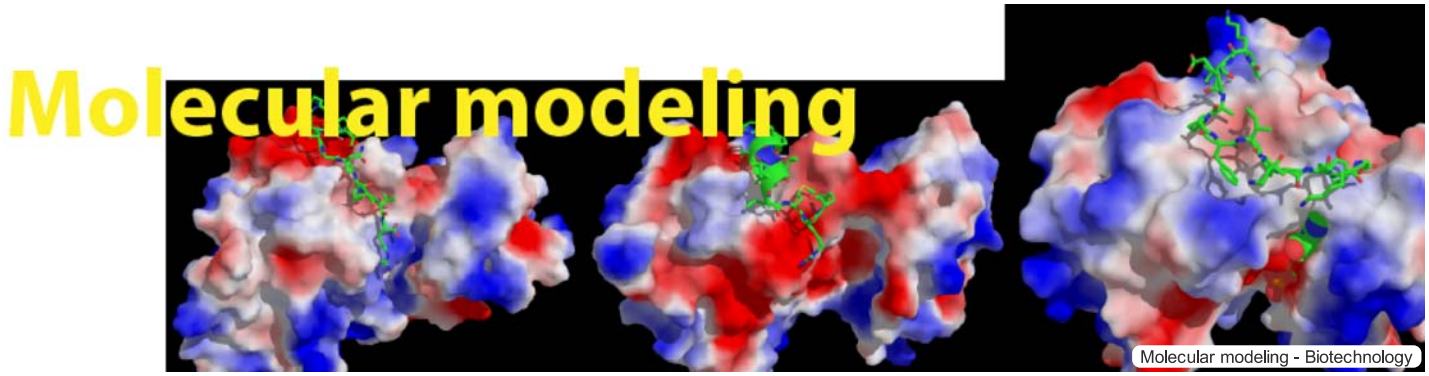


Buscar
Temas de teoría
Prácticas 1-6
Prácticas 7-12
Bibliografía
Enlaces

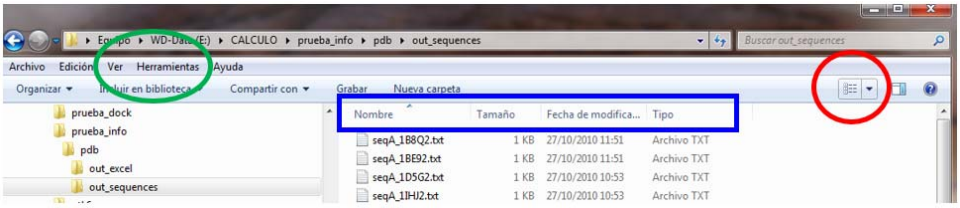
version en

Before running the programs contained in these practices, be sure of the following: (optimized for Windows 7).

1) **NEVER** use an user account containing a written accent
- Example: User account "Andrés" in C:\Users\Andrés\
(most of the programs will not work, i.e. PyMol)

2) **NEVER** use a folder containing a written accent
- Example: C:\cálculo\
(most of the programs will not work, i.e. PyMol)

3) Put the **FILE FEATURES** under your direct control. In a fresh session of the Windows File Explorer (this is not internet explorer)
- Control the file list and file **SIZE**:



Activate "Ver/Barra de estado" (green circle).
Activate "Detalles" in "Ver/Detalles" (or in the red circle).
Activate "Tamaño" (if inactive) using the contextual menu (right mouse button) in the blue square.
- Activate the file **EXTENSION**:
Open "**Herramientas**" (green circle), and press "Opciones de carpeta". Go to folder "**Ver**" and unpick "**Ocultar las extensions de archive para tipos de archivo conocidos**".
Example: The text file "seqA_1B8Q2" become "seqA_1B8Q2.txt"

4) Adjust the **REGIONAL configuration**: Most of the programs manage the dot ("**.**") as the decimal separator. As a rule to perform the practical sessions, change **ALWAYS** this regional configuration in your PC computer.

For that, go to "**Panel de control**" and select "**Configuración regional y de idioma**". Then press "**Configuración adicional**" and under the folder "**Números**" select the dot ("**.**") as "**Símbolo decimal**", the comma ("**,**") as the "**Símbolo de separación de miles**", and press **OK**.

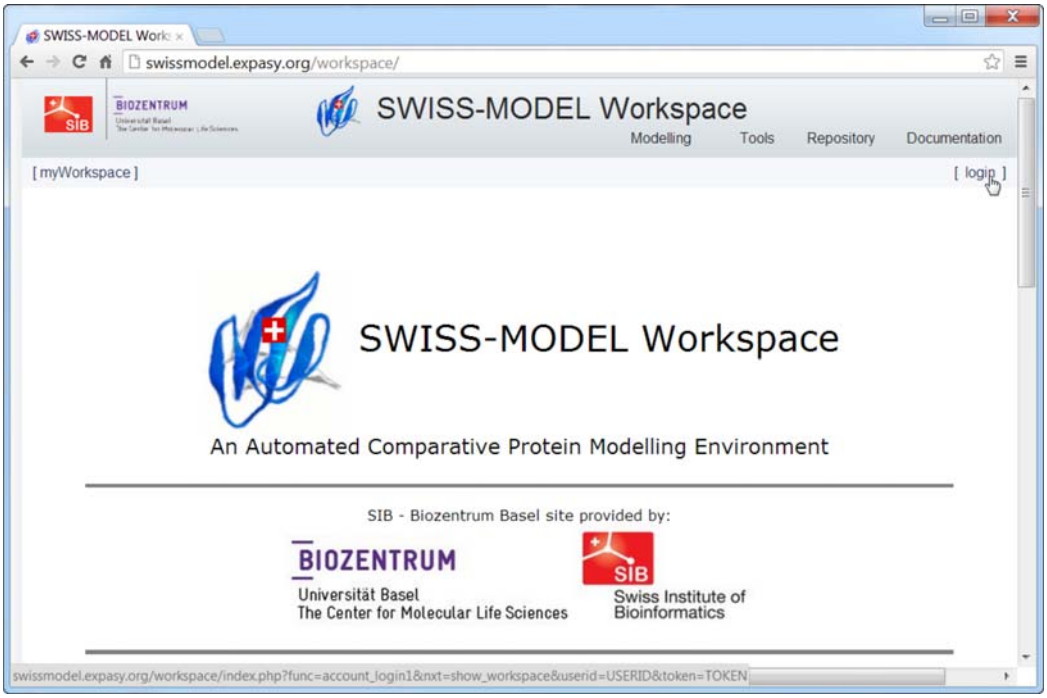
Practice 8 - HOMOLOGY MODELING.

