

# Máster Universitario en Bioinformática

## Proteómica y Bioinformática Estructural

Curso académico 2024-2025



**Universidad**  
Internacional  
de Valencia

**Dra. Magdalena Nikolaeva Koleva**

[magdalena.nikolaeva@professor.universidadviu.com](mailto:magdalena.nikolaeva@professor.universidadviu.com)

# Sesión 4

## Modelado estructural de proteínas (I): Modelado comparativo



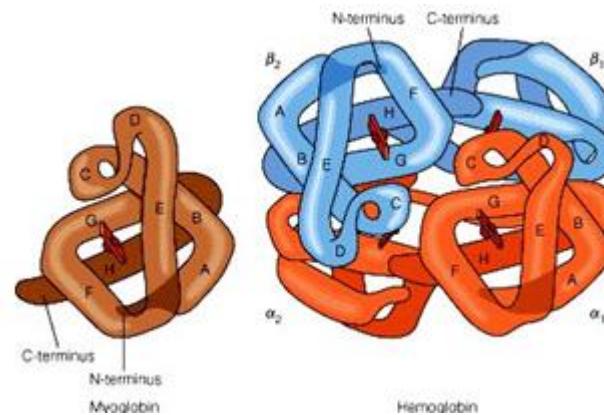
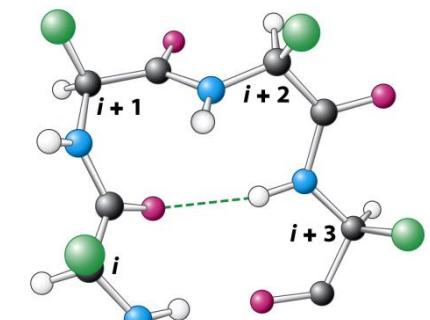
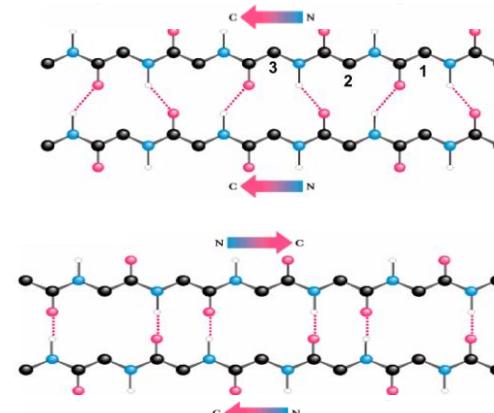
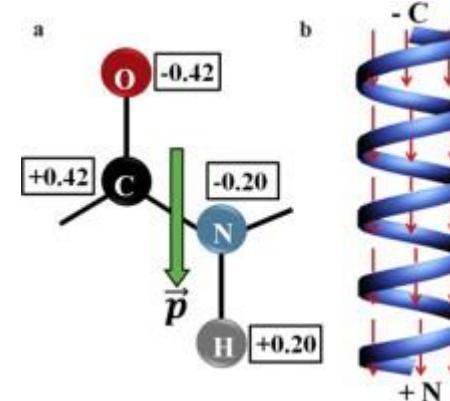
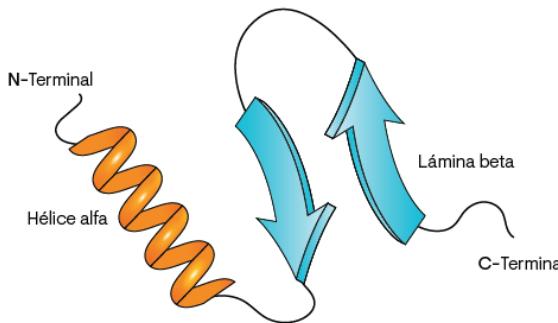
viu

**Universidad**  
Internacional  
de Valencia

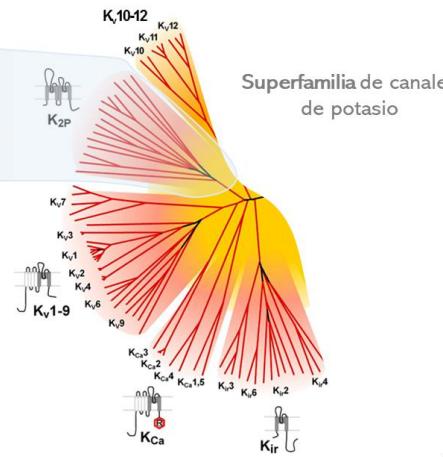
# ¿Qué vimos en la sesión anterior?

**Para que se utiliza la estructura de Ramachandran?**

- para detectar zonas estructurales resueltas incorrectamente. Se utilizan para predecir estructuras secundarias.



## Familia de canales de potasio con dos dominios de poro (K<sub>2P</sub>)



RCSB PDB  
PROTEIN DATA BANK

**UniProt**

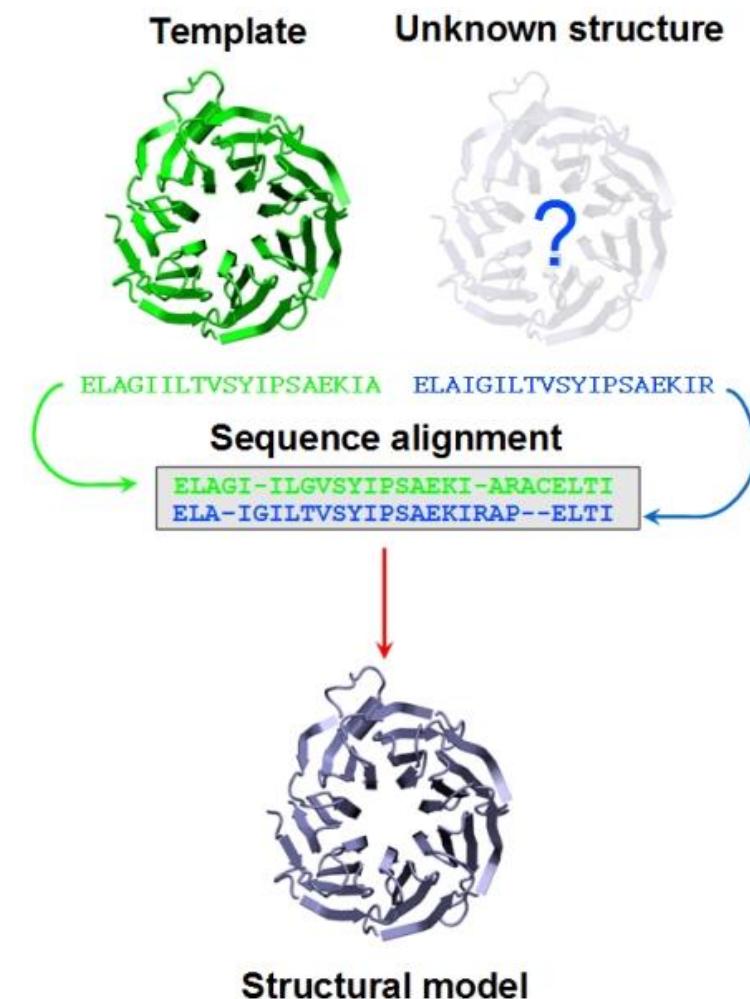
# Modelado comparativo

## Objetivo:

Inferir la estructura de una proteína (*target*) de secuencia conocida a partir de las estructuras 3D conocidas de proteínas de la misma familia evolutiva (*templates*).

