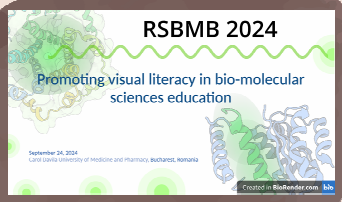


Hands-on Workshop
Tuesday,
24th Sept. 2024,
15:30-17:30

Incorporating molecular visualisation into the teaching and learning practice

Angel Herráez, PhD

Prof. of Biochemistry & Molecular Biology,
Dept. of Systems Biology, University of Alcalá.
Alcalá de Henares, Spain.

Member, FEBS Education and Training Committee

Universidad de Alcalá

Plan

Links we will use:
www.bit.ly/2024molvis

1. Come to class with a molecule in your pocket
2. Get the molecular model files
3. Use Proteopedia as a ready-made platform to support teaching
4. Try some 3D structure prediction
5. Author your own content in Proteopedia

1 Come to class with a molecule in your pocket

(demo)

- 1FCV (PDB)
  Bee venom hyaluronidase
  in complex with
  hyaluronic acid tetramer
- lactose (PubChem)
- Others

2 OK, but… Where do I get the molecule?

Macromolecules: proteins, nucleic acids, complexes:

PDB (Protein Data Bank)

`pdb.org` ➔ `rcsb.org` (USA) , `pdbe.org` (Europe)

"Small" molecules:

Chemical Identifier Resolver "CACTUS", NCI (NIH)

`https://cactus.nci.nih.gov/chemical/structure`

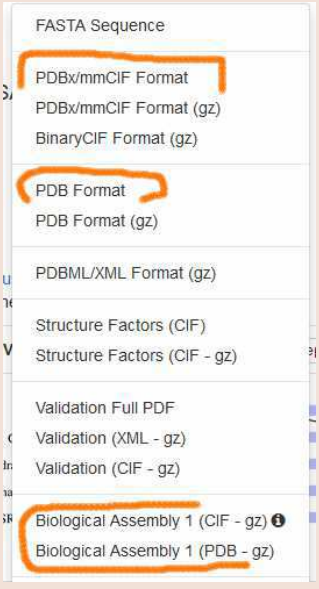
PubChem `pubchem.org`

Databases: PDB @ USA (RCSB)

https://rcsb.org/structure/8PJA

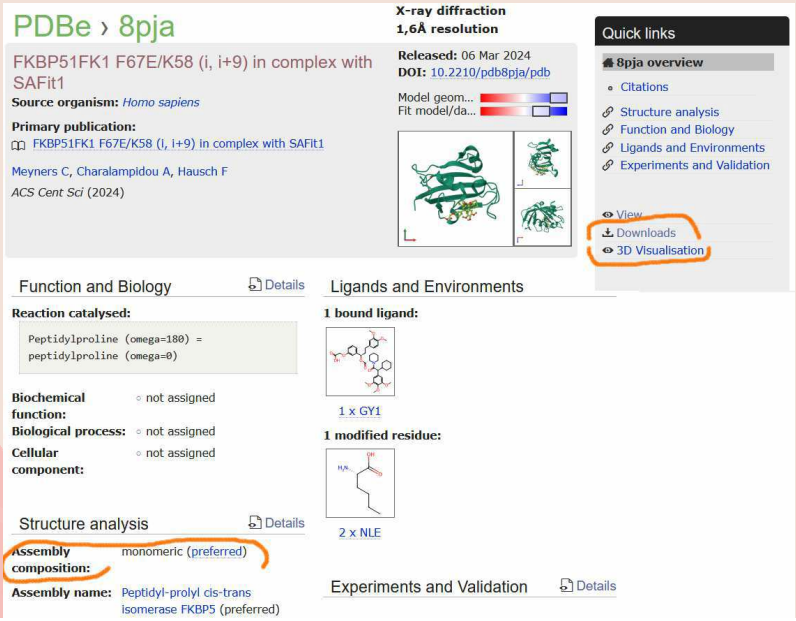
Databases: PDB @ USA (RCSB)

