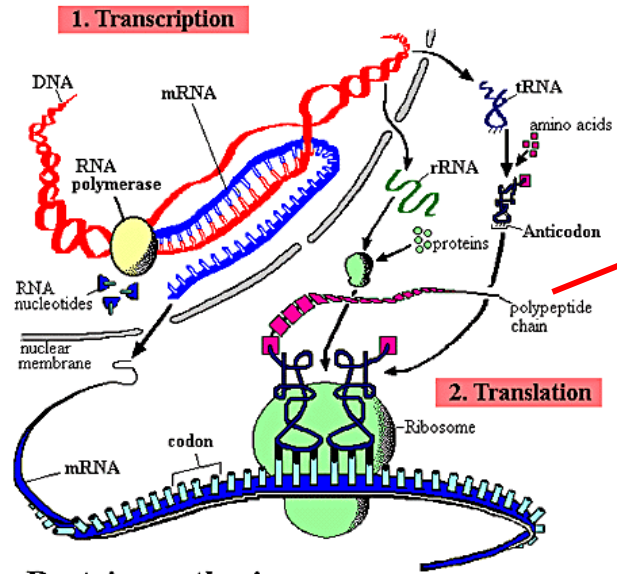

Estructura de Proteínas

Introducción

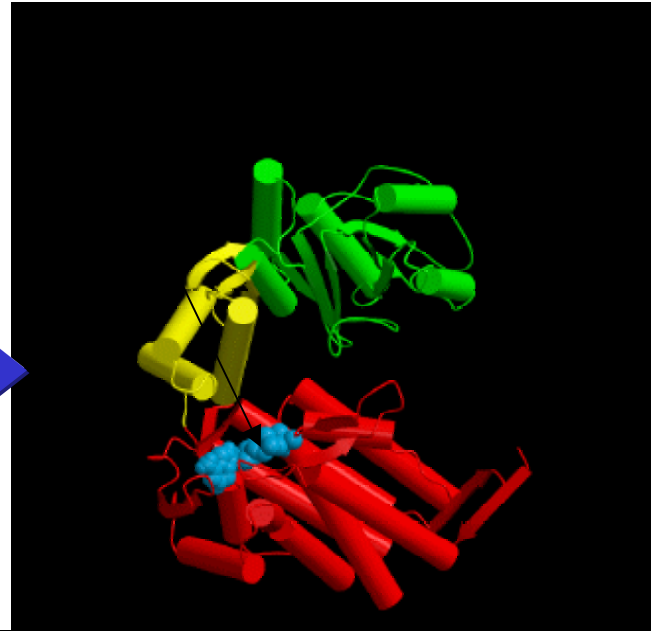
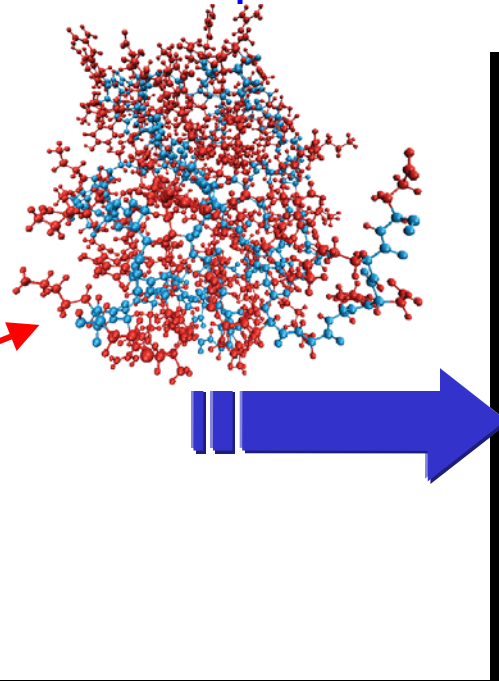
Características del Espacio de Estructuras

Florencio Pazos (CNB-CSIC)

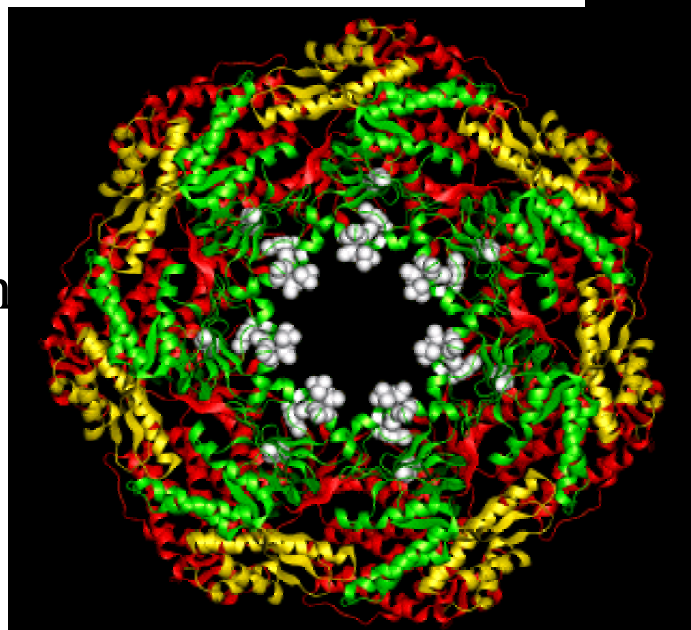
Estructura de proteínas



Protein synthesis



subunit



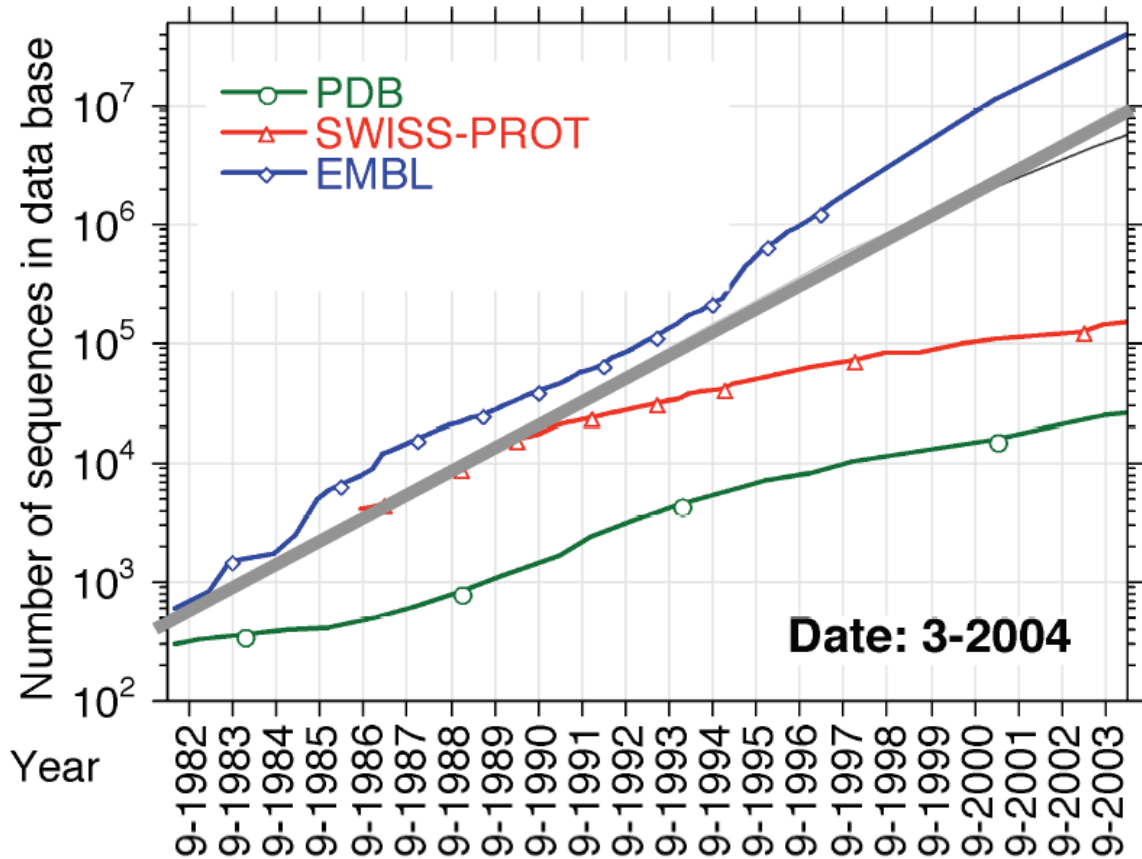
Molecular chaperonin
GroEL

heptamer

(Dr Jianpeng Ma, Harvard Univ.)