Modelización por homología o modelización basada en plantillas (*template-based*).

Se utiliza la estructura de una proteína homóloga



# Temario - Contenidos

## Tema 4. Modelización estructural de proteínas

- 4.1. Modelización por homología
- 4.2. Modelización por reconocimiento de plegamiento
- 4.3. Modelización ab initio

# Temario - Contenidos

## Tema 4. Modelización estructural de proteínas

### 4.1. Modelización por homología

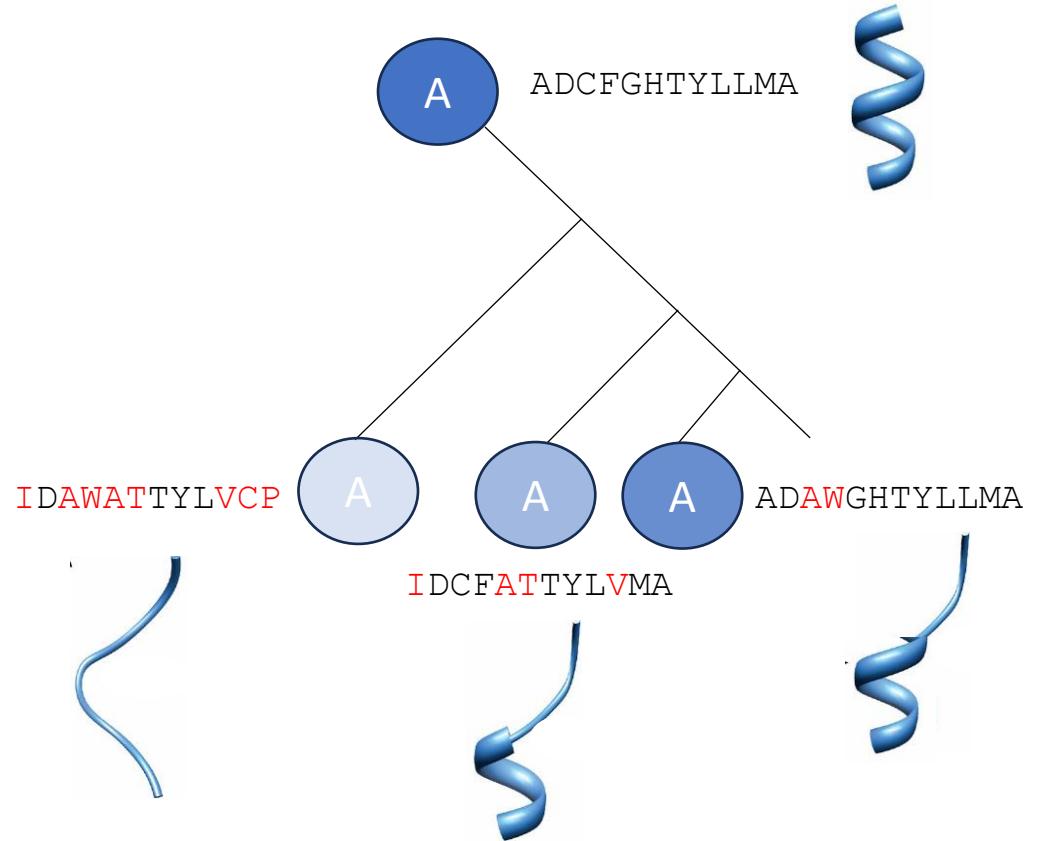
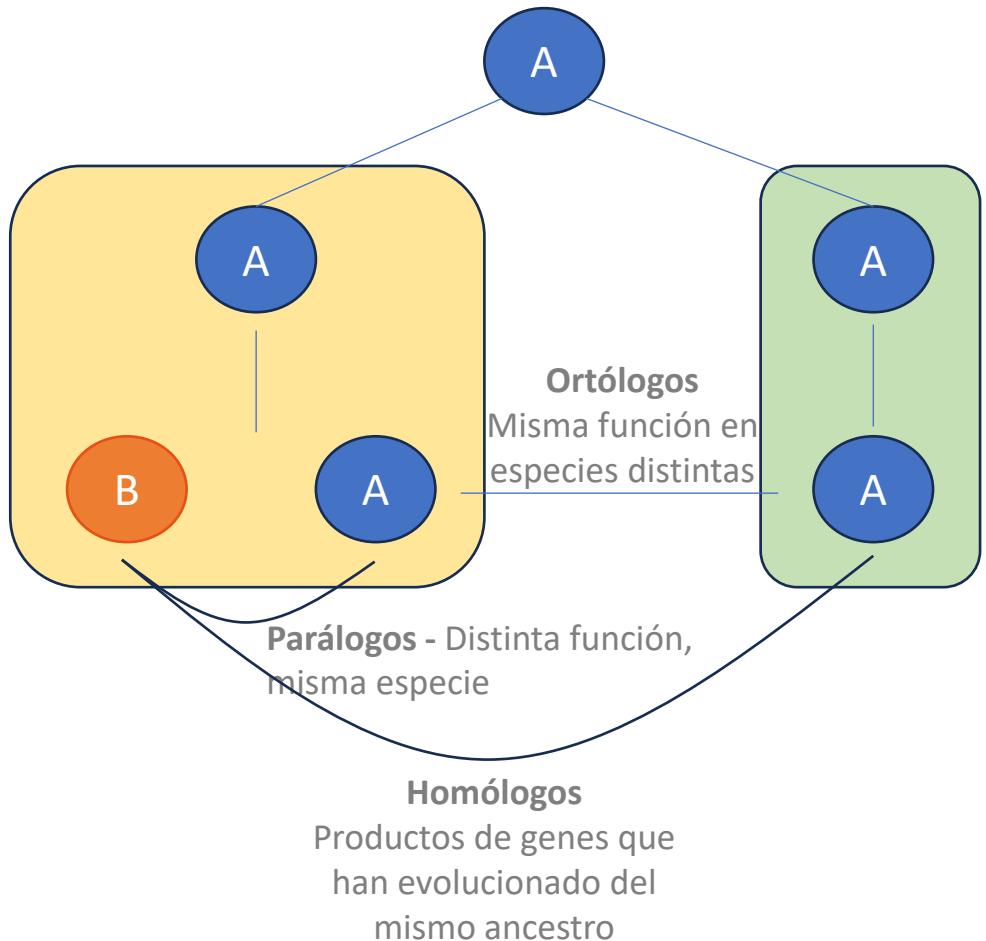
4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización ab initio

- Homología de proteínas: aspectos básicos
- Identificación de *templates*
- Construcción del modelo
- Evaluación y validación
- Aplicación a escala proteómica