https://rcsb.org/structure/8PJA



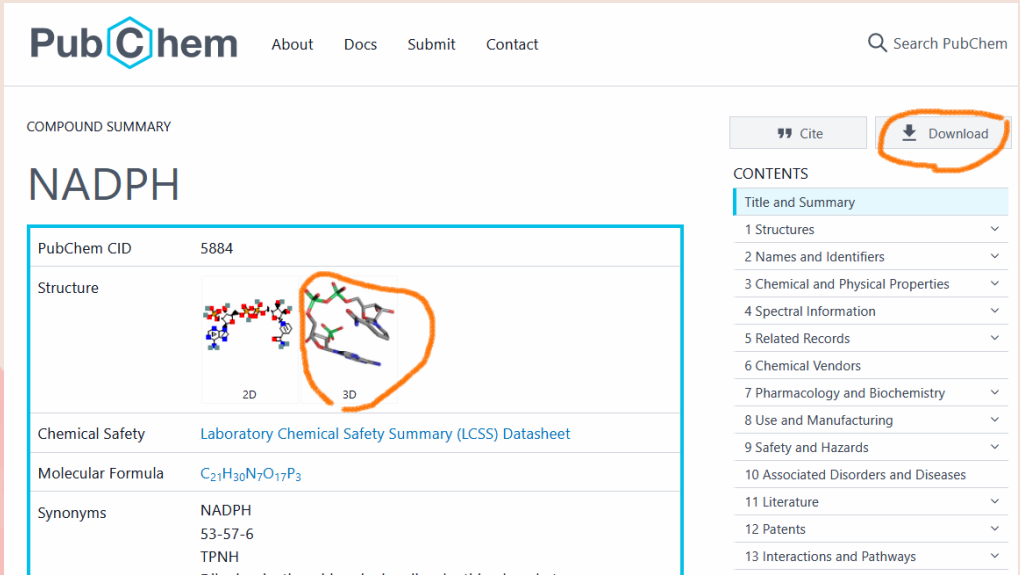
Databases: PDB @ Europe (PDBe)

https://pdbe.org/8PJA

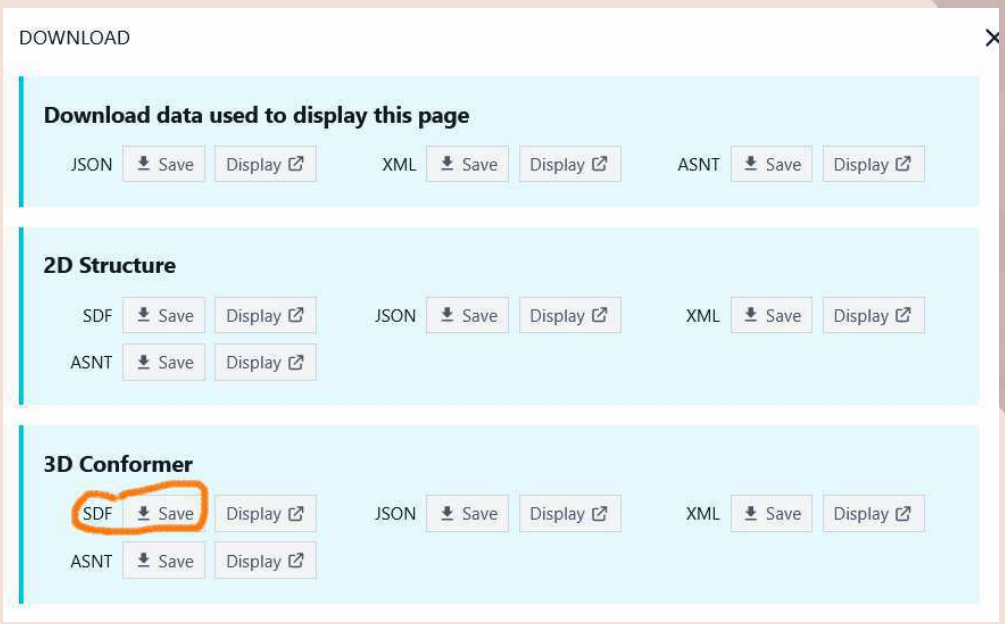


Databases: PubChem

https://pubchem.org ➔ https://pubchem.ncbi.nlm.nih.gov
Search for: NADPH ➔ https://pubchem.ncbi.nlm.nih.gov/compound/5884



