PREDICCIÓN DE ESTRUCTURA TRIDIMENSIONAL DE PROTEÍNAS

Dr. Alexis Salas Burgos
Departamento de Farmacología
Departamento de Fisiopatología
Universidad de Concepción

Junio 2012

Temario

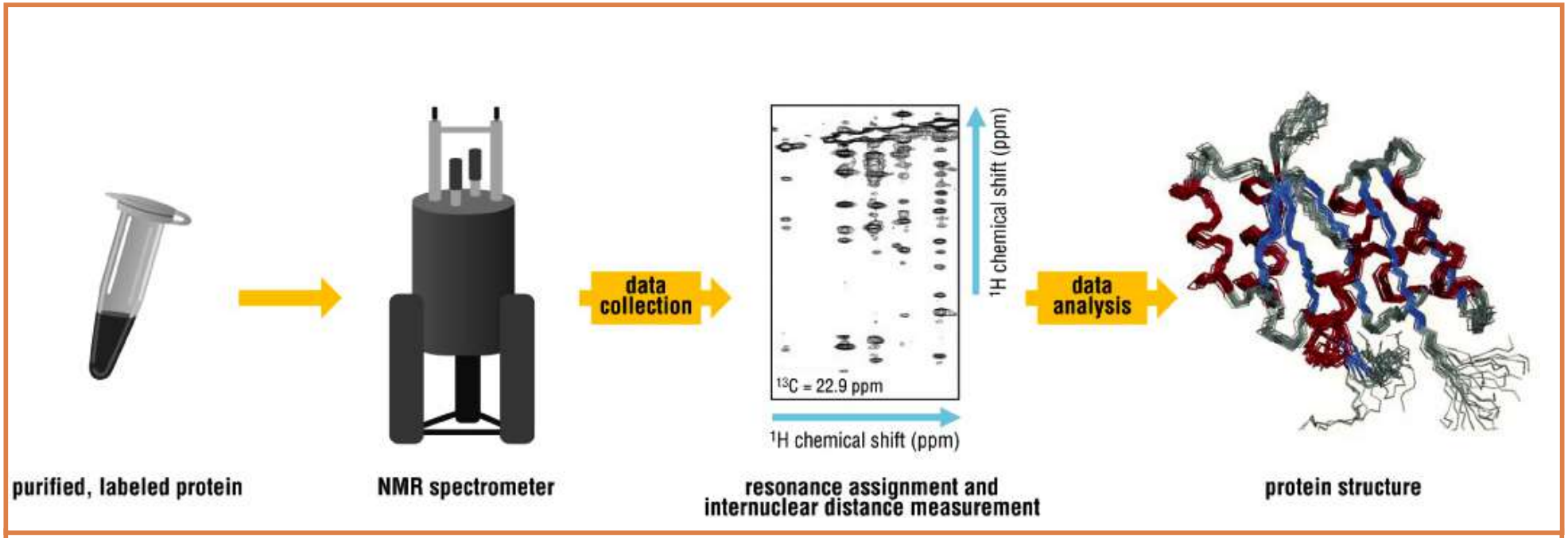
- Estructura de proteínas.
  - Reglas estructurales y excepciones.
  - Predicción de estructuras secundarias.
  - Cálculo de estructuras secundarias.
- Búsqueda de homólogos.
  - Reconocimiento de Plegamientos.
- Modelamiento Comparativo
  - Suite Modeller.
  - Suite Rossetta.
  - Análisis de Modelos.
  - Validación otras técnicas experimentales.

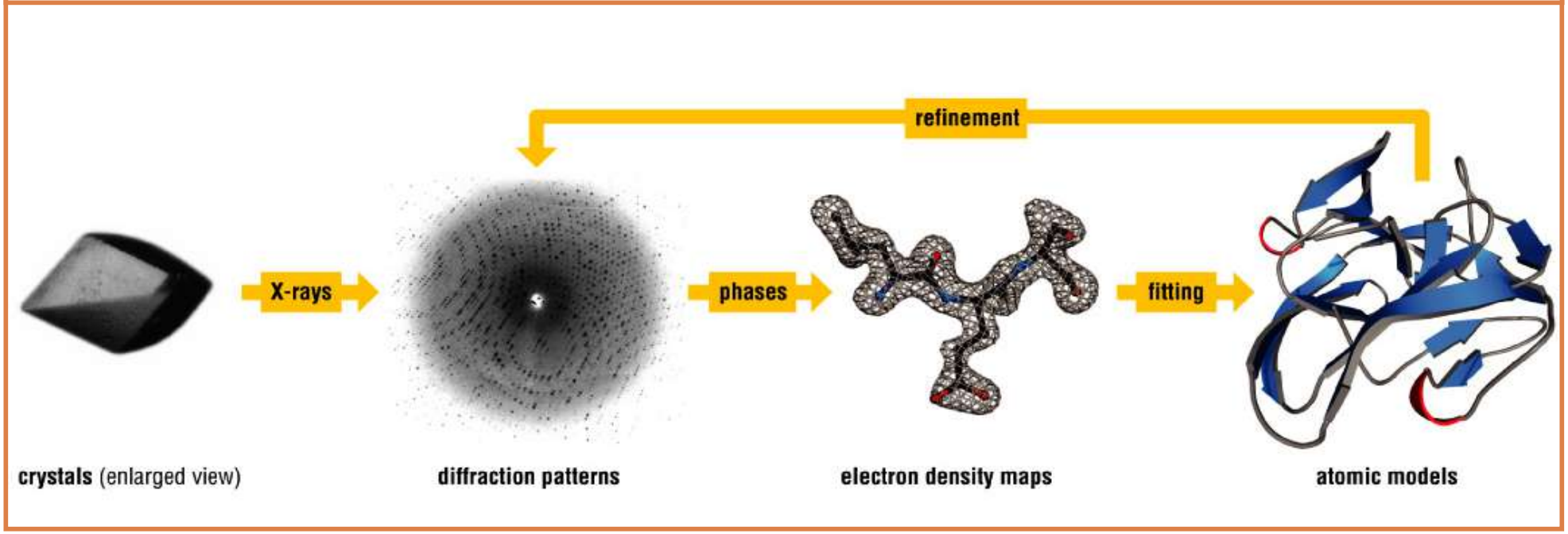
Temario

- Modelamiento ab initio.
- Modelamiento proteínas de membrana.
- Modelamiento CDR anticuerpos.
- Modelamiento de Lazos.
- Modelamiento de Mutaciones.