# 4.1. Modelización por homología

## Homología de proteínas: aspectos básicos



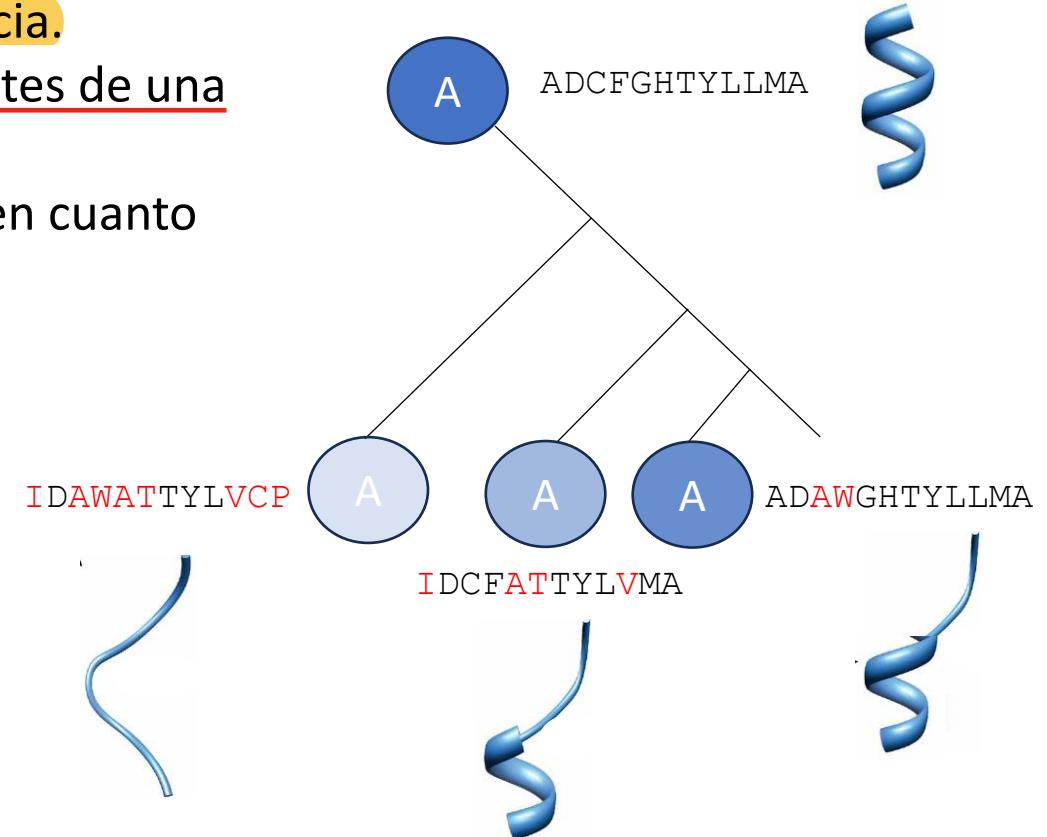
# 4.1. Modelización por homología

## Homología de proteínas: aspectos básicos

- El plegamiento está más conservado que la secuencia.
- Los elementos de estructura secundaria son las partes de una proteína más conservadas.
- Los bucles (*loops*) presentan la mayor variabilidad en cuanto a estructura.

Entre proteínas, un 20% de similitud de secuencia indica una estructura similar.

En péptidos, dicha similitud debe ser >60% para que esto se cumpla.



# Temario - Contenidos

## Tema 4. Modelización estructural de proteínas

### 4.1. Modelización por homología

### 4.2. Modelización por reconocimiento de plegamiento

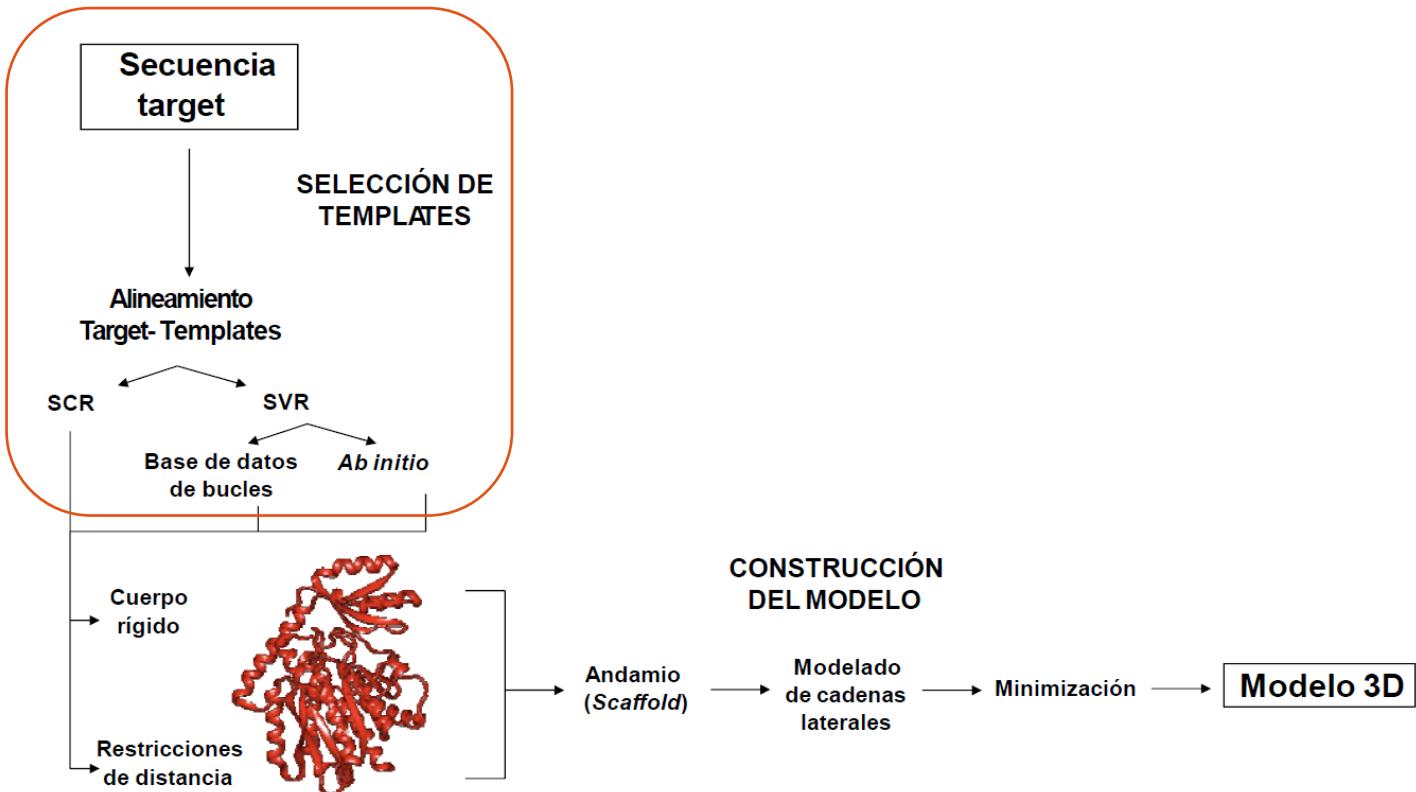
### 4.3. Modelización ab initio

- Homología de proteínas: aspectos básicos
- Identificación de *templates*
- Construcción del modelo
- Evaluación y validación
- Aplicación a escala proteómica

