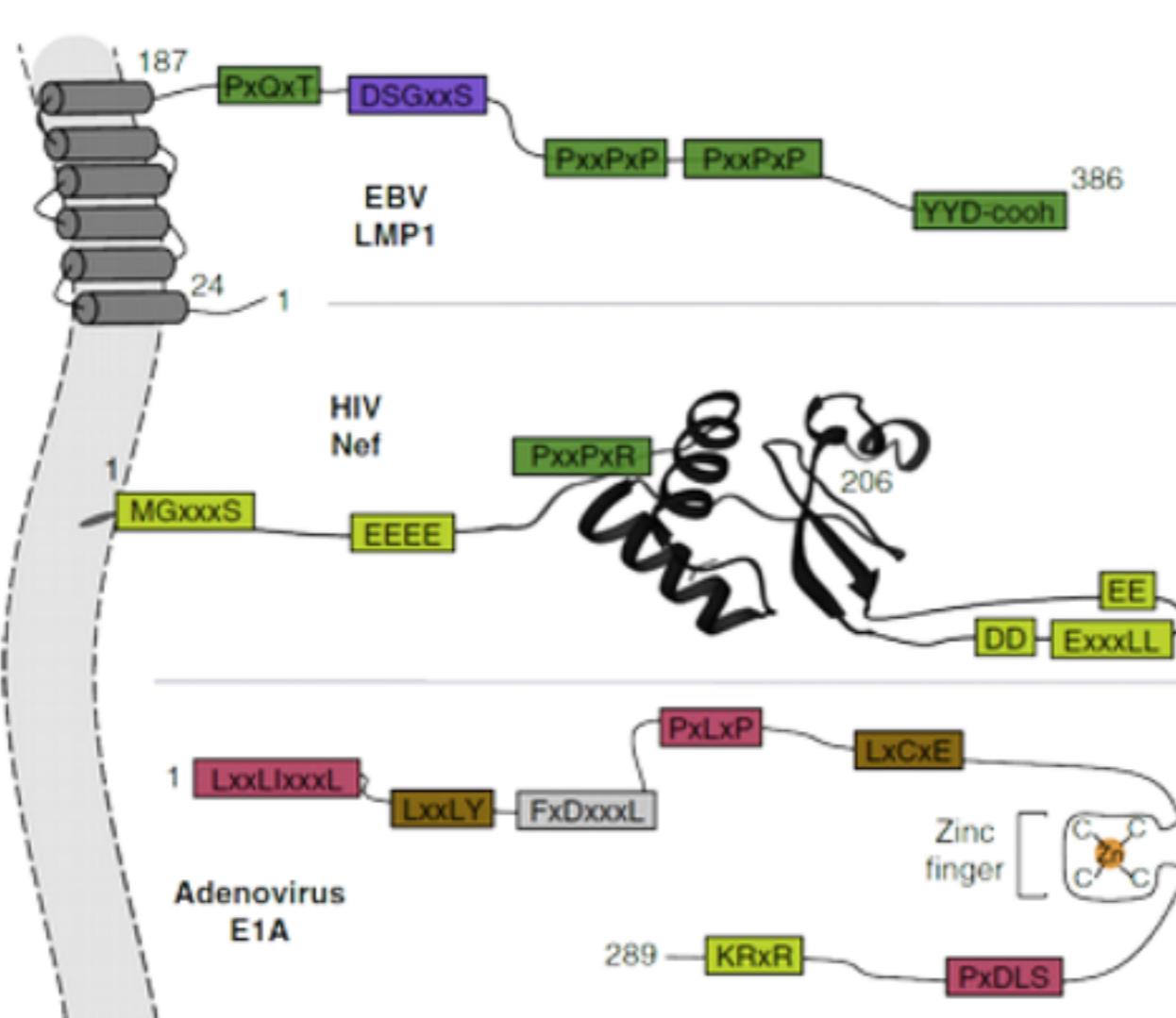
Clad: [Chordata](#) | Metazoa | QFO

12 hits in 12 proteins

Page 1 of 1 First | Prev | **1** | Next | Last

Protein Name	Peptide	Start	Stop	DeNovo score	Domain	Motif	ConScore Combined	Conserved Counter																										
								<i>Homo sapiens</i>	<i>Ovis aries</i>	<i>Equus caballus</i>	<i>Gorilla gorilla</i>	<i>Mus musculus</i>	<i>Bos taurus</i>	<i>Sacculinus hamatus</i>	<i>Pan troglodytes</i>	<i>Otoblemur garnettii</i>	<i>Loris tardigradus</i>	<i>Rattus norvegicus</i>	<i>Felis catus</i>	<i>Sus scrofa</i>	<i>Anolis platypeltis</i>	<i>Monodelphis domestica</i>	<i>Tupaia chinensis</i>	<i>Lutreola lutreola</i>	<i>Cavia tammaris</i>	<i>Xenopus tropicalis</i>	<i>Aotus carayaensis</i>	<i>Gallus gallus</i>	<i>Oryctolophus anatinus</i>	<i>Danio rerio</i>	<i>Taraxipus rubripectus</i>	<i>Tetradon nigroviridis</i>	<i>Gasterosteus aculeatus</i>	<i>Oryzias latipes</i>
Protein disulfide-isomerase (P4HB)	dqkavKDEL	505	508	0.971		1	-0	36	C	C	C	C	C	C	C	C	C	C	C	X	C	C	C	N	C	C	N	N	N	N	N	N		
78 kDa glucose-regulated protein (HSP90AA1)	edtaKDEL	651	654	0.962		1	-0	42	C	C	C	C	C	C	C	C	C	C	C	N	C	C	C	C	N	C	N	N	N	N	N	N		
Endoplasmin (HSP90AB1)	estaKDEL	800	803	0.973		1	-0	47	C	C	C	C	C	C	C	C	C	C	C	N	C	C	C	C	C	C	N	C	N	X	N	N		
Calreticulin (CALR)	vpgquKDEL	414	417	0.999		1	-0	48	C	C	C	C	C	C	C	C	C	C	C	N	N	C	C	N	C	C	N	N	C	C	C	C		
Protein disulfide-isomerase A6 (PDIA6)	lddlgKDEL	437	440	0.842		1	-0	44	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	N	C	N	C	N	X	N	N		