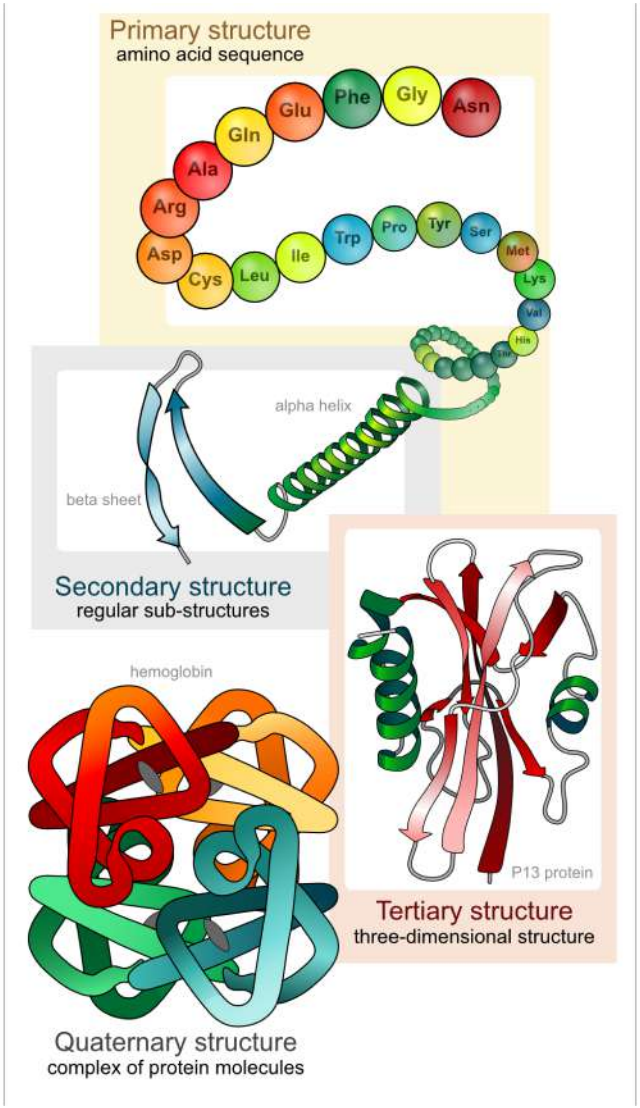
Estructura de Proteínas

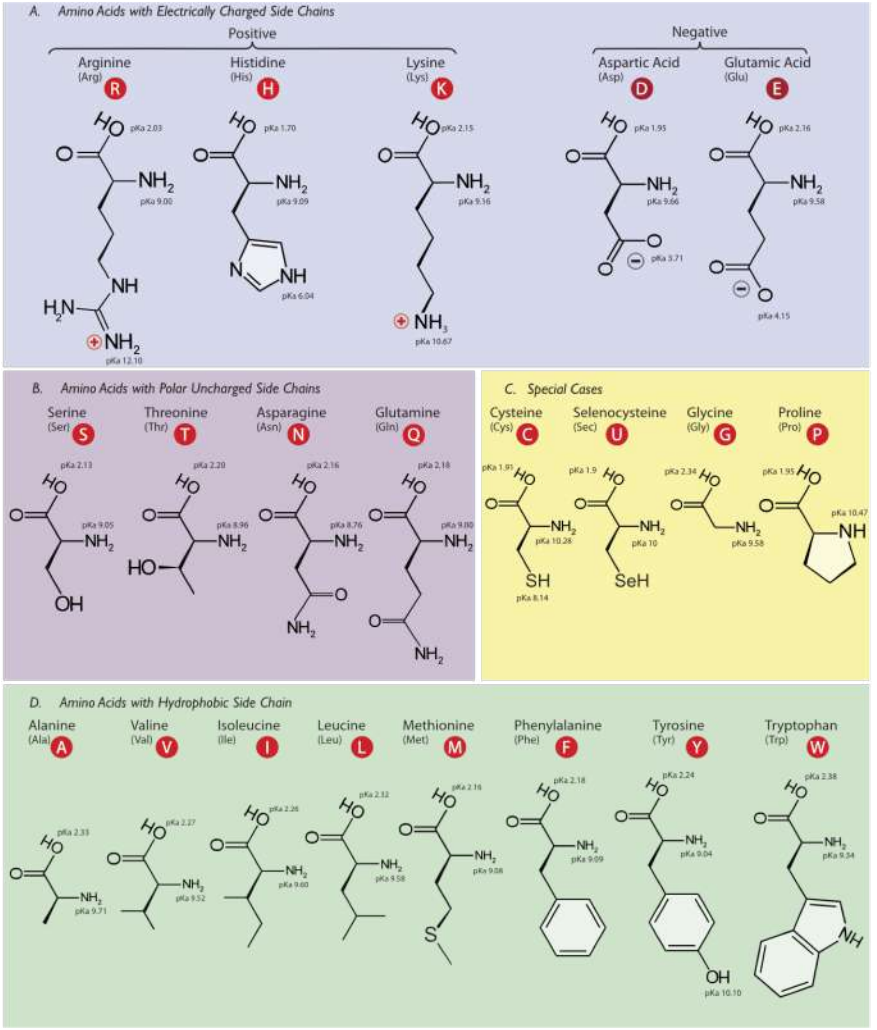
METODOS EXPERIMENTALES DE DETERMINACION DE ESTRUCTURA DE PROTEINAS





Estructura de Proteínas: Niveles

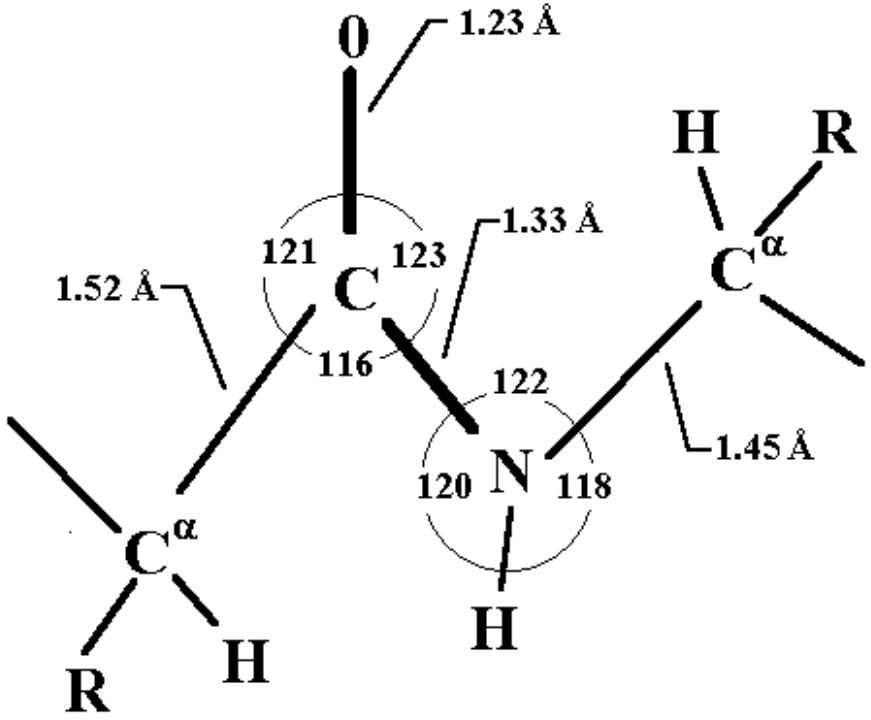


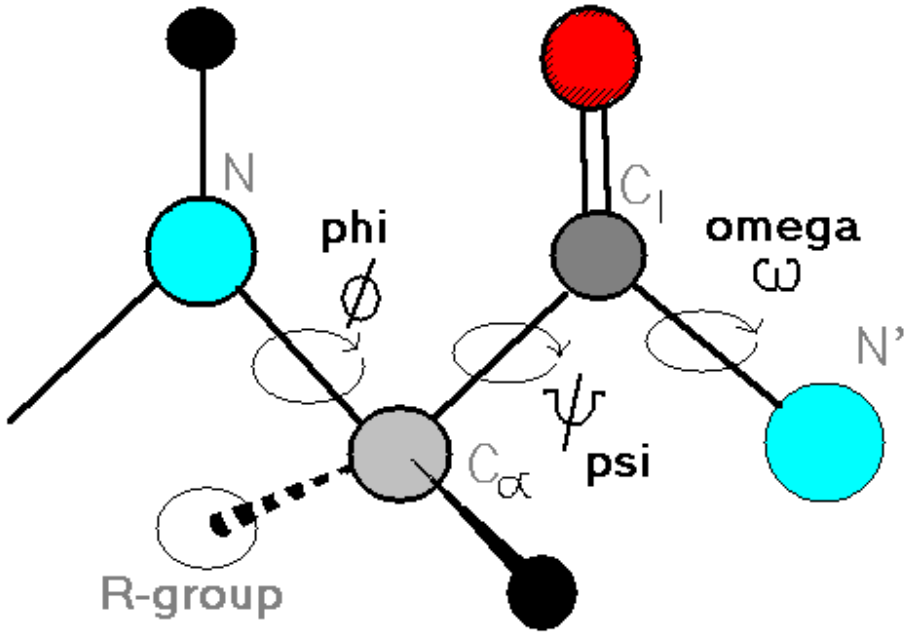


Estructura de Proteínas: Estructura Secundaria

- Alfa Hélices.
  - Posee 3.6 aminoácidos por vuelta con puentes H cada 4 residuos.
  - El largo promedio es de 10 aminoácidos o 10 Å y van desde 5 a 40 (1.5 a 11 vueltas).
  - Rica en LEMA / pobres en PSGY.
- Hojas Beta.
  - 5 a10 aminoácidos consecutivos que interaccionan con otros 5 a 10 en forma paralela o antiparalela.
- Lazos.
  - Son regiones de una cadena proteica que están entre α hélices y hojas β, de longitudes variadas y configuraciones 3D y en la superficie de la estructura. Pueden ser tan cortos como 2 aminoácidos (horquillas).
- Vueltas.
  - Son tipos de lazos con configuraciones determinadas.

Ángulos y Largos del Esqueleto Proteico





Gráficos de Ramachandran

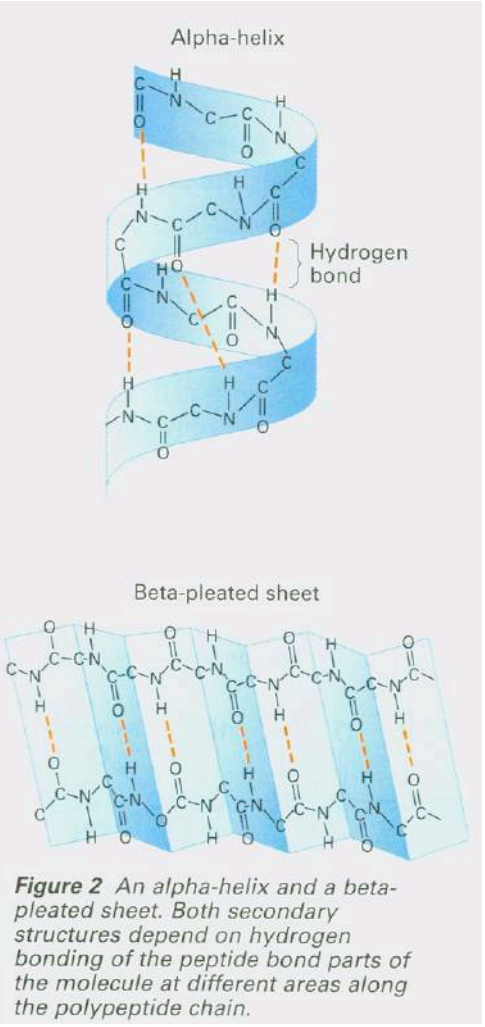
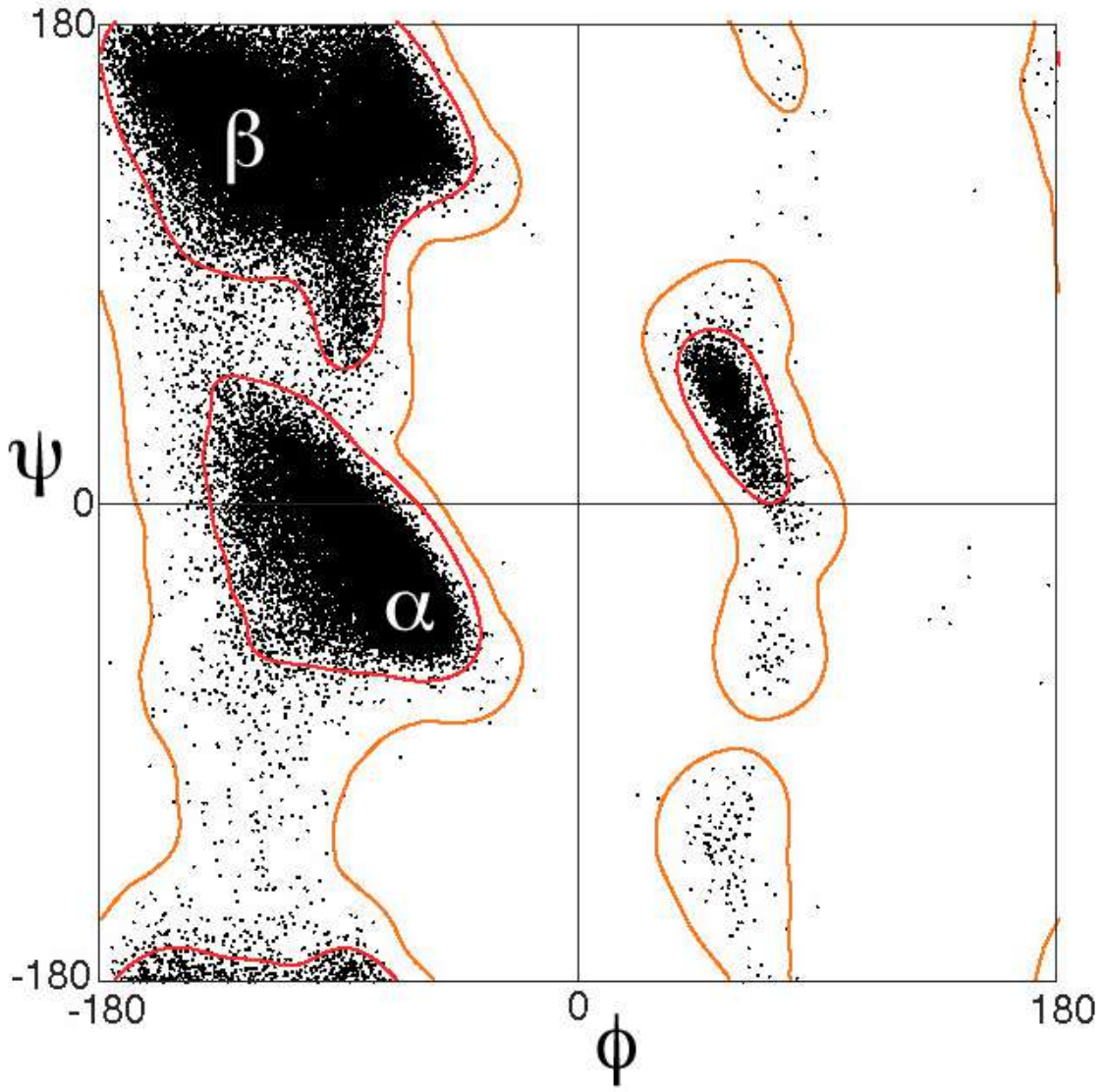