# 4.1. Modelización por homología

## Identificación de *templates*

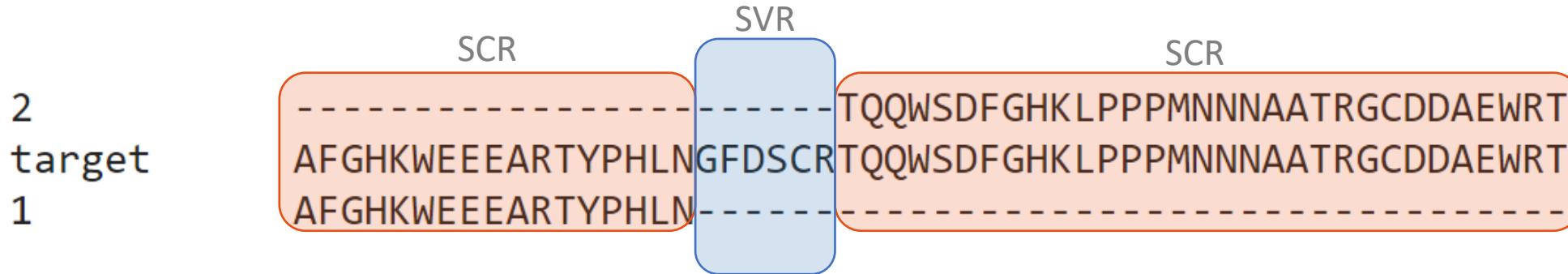
- En función de similitud de **secuencia**: BLAST, alineamientos múltiples de secuencia - ClustalW, T-coffee, HMMER.
- En función de similitud de **motivos estructurales**: alineamientos múltiples estructurales - STAMP



# 4.1. Modelización por homología

## Identificación de *templates*

- *Template* homólogo más cercano a la proteína diana.
- Dependiendo de cobertura, limitar número de *templates* (número bajo).
- Determinar regiones estructuralmente conservadas (SCR) y variables (SVR)

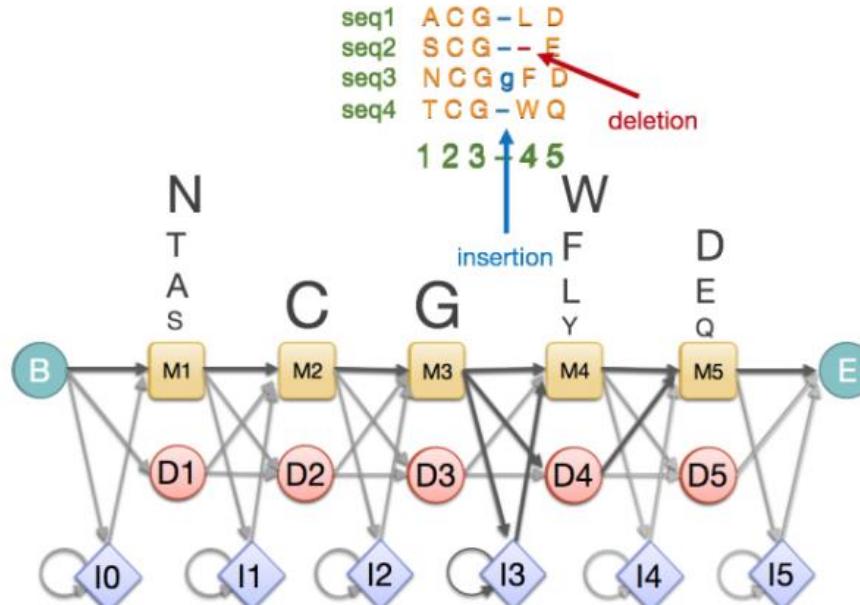
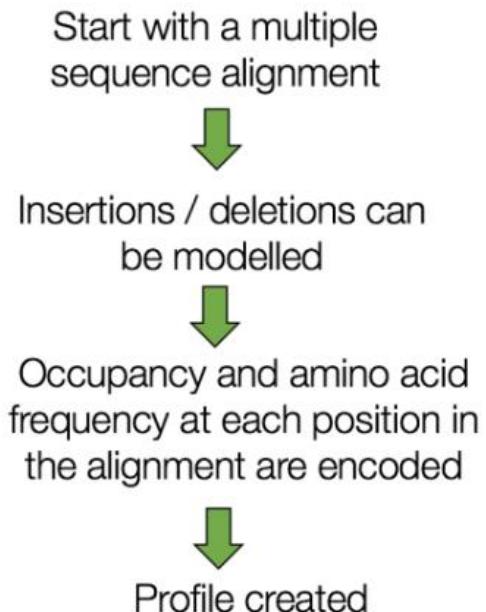


# 4.1. Modelización por homología

## Identificación de *templates*

### ○ *Profile hidden Markov models (HMM)*

- Se trata de un modelo probabilístico que **encapsula los cambios evolutivos que han ocurrido en un grupo de secuencias relacionadas (alineamiento múltiple)**.
- Capturan información específica de posición de cómo de conservado es un aminoácido en cada columna del alineamiento.
- Determina el grado de variación con el que ocurren inserciones o delecciones.
- Usa penalizadores para espacios y probabilidades de sustitución que reflejan mejor la realidad biológica.



