

Development of novel knowledge-based methodologies for the refinement of target protein models and its application to the directed design of molecules

June 25 2007

Doctoral Consortium Barcelona

Xavi Jalencas.

Chemogenomics Laboratory
Research Unit on Biomedical Informatics (GRIB)
Municipal Institute of Medical Research (IMIM)
& University Pompeu Fabra
Barcelona

cgl.imim.es

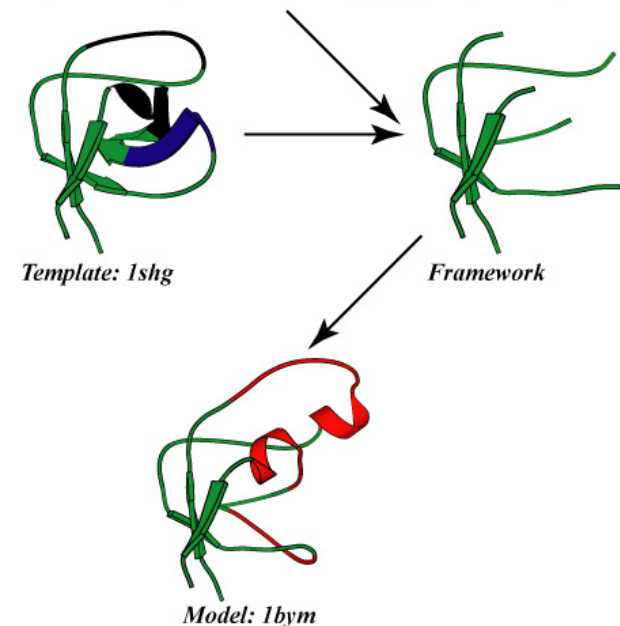
Objectives

- Incorporate systematically all available chemical information to homology modelling protocols to produce better quality protein models consistent with structural features of known ligands
- Use these ligand-adapted models in directed design of molecules to target proteins

Homology modelling:

Predict the three-dimensional structure of a protein from the sequence, based on known structures of homologue proteins, which are used as a template

```
Ishg  KELVLALYDYQE-----KSPREVTMKGKDILTLNSTNKDWWKVEVNDRCQFV---PAALVKKLD  
Ibym  RKVRIVQINIEIFQVETDQPTQLDADIRVGSEVEIVDRDGHII--TISHNGKIPELLDDLAEIRIEE
```



Justification

- Modelling protein structures is often the only source of structural information for directed design of molecules to protein targets.
- Quality of model is strongly dependent on the template
- The increasing number of structures in PDB does not imply an increasing coverage of the biological space
- A huge number of molecules have been synthesized and tested on biochemical assays for their binding affinity to target proteins.

Research questions

Lock and key principle is prevalent i protein-ligand binding.

- Can protein-ligand complementarity be reproduced at the surface level ?

- Are there privileged protein environments ?

- Can protein-ligand docking be avoided by pre-annotations of both entities ?

Methods

For the integration of information about binding site characteristics, a common set of descriptors for both ligands and proteins, is needed in order to predict if a protein model is likely to bind a certain ligand.

Proteins:

- Detection of ligand binding pockets
- Molecular surface of the cavity is patched in regions according to physiochemical features of their underlying atoms.
- Centroids from each patch compose a set of descriptors for the cavity

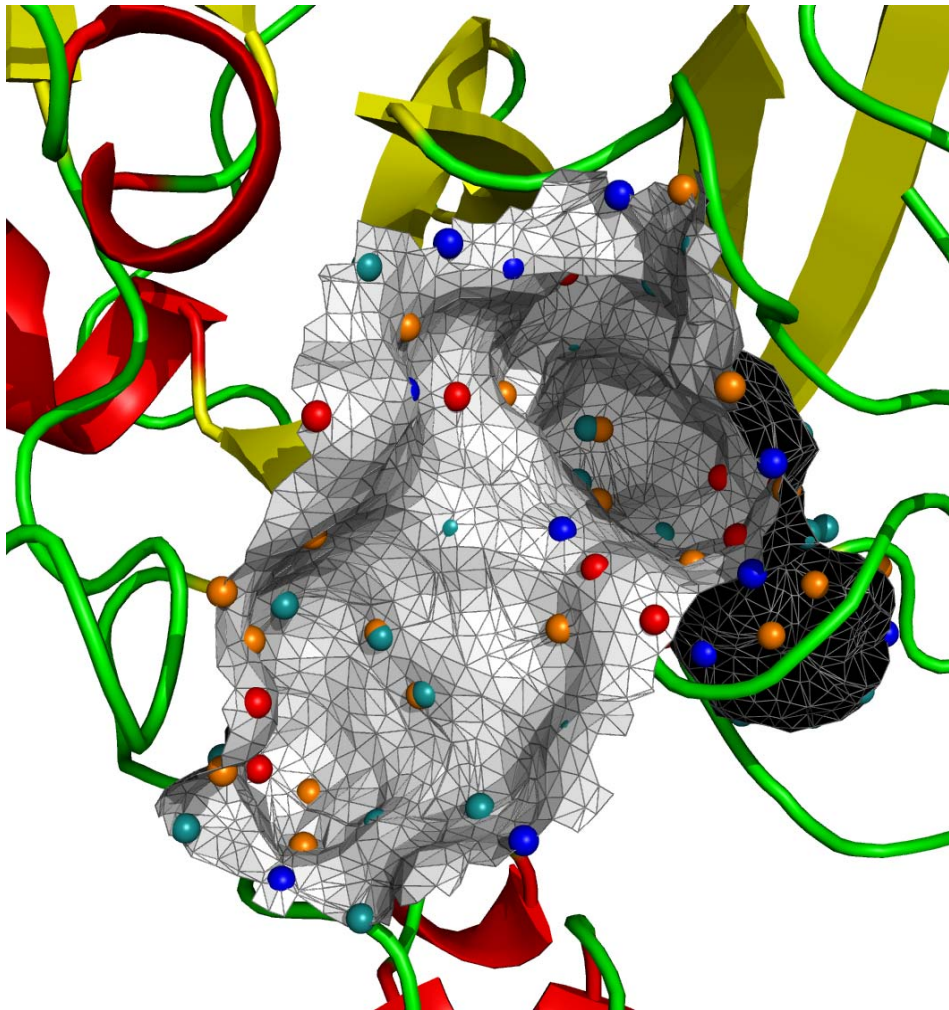
Validation:

Use of these descriptors to classify proteins according to the shape and properties of binding sites

Results to date

Protein cavity descriptors

1a5h: Plasminogen activator



Protein classification