Figure 2 *An alpha-helix and a beta-pleated sheet. Both secondary structures depend on hydrogen bonding of the peptide bond parts of the molecule at different areas along the polypeptide chain.*



Estructura de Proteínas: Definiciones

- **Dominio.**
  - **Estructural.** Una región compacta y globular que es semi-independiente del restos de la cadena polipeptídica; esta región puede ser uno o más segmentos de una cadena de AA, la cadena entera o varias cadenas.
  - **Evolutivo.** Una región de la proteína que se encuentra en la naturaleza aislada o en más de un contexto de múltiples dominios.
  - **Funcional.** Una región de la estructura de una proteína que esta asociada con una determinada función.
- **Motivo.** Región contigua dentro de una proteína que se distingue por propiedades bien definidas:
  - **Secuencial.** Patrón de AA conservados que es común a un grupo de proteínas.
  - **Estructural**. Es una combinación de elementos estructurales secundarios con un arreglo geométrico específico. Por ejemplo: b-sandwiches, helix-turn-helix (HTH), helix-hairpin-helix, etc.
- **Repeticiones.** Simetría y duplicación estructural son muy comunes en proteínas. Eventos de duplicación y re-arreglos genómicos.
- Complejos. 50-70% de las proteínas se encuentras en complejos homoméricos.
  - Superficies isologas (homologas).
  - Superficies heterologas.

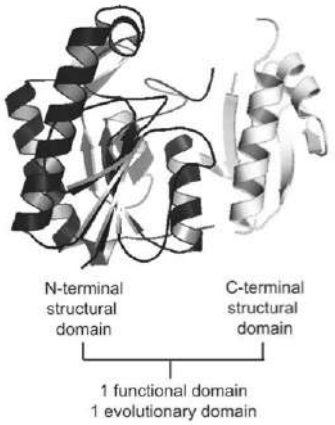


Fig.1. Domains in the structure of α-aminopeptidase DppA (pdb 1hi9).

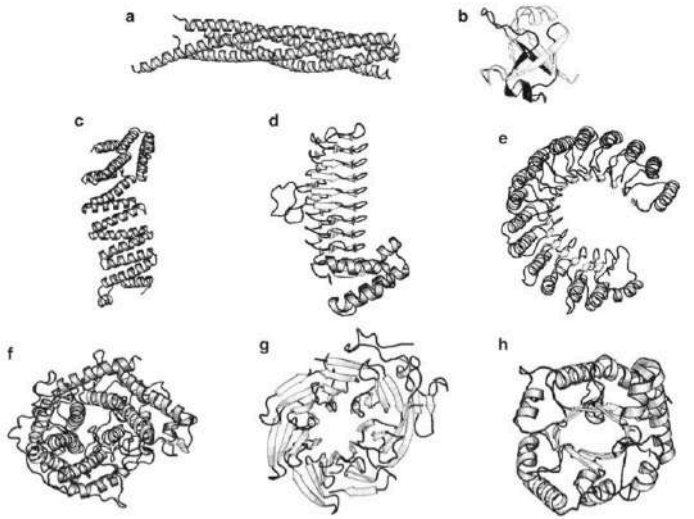


Fig. 4. Representative repetitive structures. (a) Coiled coil (pdb 1n7s), (b) structural repeats in globular domain (pdb 1cz4), (c) α-solenoid (pdb 1qqe), (d) β-solenoid (pdb 2jf2), (e) βα-solenoid (pdb 2bnh), (f) α-toroid (pdb 1gai), (g) β-toroid (pdb 1erj), and (h) βα-toroid (pdb 2jk2).

Reglas estructurales y excepciones.

- **Una secuencia muchos plegamientos.** Equilibrio entre al menos dos estados conformacionales.
- **Secuencias camaleónicas.** Cadenas de aminoácidos idénticos adoptan estructuras secundarias alternativas (a-hélice, hoja-b, vuelta).
- **Decaimiento del plegamiento.** Es un evento de deleción que afecta el plegamiento común de la proteína.
- **Transiciones de plegamiento.** Una parte de una proteína puede tener u a lata similaridad pero se pliega en una estructura secundaria diferente.
- **Transiciones de arquitectura.** Inserción de estructura secundaria adicional a plegamiento común puede desencadenar una nueva arquitectura.
- **Permutaciones circulares.** Cambios en el orden secuencial de los extremos de una proteína.
- **Intercambio y vuelta de hojas beta.** Cambios en la orientación de una hoja beta con respecto al *core*.

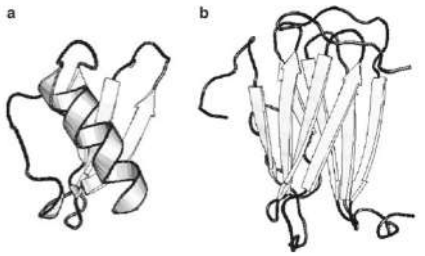


Fig. 5. The structures of two alternative folds of lymphotactin (Ltn10). (a) Monomeric Ltn10 (pdb 1j8i) and (b) dimeric Ltn10 (pdb 2jp1).

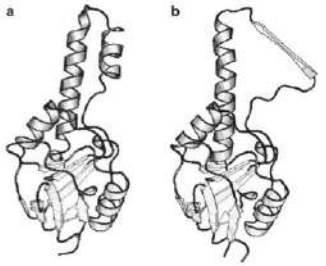


Fig. 7. α-Apical domain of thermosome. (a) Structure of isolated domain, (b) structure of a subunit in the closed thermosome.

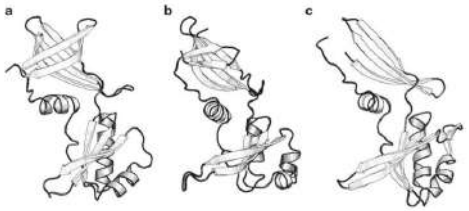


Fig. 9. Fold decay. Structures of exonuclease domains of (a) *Escherichia coli* DNA polymerase (pdb 1q8l), (b) *Sulfolobus solfataricus* DNA polymerase (pdb 1s5j), (c) *Thermococcus gorgonarius* DNA polymerase (pdb 1tgo).

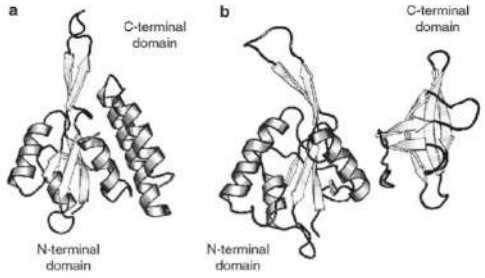


Fig. 10. Fold transition. Structures of (a) RfaH and (b) NusG.

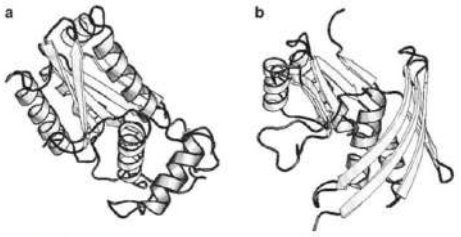


Fig. 11. Architecture transition. Structures of (a) restriction endonuclease *Bam*HI (pdb 1bam) and (b) *Yae*Q (pdb 2g3w).