# Temario - Contenidos

## Tema 4. Modelización estructural de proteínas

### 4.1. Modelización por homología

### 4.2. Modelización por reconocimiento de plegamiento

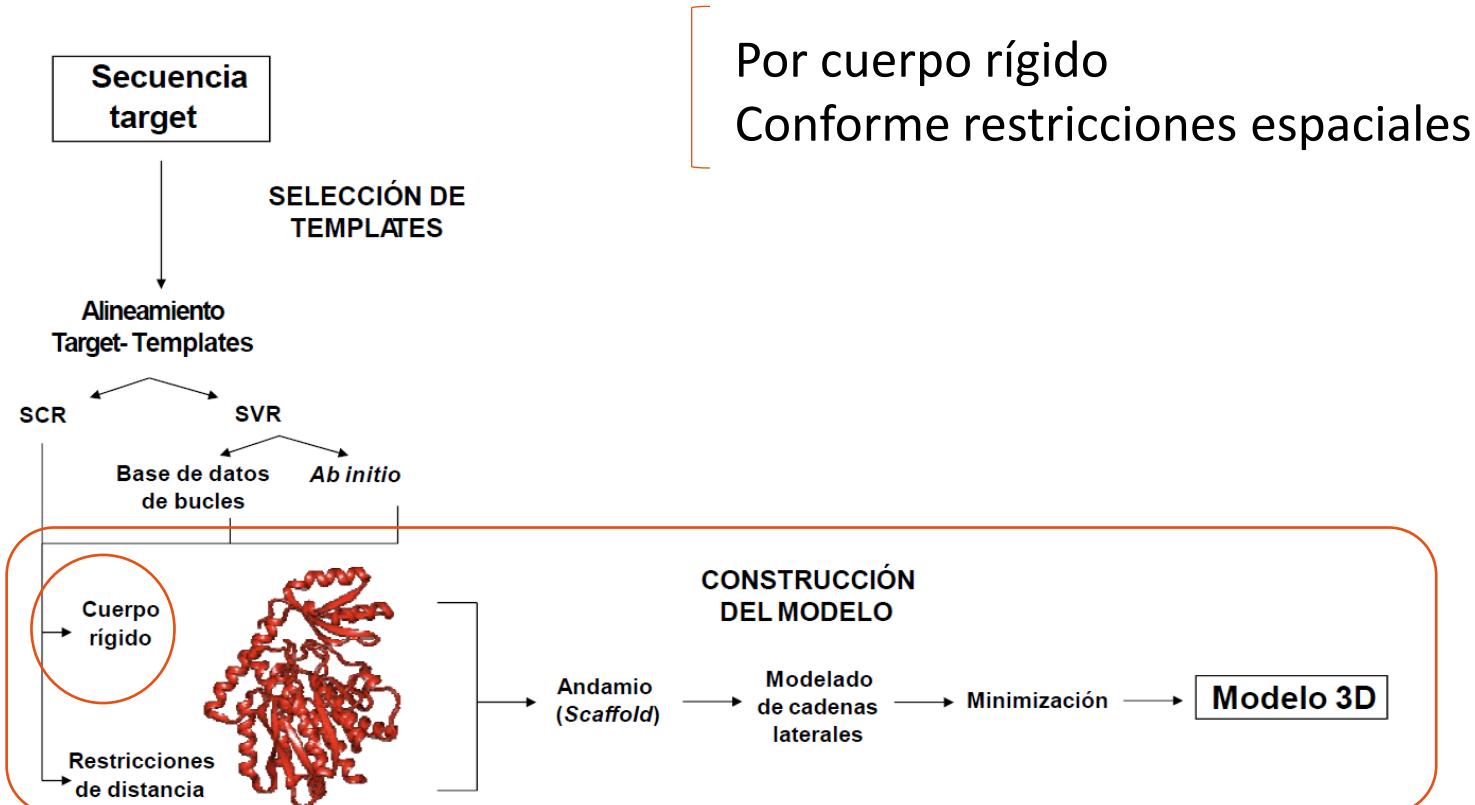
### 4.3. Modelización ab initio

- Homología de proteínas: aspectos básicos
- Identificación de *templates*
- Construcción del modelo
- Evaluación y validación
- Aplicación a escala proteómica

# 4.1. Modelización por homología

## Construcción del modelo – cadena principal (*backbone*)

- Elementos de estructura secundaria altamente conservadas entre proteínas homólogas
- *Loops* que las conectan, baja conservación. Corresponden a huecos en los alineamientos  
➤ DIFICULTAD

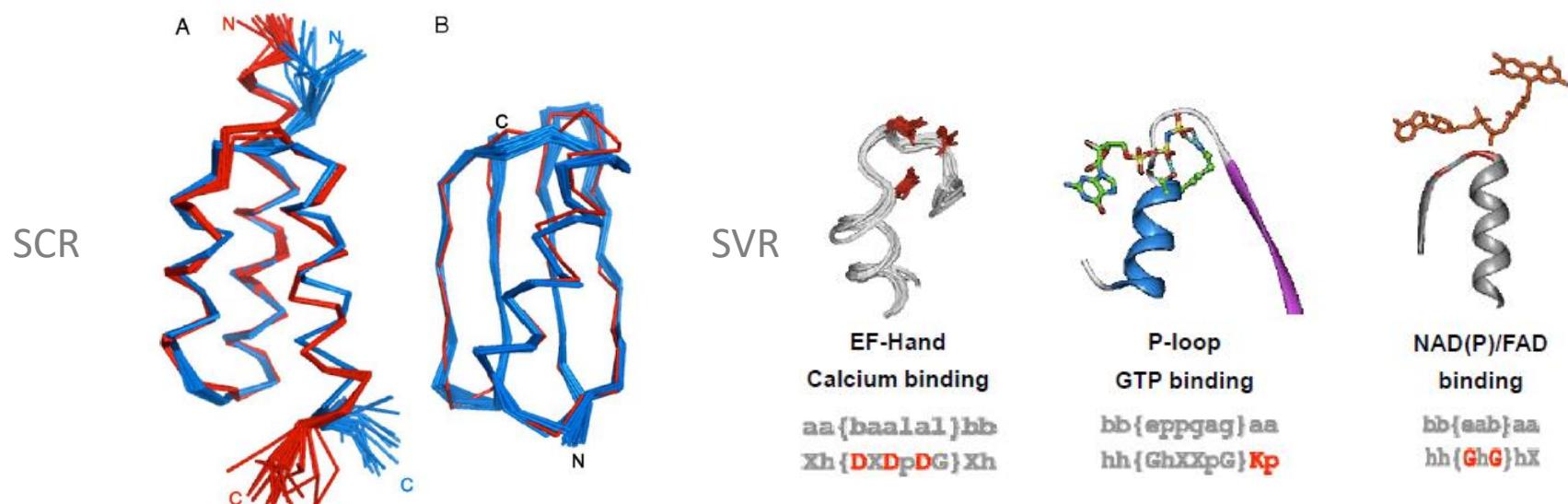


# 4.1. Modelización por homología

## Construcción del modelo – cadena principal (*backbone*)

### ○ Ensamblado por cuerpo rígido

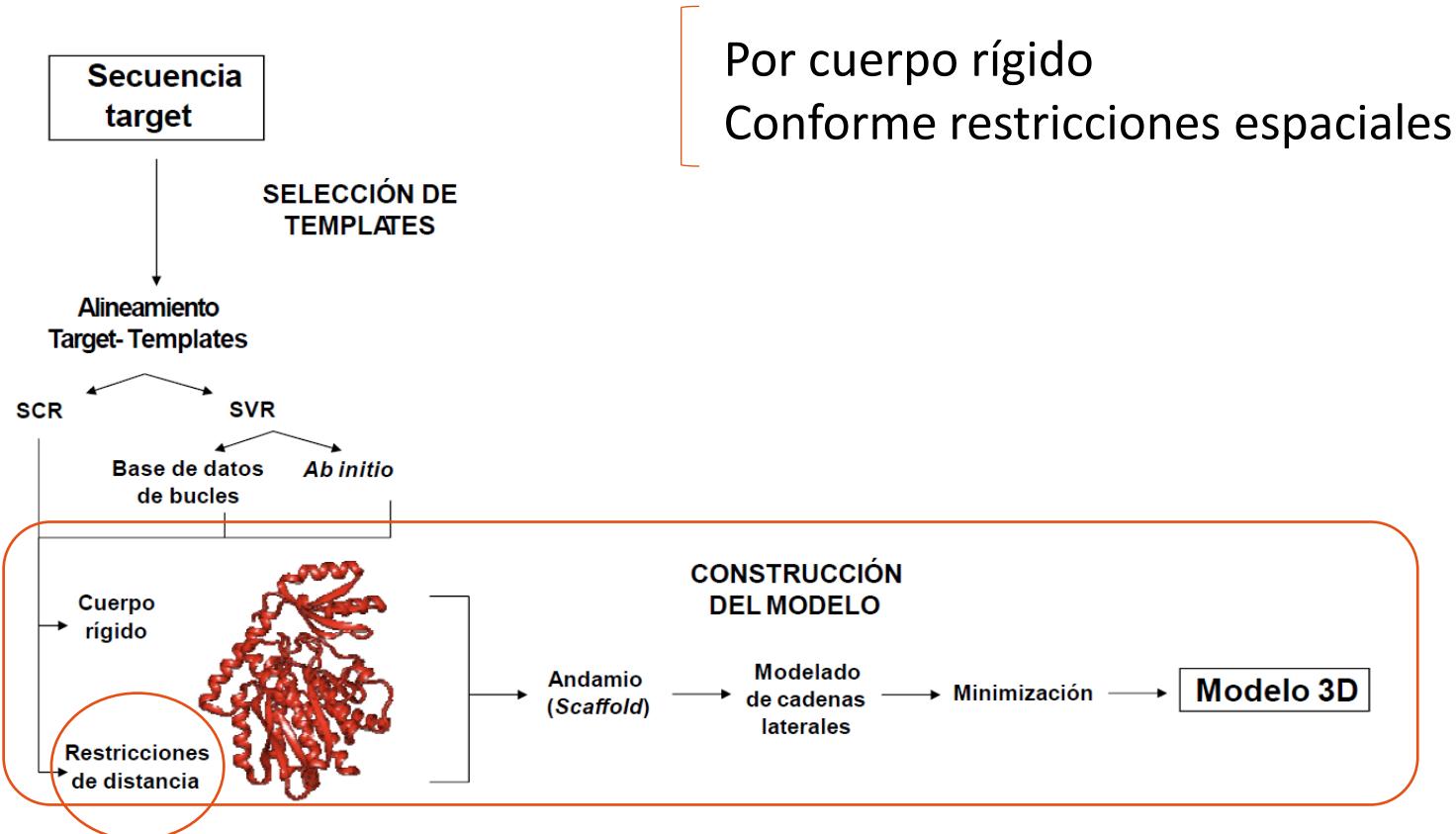
- Se realiza la superposición de la **cadena principal (SCR)** de las estructuras homólogas cercanas
- Las **regiones mejor alineadas estructuralmente** son seleccionadas para estimar la posición **promedio de cada átomo target**.
- Las **regiones altamente variables (SVR)** como bucles son modelados a partir de **librerías de bucles** (ArchPred, LoopX) o ***ab initio*** (dinámica molecular).
- **Refinar posiciones atómicas mediante optimización energética.**