Overview

DeepView offers a series of commands that let you model new structures by submitting modeling requests to Swiss-Model, a server for automated homology modeling. Homology modeling, also called comparative protein modeling or knowledge-based modeling, is the process by which a 3D model of a target sequence is built based on an homologue experimentally solved structure (experimental processes include X-ray crystallography and solution nuclear magnetic resonance).

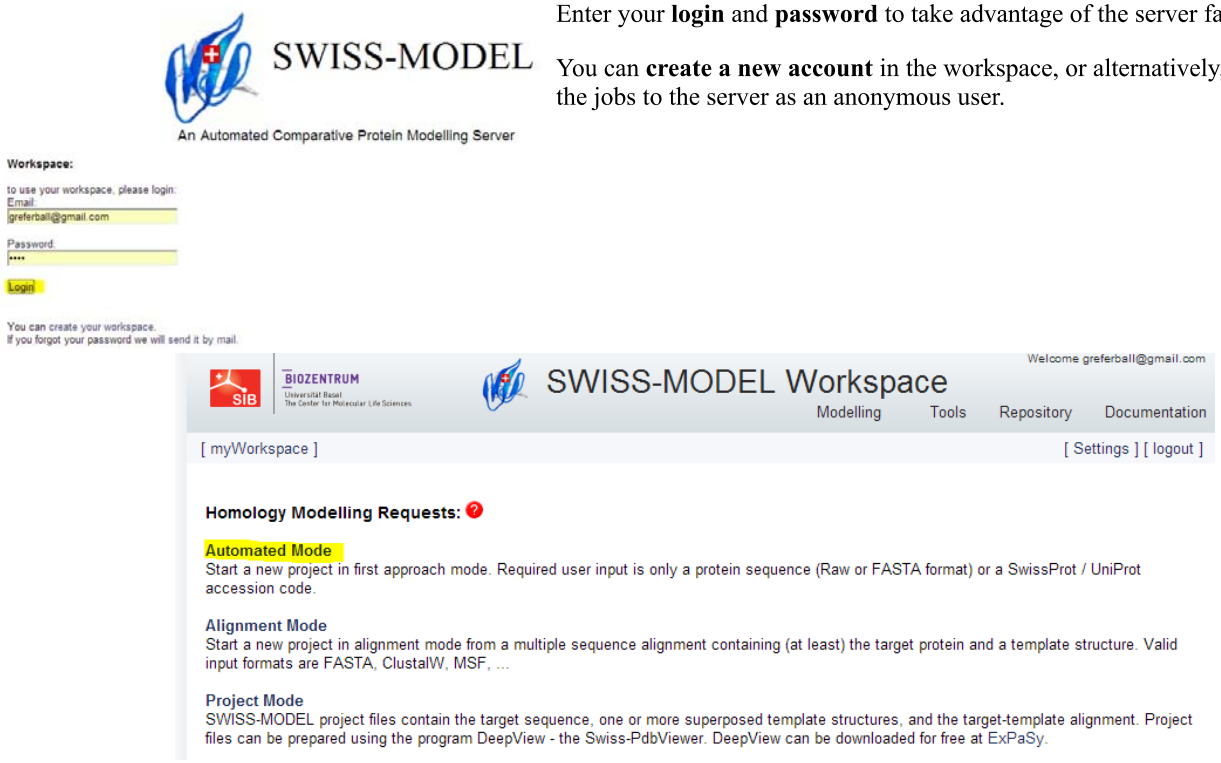
A **target sequence** is the primary sequence of a protein whose structure has to be modeled. When first loaded in the workspace, it is provisionally drawn as a long helix. A **template structure**, or simply a template, is an experimentally solved structure used as a scaffold to model the structure of the target sequence. Template sequence is the primary sequence of a template.

Swiss-Model



Swiss-Model is a server for automated comparative protein modeling. It is available free of charge at the ExPASY (Expert Protein Analysis System) site, where extensive documentation on the architecture and use of Swiss-Model can be found.

The ExPASY site is the proteomics server of the Swiss Institute of Bioinformatics (SIB). The server is devoted to the analysis of protein sequences and structures. Amongst other documentation, it comprises several protein databases such as SWISS-PROT, TrEMBL, and PROSITE, and provides links to many other molecular biology databases, such as PDB.



Enter your **login** and **password** to take advantage of the server facilities.

You can **create a new account** in the workspace, or alternatively, you can send the jobs to the server as an anonymous user.

Several modeling modes are currently available at the Swiss-Model server: Automated, Alignment and Project modes.

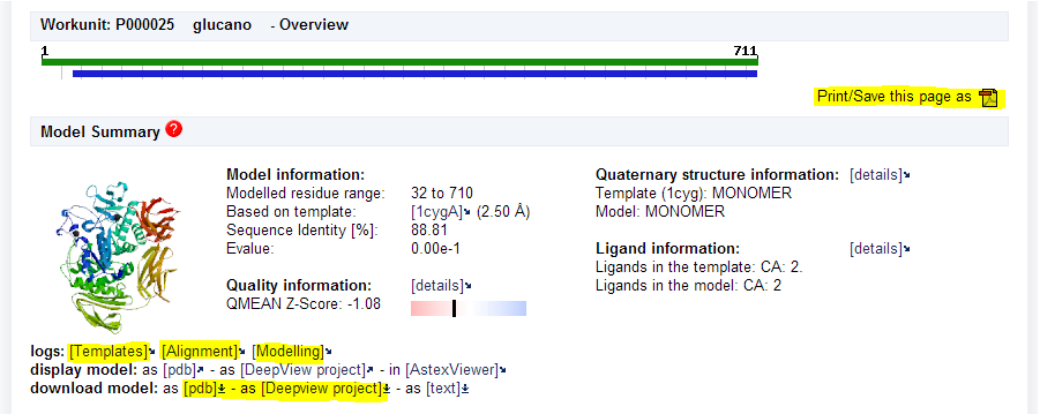
Automated mode

The "automated mode" is suited for cases where the target-template similarity is sufficiently high to allow for fully automated modeling. As a rule of thumb, **automated sequence alignments are sufficiently reliable when target and template share more than 50% percent of sequence identity**.