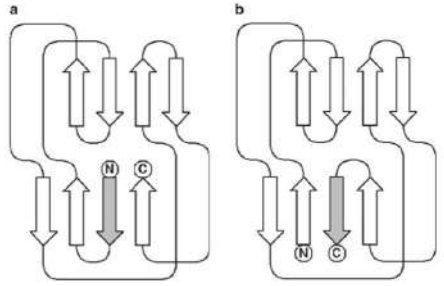
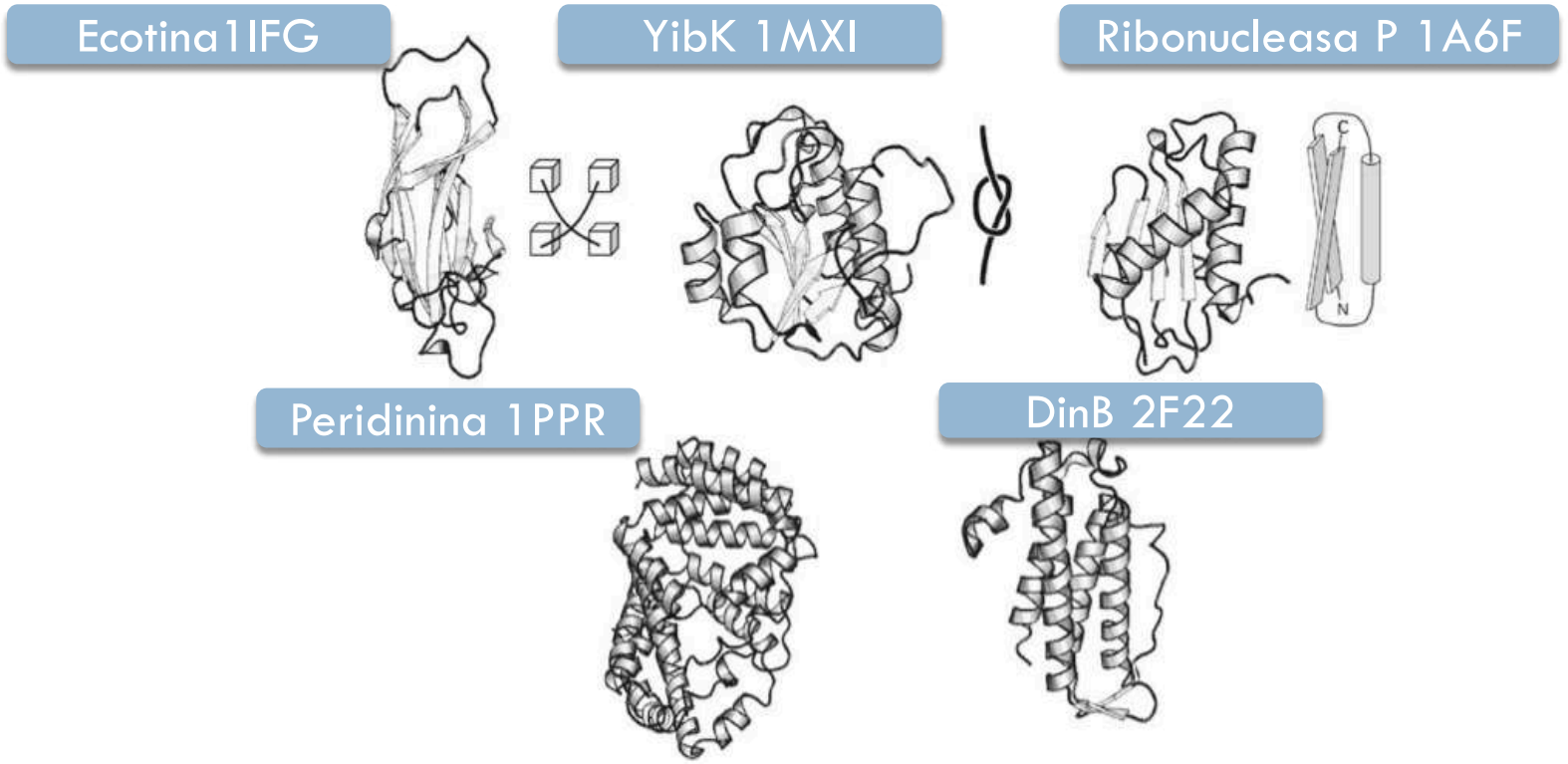


Fig. 12. Circular permutation. Topology diagram of (a) synaptotagmin C2-domain, (b) phospholipase Cδ C2-domain. Circularly permuted strand is shown in *grey*.

Reglas estructurales y excepciones.



- Conexiones entre estructuras secundarias nunca se cruzan y hacen nudos.
- Conexiones de beta-X-beta son hacia la derecha.
- La relación de hojas beta o hélices es compacta en un core hidrofóbico.
- Las piezas de estructura secundaria adyacentes están en estrecho contacto.

Predicciones Estructurales de Proteínas.

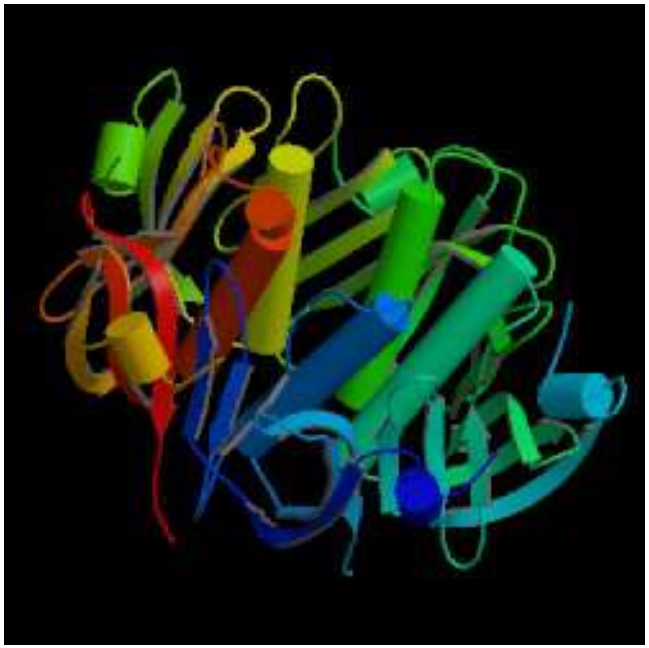
- The Critical Assessment of protein Structure Prediction (CASP) experiments.

  http://predictioncenter.org

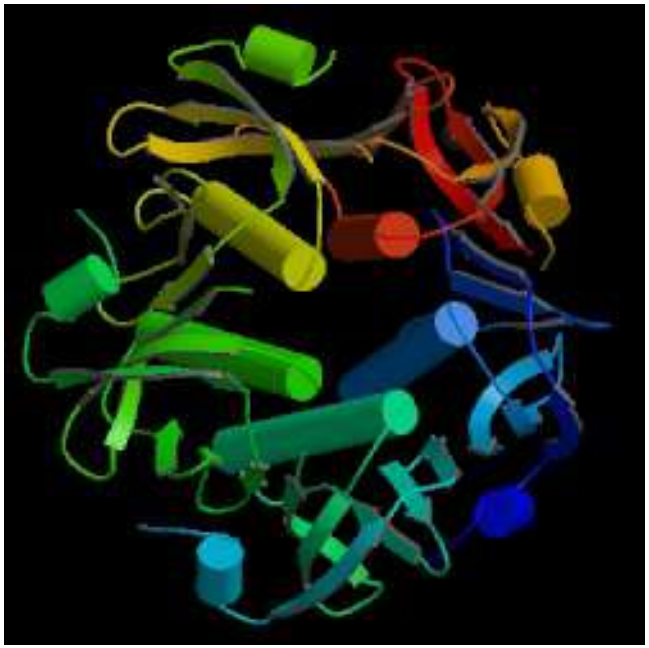
Similaridad de Secuencias

Similaridad de secuencias implica rasgos comunes estructurales, funcionales y evolutivos. Es cuantificable y es uno de los parámetros para determinar homología entre proteínas.