# 4.1. Modelización por homología

## Construcción del modelo – cadena principal (*backbone*)

- Elementos de estructura secundaria altamente conservadas entre proteínas homólogas
- *Loops* que las conectan, baja conservación. Corresponden a huecos en los alineamientos  
➤ DIFICULTAD



# 4.1. Modelización por homología

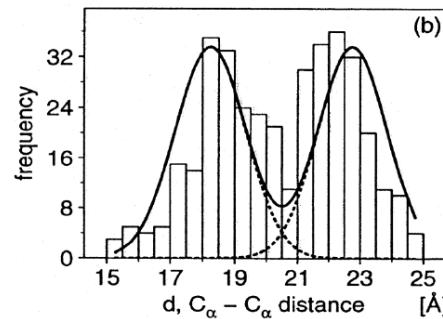
## Construcción del modelo – cadena principal (*backbone*)

- Conforme con restricciones espaciales

- Se basa en funciones de densidad de probabilidad para parámetros estructurales (Ca).
- Distribución de ángulos *phi* y *psi* no adoptan cualquier valor para una estructura secundaria específica (Ramachandran).
- Extraer valores de frecuencias de *templates* para construir la función de probabilidad y transformarla en energía empírica mediante ecuación de Boltzmann.

Ejemplo:

- Distribución of distancias Ca-Ca



Una restricción se define como una función de densidad de probabilidad (PDF),  $p(x)$ :

$$p(x_1 \leq x < x_2) = \int_{x_1}^{x_2} p(x) dx \quad \text{con} \quad p(x) > 0$$

$$\int p(x) dx = 1$$

Y puede expresarse en unidades de energía:

$$E_{pdf}(x) = -RT \log(p(x))$$

## 4.1. Modelización por homología

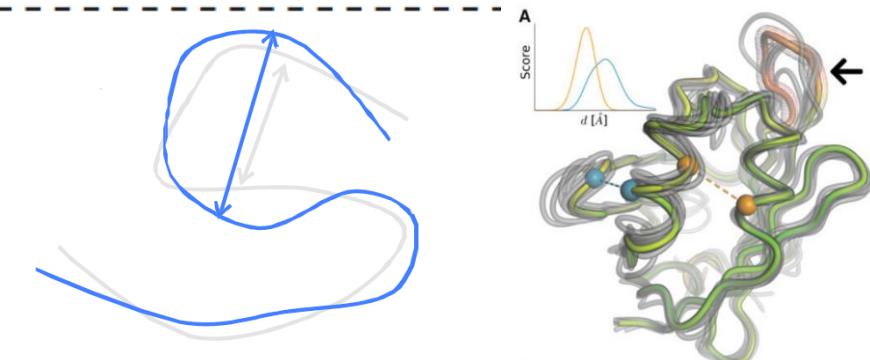
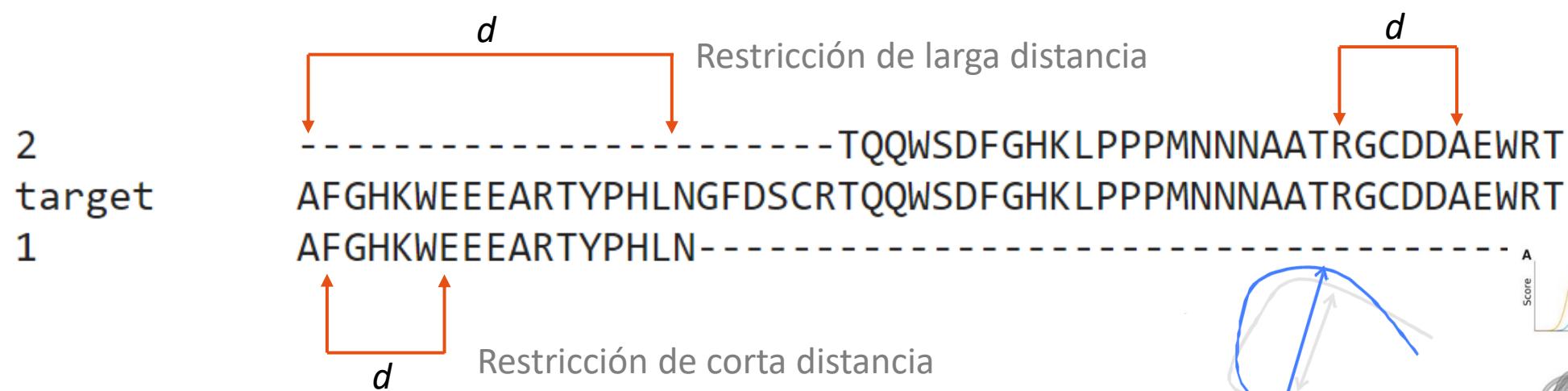
### Construcción del modelo – cadena principal (*backbone*)

- Conforme con restricciones espaciales

#### *Simulated annealing*

- Conformación del modelo al azar usando restricciones de corta distancia.
- Minimizar energía
- Repetir iterativamente, con restricciones cada vez más largas.

lo que buscamos es un mínimo energético que garantice la secuencia más apropiada para el modelo que estamos modelando



# 4.1. Modelización por homología

## Construcción del modelo – cadena principal (*backbone*)

- Conforme con restricciones espaciales

- Se basa en funciones de densidad de probabilidad para parámetros estructurales (Ca).
- Distribución de *ángulos phi y psi* no adoptan cualquier valor para una estructura secundaria específica (Ramachandran).
- Extraer valores de frecuencias de *templates* para construir la función de probabilidad y transformarla en energía empírica mediante ecuación de Boltzmann.
- Esta energía empírica se suma al resto de términos energéticos que describen el modelo.
  - Objetivo: encontrar la conformación que minimice la energía global.