This submission requires only the amino acid sequence (FASTA format or single letter raw sequence) or the UNIPROT accession code of the target protein as input data. The modeling pipeline automatically selects suitable templates based on a BLAST E-value limit,

which can be adjusted upon submission. The automated template selection will favor high-resolution template structures with reasonable stereochemical properties as assessed by ANOLEA mean force potential and Gromos96 force field energy. Follow the example in the exercise P08-01.

Model coordinate file PDB format or DeepView project format can be downloaded and viewed using Deepview or other tools for molecular visualization.



Logfile PDF: The content of the web page (including images and logfiles, but not the model coordinates) can be downloaded and saved as PDF. See "Print/Save this page as [pdf]" at the top of the page.

The modeling process and quality estimation can be viewed in the same page with the following drop-down menus [+/-]

P08-01) Work on this task: Model the protein glucanotransferase from *Bacillus stearothermophilus* in **automate mode in Workspace server**.

```
>glucano
MKRWLSVVLSMSLVFSAFFLVSDTQKVTVEAAGNLNKVNFTSDIVYQIVVDRFVD
GNTSNNPSGSLFSSGCTNLRKYCGGDWQGIINKINDGYLTEMGVTAIWISQPVEN
VFAVMNDADGSTSYHGYWARDFKKTNPFFGTLSDFQRLVDAAHAKGIKVIIDFAP
NHTSPASETNPSYMENGRLYDNGTLIGGYTNDTNSYFHHNGGTTFSNLEDGIYRN
LFDLADFNHQNQFIDKYLKDAIKLWLDMGIDGIRMDAVKHMPFGWQKSFMDEVYD
YRPVFTFGEWFLSENEVDSNNHFFANESGMSLLDFRFGQKLRQVLRNNSDDWYGF
NQMIQDTASAYDEVIDQVTFIDNHDMDRFMADEGDPRKVDIALAVLLTSRGVPNI
YYGTEQYMTGNGDPNNRKMMTSFNKNTRAYQVIQKLSSLRRSNPALSYGDTEQRW
INSDVYIYERQFGKDVVLVAVNRSLSKSYSITGLFTALPSGTYTDQLGALLDGNT
IQVGSNGAVNAFNLGPGEVGVWTYSAAESVPIIGHIGPMMGQVGHKLTIDGEGFG
TNVGTVKFGNTVASVVSWSNNQITVTVPNIPAGKYNITVQTSGGQVSAAYDNFEV
LTNDQVSVRFVVNNANTNWGENIYLVGNVHELGNWNTSKAIGPLFNQVIYSYPTW
YVDVSVPEGKTIEFKFIKKDGSGNVIWESGSNHVYTTPTSTTGTVNVNWQY
```

Download the **pdb model** and the **Deepview model**. Rename as **P08-01&auto.pdb** and **P08-01&autoproj.pdb** and send them to the server. Download the **pdf report** and have a look; DO NOT send the pdf file to the server. Annotate comments you consider interesting in your **report file**.

P08-02) Work on this task: Answer this question regarding the modeling of glucanotransferasa. Annotate the results in your **report file** and answer the question in the task web page "Cuestionarios ON-LINE":

P08-02. ¿Qué estructura pdb se ha tomado como plantilla para el modelado? En relación con la puntuación global QMEAN4 obtenida para el modelo de la glucanotransferasa, ¿qué valores de Z-score se obtuvieron para las energías de solvatación y de ángulo de torsión?

- ☐ 1CYGA; -1.08 y -0.10, respectivamente.
- ☐ 1CYG; -0.41 y -0.86.
- ☐ 1CYG; -1.51 y -0.10.
- ☐ 1CYG; -1.51 y -0.86
- ☐ 1CYGA; -0.86 y -0.41

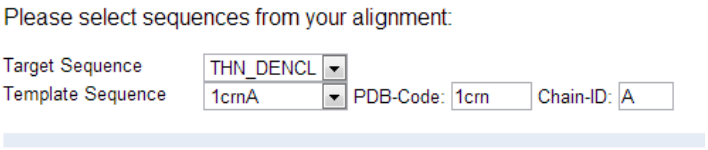
Alignment Mode

Multiple sequence alignments are a common tool in many molecular biology projects. If the three-dimensional structure is known for at least one of the members, this alignment can be used as a starting point for comparative modeling using the "**alignment mode**".

The "**alignment mode**" allows the user to test several alternative alignments and evaluate the quality of the resulting models in order to achieve an optimal result.

In order to facilitate the use of alignments in different formats, the submission is implemented as a four step procedure:

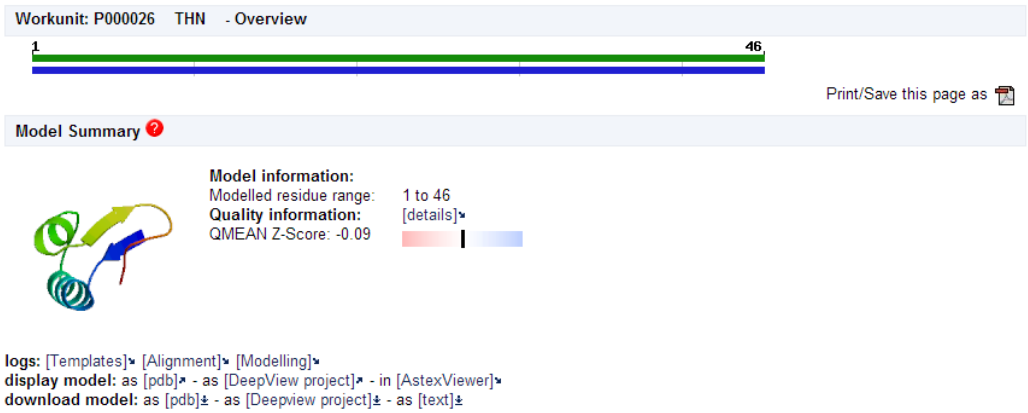
1. Prepare a multiple sequence alignment.
    - It must contain at least your target sequence and the template sequence.
    - Use any of your favorite alignment tools (i.e.: CLUSTALW).
    - Make sure the sequence names are "reasonably short".
2. Submit your alignment to the Workspace Alignment Mode.
    - Possible formats are: FASTA, MSF, CLUSTALW, PFAM and SELEX.
    - You may either upload your file or cut & paste.
    - Don't forget to specify the correct alignment format.
    - **The best way to test the procedure is with a small exercise (see below).**
3. Select Target and Template.
    - The alignment (as it was interpreted by the server) should now be displayed at the bottom part of the page.
    - The script will try to make a good guessing for the correct names based on your submission.
    - Select the sequence name of the target sequence (e.g. THN_DENCL).
    - Select the sequence of the template structure (e.g. 1crnA). You don't need to use PDB IDs, you may use any name you like.
    - Specify the template structure to which this sequence belongs. This template MUST be part of the ExPDB template library. Please use the SWISS-MODEL Template library tool to check the template.
    - Don't forget to specify the correct CHAIN ID. Note that PDB's chain IDs are normally in capital letters.