Proteínas Similares: Enterotoxina y Toxina del Cólera



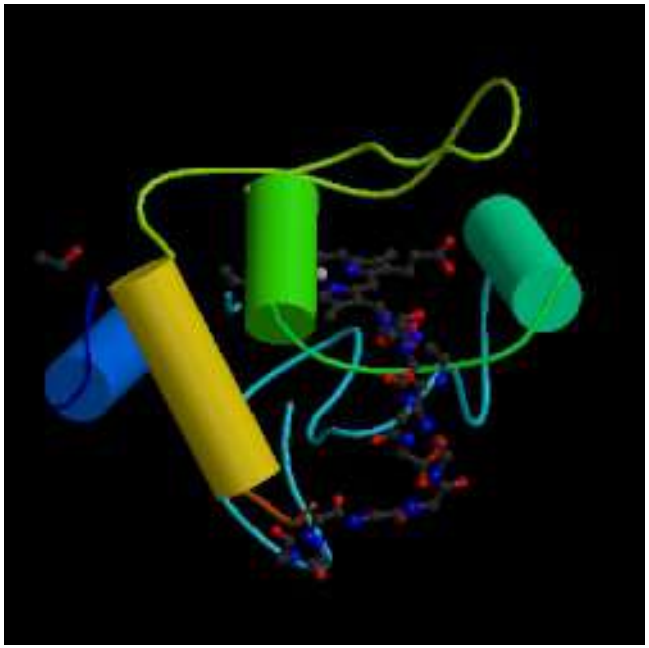
Enterotoxina



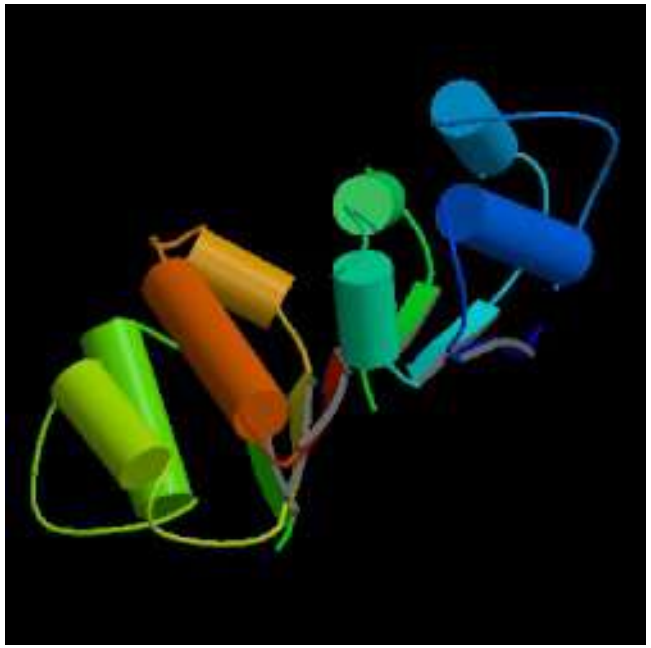
Tóxina del cólera

80% de identidad

Proteínas No Similares: Citocromo y Barstar



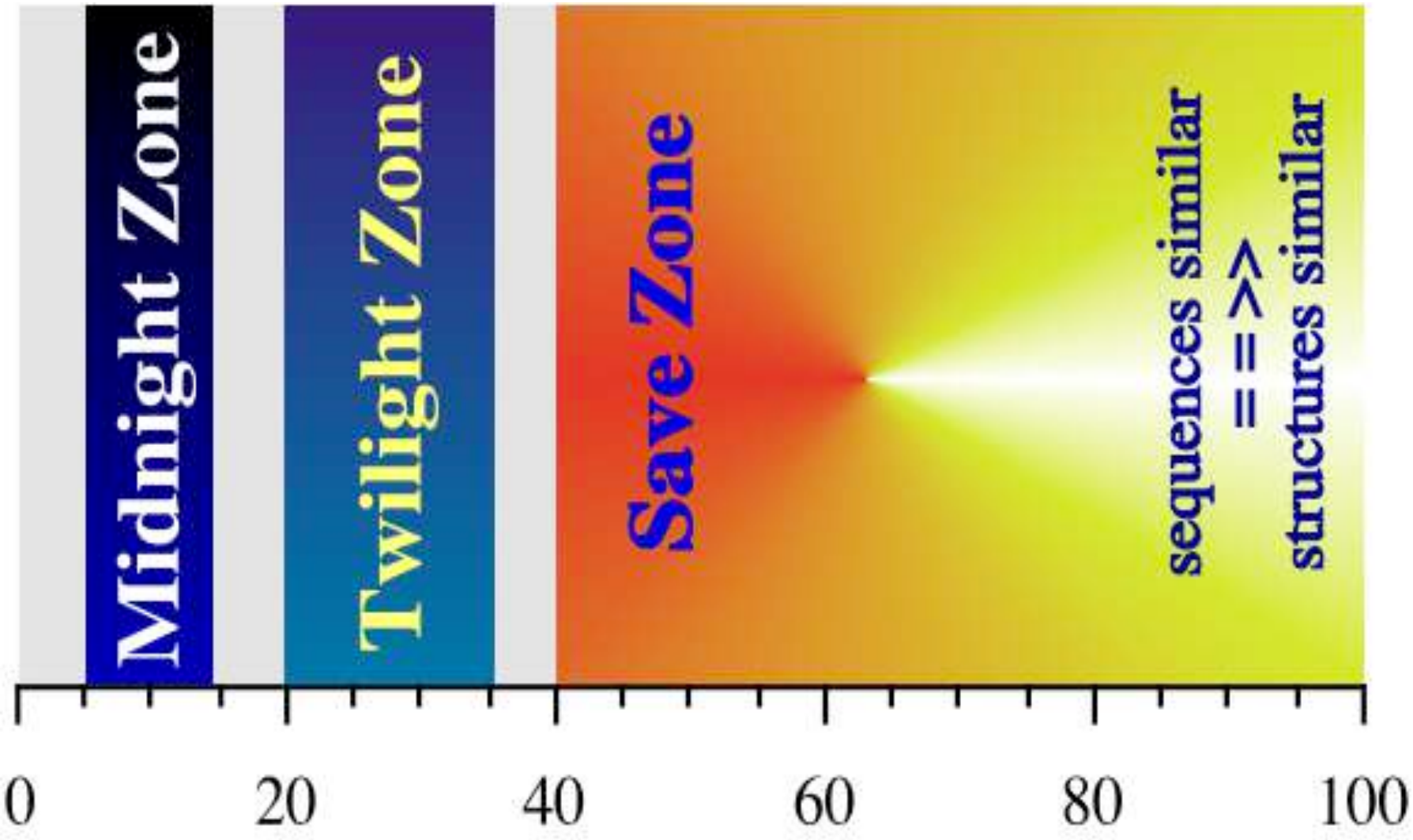
Citocromo

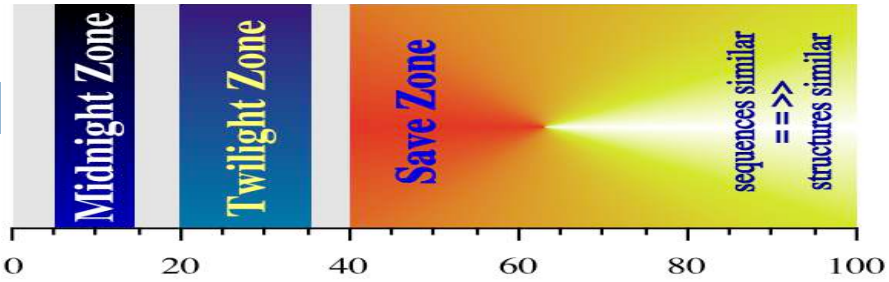


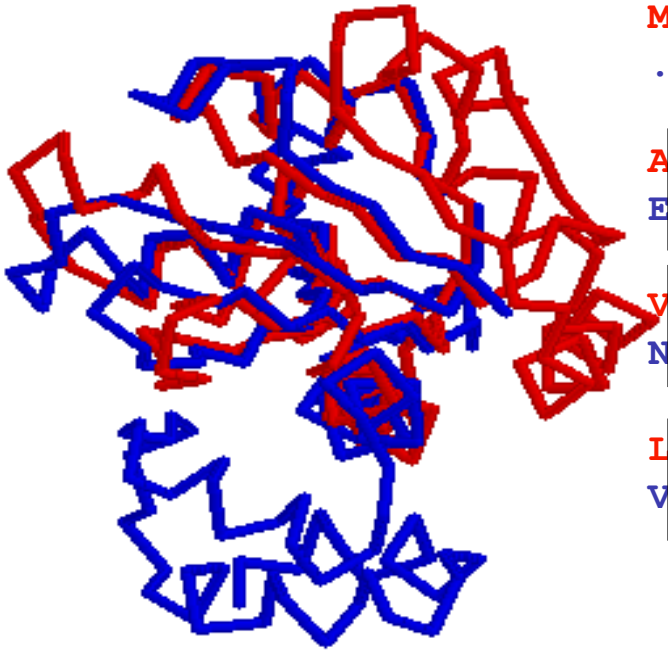
Barstar

Menos de un 20% de identidad

Zonas de Similaridad





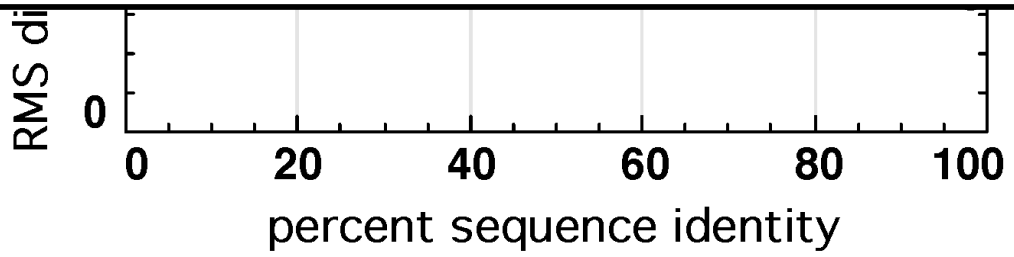


```
MHPRFQTAFAQLADNLQSALEPILADKYFPALLTGEQVSSLKSATGLDED
.....................................................MNRQ

ALAFALLPLAAACARTPLSNFNVGAIARGVSGTWYFGANMEFIGATMQQT
EL-ITEALKARDMAYAPYSKFQVGAALLTKDGKVYRGCNI--ENAAYSMC

VHAEQSAISHAWLSGEKAL----AAITV--NYTPCGHCRQFMNELNSGLD
NCAERTALFKAVSEGDTEFQMLAVAADTPGPVSPCGACRQ-VISELCTKD

LRIHL.....................
VIVVLTNLQGQIKEMTVEELLPGAFSSEDL
```



Búsqueda de homólogos: Métodos para detección de homología.