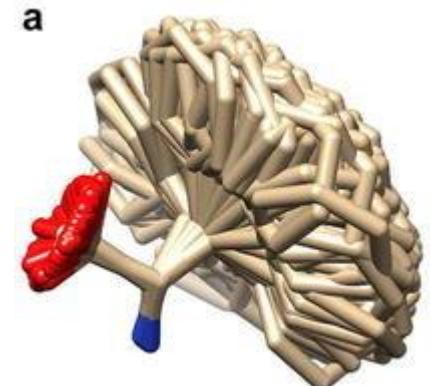
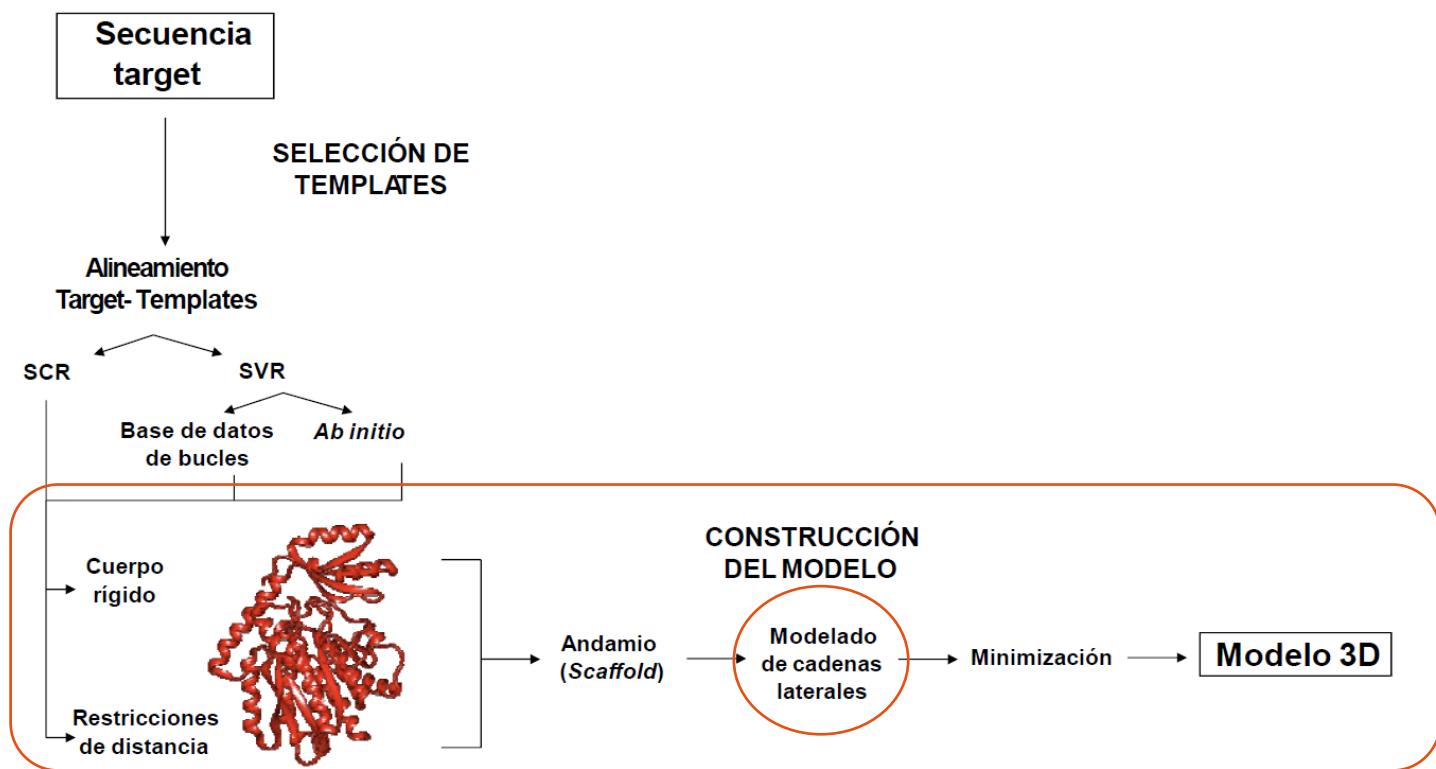
*Simulated annealing* o enfriamiento simulado

Programas: MODELLER, Rosetta, YASARA

# 4.1. Modelización por homología

## Construcción del modelo – cadenas laterales

- Librerías de rotámeros con **frecuencias precalculadas** para cada residuo en distintas **estructuras**.
- Cálculo de la **probabilidad** de cada rotámero según **energía del modelo** (Boltzmann).
- Optimización energética por dinámica molecular o *simulated annealing*.



# Temario - Contenidos

## Tema 4. Modelización estructural de proteínas

### 4.1. Modelización por homología

4.2. Modelización por reconocimiento de plegamiento

4.3. Modelización ab initio

- Homología de proteínas: aspectos básicos
- Identificación de *templates*
- Construcción del modelo
- **Evaluación y validación**
- Aplicación a escala proteómica

# 4.1. Modelización por homología

## Evaluación y validación del modelo

- Errores frecuentes en modelos:
  - Desplazamiento de residuos correctamente alineados
  - Empaquetamiento erróneo de las cadenas laterales
  - Regiones sin *template*
  - Errores de alineamiento. Difícil de determinar.
  - *Templates* incompletos

Errores de alineamiento incorrecto

```
GARFIELD THE CAT ...
GARFIELD THE FAT CAT
```

Solución:

```
GARFIELD THE ... CAT
GARFIELD THE FAT CAT
```

# 4.1. Modelización por homología

## Evaluación y validación del modelo

- Detección de errores:
  - Comparar el modelo con todos los *templates*
  - Comprobar que los elementos de estructura secundaria no estén rotos
  - Comprobar que la predicción de estructura secundaria coincide con la observada en modelo
  - Comprobar similitud de bucles del modelo con bases de datos
  - Comprobar inicio y final de hélices
  - Comparar energías entre *templates* y *target*

# 4.1. Modelización por homología

## Evaluación y validación del modelo

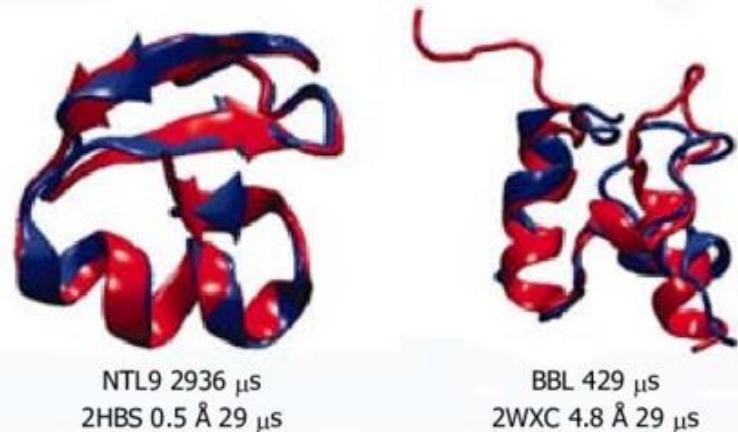
- Mejorar modelo:
  - Modificar manualmente alineamiento según predicciones de estructura secundaria
  - Cambiar *template* principal
  - Incluir nuevos *templates*
  - Comparar energía de nuevo modelo vs antiguo

# 4.1. Modelización por homología

## Evaluación y validación del modelo

- Control del método predictivo:
  - Comparar RMSD entre estructura real y modelo. Importante el método de superposición.
  - Comprobar que los elementos de estructura secundaria estén bien alineados.
  - Calcular fracción de contactos nativos.
  - Calcular porcentaje de residuos correctamente alineados.
  - Comparar energía entre modelo y estructura real.