4. Check Alignment and Submit.
    - The alignment at the bottom of the page should represent the correct mapping of the template structure on the target sequence. Please check carefully before submission.
    - As usual, please provide name and e-mail for the SWISS-MODEL submission.
    - Good luck with you model ...

The server pipeline will build the model purely based on this alignment. During the modeling process, implemented as rigid fragment assembly in the SWISS-MODEL pipeline, the modeling engine might introduce minor heuristic modifications to the placement of insertions and deletions.

You will get the following information (see figure). You can download exactly the model you requested to the server, as well as the quality parameters, the alignments, etc.



P08-03) Work on this task: Send the following alignment in CLUSTALW format to the server under alignment mode.

```
CLUSTAL W (1.82) multiple sequence alignment
THN_DENCL       KSCCPTTAARNQYNICRLPGTPRPVCAALSGCKIISGTGCPPGYRH- 46
THNX_TEST       KSCCPDTTGRDIYNTCRFGGGSRQVCARISGCKIISASTCPS-YPNK 46
1crnA           TTCCPSIVARSNFNVCRLPGTPEALCATYTGCIIIPGATCPGDYAN- 46
                .:***  ..*  :  **: * .. :**  :** **..: **  *
```

Download the **pdb model** and the **Deepview project model**. Rename as **P08-03&align.pdb** and **P08-03&alignproj.pdb** and send them to the server. Download the **pdf report** and have a look; DO NOT send the pdf file to the server. Annotate comments you consider interesting in your **report file**.

P08-04) Work on this task: Answer this question regarding the modeling of THN_DENCL protein. Annotate the results in your **report file** and answer the question in the task web page "Cuestionarios ON-LINE":

P08-04. ¿Qué estructura pdb se ha tomado como plantilla para el modelado de la proteína THN_DENCL?

- ☐ No hay ninguna estructura que se haya usado como plantilla.
- ☐ 1CRNA.
- ☐ El servidor elige una estructura adecuada de forma automática.
- ☐ No hace falta ninguna plantilla para modelar ya que tenemos un alineamiento.
- ☐ 1CRN.

Project mode (Main application)

In difficult modeling situations, where the correct alignment between target and template cannot be clearly determined by sequence based methods, visual inspection and manual manipulation of the alignment can significantly help improving the quality of the resulting model. Project files containing the superposed template structures, and the alignment between the target and the template can be generated using the program DeepView (Swiss-PdbViewer).The user has therefore full control over essential modeling parameters, i.e. the choice of template structures, the correct alignment of residues, and the placement of insertions and deletions in the context of the three-dimensional structure. Modeling of oligomeric proteins with Swiss-Model Workspace can be done using the Project Mode.

Modeling a dimeric protein. In order to demonstrate oligomer-modeling, we are going to build a model of the **protease of murine leukemia virus** based on the structure of a **HIV-1 protease**.

1. Get the template in the correct quaternary state. First, check the correct biological assembly of your template protein. Download "dimer.pdb" and save the template in your local disk.
2. Remove all non-aminoacid residues. Open the file in DeepView and remove all non-aminoacid groups such as ions, ligands, etc. from the template. You can do this by selecting the groups in the control panel of DeepView and Remove the selected residues ("**Build**" menu).
3. Ensure Unique Chain IDs. Make sure each chain has a unique name, e.g. "A","B", etc. Overwrite the modified structure as "dimer.pdb".
4. Target Sequence. Create a FASTA file (text file) with your target sequences for each chain in the SAME order as in the template, i.e. "A", then "B" etc. separated by semicolons. Name as "**target.txt**". (See the work task below).
5. Load the target sequence into DeepView. Start a fresh session of DeepView. Please make sure to start with loading the amino acid sequence of your target protein *first* using the SwissModel menu (before loading any template structures). Use **Load Raw Sequences from Amino Acids**. A long helix appears.
6. Load the template structure "dimer.pdb" into DeepView. Generate a preliminary target-template alignment using Menu: **Fit - Fit raw sequence**. Center view and now activate the layer of "dimer.pdb" and "color by chains", which helps to see the chain boundaries in both sequences.
7. Adjust target-template Alignment in DeepView. Open the alignment window and adjust alignment. Make sure NOT to align residues of different chains. Do not align to "non aminoacid residues" like het groups, OXT, etc. Make sure all insertions & deletions are correctly positioned in the structural context. In order to make the two alignments identical, in the alignment window, move the residue Q99 (bad position in the left figure) to the correct position (right figure). For this, pick on Q99 in the alignment window, press "**Ctrl + space bar**" three times (this means press Ctrl, and without releasing, press space bar). This moves also the remaining sequence to C-terminal. To recover the correct alignment, pick on P107 and press "**Ctrl + backspace**" three times.

8. SWISS-MODEL Submission. Save the project to your local disk [e.g.: "dimer_proj.pdb") and submit the file to the project mode of SWISS-MODEL workspace for model building. A new work unit will be created, containing the modeling results, including log file, ANOLEA evaluation, and model project file of the modeled dimer.

P08-05) Work on this task: Model the HIV protease as a dimer of identical subunits. Create the file "**target.txt**", including the text en the correct fasta format below: identical sequences for chain A and B, separated by a semicolon. Follow the steps explained above and

send the "**dimer_proj.pdb**" to the server in project mode..

```
>TARGET
QGQEPPPEPRITLTVGGQPVTFLVDTGAQH
SVLTQNPGPLSDRSAWVQGATGGKRYRWTT
DRKVHLATGKVTHSFLHVPDCPYPLLGRDL
LTKLKAQI;
QGQEPPPEPRITLTVGGQPVTFLVDTGAQH
SVLTQNPGPLSDRSAWVQGATGGKRYRWTT
DRKVHLATGKVTHSFLHVPDCPYPLLGRDL
LTKLKAQI
```

Rename the "**dimer_proj.pdb**" to **P08-05&input.pdb**. Download the Deepview project model in the result page, and rename as **P08-05&dimerproj.pdb**. Open the project with Deepview, extract the model (first layer) and rename as **P08-05&dimer.pdb**. Send all three files to the server. Download the pdf report and look. DO NOT send the pdf file to the server. Annotate comments you consider interesting in your **report file**.