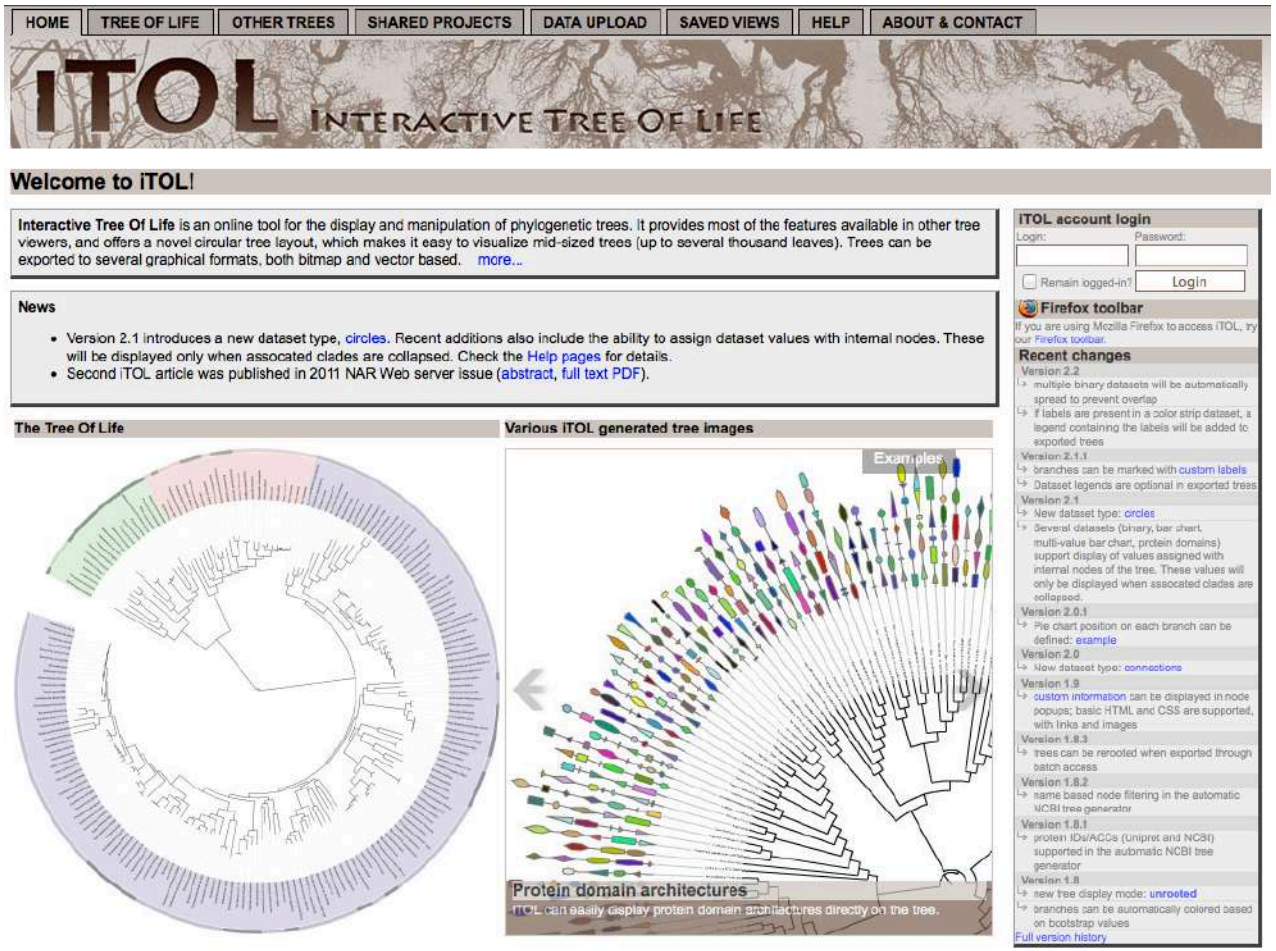
BLAST	Sequence–Sequence	http://blast.ncbi.nlm.nih.gov/
FASTA/Ssearch	Sequence–Sequence	http://fasta.bioch.virginia.edu/ http://www.ebi.ac.uk/Tools/sss/fasta/
CS-BLAST	Sequence (profile)–Sequence	http://toolkit.lmb.uni-muenchen.de/cs_blast/
PSI-BLAST	Profile–Sequence	http://blast.ncbi.nlm.nih.gov/
CSI-BLAST	Profile–Sequence	http://toolkit.lmb.uni-muenchen.de/cs_blast/
HMMER	HMM–Sequence	http://hmmer.org/
SAM	HMM–Sequence	http://compbio.soe.ucsc.edu/HMM-apps/
COMPASS	Profile–Profile	http://prodata.swmed.edu/compass/
PROCAIN	Profile–Profile + additional sequence features + SS[a]	http://prodata.swmed.edu/procain/
COMA	Profile–Profile	http://www.ibt.lt/bioinformatics/coma/
HHsearch	HMM–HMM + SS[a]	http://toolkit.lmb.uni-muenchen.de/hhpred/
PRC	HMM–HMM	http://supfam.org/PRC http://www.ibi.vu.nl/programs/prcwww/

Búsqueda de homologos: Métodos MSA.

1	ClustalW	Método de alineamiento progresivo Incluye procesamiento multicore	http://www.ebi.ac.uk/clustalw	(Thompson y col. 1994)
2	Muscle	Refinamiento sucesivos con árbol calculado	http://www.drive5.com/muscle	(Edgar 2004)
3	T_Coffee	Preprocesamiento de Pares de alineamiento Incluye procesamiento multicore	http://www.tcoffee.org	(Notredame y col. 2000)
4	Mafft	Transformada rápida de Fourie para identificar regiones homologas (FFT).	http://align.genome.jp/mafft	(Katoh y col. 2002)
5	Mafft ginsi	MAFFT Alineamiento global	http://align.genome.jp/mafft	(Katoh y col. 2002)
6	Mafft linsi	MAFFT Alineamiento local	http://align.genome.jp/mafft	(Katoh y col. 2002)
7	Mafft einsi	MAFFT Para secuencias con zonas extensas no alineadas	http://align.genome.jp/mafft	(Katoh y col. 2002)
8	Kalign	Implementación de algoritmo de Wu-Manber	http://msa.cgb.ki.se	(Lassmann y col. 2009)
9	ProbCons	Incluye consistencia probabilística	http://probcons.stanford.edu	(Do y col. 2005)
10	Dialign-TX	Zonas altamente similares son utilizadas como semillas de alineamiento.	http://dialign-tx.gobics.de	(Subramanian y col. 2008)
11	Probalign	Partición de las probabilidades de función. Pares HMM	http://probalign.njit.edu	(Chikkagoudar y col. 2007)
12	MsaProbs	Incluye consistencia probabilística y calculo multicore	http://sourceforge.net/projects/msaprobs	(Liu y col. 2010)
13	Prank	Usa filogenia y reconoce inserciones y deleciones	http://www.ebi.ac.uk/goldman-srv/webPRANK	(Löytynoja y Goldman 2005; Löytynoja y Goldman 2008b)
14	Picxaa PF	Probabilistico y No progresivo, metodo de evaluación de paridad factor	http://www.ece.tamu.edu/~bjyoon/picxaa	(Sahraeian y Yoon 2010)

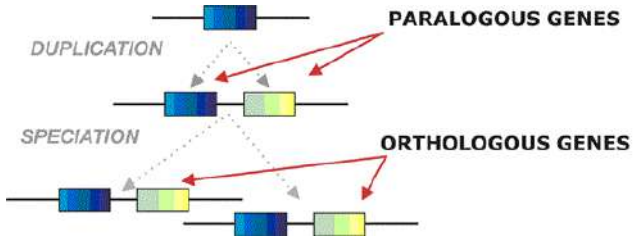
Filogenia

http://itol.embl.de



Ortólogos.
Parálogos.

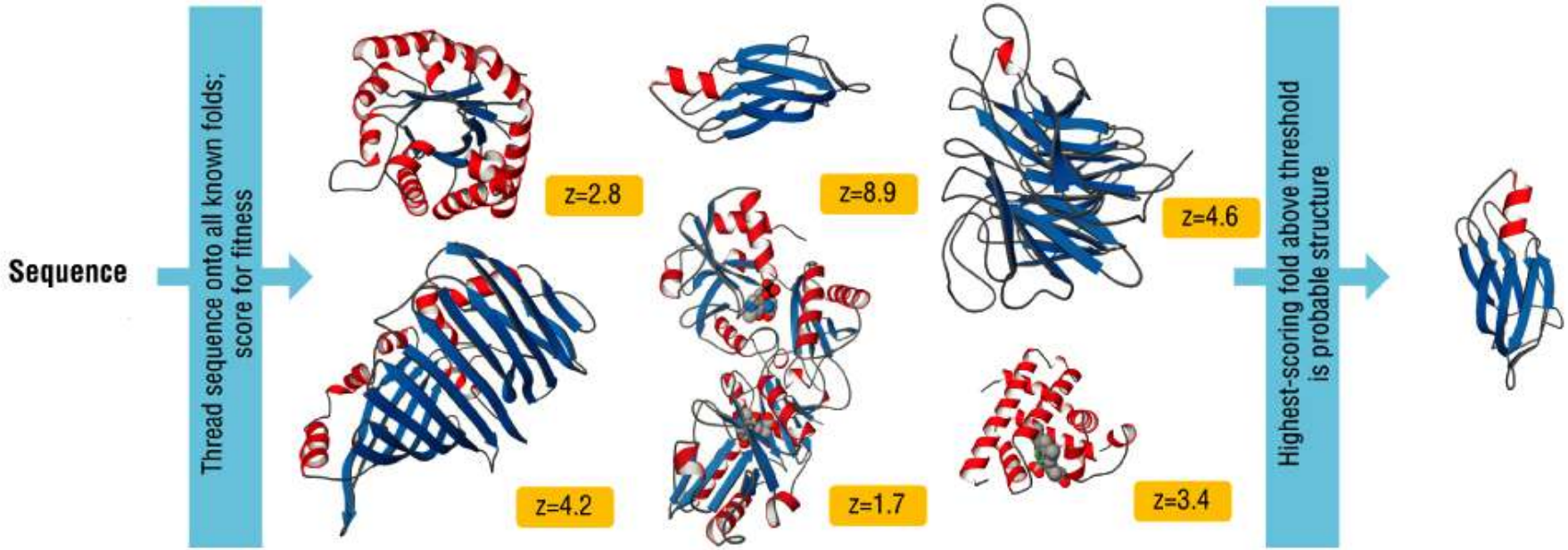
INPARANOID7
http://inparanoid.sbc.su.se



Letunic, I. & Bork, P. (2011)
'Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.',
Nucleic Acids Res ***39(Web Server issue), W475-W478.***