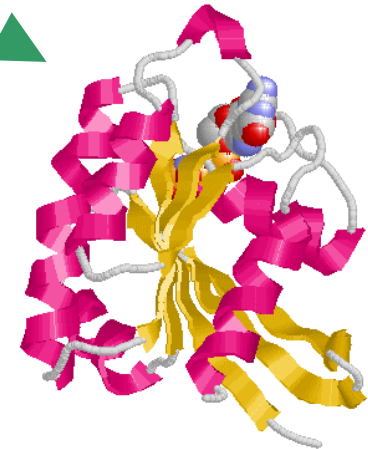
Figure: P. Gomez-Puertas

Conocimiento experimental de secuencias, funciones y estructuras de proteínas

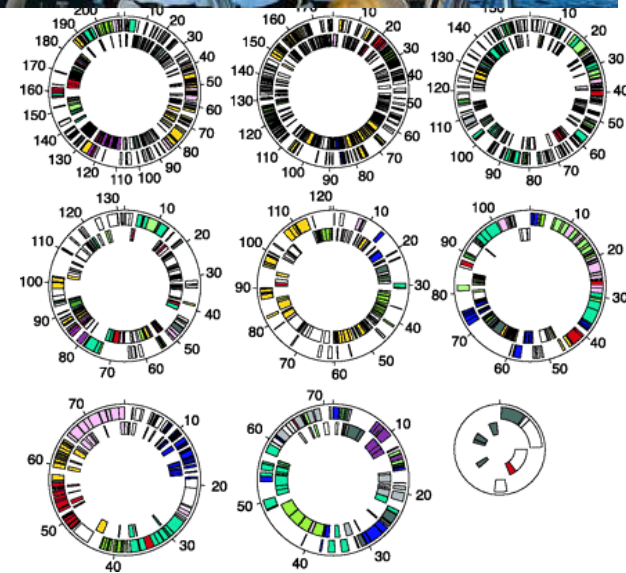
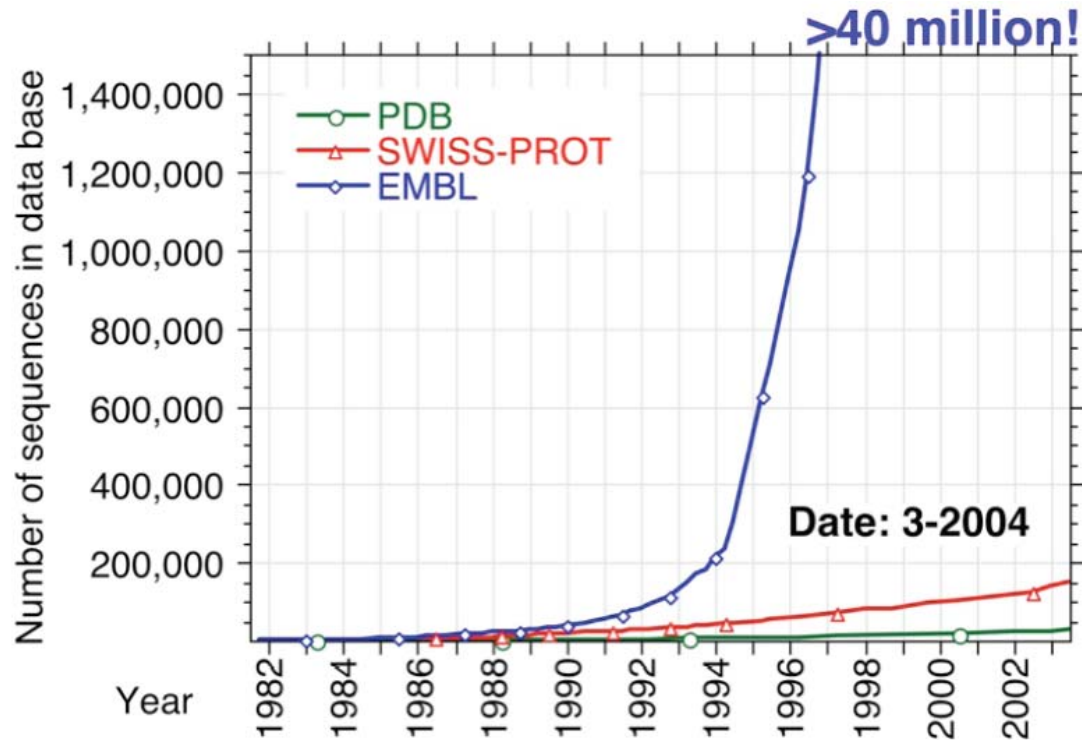
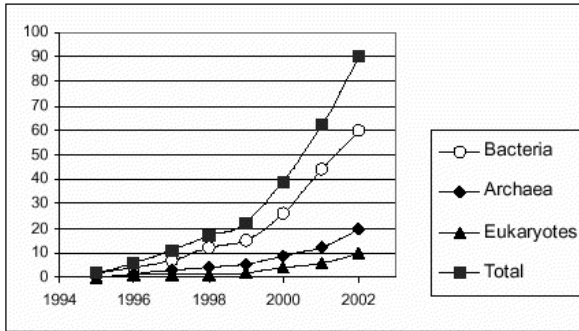


MREYKLVVLGSGGVGKSALT VQFVQGIFVDEYDPTIEDSY
 RKQVEVDCQCMLLEILD TAGTEQFTAMRDLYMKNQGQFA
 VYSITAQSTFNDLQDLREQILRVKDTEDVPMILVGNKCDL
 EDERVVGKEQQNLARQWCNCAFLESSAKSKINVNEIFYD
 LVRQINR