P08-06) Work on this task: Answer this question regarding the modeling of HIV protease. Annotate the results in your **report file** and answer the question in the task web page "Cuestionarios ON-LINE":

P08-06. Si miramos el alineamiento entre la secuencia de partida ("TARGET") y el archivo pdb que sirve de plantilla ("dimer"), ¿Dónde encontramos inserciones o delecciones?

- ☐ No hay inserciones o delecciones. Toda la secuencia encaja perfectamente, excepto por los dos últimos aminoácido en el C-terminal.
- ☐ Hay una inserción de la secuencia de partida con respecto a la plantilla. Es una Cys extra.
- ☐ Hay una delección de la secuencia de partida con respecto a la plantilla. Es una Cys extra.
- ☐ La inserción se encuentra en el C-terminal.
- ☐ La delección se encuentra en el C-terminal.

Modeling proteins in complex with ligands

This exercise should be done by using all the abilities acquired in the former exercises and in other practical sessions.

P08-07) Work on this task: Model a yeast SH3 domain in complex with its ligand by using the template 1cka-temp.pdb.

```
Yeast SH3 sequence:
DKLYALYAFNGHDSSHCQLGQDEPCILLNDQDAYWWLVKRITDGKIGFAPAEILETF
Yeast ligand:
RLPVLPLY
```

Send the input project as "**P08-07&input.pdb**". Download the Deepview project model in the result page, and rename as **P08-07&SH3proj.pdb**. Open the project with Deepview, extract the model (first layer) and rename as **P08-07&SH3.pdb**. Send all three files to the server. Download the pdf report and look. DO NOT send the pdf file to the server. Annotate comments you consider interesting in your **report file**.

Modeling proteins locally

There are many ways to model a protein locally. Today we will learn how to use the program FoldX to perform easy models, in protein alignments that dont have insertions or deletions.

Using FoldX for homology modeling:

1. Prepare a folder for calculations. Save and unzip in this folder the file FoldX, containing the executable "**FoldX_28.exe**", as well as the files "**rotabase.txt**", "**template.txt**", "**mutant_file.txt**", and the FoldX script "**hmodel.txt**"
2. Edit "**template.txt**" and write the name of your template pdb file (i.e. 1abo2.pdb) in a single line, and save the file.
3. Edit "**mutant_file.txt**" and write in the first line the template sequence, and in the second line the query sequence. Both sequences must be equal in length. No insertions or deletions are allowed. No symbols (i.e.: '-') different from single amino acid code letters are allowed.
4. Open a DOS shell window. Go to the folder containing the executable FoldX_28.exe and write:
5. >**FoldX_28.exe -runfile hmodel.txt** (and press enter). ['>' is the prompt of the operative system, so you don't need to type it].

6. After homology modeling calculations you will have finally your model (i.e. 1abo2_1.pdb). Visualize the model with DeepView by superimposing the model with the template used for modeling. Check the alignment window.
7. Minimize energy with DeepView.

P08-08) Work on this task: Model the following sequence with FoldX. Use the template 1sem, chains A and C.

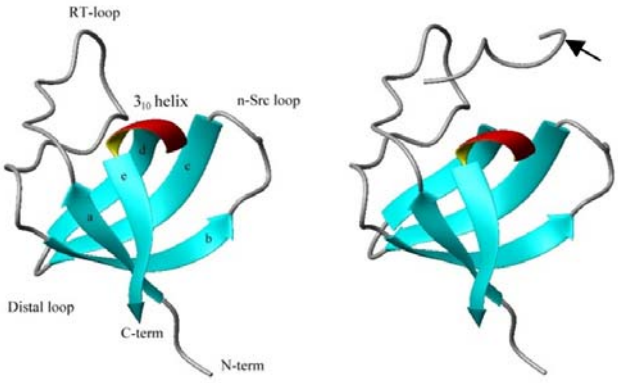
```
>protein
RRVRALYDLTTNEPDELSFRKGDVITVLEQVYRDWWKGALRGNMGIFPLNYVTPI
```

Clean the **1sem.pdb** template by deleting **chains B** and **D**, and save the pdb file as "**1sem2.pdb**". This is your template. Perform a sequence alignment to prepare the "**mutant_file.txt**" correctly, as explained above, and run the program FoldX. This is the template to use. Prepare the "**mutant_file.txt**" correctly, and run the program FoldX as explained above.
Once the calculations have finished, prepare the following files to send to the server: Rename your "**template.txt**" file to **P08-08&templ.txt**; rename your "**mutant_file.txt**" to **P08-08&mutant.txt**; rename the output model to **P08-08&model.pdb**. Annotate comments you consider interesting in your **report file**.

Selecting templates for homology modeling

Most of the times, the selection of the correct template is much more complicated than the modeling itself. In this series of exercises there are several examples of template selection for standard SH3 domain. All features commented in this session are applicable to any domain, soluble or membrane protein, complexes, etc., but taking into account the particular features (structure, activity, cofactors, etc.) of each domain or protein.

SH3 domain description



Size. SH3 (src Homology-3) domains are small protein modules containing approximately 50 amino acid residues

Location. They are found in a great variety of intracellular or membrane-associated proteins, for example, in a variety of proteins with enzymatic activity, in adaptor proteins that lack catalytic sequences and in cytoskeletal proteins.

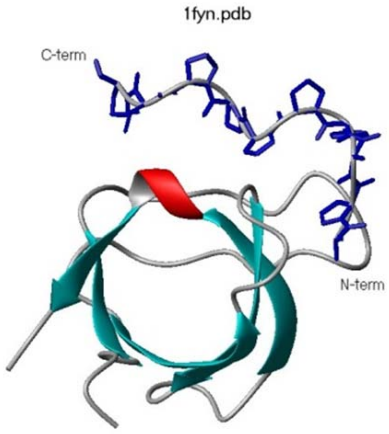
Activity. SH3 domains bind to target proteins through sequences containing proline and hydrophobic amino acids. Pro-containing polypeptides may bind to SH3 domains in 2 different binding orientations. The ligand binds with low affinity but this may be enhanced by multiple interactions. The region bound by the SH3 domain is in all cases proline-rich and contains PXXP as a core-conserved binding motif.

