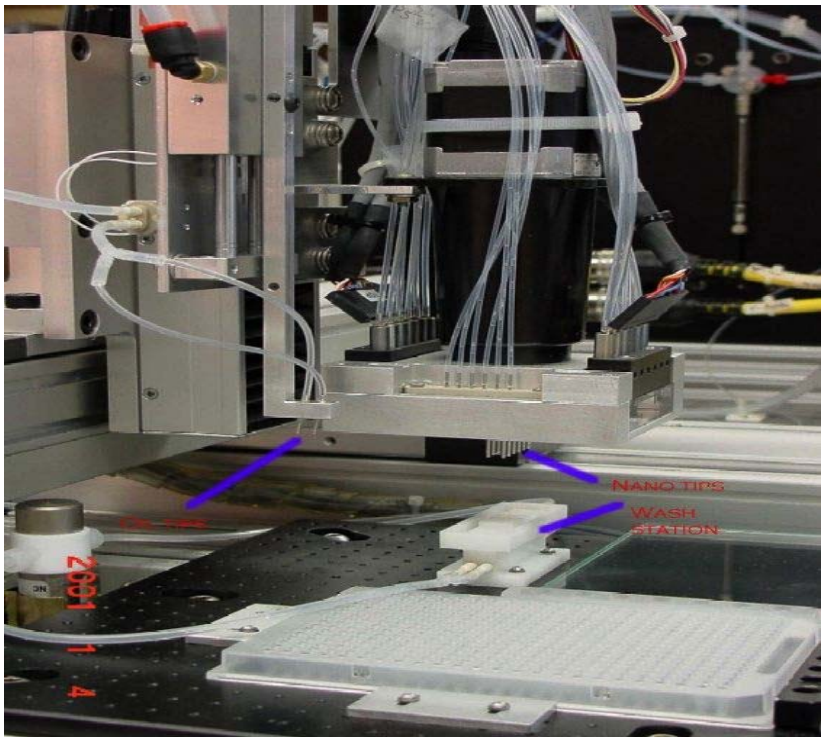
Estructura de Proteínas

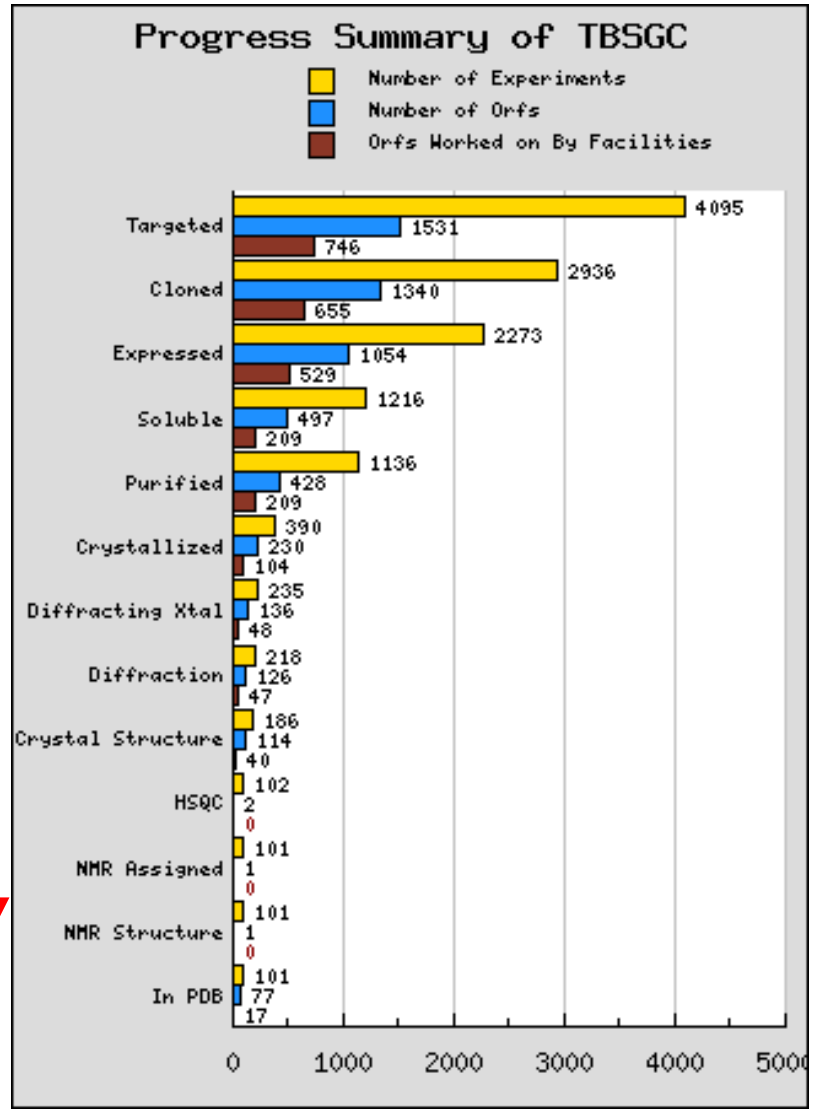
Predicción de Estructura Tridimensional de Proteínas

Florencio Pazos (CNB-CSIC)

Florencio Pazos Cabaleiro
Computational Systems Biology Group (CNB-CSIC)
pazos@cnb.csic.es

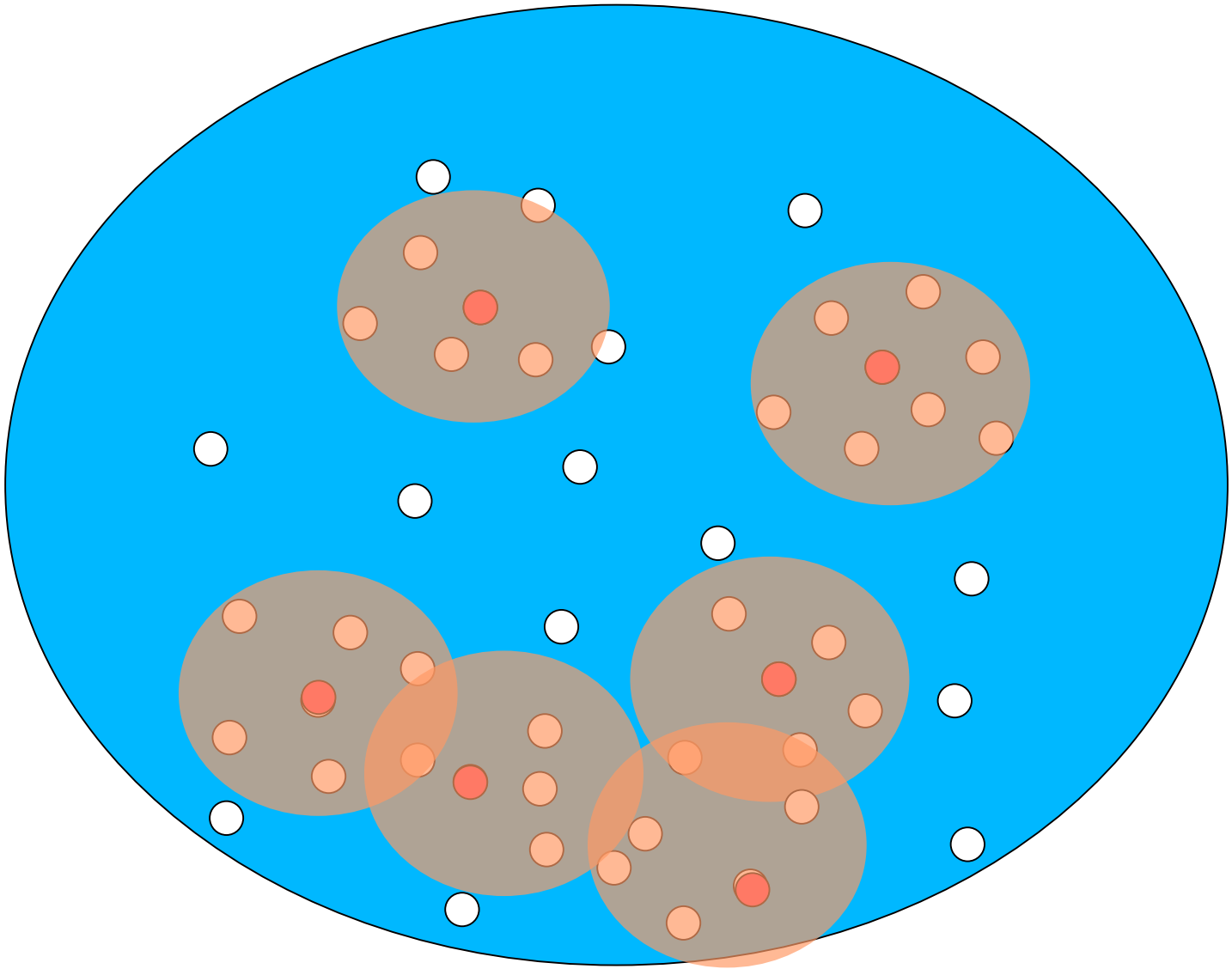
Genómica Estructural



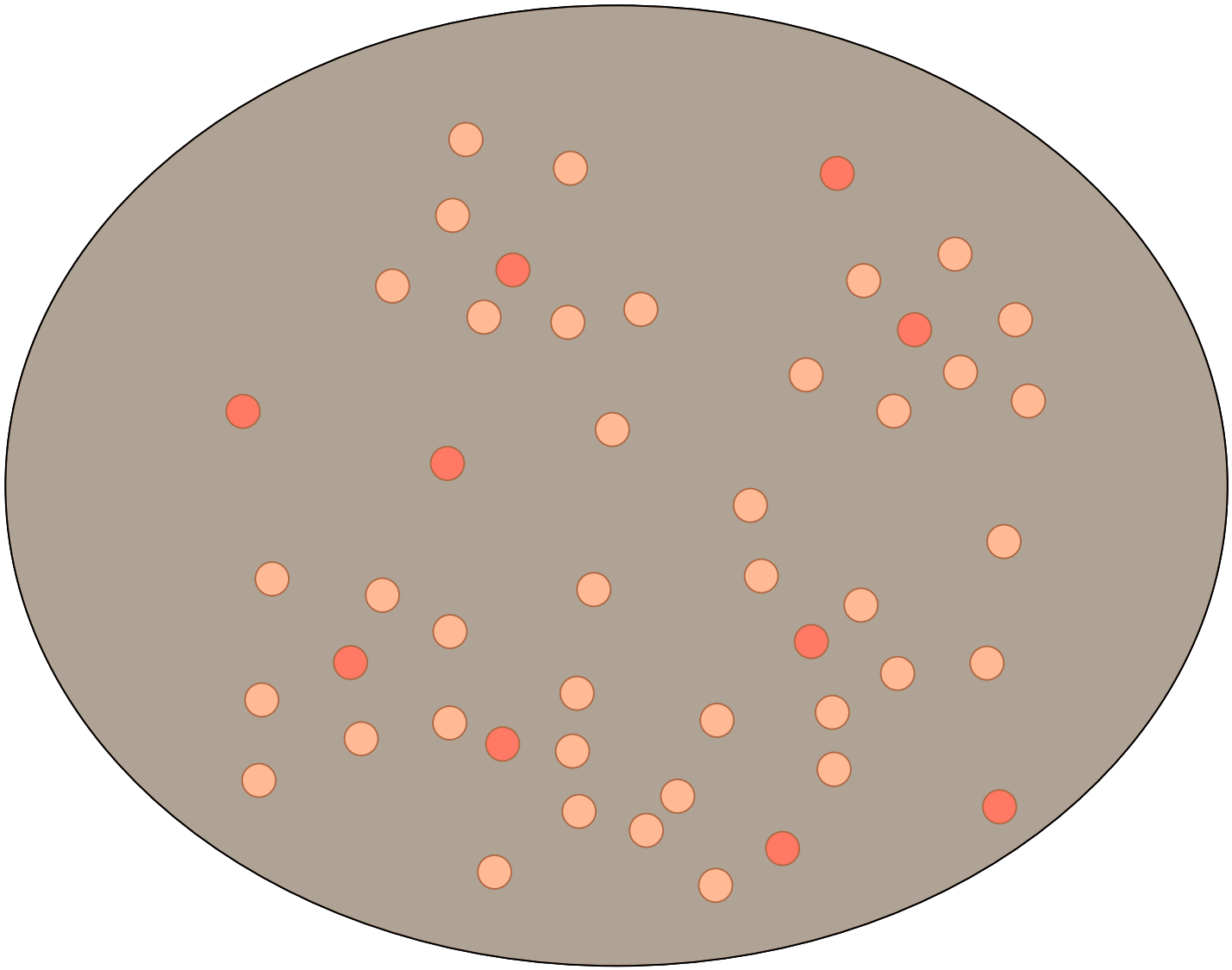


Vitkup, D., Melamud, E., Moult, J. and Sander, C. (2001) Completeness in structural genomics. *Nat Struct Biol*, **8**, 559-566.

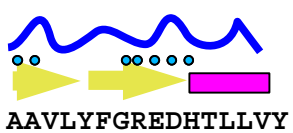
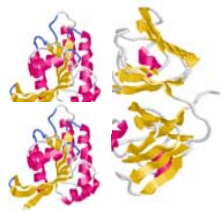
Genómica estructural y predicción de estructura



Genómica estructural y predicción de estructura



Predicción de estructura de proteínas
Clasificación de los métodos de predicción

Nivel estructura proteínas / Representación de la proteína / Uso de información extra	Secundaria	--------	terciaria	cuaternaria
	1D 	2D 	3D 	4D 
Ab Initio	pred. str. secundaria	mutaciones correlacionadas	- dinámica molecular - minimización de energía	*docking*
No Ab-Initio	pred. str. secundaria		- modelado por homología - *threading*	*docking con filtros*

Predicción de Estructura de Proteínas
Métodos 3D

- *Ab initio*

- Diseño por homología/Modelado comparativo.