MLEILD TAGTEQFTAMRDLYMKNQGQFAL
 VYSITAQSTFNDLQDLREQILRVKDTEDVPMIL
 VGNKCDLEDERV

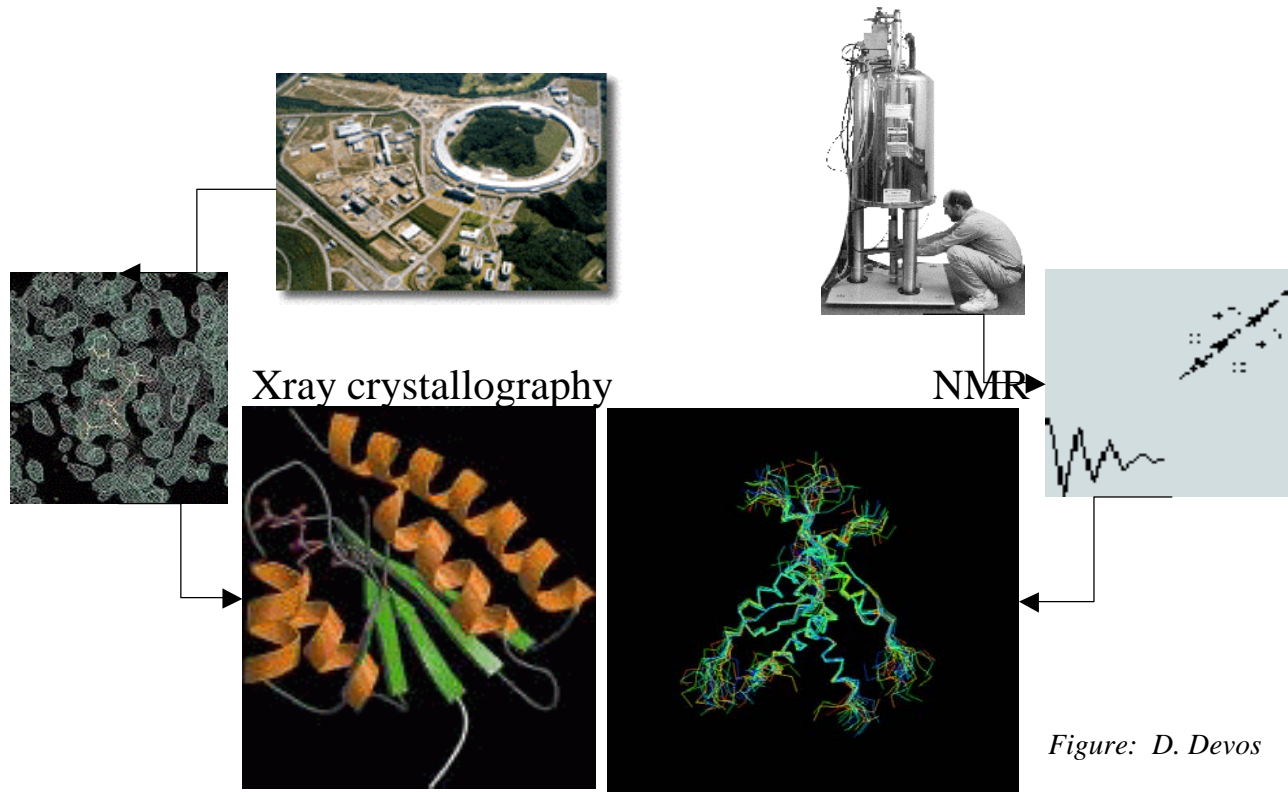


Secuenciación de Genomas Completos y “Metagenomas”

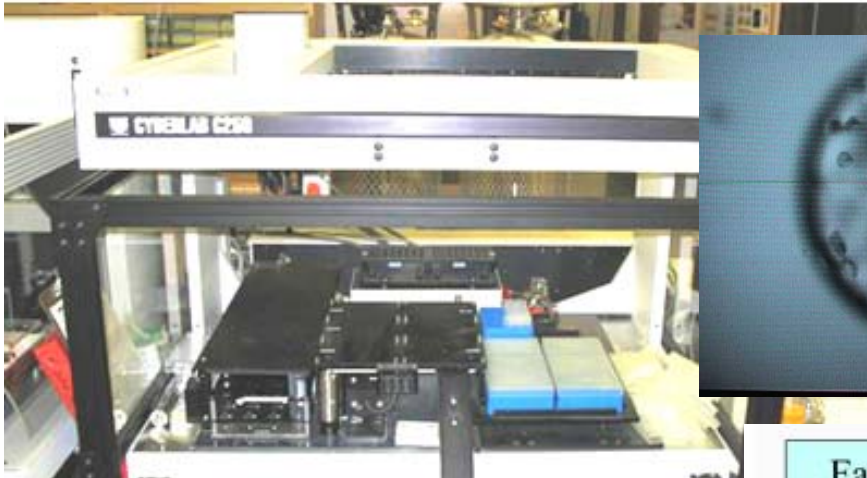


- Collins, F.S., Green, E.D., Guttmacher, A.E. & Guyer, M.S. (2003) A vision for the future of genomic research. *Nature*, **422**, 835-847.
- Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. (2004). *Science* **304**, 66-74.

Determinación experimental de estructura de proteínas



Genómica estructural

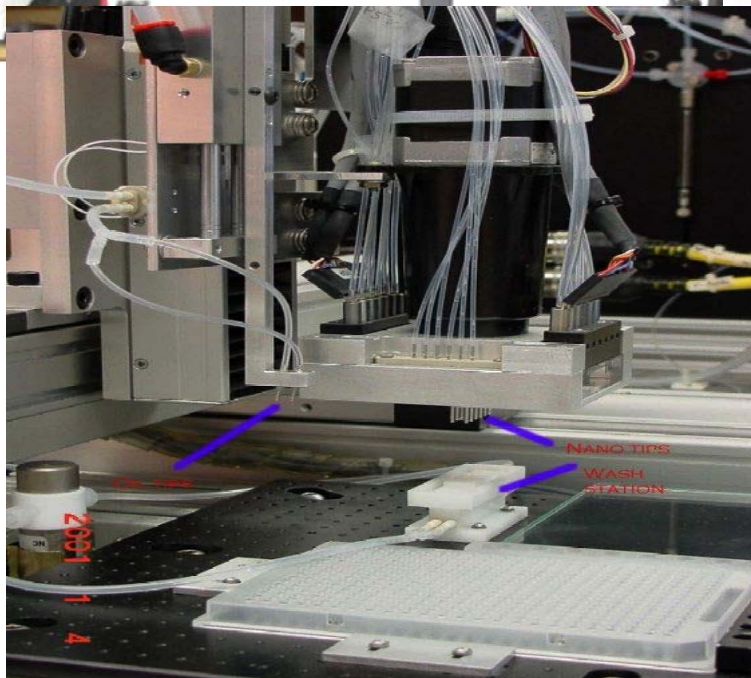
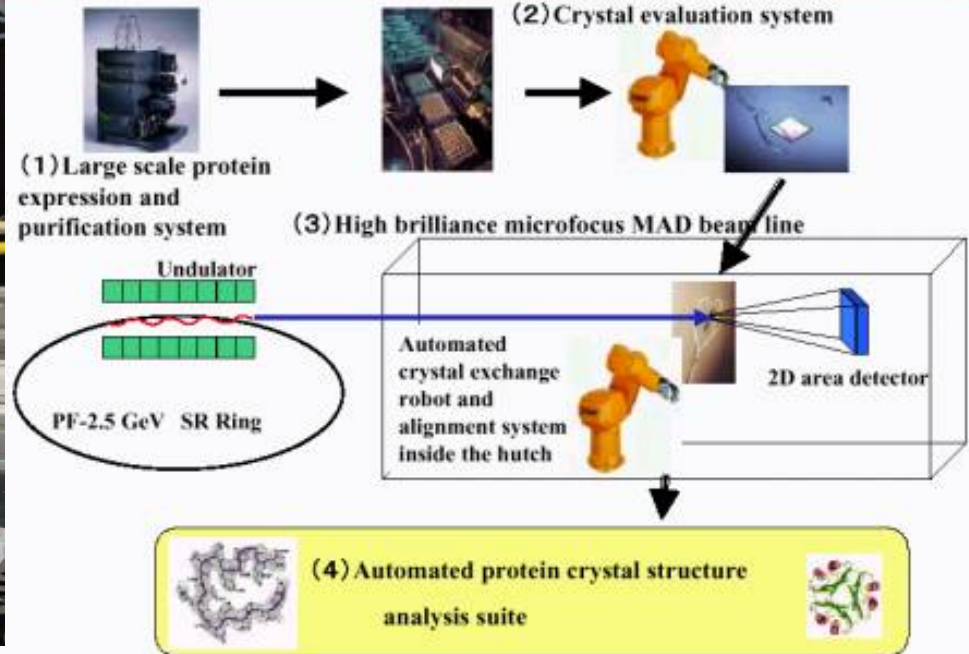


ment of two minimal genomes, *Mycoplasma* crystallography are being used for structural