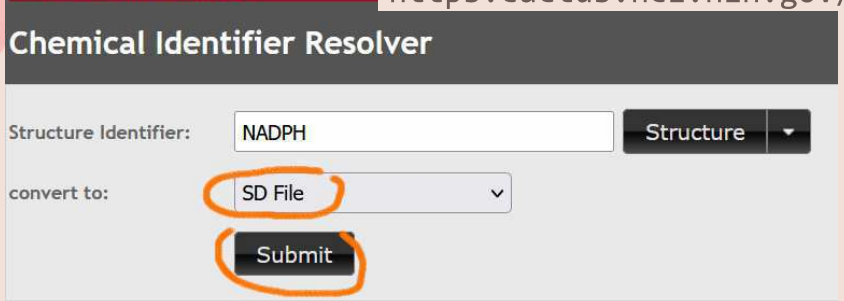
Databases: PubChem



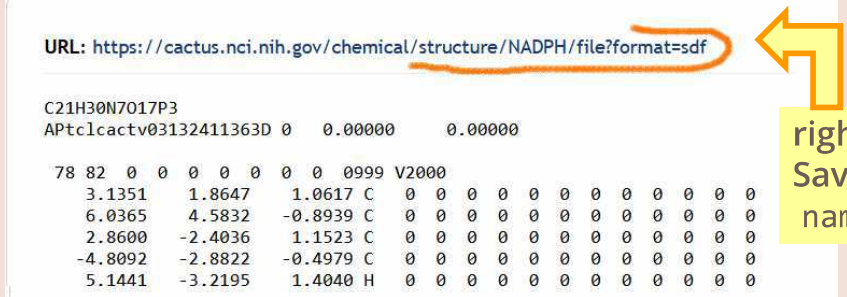
Server: CACTUS

https://cactus.nci.ih.gov

https:cactus.nci.nih.gov/chemical/structure



by
name
other ID
SMILES



right-click
Save link as...
name.sdf

Database connection: directly from Jmol

- from PDB: top menu > File > Get PDB > `8PJA`
- from PDB: console > `load =8PJA`
- from PDBe: console > `load *8PJA`
- from PubChem: console > `load :NADP`
- from CACTUS: top menu > File > Get MOL
- from CACTUS: console > `load $NADP`

RCSB PDB	RCSB Ligands	EBI PDB Europe	NCI (Cactus)	PubChem	CrystOD	AMCSD	Materials Project
=	==	*	$	:			
=pdb/	=ligand/	=pdbe/	=nci/	=pubchem/	=cod/	=ams/	=mp/

https://wiki.jmol.org/index.php/Database_Connection

1b

So, now, let's do it yourselves

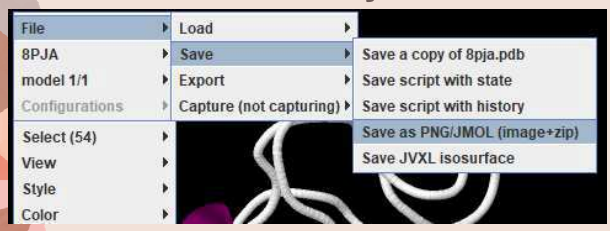
1. Locate the molecular file you downloaded
2. Open your copy of the Jmol application (Jmol.jar file, will open using Java)
3. top menu > File > Open
4. Adjust the display / rendering a right-click opens the pop-up menu
   - Select > Protein > all
   - Style > Scheme > Cartoon
   - Select > Hetero > All PDB "HETATM"
   - Style > Scheme > Ball & stick
   - Select > Hetero > All solvent
   - Style > Atoms > Off

how to save to disk...