- Reconocimiento de plegamiento/
  diseño por homología remota/*threading*

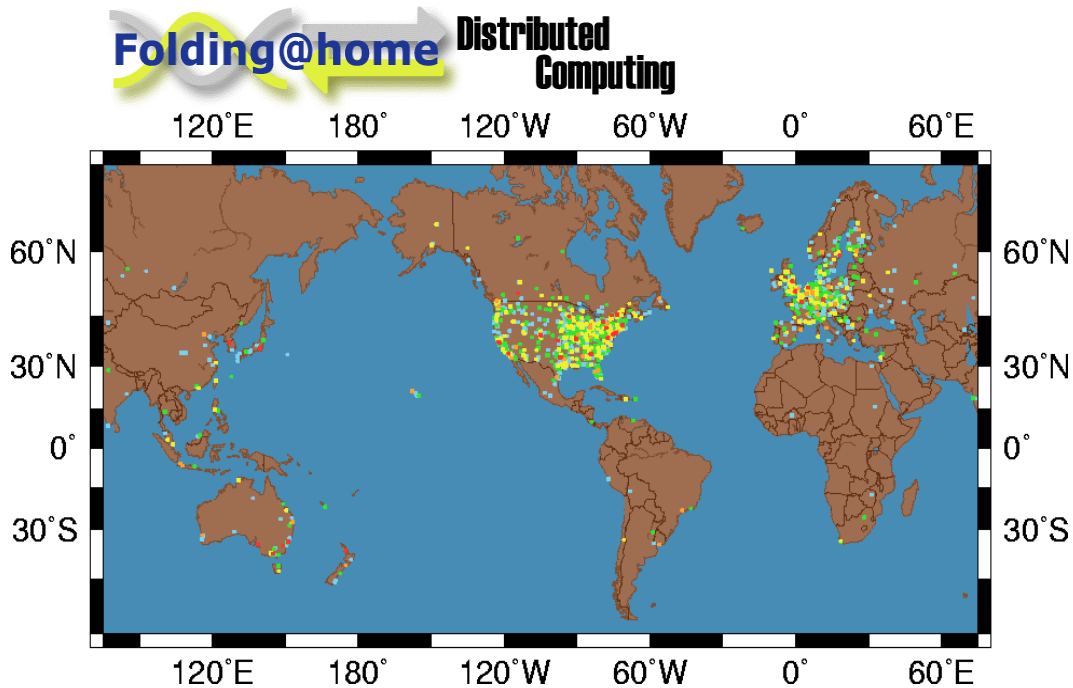
Ab Initio

Basados en principios físico/químicos básicos (potenciales de interacción, ...).

Solo usan la secuencia primaria como entrada.

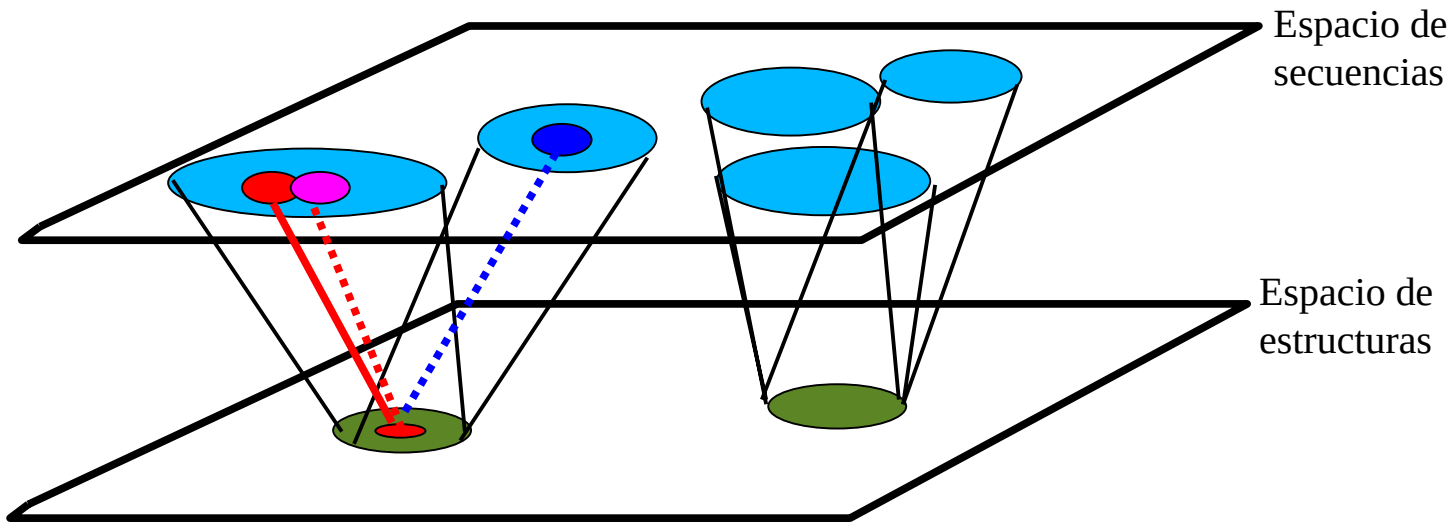
Interesantes pues son los que más conocimiento aportan sobre el proceso de plegamiento.

No usables aún (en general) para predicción 3D por
- Potenciales empíricos y semiempíricos con pequeñas inexactitudes que se acumulan para proteínas grandes y/o simulaciones largas.
- Requieren muchísima potencia de calculo.



http://folding.stanford.edu/

Modelado por Homología *vs. Threading*

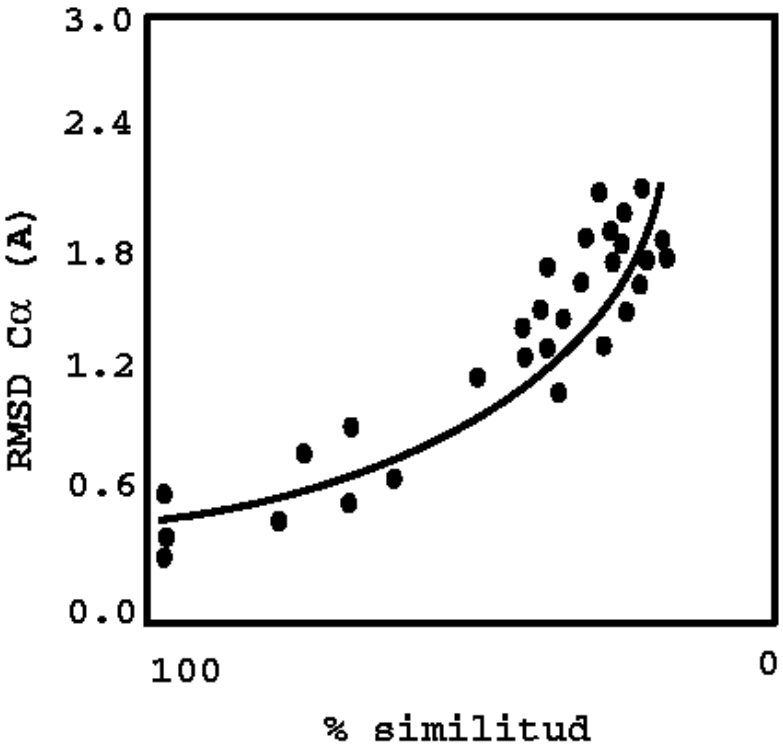


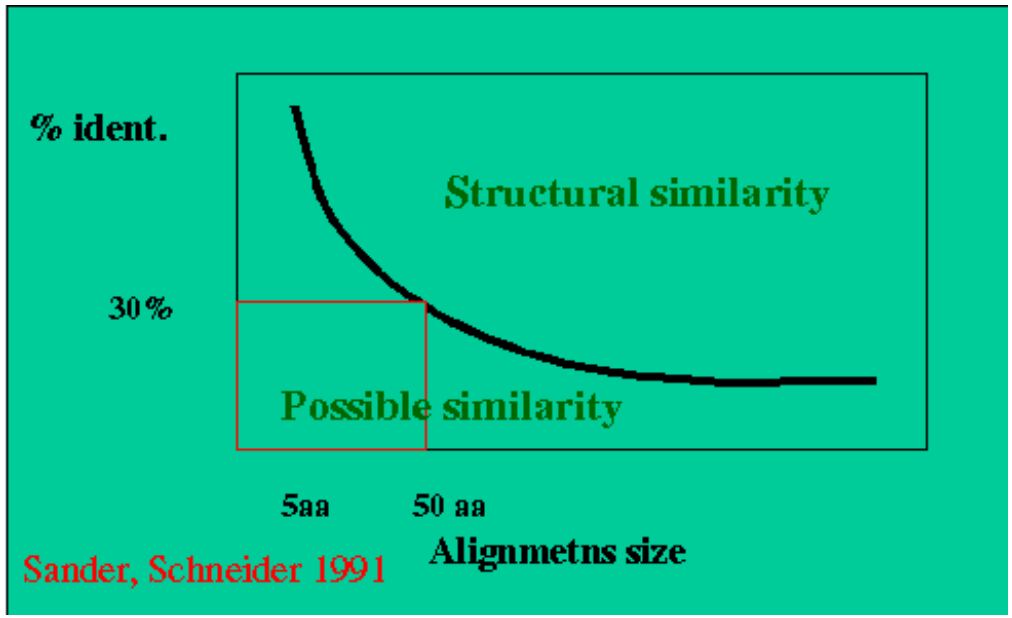
Diseño por homología

threading

Modelado por homología

Relación entre parecido estructural y parecido en secuencia





Chothia, C. & Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823-826.

Sander, C. & Schneider, R. (1993) The HSSP data base of protein structure-sequence alignments. *Nucleic Acids Res.*, **21**, 3105-3109.

Modelado por Homología. Estrategia general

- Localizar molde
- Generar alineamiento entre secuencia problema y molde
- Para átomos del esqueleto tomar las coordenadas del molde.
- Para aa. conservados tomar mismas coordenadas para las cadenas laterales
- Cadenas laterales de otros aa.
  - usar librerías de rotámeros
  - usar la mayor cantidad de átomos posibles (Cβ -> C γ, ...)
- Modelar loops (inserciones y delecciones)
- Optimizar empaquetamiento (MD, ...)
- Evaluar calidad del modelo

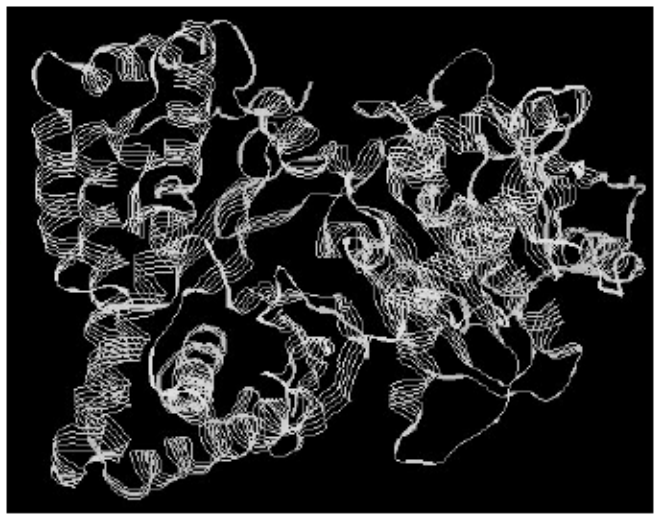
Modelado por Homología. Alineamiento molde-proteína problema

```
AHPLTSDFGGHTERDLHA
||  |||:  |||||  |:||
AHTLTSEGGGHTEADVHA
```

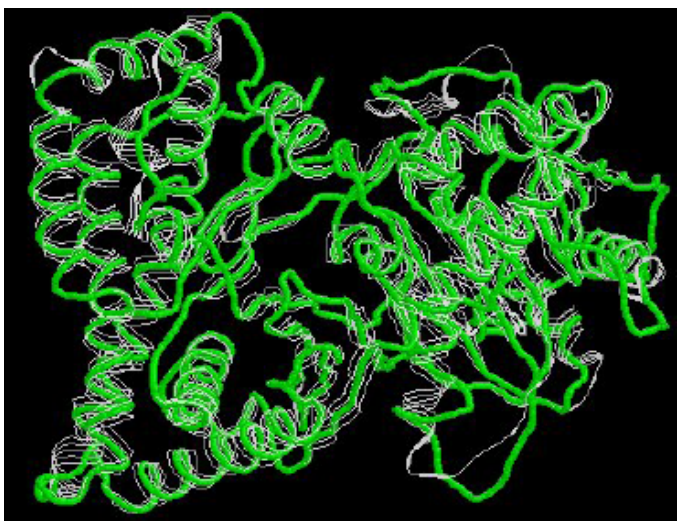
target

?

template



template (1ndb)



model

Modelado por Homología. Alineamiento molde-proteína problema

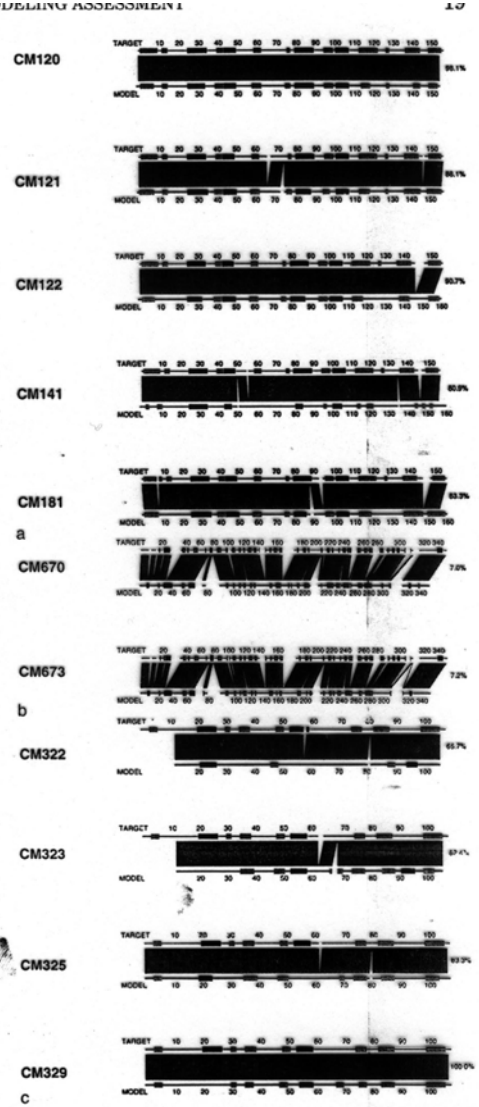
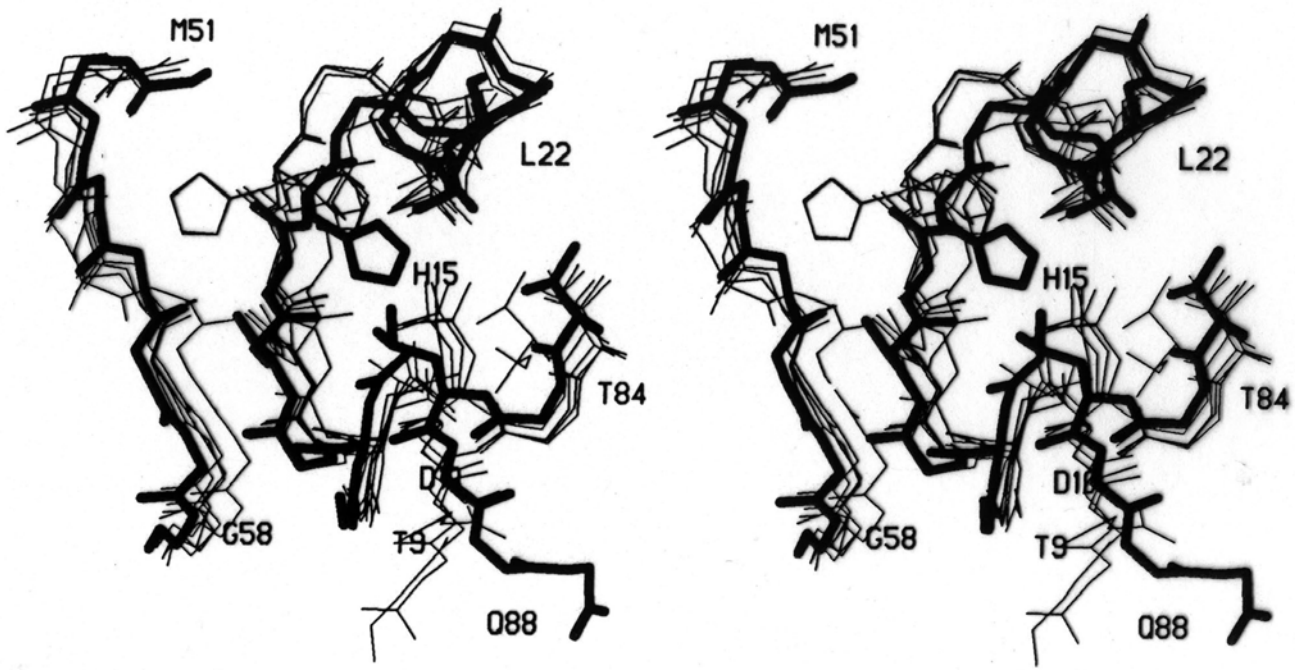