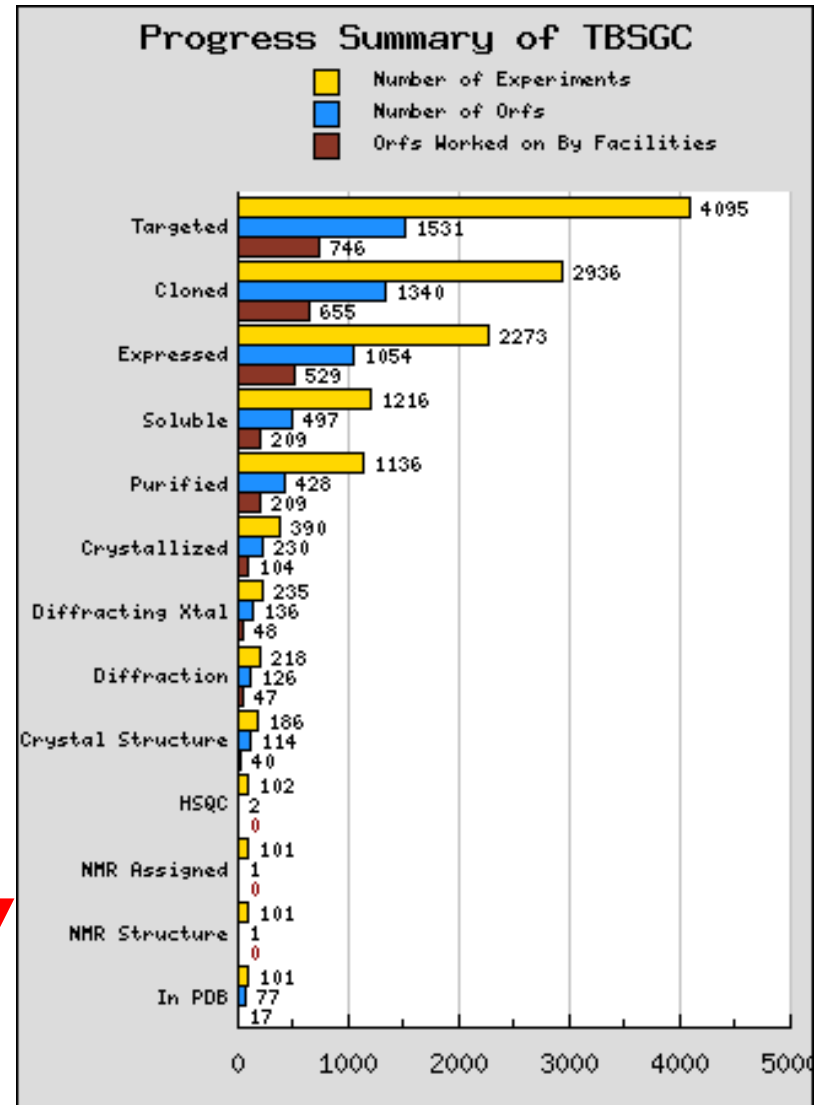
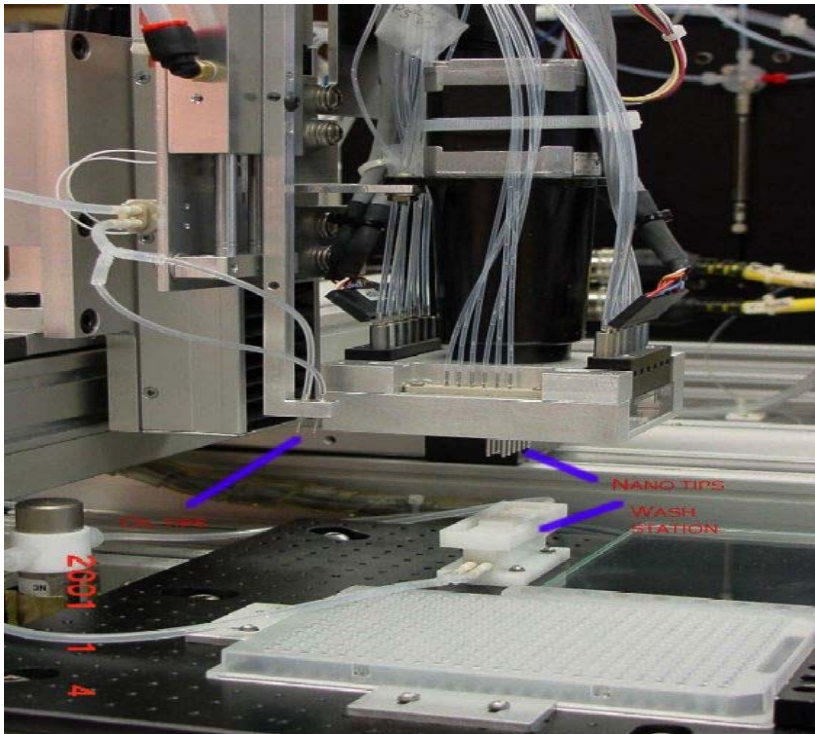
ture determination of biologically important biologically important proteins in *Arabidopsis*. The

l the main proteins of interest are signaling *Thermotoga maritima*, and creating a high-resolution structural determination.

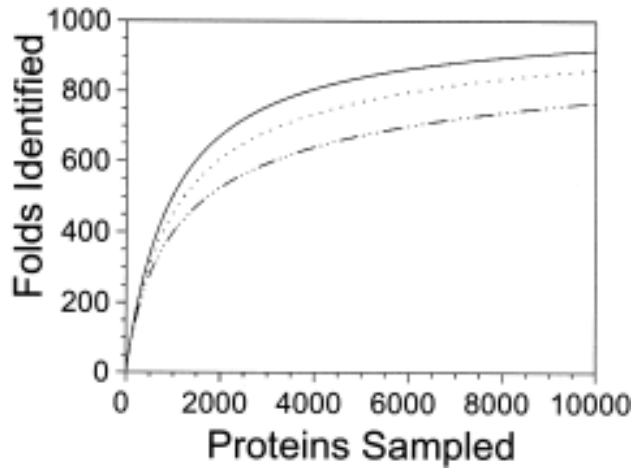
Facilities for enhancement of protein structure analysis



Genómica Estructural



Características del Espacio de Estructuras



Leonov, H., Mitchell, J.S. & Arkin, I.T. (2003) Monte Carlo estimation of the number of possible protein folds: effects of sampling bias and folds distributions. *Proteins*, **51**, 352-359.