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (\vec{x}_{1,i} - \vec{x}_{2,i})^2}{n}}$$



# 4.1. Modelización por homología

## Evaluación y **validación** del modelo

- **Control** del método predictivo:
  - Desarrollo de nuevos métodos predictivos objetivamente evaluados
  - CASP (*Critical Assessment of Protein Structure Prediction*)  
[\(https://predictioncenter.org/casp15/index.cgi\)](https://predictioncenter.org/casp15/index.cgi)
    - Modelado por homología, *ab initio*, refinado, modelado híbrido, modelado de ensamblados multimoleculares
  - CAMEO ([www.cameo3d.org](http://www.cameo3d.org))

# Temario - Contenidos

## Tema 4. Modelización estructural de proteínas

### 4.1. Modelización por homología

### 4.2. Modelización por reconocimiento de plegamiento

### 4.3. Modelización ab initio

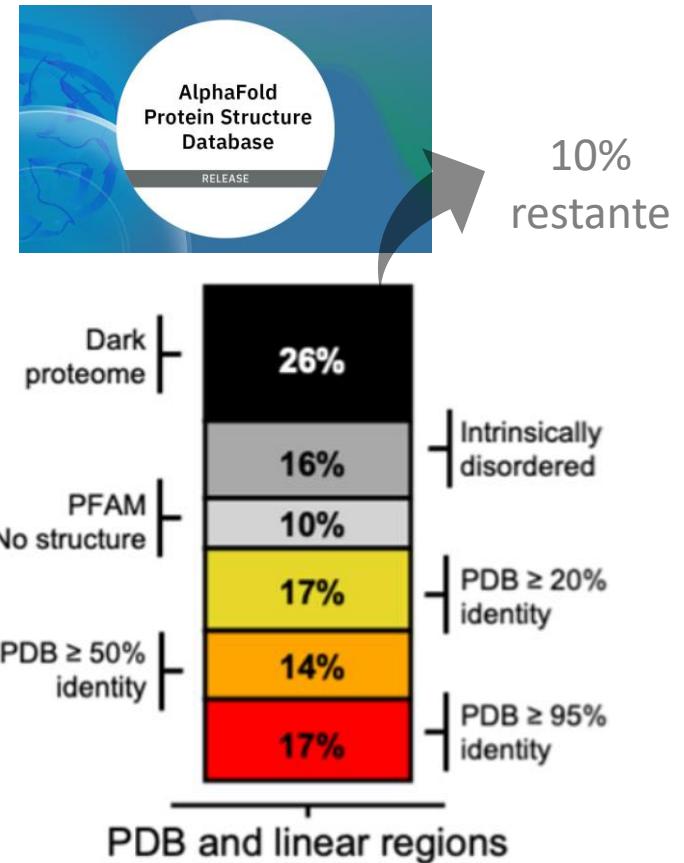
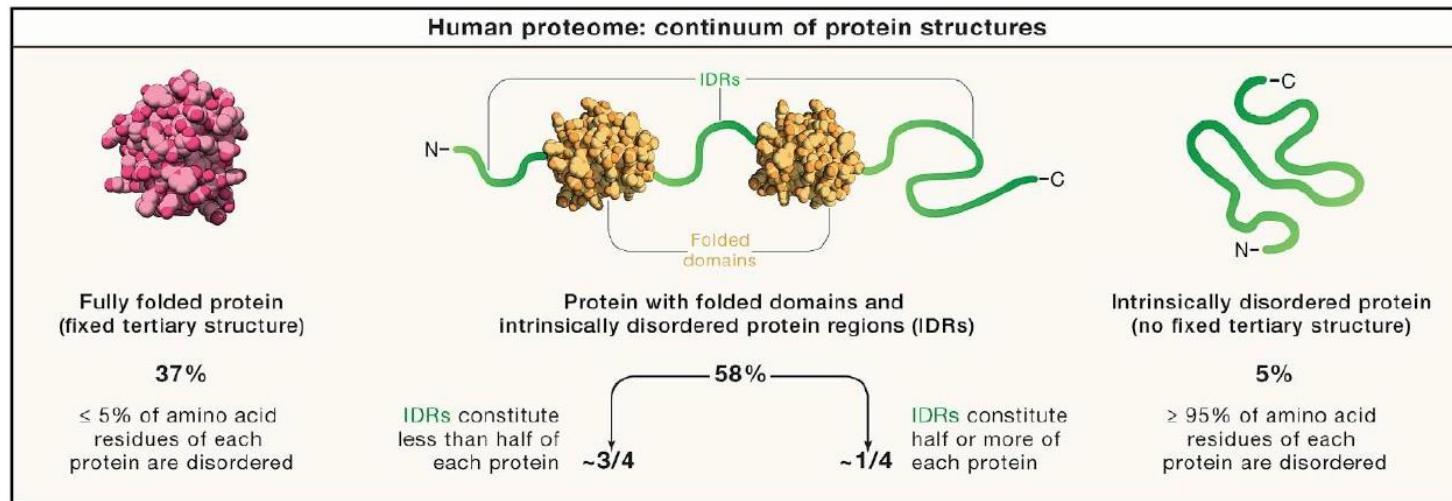
- Homología de proteínas: aspectos básicos
- Identificación de *templates*
- Construcción del modelo
- Evaluación y validación
- **Aplicación a escala proteómica**

# 4.1. Modelización por homología

## Aplicación a escala proteómica

Se estima que existen unas **20.000 proteínas humanas** (no redundantes)

En los últimos **10 años**, los proyectos de proteómica han identificado unas **18.000 proteínas** (!!)



# 4.1. Modelización por homología

## Herramientas para modelado comparativo

### SWISS MODEL

<http://www.expasy.ch/swissmod/>

### Deep View - SPDBV

homepage: <https://spdbv.unil.ch/>

Tutorials [https://spdbv.unil.ch/main\\_tut.html/](https://spdbv.unil.ch/main_tut.html/)

### WHATIF

homepage: <https://swift.cmbi.umcn.nl/whatif/>

Gert Vriend's suite of protein structure modeling & analysis programs

### Modeller

homepage: <https://salilab.org/modeller/>

Andrey Sali's homology protein structure modelling by satisfaction of spatial restraints



viu

**Universidad**  
Internacional  
de Valencia

[universidadviu.com](http://universidadviu.com)

De:  
 Planeta Formación y Universidades