Fig. 3. Comparisons of sequence alignment with structural alignment for selected examples. Complete sets can be viewed on our Web page. The horizontal axes show the structural alignment with helices indicated in black and strands in grey; the lines between the axes indicate the sequence aligned residues. Thus, if the alignment is 100% correct, all the lines will be vertical; lines at an angle indicate errors in the alignment. **a:** T0001 (dihydrofolate reductase). **b:** T0027 (pectate lyase). **c:** T0009 (stellacyanin).

MOLECULAR MODELING OF PROTEIN TERTIARY STRUCTURES



Modelado por Homología. Cadenas laterales

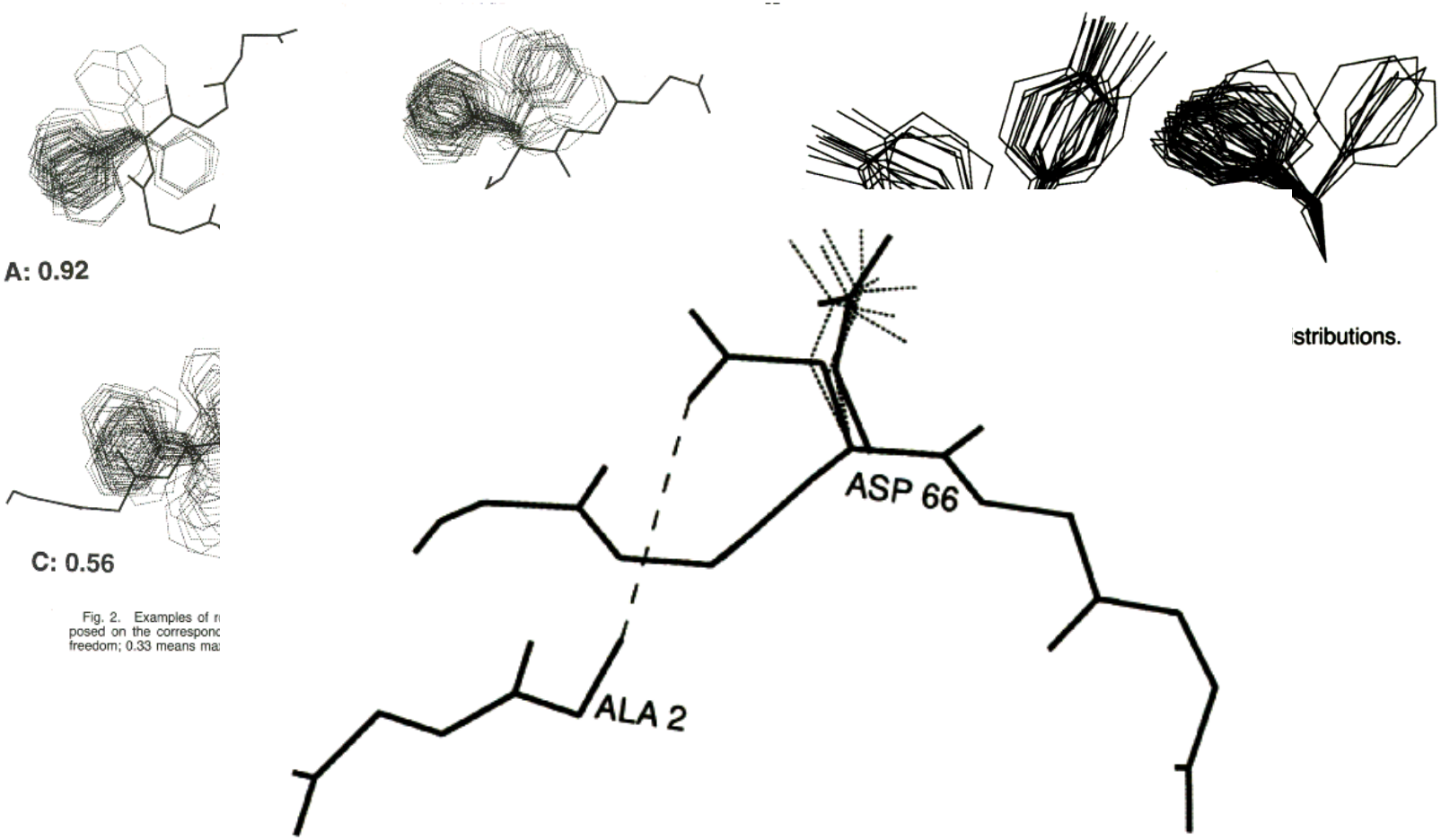


Fig. 2. Examples of r
posed on the correspond
freedom; 0.33 means ma

stributions.

Fig. 6. Rotamer distribution and real position for aspartic acid-66 in HPR. The hydrogen bond between residue 66 and the backbone of residue 2 is indicated. The position-specific rotamers for residue 66 are shown in dashed lines.

Modelado por Homología. Nucleo Estructural

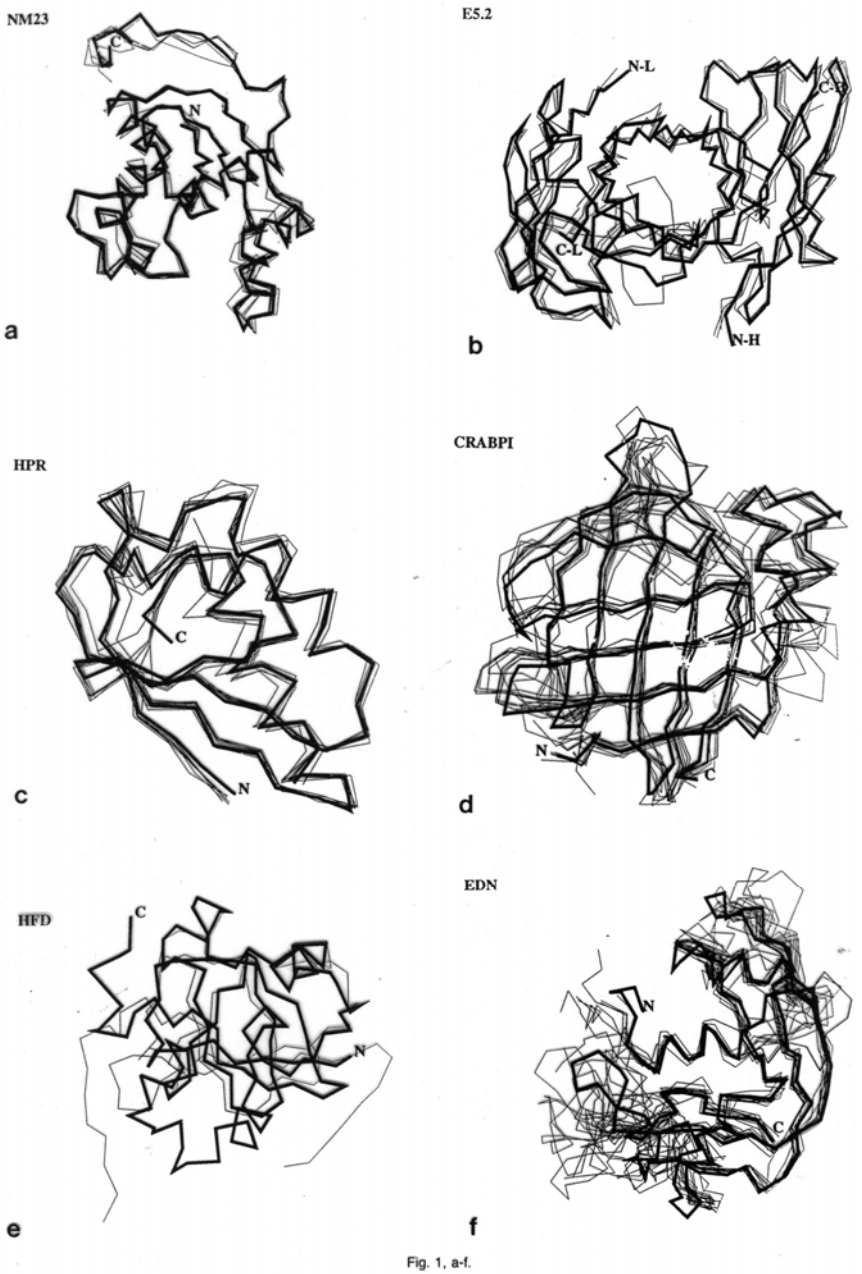
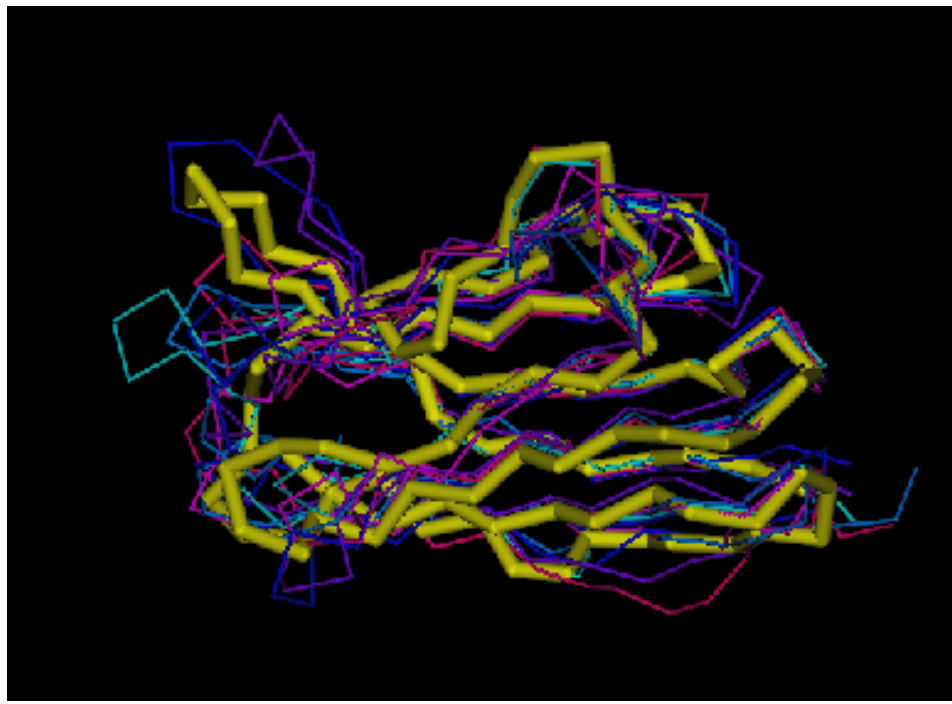
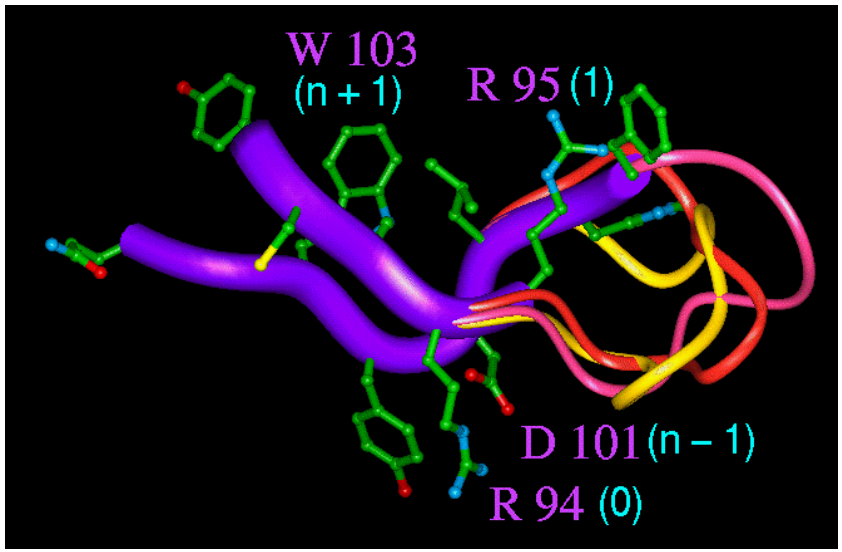
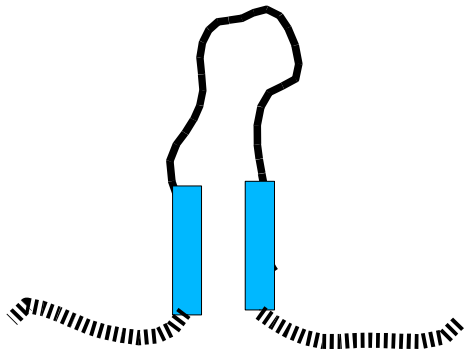


Fig. 1, a-f.