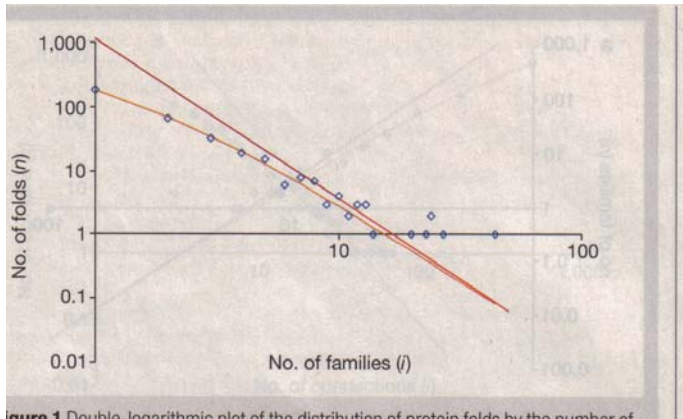
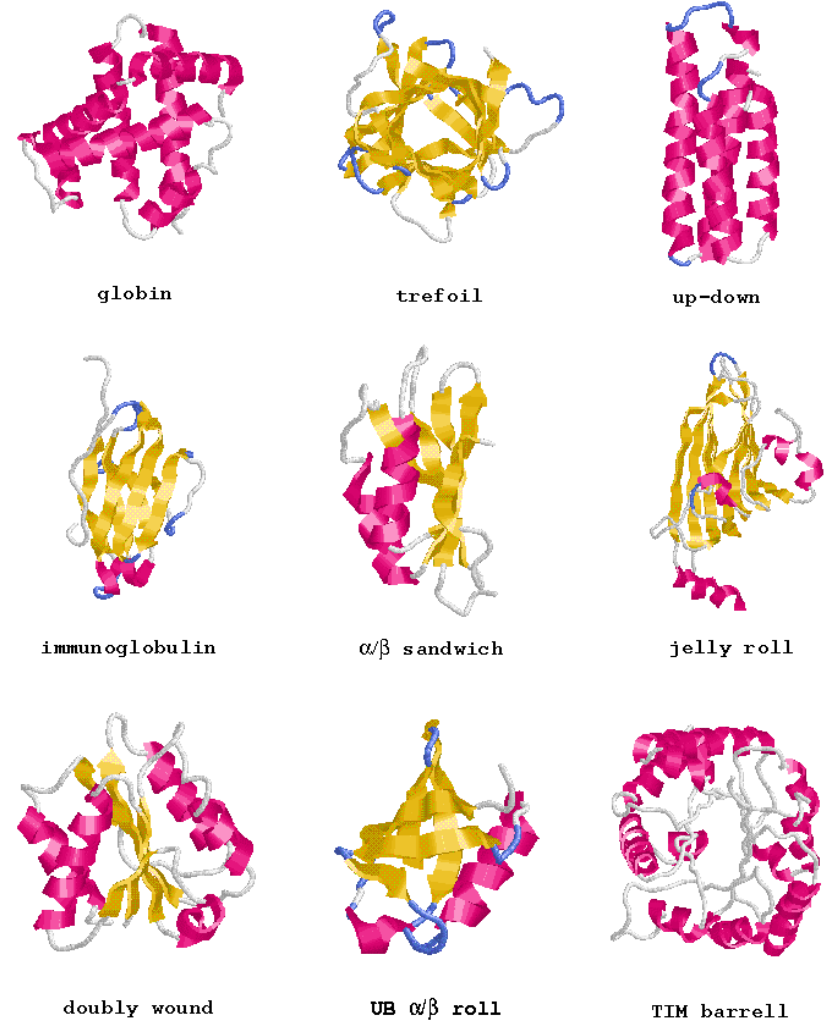


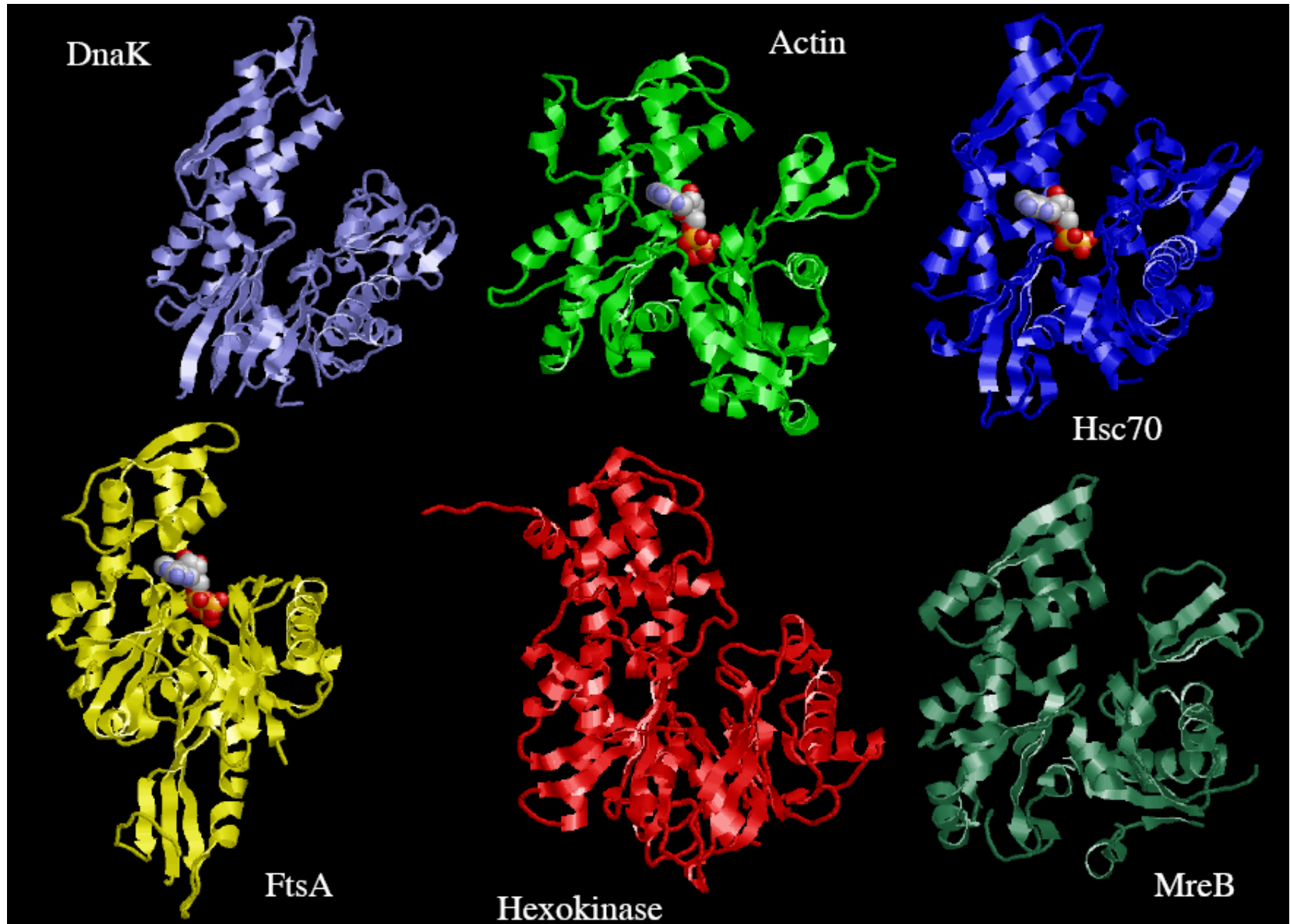
Figure 1 Double-logarithmic plot of the distribution of protein folds by the number of

Koonin, E.V., Wolf, Y.I. & Karev, G.P. (2002) The structure of the protein universe and genome evolution. *Nature*, **420**, 218-223.