Prolyl 3-hydroxylase 1 (LEPRE1)	seskpKDEL	733	736	0.964		1	-0	29	C	C	C	C	C	C	N	C	C	C	C	N	N	N	C	N	X	C	N	X	N	N	N	N		
Putative endoplasmic-like protein (HSP90BP1)	estaKDEL	396	399	0.945			-0	2	C	X	X	C	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
KDEL motif-containing protein 1 (KDELC1)	hrkkttKDEL	499	502	0.766	1	1	-0	35	C	C	C	C	C	C	C	C	C	C	C	C	N	N	C	C	N	N	N	N	C	C	N	N	N	
Prolyl 3-hydroxylase 2 (LEPREL1)	IndnpKDEL	705	708	0.86		1	-0	41	C	C	C	C	C	C	C	C	C	C	C	C	N	C	C	N	X	C	C	C	X	X	N	N	N	
DnaJ homolog subfamily C member 10 (DNAJC10)	qgkrnKDEL	790	793	0.859		1	-0	22	C	C	C	C	C	N	C	N	C	C	C	N	N	X	N	C	N	N	N	N	N	N	N	N	N	
Thioredoxin-domain-containing protein 5 (TXNDC5)	vlsqKDEL	429	432	0.695	2	1	-0	27	C	C	C	X	C	C	C	C	C	C	C	N	X	N	C	C	N	N	C	N	N	N	N	N	X	
ER degradation-enhancing alpha-mannosidase-like protein 3 (EDEM3)	fennetKDEL	929	932	0.796		1	-0	46	C	C	C	C	C	C	C	C	C	C	C	N	C	N	C	C	C	C	C	C	C	N	C	N	X	N

IMPORTANCE OF MOTIFS



Motif	Host interactor
204 PxQxT ₂₀₉	TRAFs
210 DSGxxS ₂₁₆	βTrCP
275,302 PxxPxP _{281,308}	JAK3
384 YYD ₃₈₆	TRADD
1 MGxxxS ₇	NMT1
62 EEEE ₆₆	PACS1
72 PxxPxR ₇₈	Hck
154 EE ₁₅₆	β-COP
160 ExxxLL ₁₆₆	AP-1/2/3
174 DD ₁₇₆	AP-2
20 LxxLxxxxL ₂₉	TR
43 LxxLY ₄₈	RB
66 FxDxxxL ₇₃	p300
113 PxLxP ₁₁₈	BS69
122 LxCxE ₁₂₇	RB
279 PxDLS ₂₈₄	CtBP
286 KRxR ₂₈₉	Importin α

TIBS-817; No. of Pages 11

ARTICLE IN PRESS

Review

Cell
PRESS

How viruses hijack cell regulation

Norman E. Davey¹, Gilles Travé² and Toby J. Gibson¹

¹ Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

² Equipe Oncoprotéines, FRE CNRS 3211, ESBS, 1, Bld Sébastien Brandt, BP10413, 67412 Illkirch, France

IMPORTANCE OF MOTIFS