Modelado por Homología. *Loops*





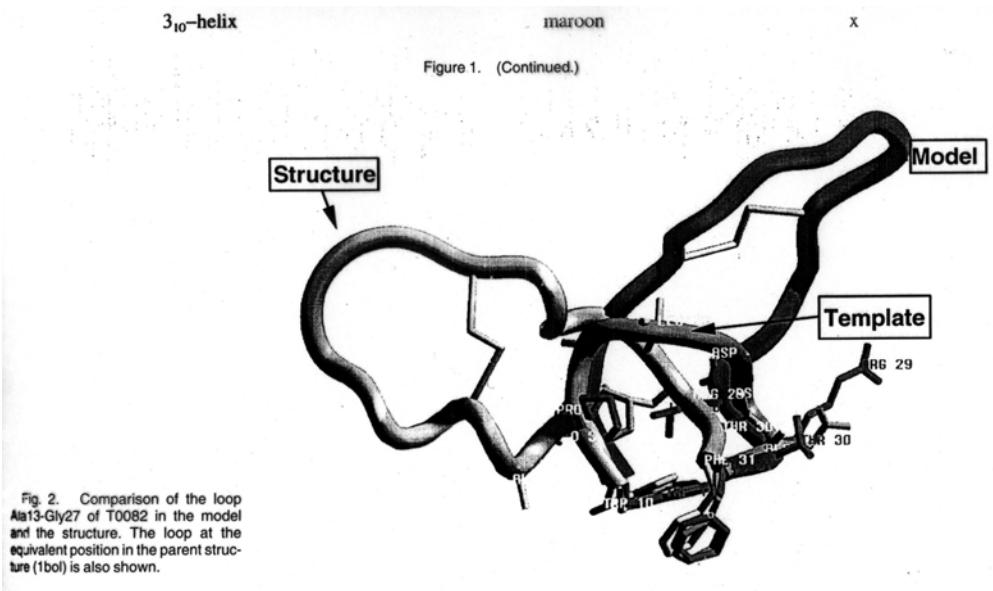
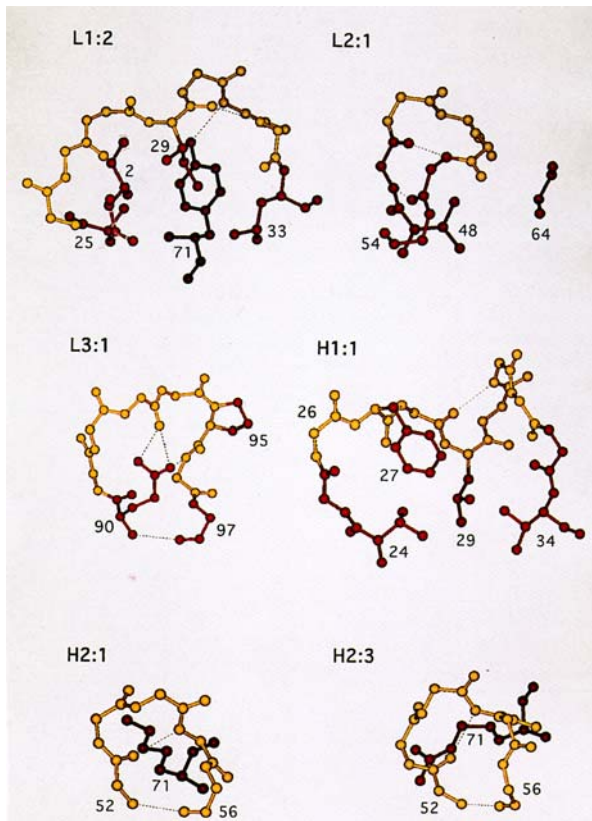
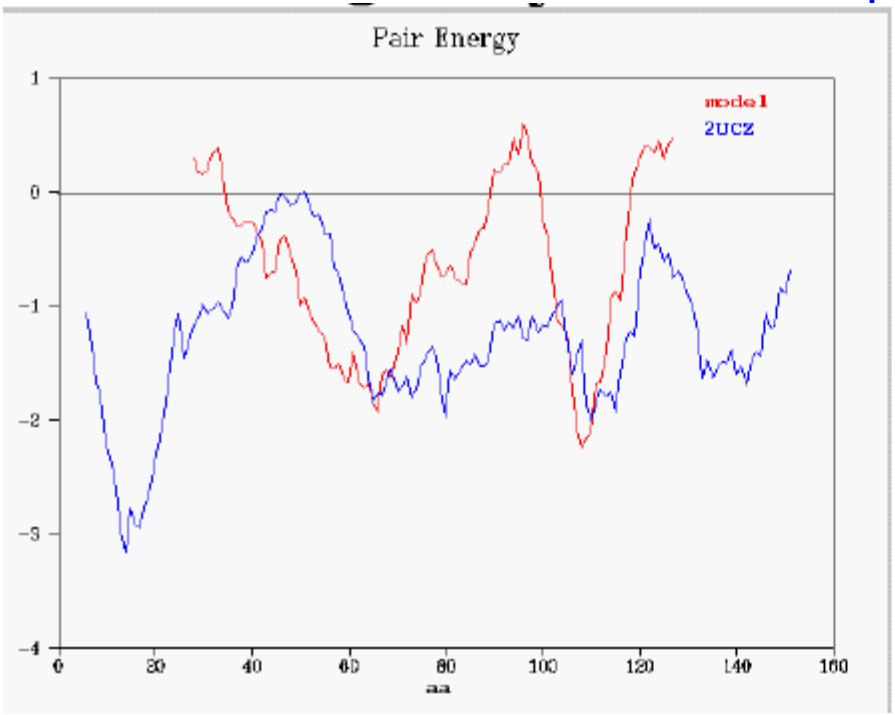
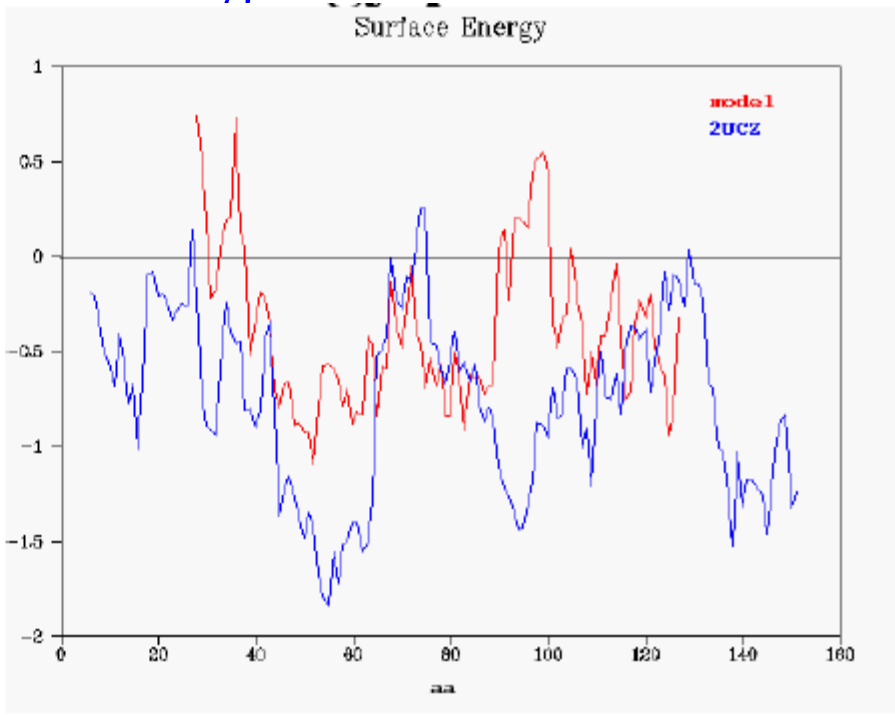


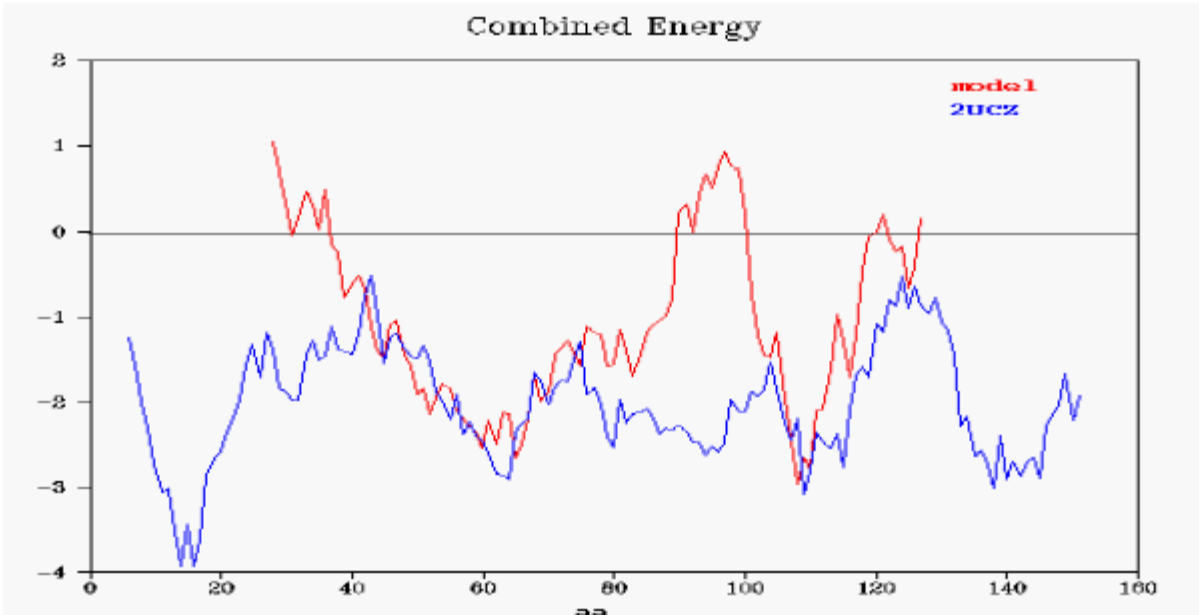
Fig. 2. Comparison of the loop Ala13-Gly27 of T0082 in the model and the structure. The loop at the equivalent position in the parent structure (1bol) is also shown.



Modelado por Homología







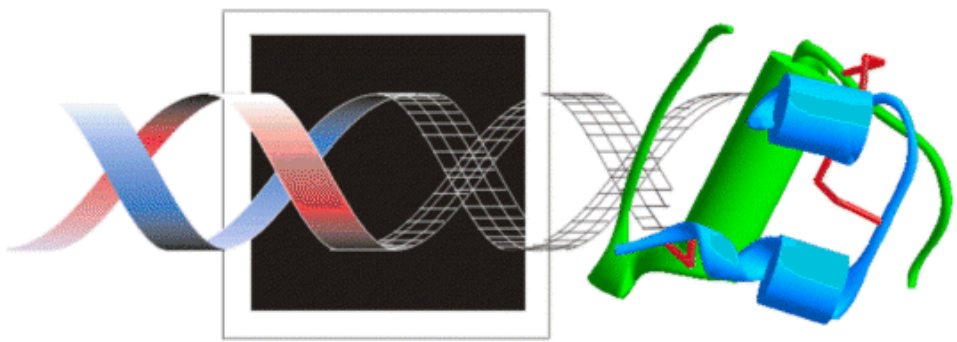
Modelado por Homología
Servidores Públicos y Repositorios de Modelos

SWISS-MODEL - *www.expasy.ch/swissmod/SWISS-MODEL.html*
An automated comparative modelling server (ExPASy, CH)

CPHmodels - *www.cbs.dtu.dk/services/CPHmodels/*
Server using homology modelling (BioCentrum, Denmark)

SDSC1 - *cl.sdsc.edu/hm.html*
Protein structure homology modeling server (San Diego, USA)

3D-JIGSAW - *www.bmm.icnet.uk/servers/3djigsaw/*
Automated system for 3D models for proteins (Cancer Research UK)



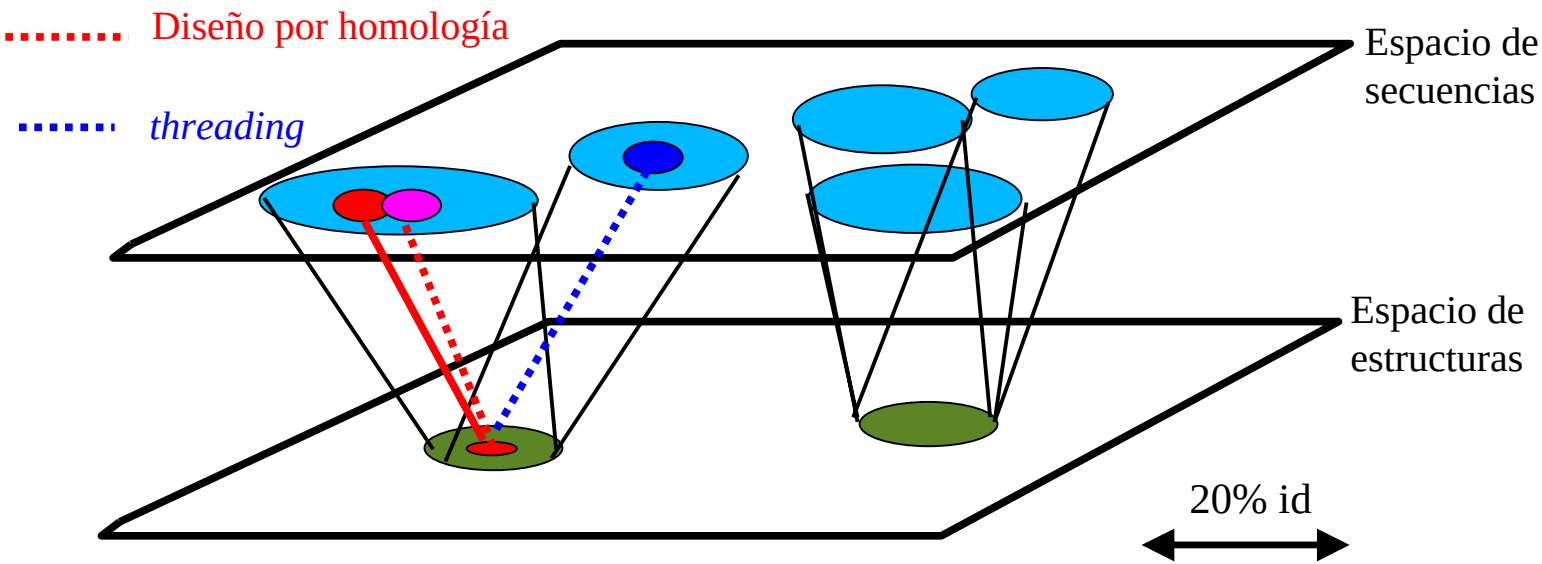
http://www.expasy.ch/swissmod/SM_3DCrunch.html