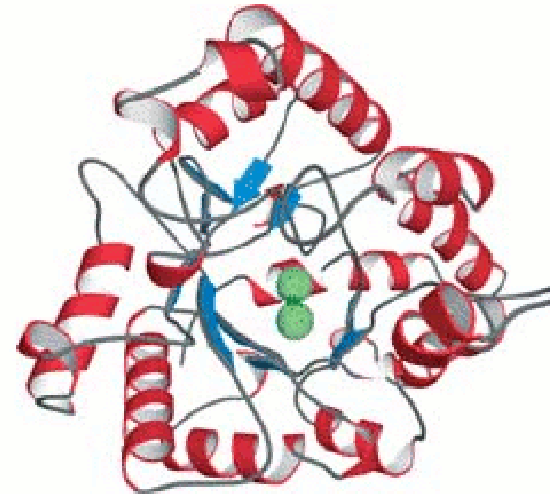
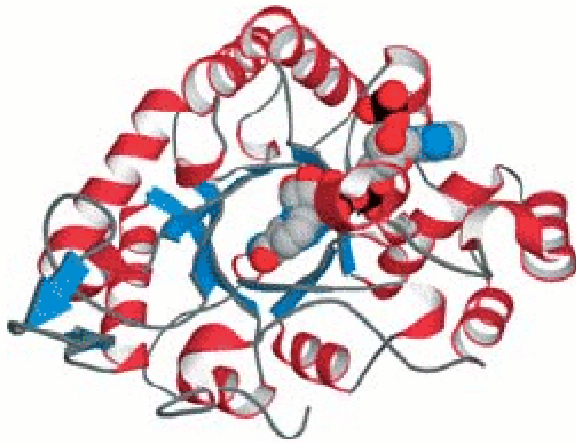
Orengo, C.A., Jones, D.T. & Thornton, J.M. (1994) Protein superfamilies and domain superfolds. *Nature*, **372**, 631-634.

Características del Espacio de Estructuras Relación con el espacio de secuencias



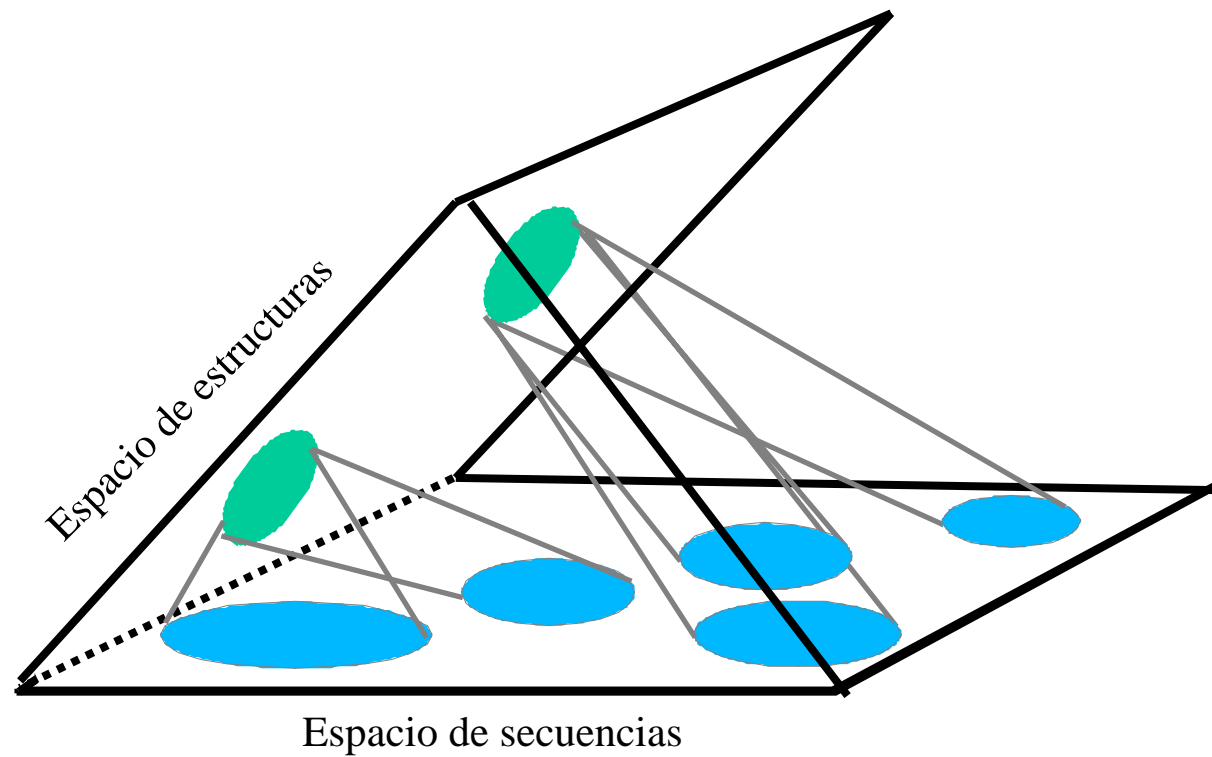
Características del Espacio de Estructuras