Function. Is not well understood but they may mediate many diverse processes such as increasing local concentration of proteins, altering their subcellular location and mediating the assembly of large multiprotein complexes.

Folding. The SH3 domain has a characteristic fold which consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices. The surface of the SH3-domain bears a flat, hydrophobic ligand-binding pocket which consists of three shallow grooves defined by conservative aromatic residues in which the ligand adopts an extended left-handed helical arrangement. The binding pocket is formed between the RT-loop and the n-Src loop.

SH3 domain in the laboratory of Molecular Biology

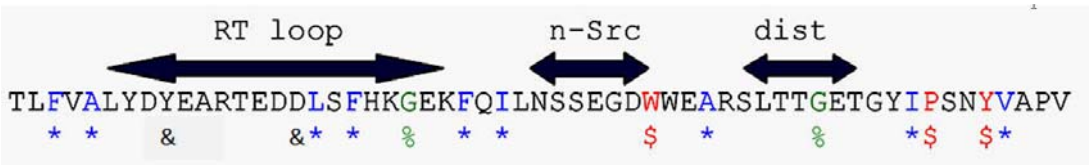
The two-hybrid experiments done with our protein of interest gave several putative interacting proteins. All these proteins have at least one SH3 domain. It would be quite interesting to model the SH3 domains and propose a mechanism of interaction.

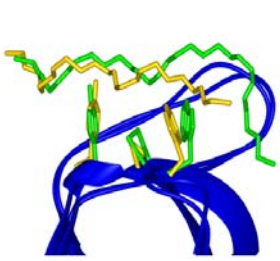


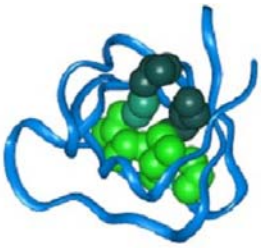
To model a protein, first we need to look for a template. This can be done using different strategies, tools and web servers. Here we propose the use of NCBI BLAST at EBI, the SWISS-MODEL Workspace, and FoldX.

There are a lot of SH3 crystals in the protein structure databases, and so, it is expected to find a suitable template for the proposed SH3 domains.

In addition we have detailed information on SH3 architecture, motifs, sequence determinants, core positions, loop length, binding pocket properties, etc. This information can be used as rules to select the appropriate template. In the figure there is depicted a scheme of a typical SH3 domain from a proto-oncogene tyrosine kinase (1fyn.pdb).







* The arrows show the length of the 3 loops in the SH3 domain of 1fyn (our reference).

* There are no insertion or deletions in the loops forming the binding pocket (RT and n-Src). The distal loop can be modeled if necessary. This will be important in the alignments with others SH3 domains.

* The three conserved residues in the binding pocket (WPY, marked with a $ in red) should match in the alignment. Some weird SH3 domains are WPF or WPR.

* The hydrophobic core residues (marked with an asterisk, in blue) should preserve hydrophobicity. No options for putative SH3 having polar or charged residues in these positions.

* Conserved Gly positions (marked with a % in green). The second one is quite conserved.

* Aromatic and polar motif conservation in the RT-loop (marked with an & in black). This is to refine selection in the SH3 domains.

To proceed with the template search, copy the query sequence and paste it in NCBI BLAST at EBI.

1. Step 1. In PROTEIN DATABASE select "only" Structure // Protein Structure Sequences. This will search for homologue proteins which structures already known. Other options remain as default.
2. Step 2. Paste query sequence.
3. Step 3. Select blastp.
4. Run BLAST.
5. Check the E value (expectation value, see glossary) at the top of the list. The lower the E value, the more significant the score.
6. Check the percentage of identity, similarity and gaps.
7. Show the alignments of the top matches (Tool Output) and look at the rules mentioned above for SH3 domains. Consider those alignments that span as much sequence as possible.
8. DECIDE which template will be used for modeling.
9. X-Ray better than NMR. Between X-Ray templates, choose the one with better resolution.

Read and learn the strategies to select templates with these two examples.

Example 1.

```
>SH3-example1
FLAKVVHPFDAQAPGELSLAVDDYVIVRQVAGTGWSEGEYKGKAGWFPSAYVEKQE
```

Solution: The following figure is part of the solutions. Structure **2drm** fulfills most of the requirements: three conserved residues (**WPY**), no insertions and deletions. A **Gly** is conserved, etc. etc.

```
 pdb|2DRM|D  Chain D, Acanthamoeba Myosin I Sh3 Domain Bound To Acan125
Length=58

 Score = 46.6 bits (109),  Expect = 2e-08, Method: Compositional matrix adjust.
 Identities = 22/50 (44%), Positives = 29/50 (58%), Gaps = 0/50 (0%)

Query  4   KVVHPFDAQAPGELSLAVDDYVIVRQVAGTGWSEGEYKGKAGWFPSAYVE  53
           K ++ +DAQ   EL+    D +IV Q    GW EGE  GK GW P+ YV+
Sbjct  7   KALYDYDAQTGDELTFKEGDTIIVHQKDPAGWWEGELNGKRGWVPANYVQ  56


>pdb|2DRK|A  Chain A, Acanthamoeba Myosin I Sh3 Domain Bound To Acan125
Length=59

 Score = 46.6 bits (109),  Expect = 2e-08, Method: Compositional matrix adjust.
 Identities = 22/50 (44%), Positives = 29/50 (58%), Gaps = 0/50 (0%)

Query  4   KVVHPFDAQAPGELSLAVDDYVIVRQVAGTGWSEGEYKGKAGWFPSAYVE  53
           K ++ +DAQ   EL+    D +IV Q    GW EGE  GK GW P+ YV+
Sbjct  8   KALYDYDAQTGDELTFKEGDTIIVHQKDPAGWWEGELNGKRGWVPANYVQ  57


>pdb|2NWM|A  Chain A, Solution Structure Of The First Sh3 Domain Of Human
Vinexin And Its Interaction With The Peptides From Vinculin
Length=65

 Score = 46.2 bits (108),  Expect = 3e-08, Method: Compositional matrix adjust.
 Identities = 21/51 (41%), Positives = 30/51 (59%), Gaps = 0/51 (0%)

Query  3   AKVVHPFDAQAPGELSLAVDDYVIVRQVAGTGWSEGEYKGKAGWFPSAYVE  53
           A++   F AQ+P EL+L   D V + +     W EGE+ G+ G FP+ YVE
Sbjct  4   ARLKFDFQAQSPKELTLQKGDIVYIHKEVDKNWLEGEHHGRLGIFPANYVE  54
```

2drm is X-Ray with 1.35 Å; **2drk** is 1.42 Å; **2nwm** is NMR. We will select **2drm**, having the best scores, and best resolutions in X-Ray.