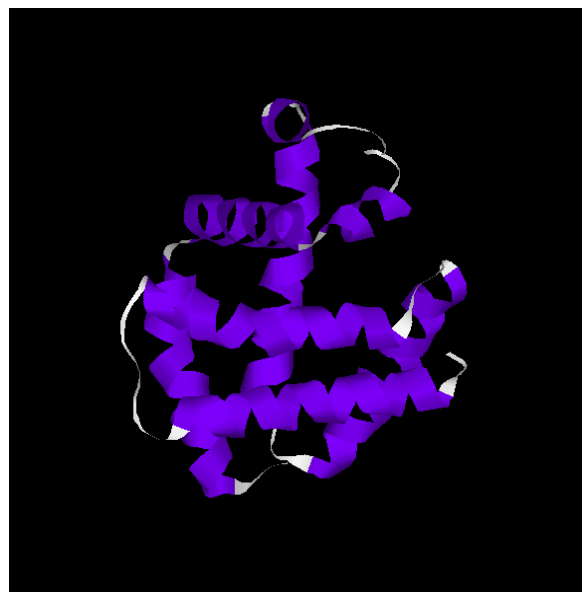
MOD BASE

Database of Comparative Protein Structure Models

http://pipe.rockefeller.edu/modbase

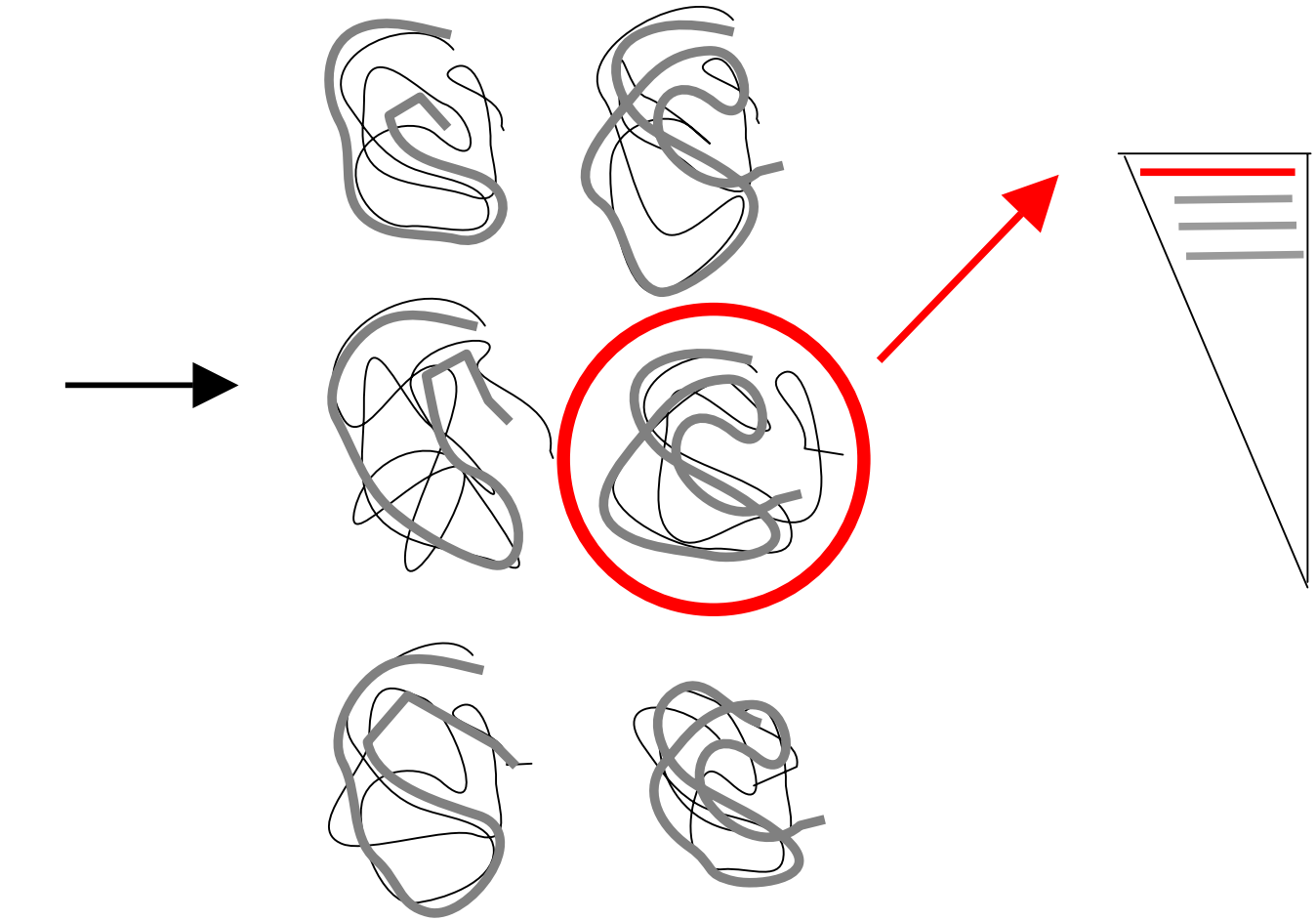
Modelado por Homología *vs. Threading*



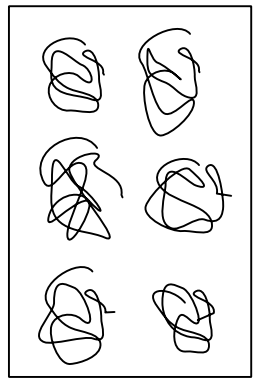


Threading. Estrategia general

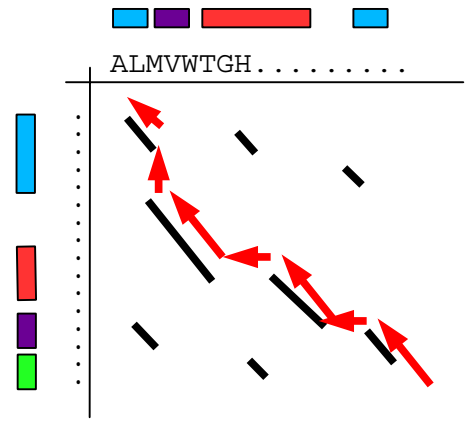
Secuencia problema



Algoritmos de *threading*. General.



1. Library of protein structures (**fold library**)
   - all known structures
   - representative subset (seq. similarity filters)
   - structural cores with loops removed



2. Binary alignment algorithm with **Scoring function**
   - contact potential
   - environments

   Instead of aligning a sequence to a sequence, align strings of descriptors that represent 3D structual features.
   Usual Dynamic Programming: score matrix relates two amino acids
   Threading Dynamic Programming: relates amino acids to environments in 3D structure

3. Method for generating models via alignments (same as homology modl.)

Algoritmos de *threading*
Función de Puntuación

- Aminoácido en ambiente similar a como suele estar en estructuras conocidas.

- Potenciales de solvatación.

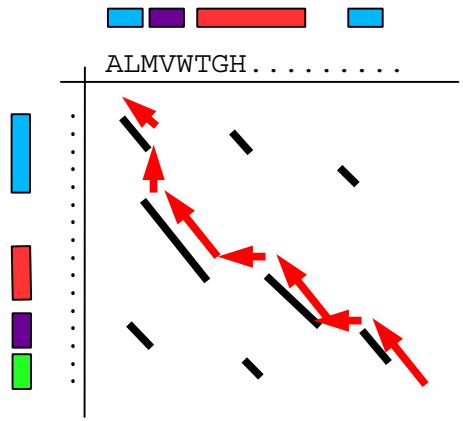
- Potenciales de contacto.

- Coincidencia de estructuras secundarias (real y predicha) y accesibilidades.

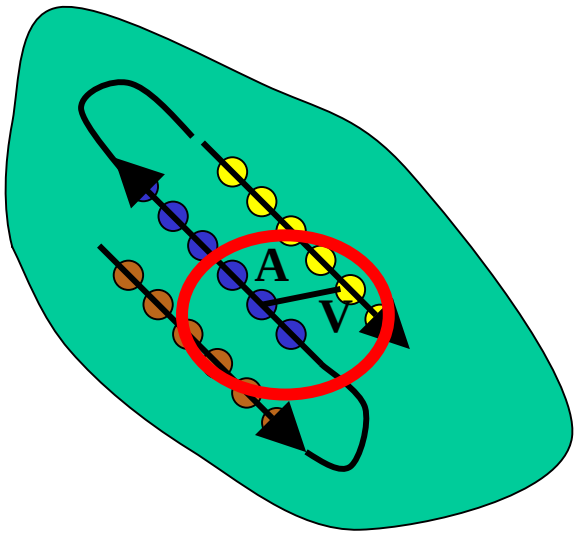
- Matrices de homología remota extraídas de alineamientos estructurales.

- .....

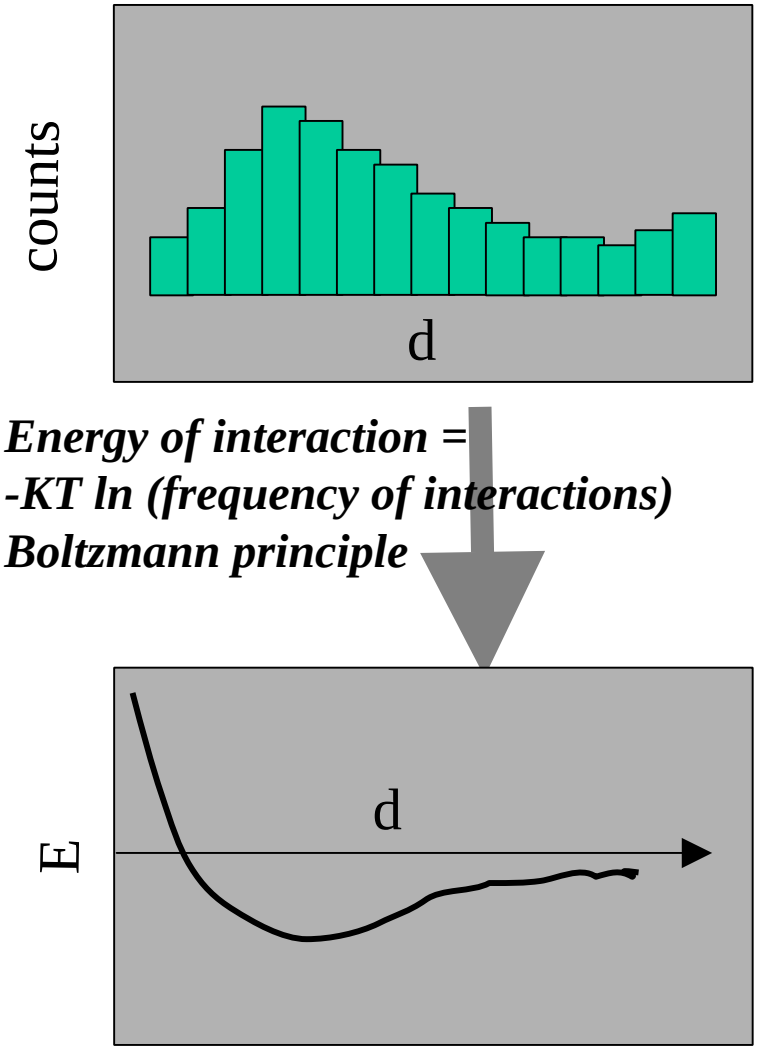
- *Búsqueda con Modelos de Markov (HMMs).*



Algoritmos de *threading*
Potenciales de contacto (potenciales estadísticos)



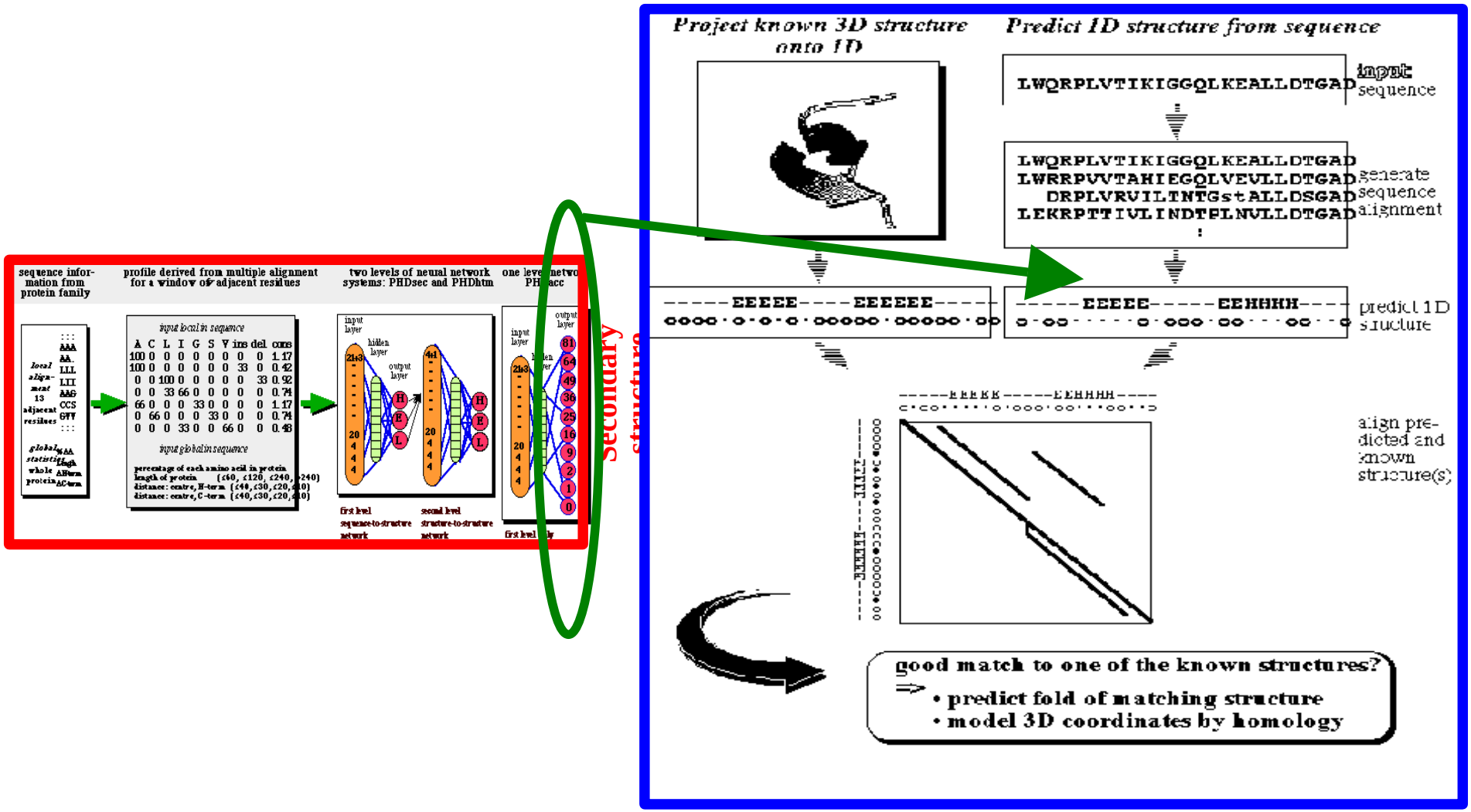
Count pairs of each residue type at different separations



Jones, D., Taylor, W. and Thornton, J. (1992) A new approach to protein fold recognition. *Nature*, **358**, 86-89.