Relación con el espacio de secuencias

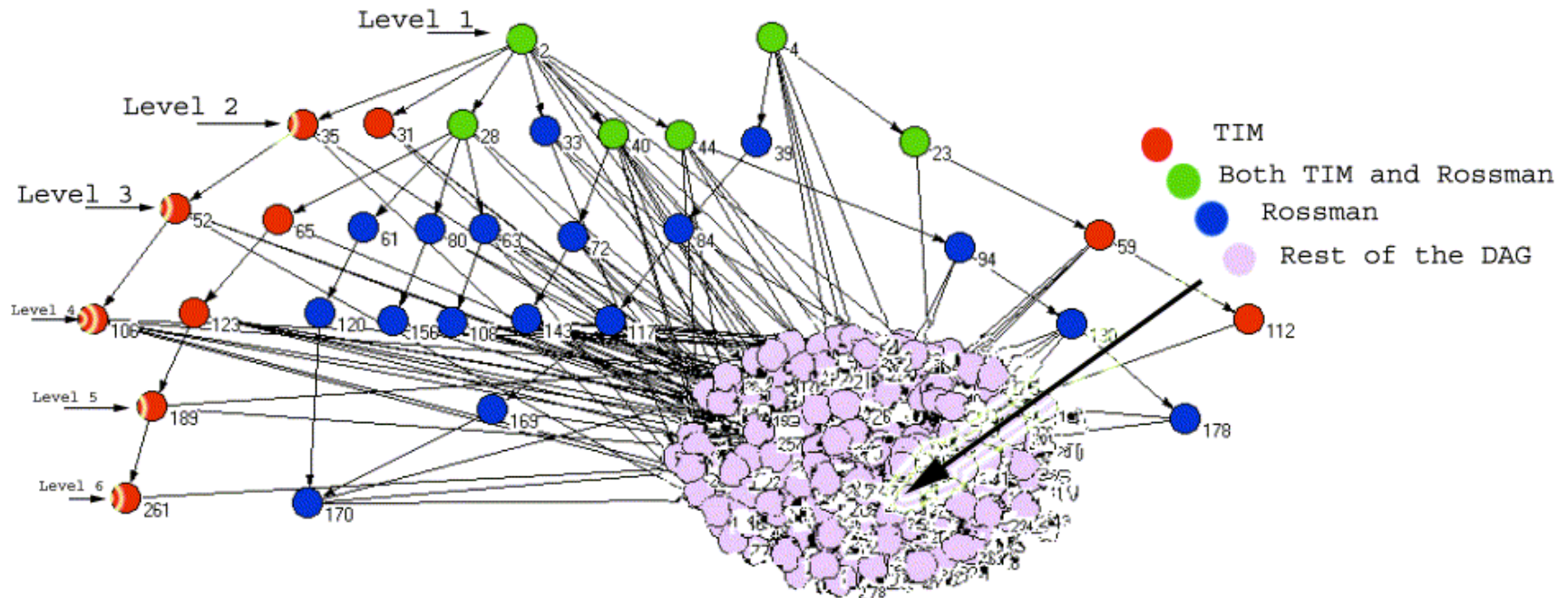


Características del Espacio de Estructuras

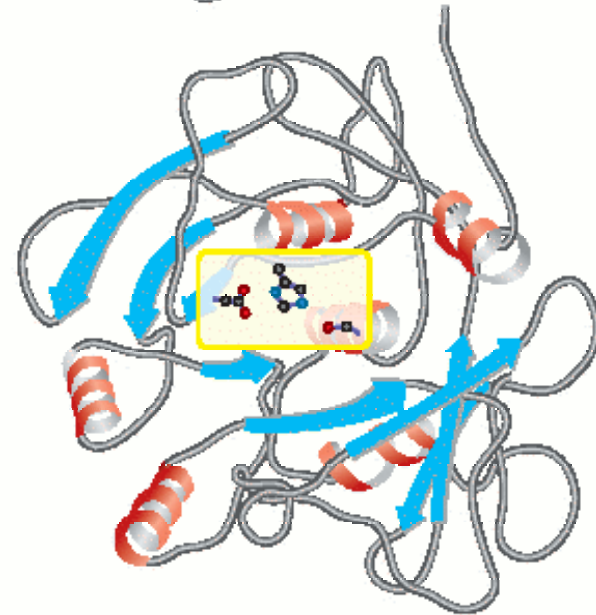
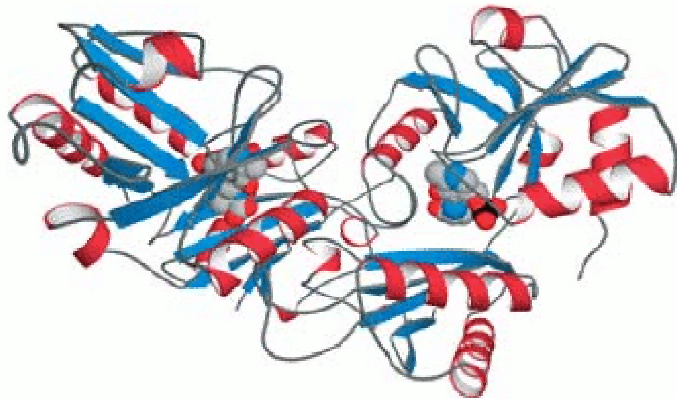
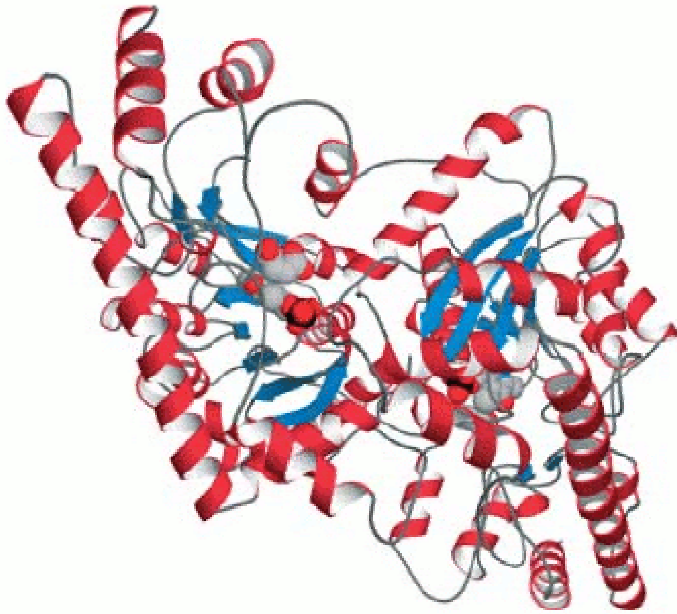
Relación con el espacio de secuencias



Características del Espacio de Estructuras Relación con el “espacio de funciones”

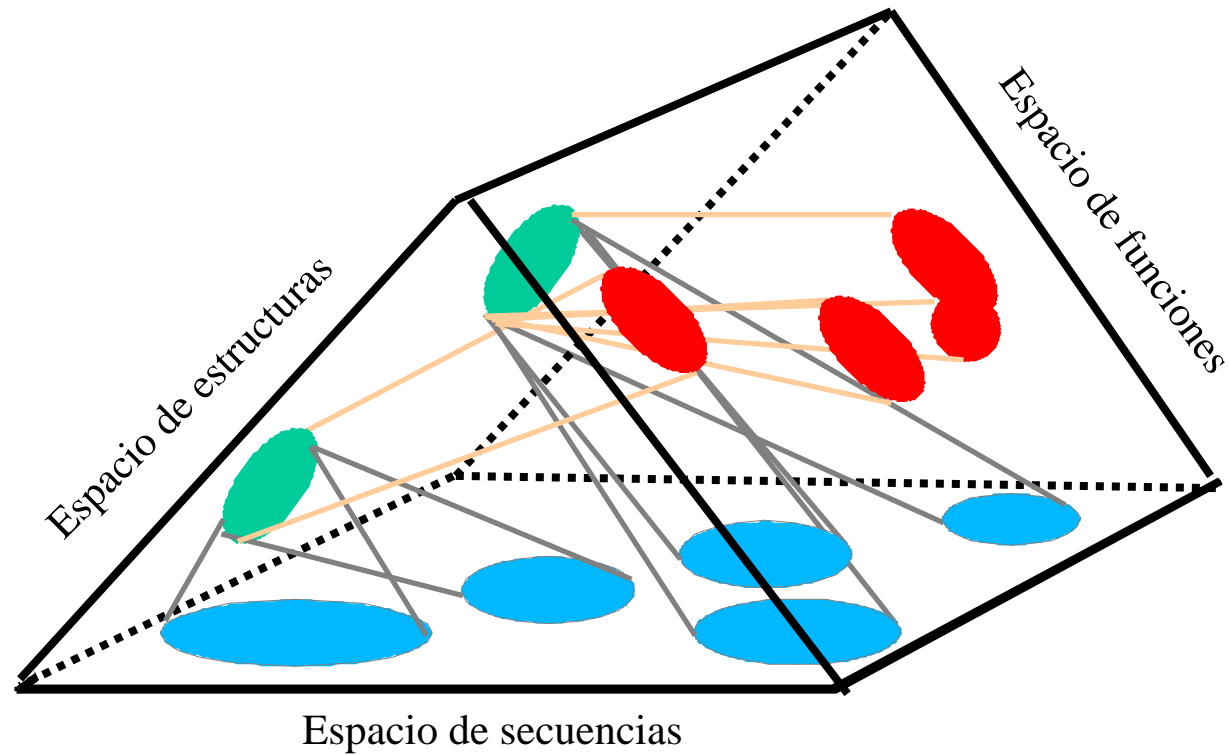


Características del Espacio de Estructuras Relación con el “espacio de funciones”