Example 2.

```
>SH3-example2
RFGTALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGRDGKMAGLVPVLYVNQS
```

Solution: The top score has an insertion (5 aa) in the distal loop. This is not a problem since the long loop can be easily modeled with no relevance for the activity of the protein. However, the deletion in the n-Src loop is a real problem since we are close to the binding pocket, and we do not know the exact conformation of the loop for activity.

```
>pdb|2DIL|A  Chain A, Solution Structure Of The Sh3 Domain Of The Human Proline-
Serine-Threonine Phosphatase-Interacting Protein 1
Length=69

 Score = 34.7 bits (78),  Expect = 5e-04, Method: Compositional matrix adjust.
 Identities = 24/56 (43%), Positives = 35/56 (63%), Gaps = 7/56 (13%)

Query  5   ALYDFTAGGDDELNLTAEEELEIEYE-VDGWFYVKKKRPGRDGKMAGLVPVLYVNQ  59
           ALYD+TA   DEL+L+A + LE+  E  D W+ V+     R+G+  G VP  Y+ +
Sbjct  13  ALYDYTAQNPDELDLSAGDILEVILEGEDGWWTVE-----RNGQR-GFVPGSYLEK  62
```

In this case, we should look for any other template. Look at these alignments:

```
>pdb|1WXB|A  Chain A, Solution Structure Of The Sh3 Domain From Human Epidermal
Growth Factor Receptor Pathway Substrate 8-Like Protein
Length=68

 Score = 28.1 bits (61),  Expect = 0.093, Method: Compositional matrix adjust.
 Identities = 18/49 (37%), Positives = 28/49 (57%), Gaps = 5/49 (10%)

Query  6   LYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGRDGKMAGLVPV  54
           LYDFTA   +EL++  +E LE+  +   W+ ++     R G+ AG VP
Sbjct  13  LYDFTARNANELSVLKDEVLEVLEDGRQWWKLRS----RSGQ-AGYVPC  56
```

```
>pdb|2JXB|A Chain A, Structure Of Cd3epsilon-Nck2 First Sh3 Domain Complex
Length=86

 Score = 26.9 bits (58),  Expect = 0.43, Method: Compositional matrix adjust.
 Identities = 16/53 (30%), Positives = 24/53 (45%), Gaps = 5/53 (9%)

Query  5   ALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGRDGKMAGLVPVLYV  57
           A +D+TA  D EL++   E L +  +   W+ V+           G VP  YV
Sbjct  36  AKWDYTAQQDQELDIKKNERLWLLDDSKTWWRVRNA-----ANRTGYVPSNYV  83


>pdb|2B86|A Chain A, Solution Structure Of The First Src Homology 3 Domain
Of Nck2
Length=67

 Score = 26.2 bits (56),  Expect = 0.51, Method: Compositional matrix adjust.
 Identities = 16/53 (30%), Positives = 24/53 (45%), Gaps = 5/53 (9%)

Query  5   ALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGRDGKMAGLVPVLYV  57
           A +D+TA  D EL++   E L +  +   W+ V+           G VP  YV
Sbjct  9   AKWDYTAQQDQELDIKKNERLWLLDDSKTWWRVRNA-----ANRTGYVPSNYV  56


 pdb|1I0C|B Chain B, Eps8 Sh3 Closed Monomer
Length=60

 Score = 25.8 bits (55),  Expect = 0.73, Method: Compositional matrix adjust.
 Identities = 11/38 (29%), Positives = 21/38 (55%), Gaps = 0/38 (0%)

Query  1   RFGTALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVK  38
           ++  + YDF A    EL++  ++ LEI  +   W+ V+
Sbjct  2   KYAKSKYDFVARNSSELSVMKDDVLEILDDRRQWWKVR  39


>pdb|1GRI|A Chain A, Grb2
 pdb|1GRI|B Chain B, Grb2
Length=217

 Score = 25.4 bits (54),  Expect = 2.5, Method: Composition-based stats.
 Identities = 15/42 (36%), Positives = 22/42 (52%), Gaps = 0/42 (0%)

Query  5   ALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGRDG  46
           A YDF A  DDEL+    + L++  E     + K +  G+DG
Sbjct  5   AKYDFKATADDELSFKRGDILKVLNEECDQNWYKAELNGKDG  46


>pdb|3HAJ|A Chain A, Crystal Structure Of Human Pacsin2 F-Bar Domain (P212121
Lattice)
 pdb|3HAJ|B Chain B, Crystal Structure Of Human Pacsin2 F-Bar Domain (P212121
Lattice)
Length=486

 Score = 24.6 bits (52),  Expect = 5.9, Method: Compositional matrix adjust.
 Identities = 20/54 (37%), Positives = 25/54 (46%), Gaps = 5/54 (9%)

Query  5    ALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGR-DGKMAGLVPVLYV  57
            ALYD+     DEL+  A +EL    + D   + K    GR D    GL P  YV
Sbjct  433  ALYDYEGQEHDELSFKAGDELTKMEDEDEQGWCK----GRLDNGQVGLYPANYV  482
```

Note that:

- The "Score" (the higher, the better) is lower than the top score, as well as the "Identities" and "Positives". In contrast, the gap in the n-Src loop disappears ("Gaps" are improved; the lower, the better).
- Our selection could be: template **1wxb_A** (chain A), **2jxb_A**, **2b86_A**, **1i0c_B**, **1gri_A** or **3haj_A**. (Note that other structures could be selected).
- Among these, which structures should be selected? X-Ray better than NMR. Then, in principle, look at the **1i0c_B**, **1gri_A** or **3haj_A**.
- And now? Better resolution is the best. Select **1i0c** (2 Å) instead of **3haj** (2.78 Å) and **1gri** (3.1 Å).
- ***ATTENTION***: **1gri** and **3haj** are not correctly aligned. The alignment is bad and **should be never selected** (Check the W in the conserved position). This was shown for pedagogical purposes.
- Once we have a putative template, make a pairwise alignment:

```
SH3-example2      -RFGTALYDFTAGGDDELNLTAEEELEIEYEVDGWFYVKKKRPGRDGKMAG-LVPVLYVNQS 60
1i0c              KKYAKSKYDFVARNSSELSVMKDDVLEILDDRRQWWKVRN-ASGDSGFVPNNILDIMRTP-- 59
                   ::..: ***.* ...**.:  :: ***  :   *: *::  .* .* :.. :: :: .
```

Here we are missing the two other residues (P and I) in the binding pocket!! **Discard this structure**. In fact this structure should be discarded previously since no proper alignment was presented in the NCBI page. This was shown for pedagogical purposes.