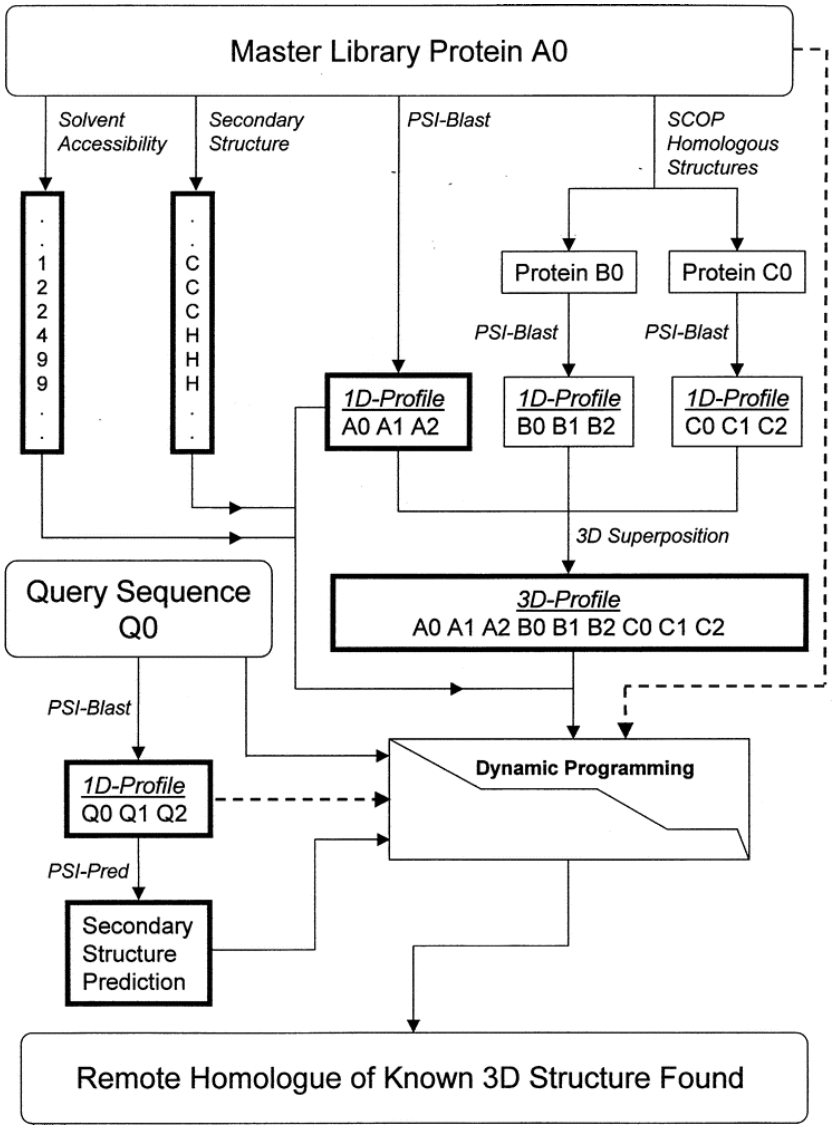
Sippl, M.J. (1995) Knowledge-based potentials for proteins. *Curr Opin Struct Biol*, **5**, 229-235.

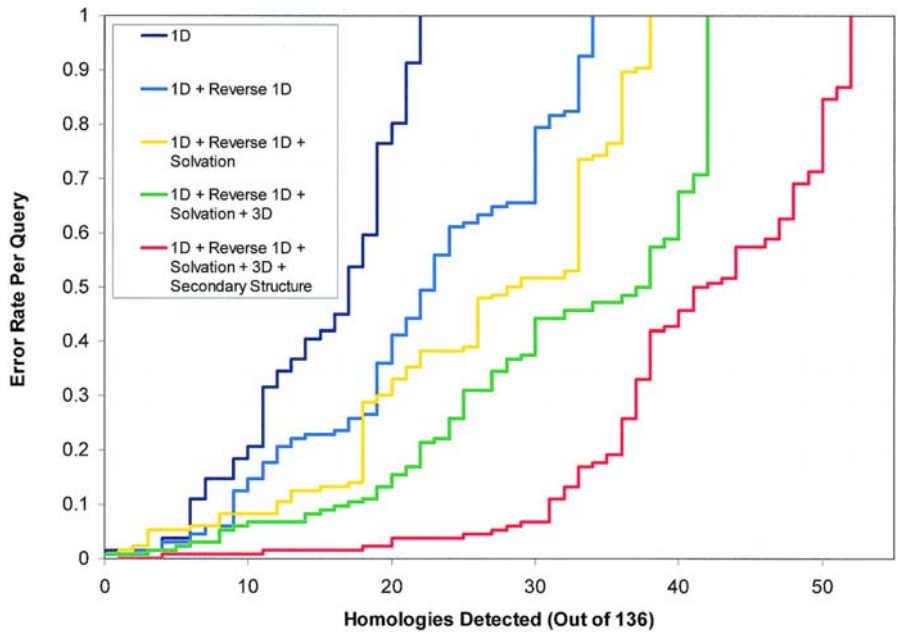
Algoritmos de *threading*
Coincidencia de estructura secundaria y accesibilidad



Rost, B. (1995) TOPITS: Threading one dimensional predictions into three dimensional structures. In Rawlings, C. and als, e. (eds.), *Third International Conference on Intelligent Systems for Molecular Biology*. AAAI Press. Menlo Park, CA. USA., England, pp. 314-321.

Algoritmos de *threading*
Perfiles de secuencia + estructura secundaria



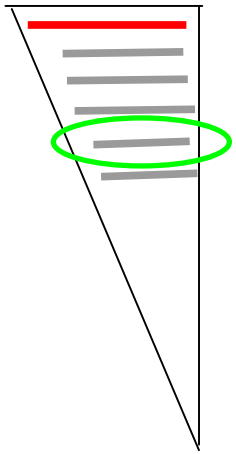


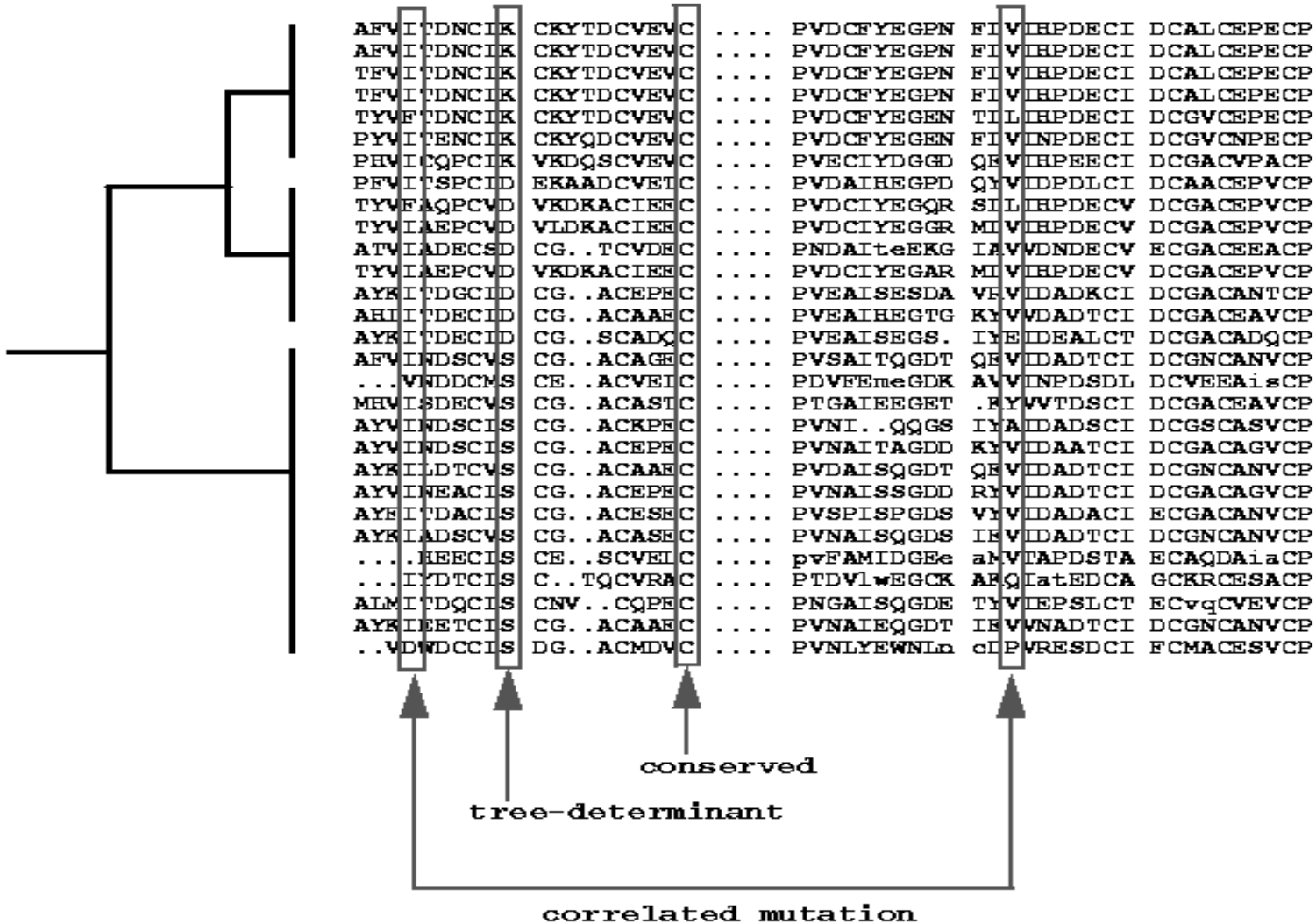
Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol*, **299**, 499-520.

Threading

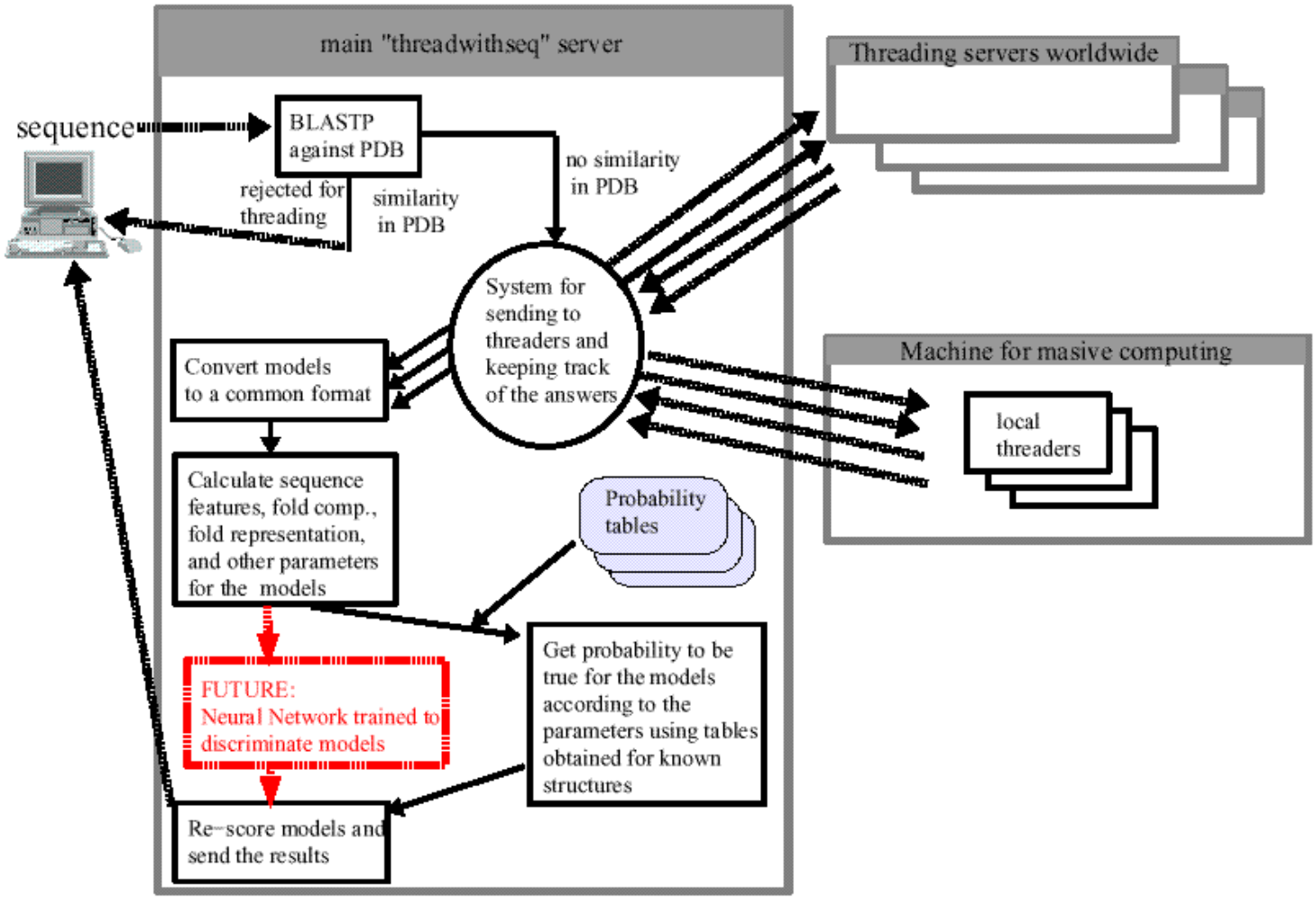
Post-procesamiento de resultados

Combinación con información adicional





Threading
Post-procesamiento de resultados
Meta-servidores



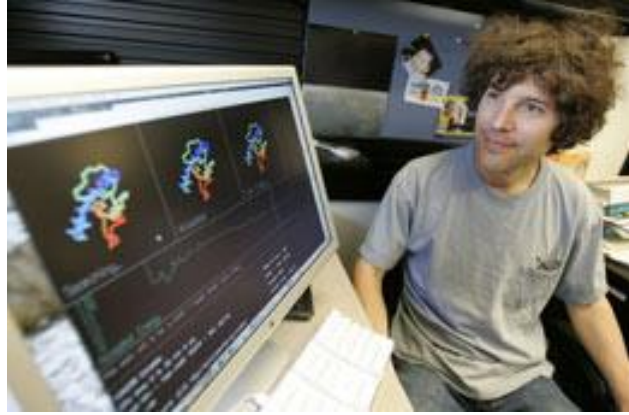
http://bioinfo.pl/meta (Metaserver + “3D-jury”)