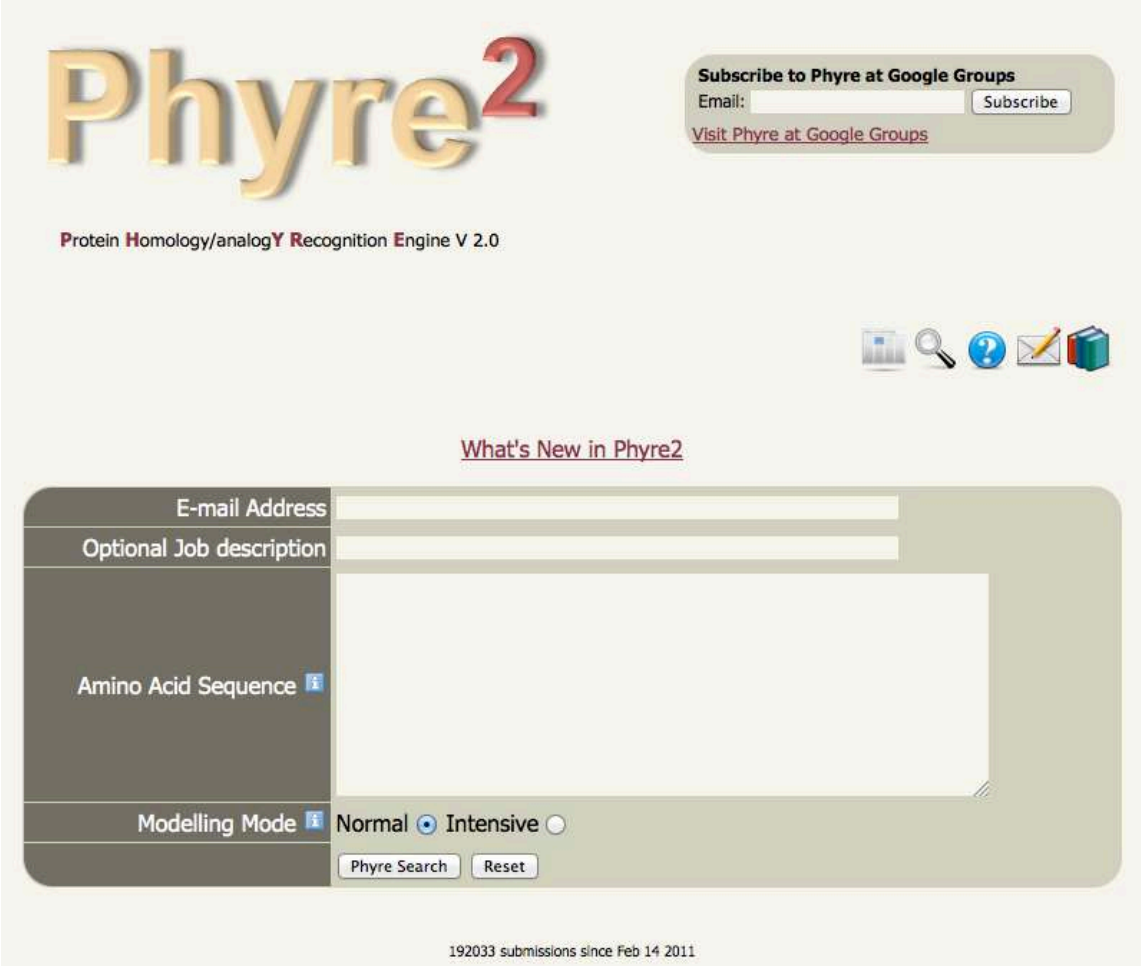
Ostlund, G.; Schmitt, T.; Forslund, K.; Köstler, T.; Messina, D. N.; Roopra, S.; Frings, O. & Sonnhammer, E. L. L. (2010)
'InParanoid 7: new algorithms and tools for eukaryotic orthology analysis.'
Nucleic Acids Res ***38(Database issue), D196-D203.***

Threading


Sequence
Thread sequence onto all known folds; score for fitness
z=2.8
z=8.9
z=4.6
z=4.2
z=1.7
z=3.4
Highest-scoring fold above threshold is probable structure

Búsqueda de moldes. Phyre2

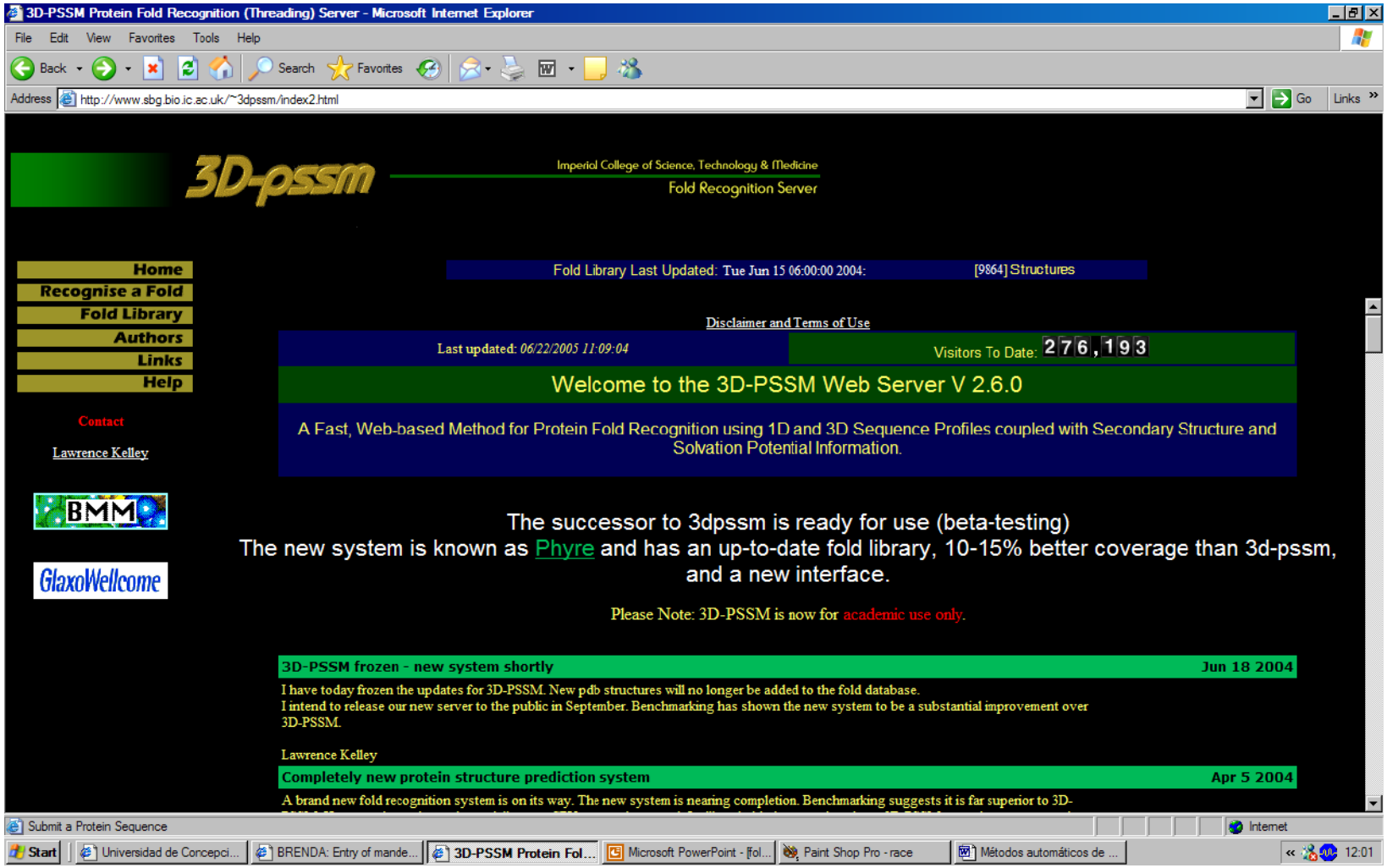
http://www.sbg.bio.ic.ac.uk/phyre2



Kelley LA and Sternberg MJE. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* 4, 363-371