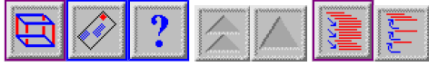


Características del Espacio de Estructuras

Relación con el espacio de secuencias y funciones



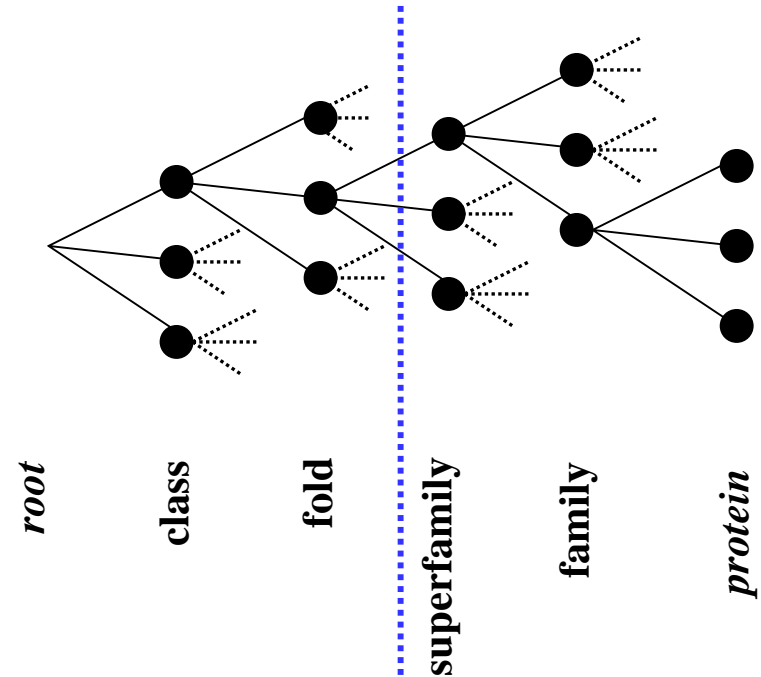
Clasificación jerárquica de estructuras de proteínas (SCOP)



Root: scop

Classes:

1. [All alpha proteins](#) [46456] (218)
2. [All beta proteins](#) [48724] (144)
3. [Alpha and beta proteins \(a/b\)](#) [51349] (136)
Mainly parallel beta sheets (beta-alpha-beta units)
4. [Alpha and beta proteins \(a+b\)](#) [53931] (279)
Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. [Multi-domain proteins \(alpha and beta\)](#) [56572] (46)
Folds consisting of two or more domains belonging to different classes
6. [Membrane and cell surface proteins and peptides](#) [56835] (47)
Does not include proteins in the immune system
7. [Small proteins](#) [56992] (75)
8. [Coiled coil proteins](#) [57942] (6)
Not a true class
9. [Low resolution protein structures](#) [58117] (24)
Not a true class
10. [Peptides](#) [58231] (116)
Peptides and fragments. Not a true class
11. [Designed proteins](#) [58788] (42)
Experimental structures of proteins with essentially non-natural sequences. Not a true class

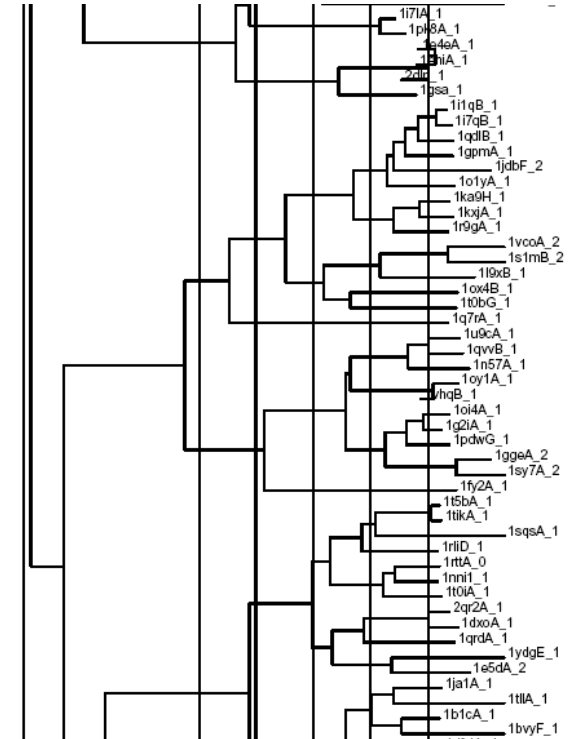
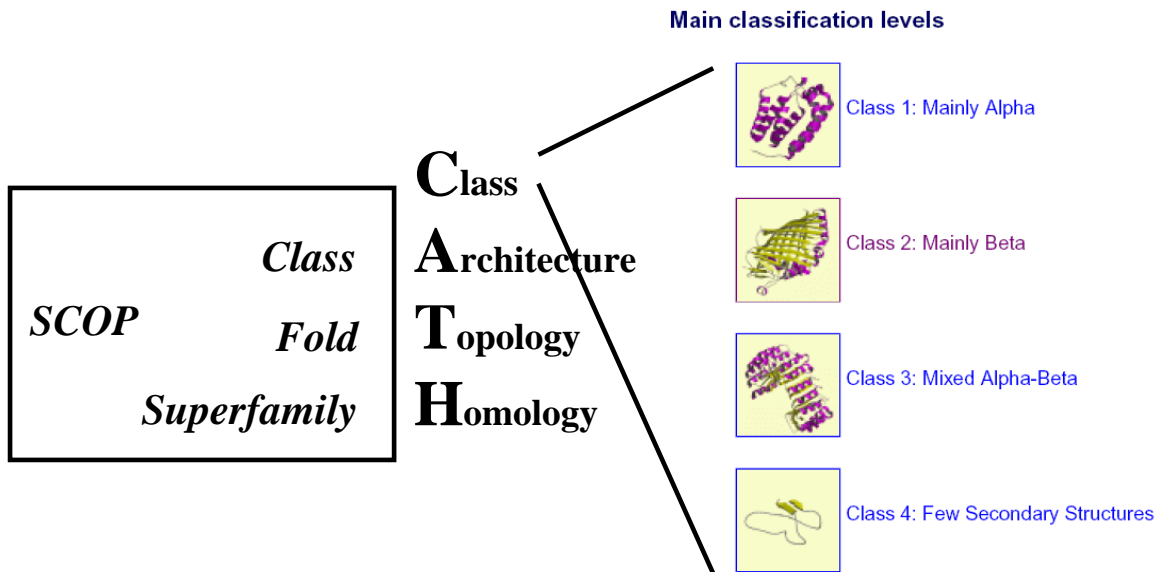


Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226-229.

Structural Classifications

CATH

Dali/FSSP



Pearl, F., Todd, A., Sillitoe, I., Dibley, M., Redfern, O., Lewis, T., Bennett, C., Marsden, R., Grant, A., Lee, D., *et al.* (2005) The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. *Nucleic Acids Res.*, **33**, D247-251.

Dietmann, S., Park, J., Notredame, C., Heger, A., Lappe, M. and Holm, L. (2001) A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3. *Nucleic Acids Res.*, **29**, 55-57.

The Protein “Universe”

10⁴⁰⁰ Possible sequences

10¹⁰ Sequences in the biosphere

10⁵ Families

10³-10⁴ Folds

10⁴ Functions (GO)

Estructura de Proteínas

Programa

Día 1

Introducción

Bases de datos con información estructural

Visualización y manipulación de estructuras

Día 2

Comparación y alineamiento de estructuras

Introducción a la predicción de estructura

Predicción de estructura secundaria y otras características 1D

Día 3

Predicción de estructura tridimensional

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

Florencio Pazos Cabaleiro

Computational Systems Biology Group (CNB-CSIC)

pazos@cnb.csic.es