Going back to the NMR structures, **2jxb** and **2b86** are identical, so we select, in principle, **2jxb**. Between **1wxb** and **2jxb**, we select **2jxb** because it spans more sequence, including the three **WPY** forming the binding pocket. The insertion in the distal loop will be modeled later. Now it is possible to download the structure, clean for other extra molecules and model by using any modeling program or server.

IMPORTANT LESSON: This selection is for SH3 domains, and for this particular sequence. There is no automatic way to do this. Criteria can change for other domains and/or sequences. The selections can be done by the modeler by using a combination of experience, common sense, and the few rules exposed above.

P08-09) Work on this task: Answer the following question regarding the sequence SH3-1.

```
>SH3-1
YKAKALYPYDADDDDAYEISFEQNEILQVSDIEGRWWKARRANGETGIIPSNYVQLI
```

Annotate the results in your **report file** and answer the question in the task web page "Cuestionarios ON-LINE":

P08-09. ¿Qué estructura usarías como plantilla para modelar la secuencia SH3-1?

- ☐ No hay forma de modelar esta secuencia.
- ☐ La mejor plantilla es 2vkn. Se debería usar esta estructura para modelar la secuencia.
- ☐ La plantilla 3ua6 es suficientemente buena para modelar la secuencia.
- ☐ La plantilla 2lqw tiene una pequeña inserción, pero el modelado es todavía posible.
- ☐ No hace falta modelado alguno. Ya tenemos disponible la estructura en el PDB.

P08-10) Work on this task: Answer the following question regarding the sequence SH3-2.

```
>SH3-2
EYVEALYDFEAQQDGDLSLKTGDKIQVLEKISPDWYRGKSNNKIGIFPANYVKPA
```

Annotate the results in your **report file** and answer the question in the task web page "Cuestionarios ON-LINE":

P08-10. ¿Qué estructura usarías como plantilla para modelar la secuencia SH3-2?

- ☐ Las plantillas 1k76 y 1kfz son muy buenas.
- ☐ Las plantillas 1io6, 1gfc y 1gfd no presentan inserciones ni delecciones.
- ☐ Aunque hay muchas estructuras que servirían de plantilla, la mejor es 2d0n.
- ☐ La plantilla 1gcq tiene peor valor-E, pero las coincidencias en secuencia son grandes.
- ☐ Las plantillas 1ynz y 1zx6 son las que mejor valor-E tienen y las que se ajustan mejor.

P08-11) Work on this task: Answer the following question regarding the sequence SH3-3.

```
>SH3-3
RRVRALYDLTTNEPDELSFRKGDVITVLEQVYRDWWKGALRGNMGIFPLNYVTPI
```

Annotate the results in your **report file** answer the question in the task web page "Cuestionarios ON-LINE":

P08-11. ¿Qué estructura usarías como plantilla para modelar la secuencia SH3-3?

- ☐ Las plantillas 2vvk y 1io6 se podrían usar indistintamente.
- ☐ La plantilla 1uti se ha de utilizar antes que la de 2vvk.
- ☐ La plantilla 1io6 es la preferida por ser de rayos X y mayor resolución.
- ☐ La plantilla 1h3h abarca más secuencia hacia N-terminal y por lo tanto es al que elegiríamos.
- ☐ La plantilla 2vvk es la mejor porque es de rayos X y tiene el valor-E mas pequeño.

P08-12) Work on this task: Model the sequence SH3-3 with the template **2d0n** locally (FoldX).

Rename your "**template.txt**" file to **P08-12&templ.txt**; rename your "**mutant_file.txt**" to **P08-12&mutant.txt**; rename the output model to **P08-12&model.pdb**. Annotate comments you consider interesting in your report file.

P08-13) Work on this task: Answer the following question regarding the sequence SH3-4.

```
>SH3-4
GIYRAVYAYEPQTPEELAIQEDDLLYLLQKSDIDDWWTVKKRVIGSDSEEPVGLVPSTYIEEA
```

Annotate the results in your **report file** and answer the question in the task web page "Cuestionarios ON-LINE":

P08-13. ¿Qué estructura usarías como plantilla para modelar la secuencia SH3-4?

- ☐ Los mejores son 3ehr o 1zlm, con una inserción en el loop distal.
- ☐ El mejor valor-E es 1x2k, y es la plantilla adecuada para el modelado.
- ☐ La 2cdt tiene una inserción mínima y es una buena alternativa.
- ☐ La 1shg presenta solo una inserción en el loop distal.
- ☐ La mejor es 2dl5, aunque no se encuentre entre las primeras.

P08-14) (OPTIONAL) Work on this task: Try to model the sequence SH3-4 with the template of your choice, in Swiss-Model Workspace Server, in project mode. If your selection is a multiple NMR structure, calculate and use the average structure.

Send the resulting Deepview project to the server, named as **P08-14&project.pdb**

Average states in pymol (help for P08-14)

- launch PyMOL as usual.
- load your NMR structure with several states as usual.
- import the "average3d.py" module by using File/Run facility
- write the function avgStates() in the PyMol prompt (PyMOL>)to average the conformers, calculate RMSDs, etc.
- save new averaged data (usually 2nd object) to a pdb file.
- you may also type help(avgStates) for info about the syntax of the avgStates() function.

Última actualización 27/11/2016

Arriba Universidad Miguel Hern&#65533;ndez de Elche
Dr. Gregorio Fernández-Ballester and Dr. José Antonio Encinar
Instituto de Biología Molecular y Celular
Telf. + 34 96-665 8441 Fax: + 34 96 665 8758
email: gregorio@umh.es and jant.encinar@umh.es