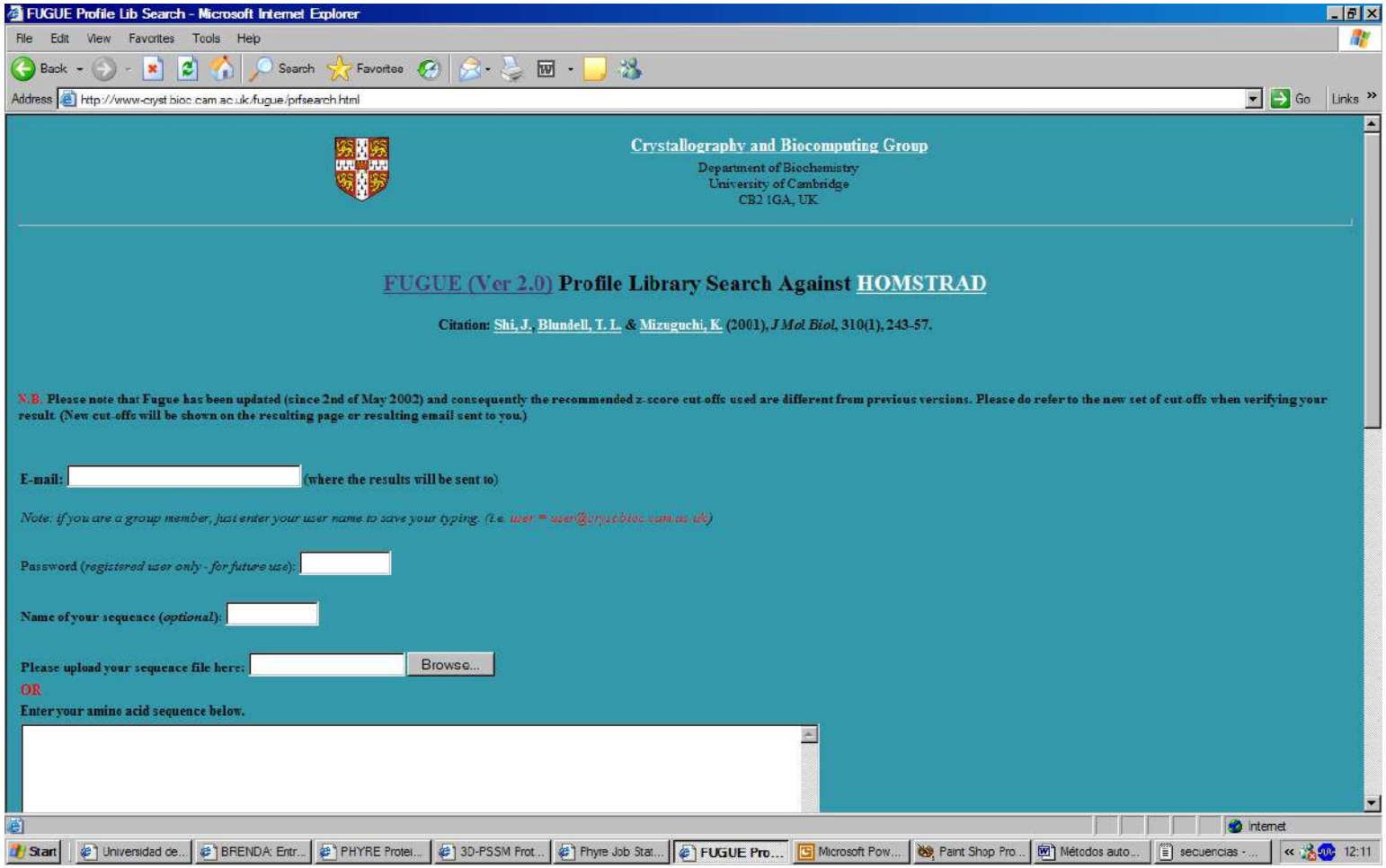
3DPSSM http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html

Busca homólogos remotos, para ello combina perfiles de secuencia con información estructural (potenciales de solvatación e información de estructura secundaria) para realizar el reconocimiento de plegamiento

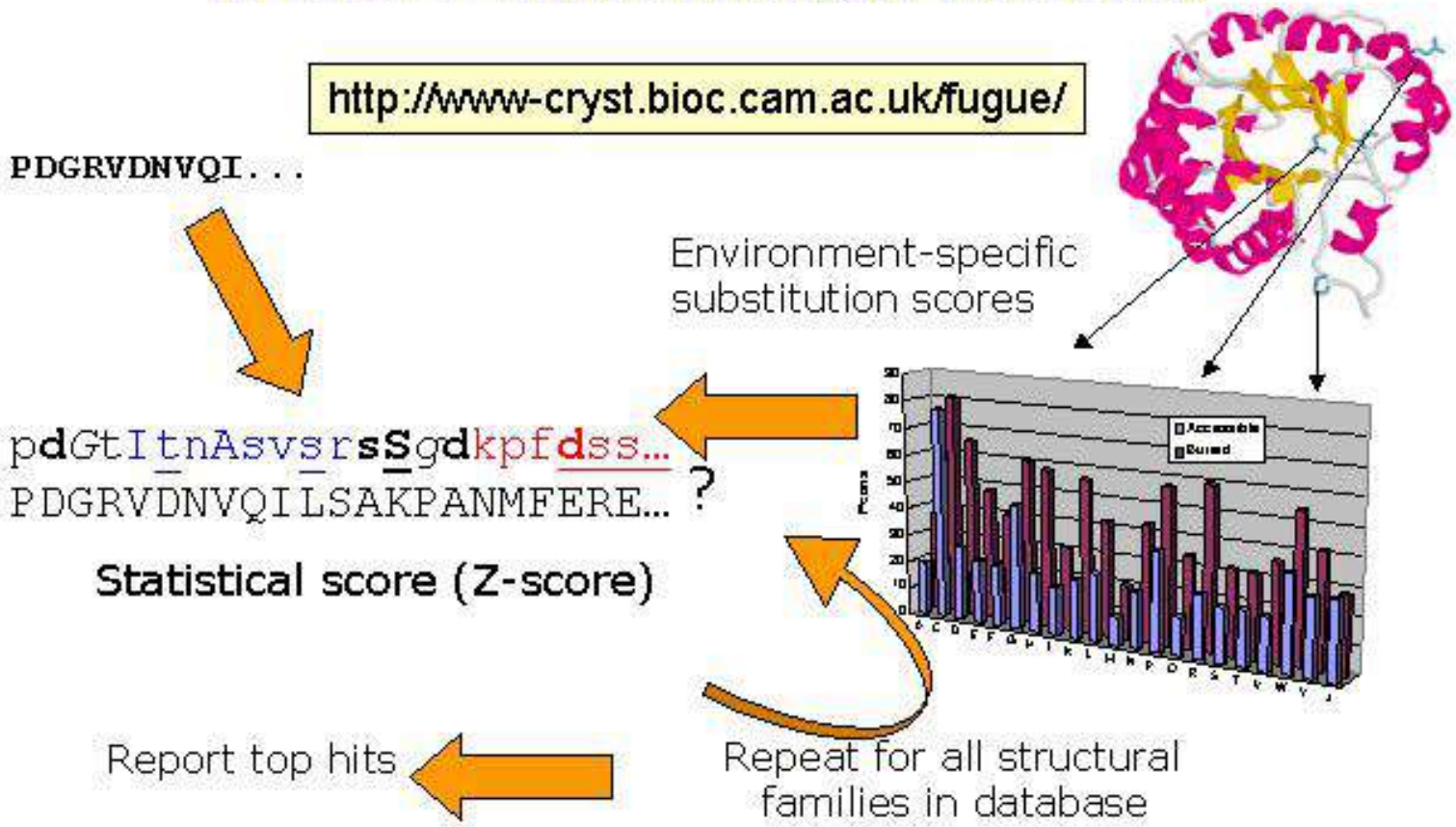


FUGUE http://www-cryst.bioc.cam.ac.uk/fugue/prfsearch.html

Emplea tablas de sustitución específicas para ambientes y penalizaciones para los gaps que son dependientes de estructura, donde los score para el emparejamiento de aminoácidos y las inserciones y delecciones son evaluadas dependiendo del ambiente local de cada aminoácido dentro de una estructura conocida



FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties

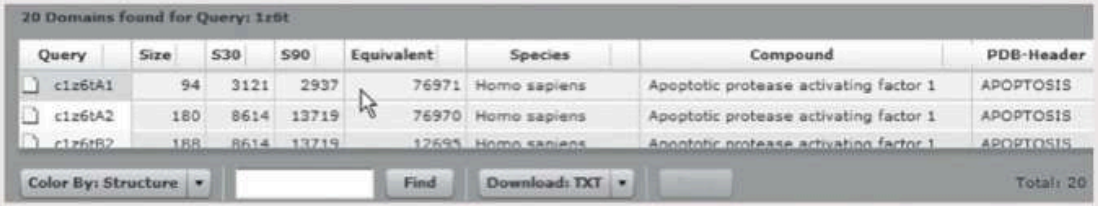


Shi, J., Blundell, T.L., Mizuguchi, K. *J. Mol. Biol.* **310**, 243-257 (2001)

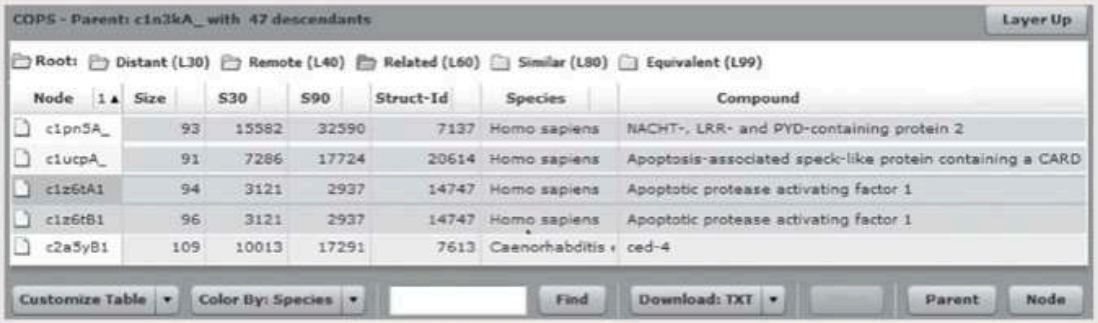
Búsqueda de homologos:
COPS: http://topsearch.services.came.sbg.ac.at

(a) Enter a PDB code like 3bey and press enter/return or the *Search* button.

(b) Search results are listed in the *Selection Widget*. The rows in the *Selection Widget* correspond to the domains of the given PDB file. The first domain is automatically selected (see below) and the respective *Equivalent* layer is opened (see (c)).



(c) After the search (a) has been finished the first domain is automatically selected (b) and its parent on the lowest COPS layer (*Equivalent*) is retrieved from the system. The *Tree Result Table* displays all descendants of the parent of a selected layer. Below, the *Related* layer is selected as indicated by the last opened (red) folder icon.



COPS provee 5 niveles de clasificación:

- *Distant* (30%)
- *Remote* (40%)
- *Related* (60%)
- *Similar* (80%)
- *Equivalent* (99%)