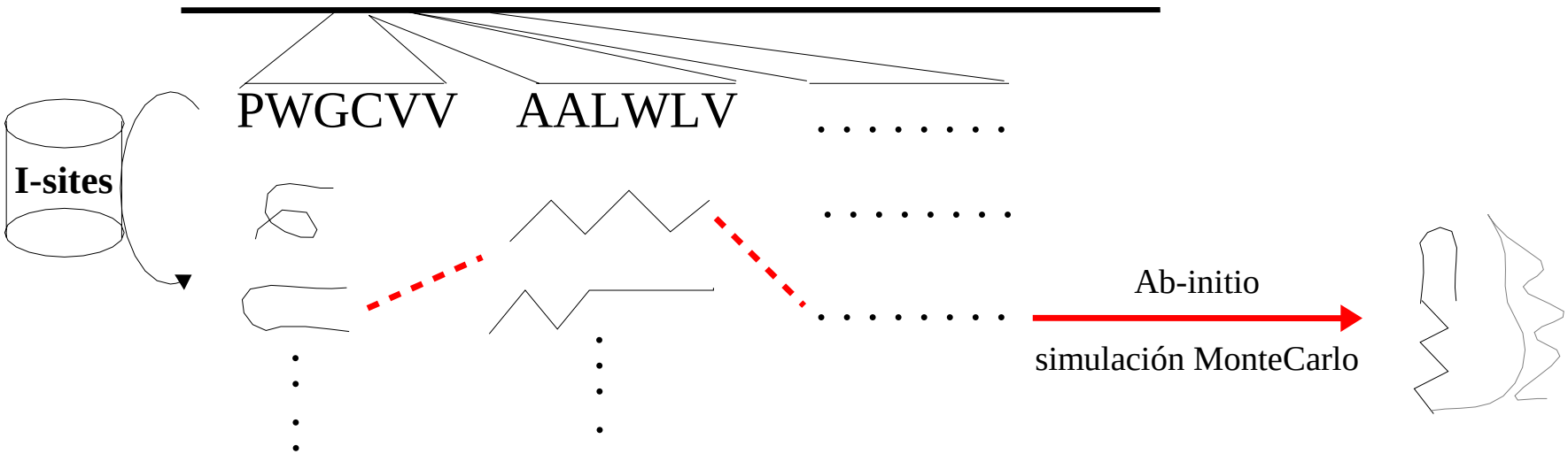
Ginalski K, Rychlewski L. (2003). Detection of reliable and unexpected protein fold predictions using 3D-Jury.*Nucleic Acids Res*. **31(13):**3291-2.

Predicción de Estructura

Mini-threading + ab-initio

Rosetta

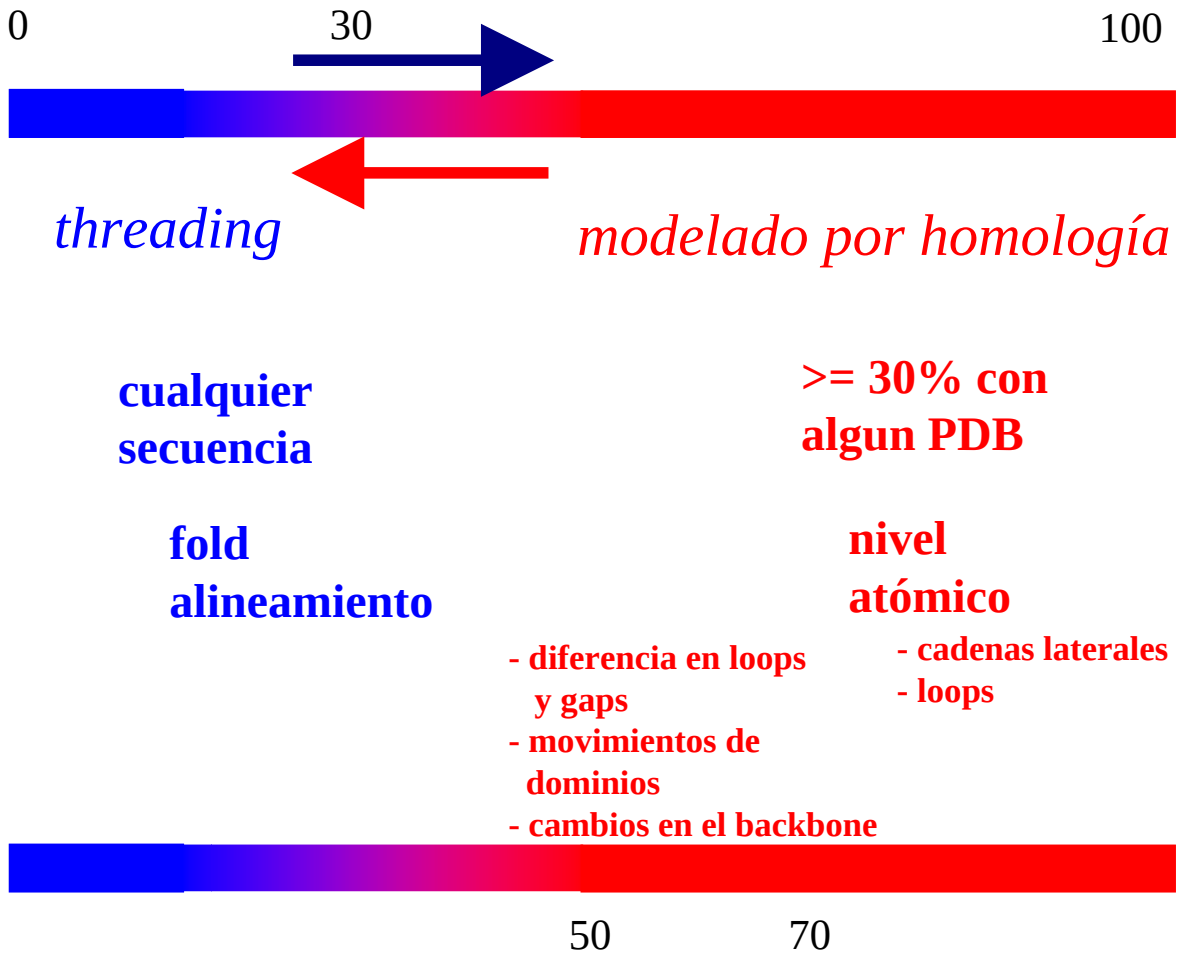


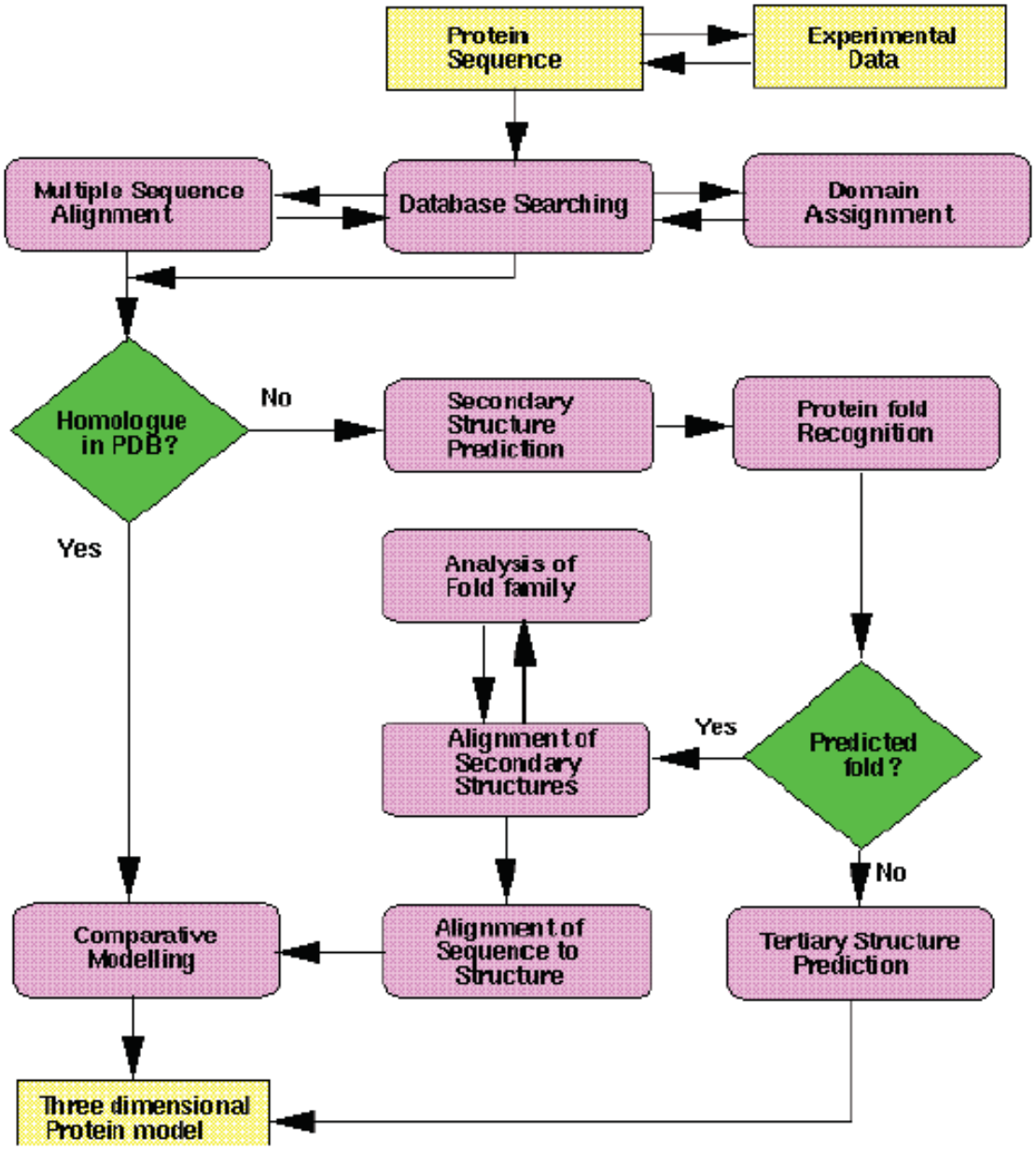


http://robetta.bakerlab.org/

Simons, K.T., Kooperberg, C., Huang, E. & Baker, D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J Mol Biol*, **268**, 209-225.

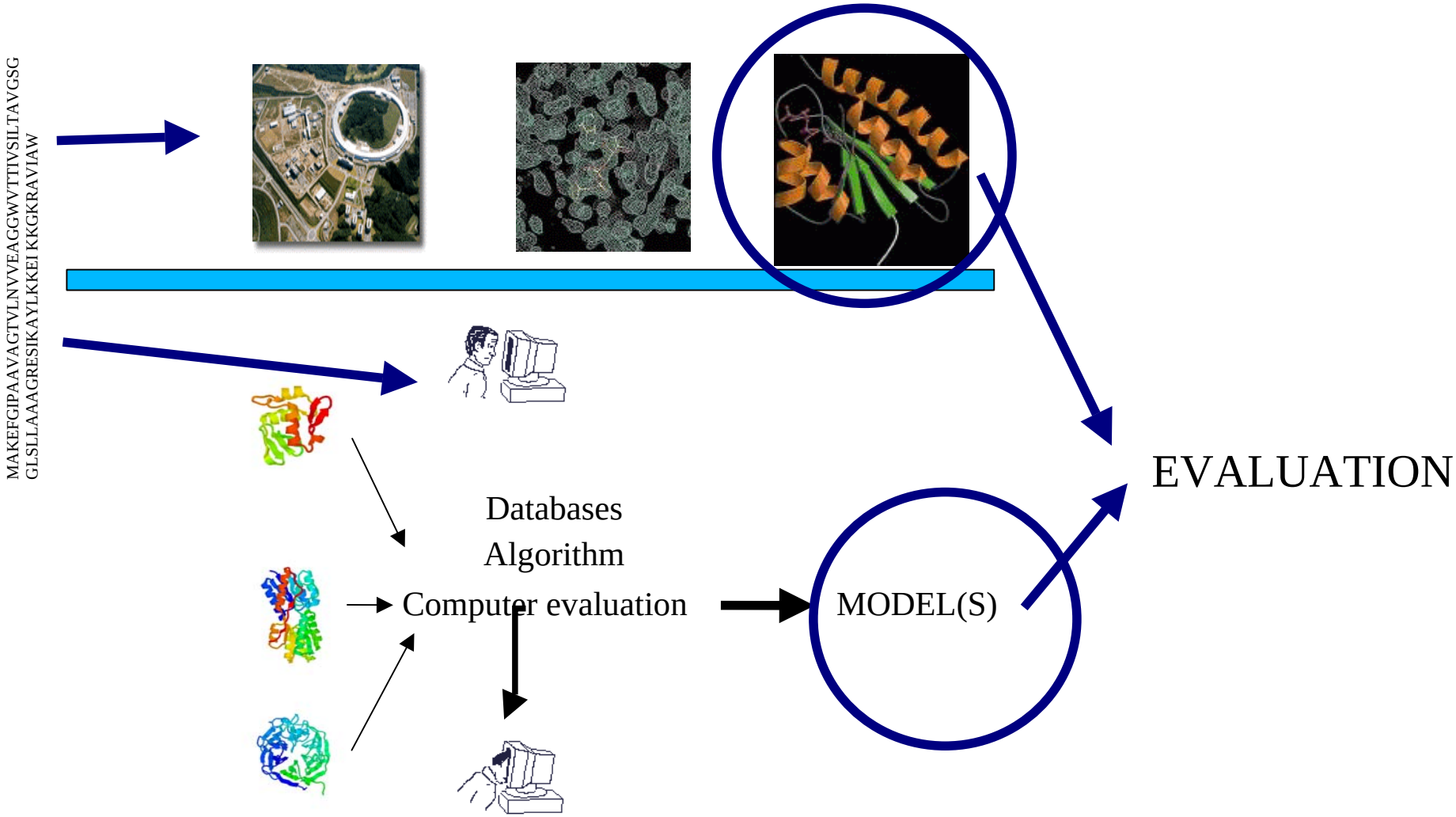
Modelado por Homología *vs. Threading*





R. Russell
EMBL

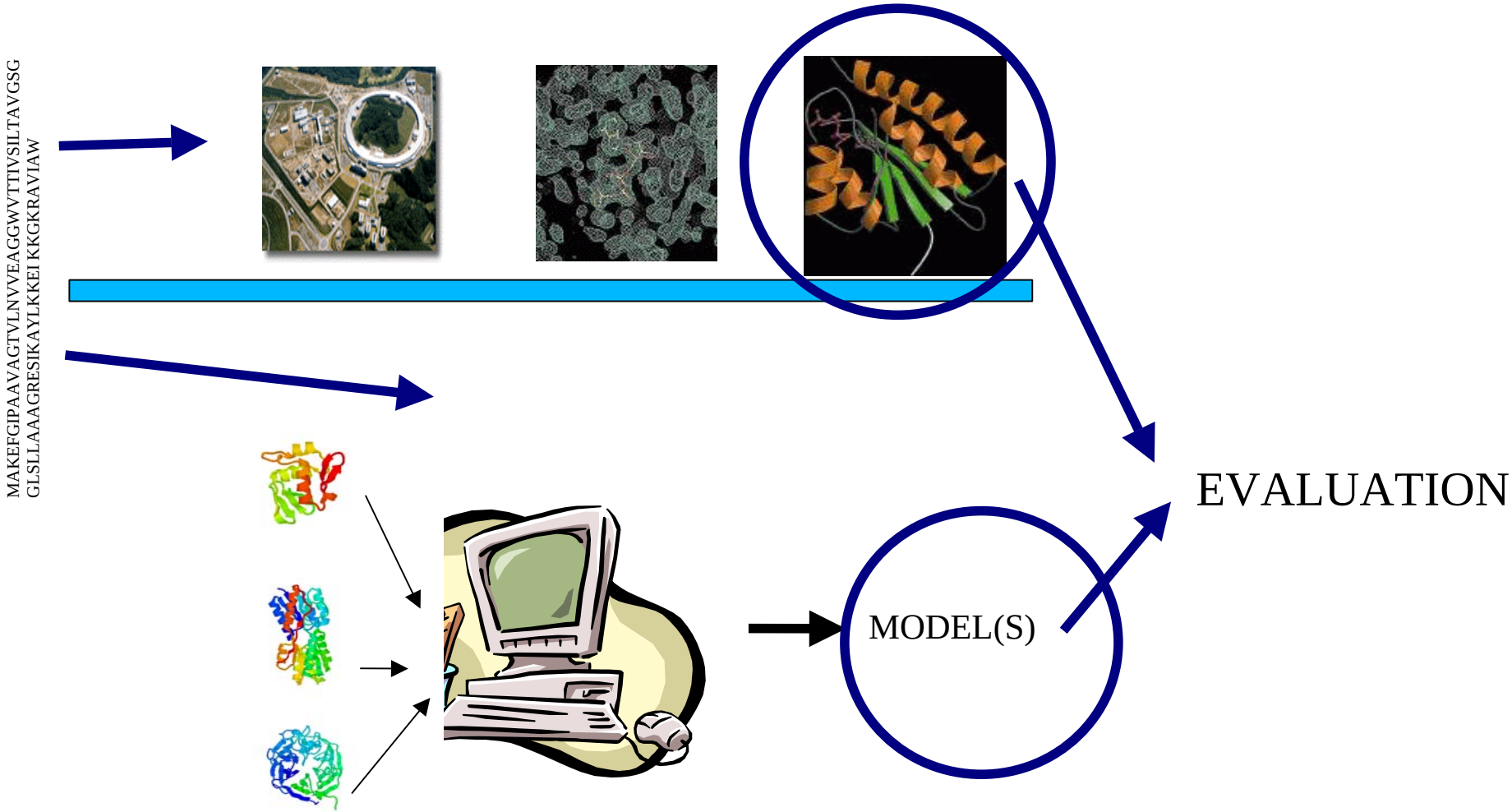
Evaluación de métodos de predicción
I) CASP (bianual 94-06)



http://predictioncenter.org/

Proteins biannual issue on CASP/CAFASP

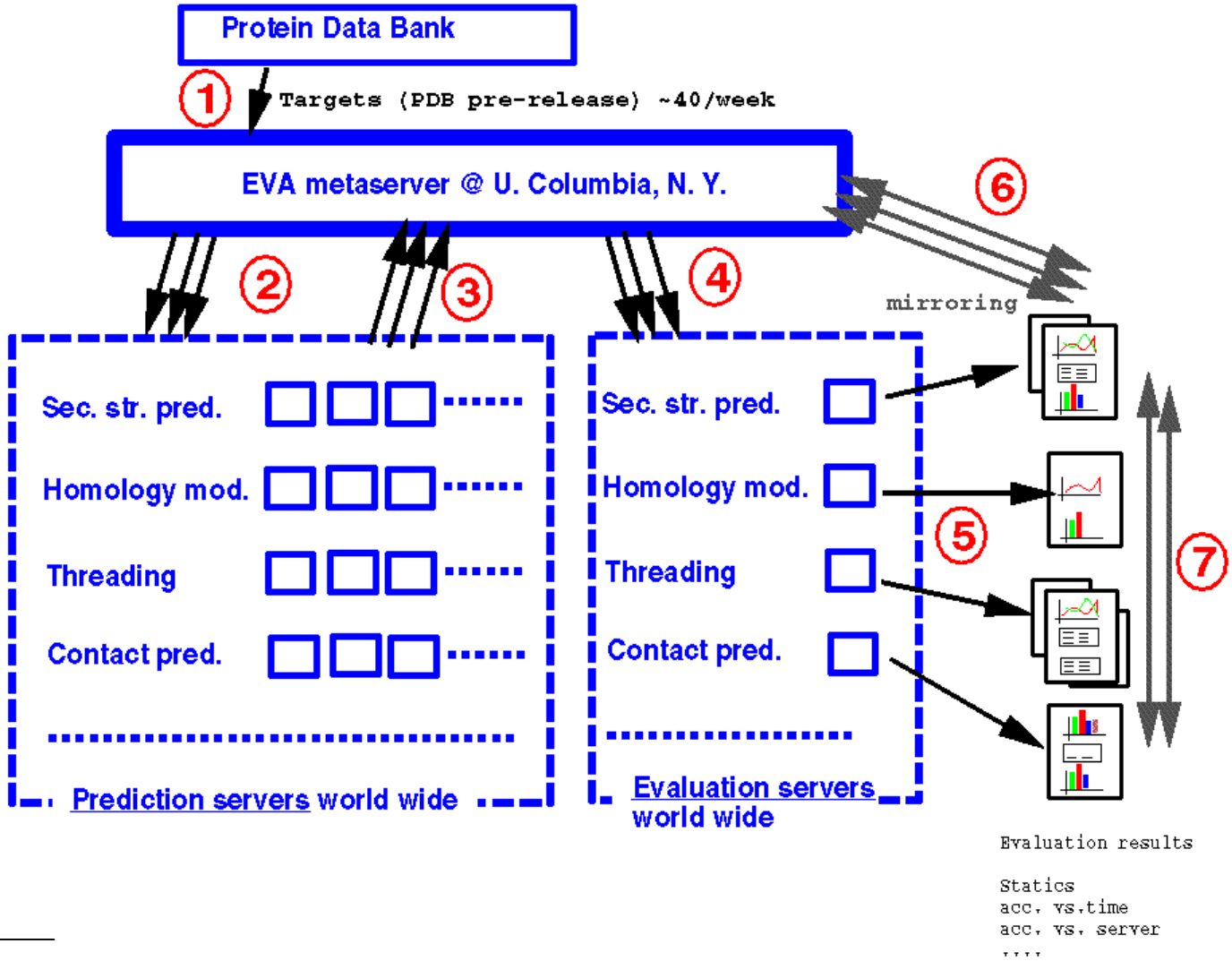
Evaluación de métodos de predicción
II) CA**FA**SP (bianual 98-06)



http://www.cs.bgu.ac.il/~dfischer/CAFASP4/

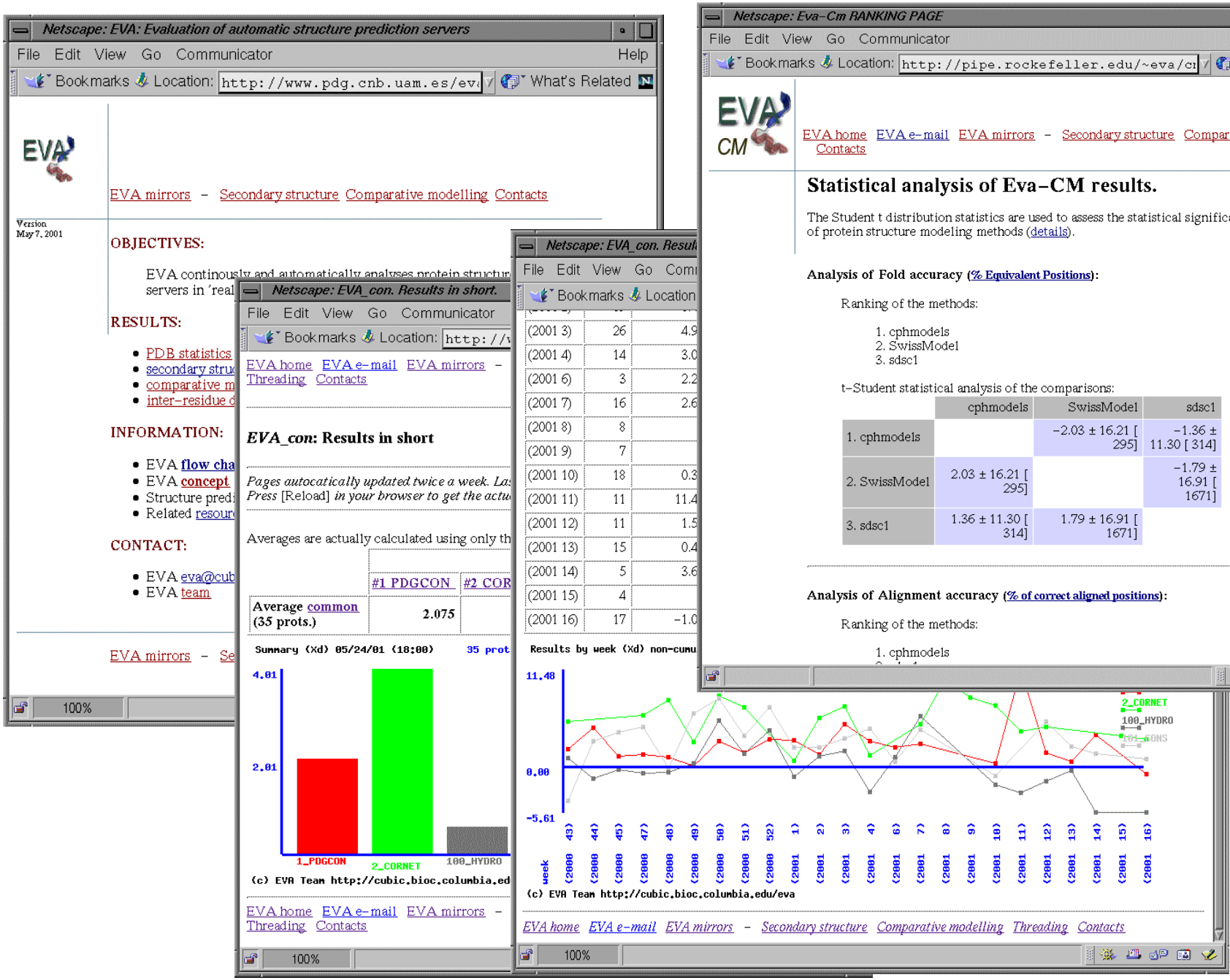
Proteins biannual issue on CASP/CAFASP

Evaluación de métodos
III) EVA/LiveBench



Eyrich, V.A., Marti-Renom, M.A., Przybylski, D., Madhusudhan, M.S., Fiser, A., Pazos, F., Valencia, A., Sali, A. and Rost, B. (2001) EVA: continuous automatic evaluation of protein structure prediction servers. *Bioinformatics*, **17**, 1242-1243.

Koh, I.Y.Y., Eyrich, V.A., Marti-Renom, M.A., Przybylski, D., Madhusudhan, M.S., Eswar, N., Grana, O., Pazos, F., Valencia, A., Sali, A., *et al.* (2003) EVA: evaluation of protein structure prediction servers. *Nucl. Acids. Res.*, **31**, 3311-3315.



4 weeks running EVA => mora targets than CASP1+2+3+4

Predicción de Estructura de Proteínas
Futuro…?

- Combinación de métodos de predicción con datos estructurales a baja resolución

- Integración en los proyectos de *structural genomics*.

- Predicción de función

• Qu, Y., Guo, J.T., Olman, V. and Xu, Y. (2004) Protein structure prediction using sparse dipolar coupling data. *Nucleic Acids Res*, **32**, 551-561. Print 2004.
• Arakaki, A.K., Zhang, Y. and Skolnick, J. (2004) Large-scale assessment of the utility of low-resolution protein structures for biochemical function assignment. *Bioinformatics*, **20**, 1087-1096.
• Meiler, J. and Baker, D. (2003) Rapid protein fold determination using unassigned NMR data. *Proc Natl Acad Sci U S A*, **100**, 15404-15409.
• Li, W., Zhang, Y., Kihara, D., Huang, Y.J., Zheng, D., Montelione, G.T., Kolinski, A. and Skolnick, J. (2003) TOUCHSTONEX: Protein structure prediction with sparse NMR data. *Proteins*, **53**, 290-306.
• Zheng, W. and Doniach, S. (2002) Protein structure prediction constrainded by solution x-ray scattering data and structural homology identification. *J Mol Biol*, **316**, 173-187.

http://pdg.cnb.uam.es/pazos/cursos/protstr_cnb