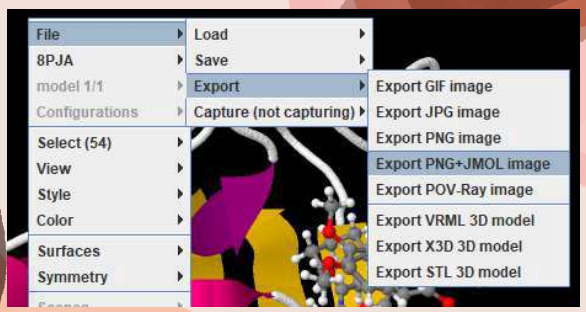
Do you like it? Save it! and do it often

1. Choose the view (rotate and zoom with the mouse)

2. right-click
   > File
   > Save
   > as PNG/Jmol



alternative:



This PNG file displays as an image in any software

but can be opened or dragged & dropped into Jmol to recover the saved scene (the "state")

3 Use Proteopedia as a ready-made platform

Proteopedia holds a copy of all files in PDB and presents them simplified and focused ("seeded pages")

https://proteopedia.org/w/8pja (lowercase PDB ID!)

You can still use the pop-up menu (right-click in the Jmol panel) to change rendering

There are also "authored pages"

enter a name and click "Search" button

in the results, go to bottom and check "only Human created pages" and click on "Search" button

Examples: https://proteopedia.org/w/Aldehyde_Dehydrogenase
https://proteopedia.org/w/Nitrogenase
Proteopedia:Table of Contents

Summary of Proteopedia capabilities

text + images + videos
(like in Wikipedia)

+ 3D interactive molecular structures
"scenes"

Author collaboration & recognition

"green links"
act on the 3D model

SAT = Scene Authoring Tool
builds scenes without programming, without Jmol commands / scripting language

Version history of both content and scenes

4 3D structure prediction

From protein sequence to tertiary structure

Databases:

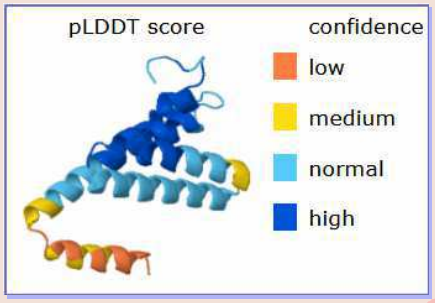
AlphaFoldDB https://alphafold.ebi.ac.uk
open access to over 200 million protein structure predictions

ESM Metagenomic Atlas https://esmatlas.com
open atlas of 772 million predicted protein structures

Predict from sequence:

AlphaFold 2 (2020) → AlphaFold DataBase
AlphaFold 3 ← AlphaFold Server
ESM "Fold Sequence"
Proteopedia using ESMfold

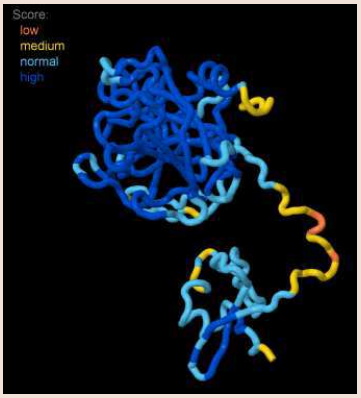
more about AlphaFold

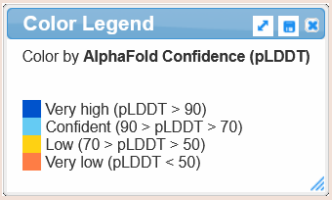


3D structure prediction

Advice:

address what the students can learn from testing prediction, rather than mere knowledge of the possible structure.



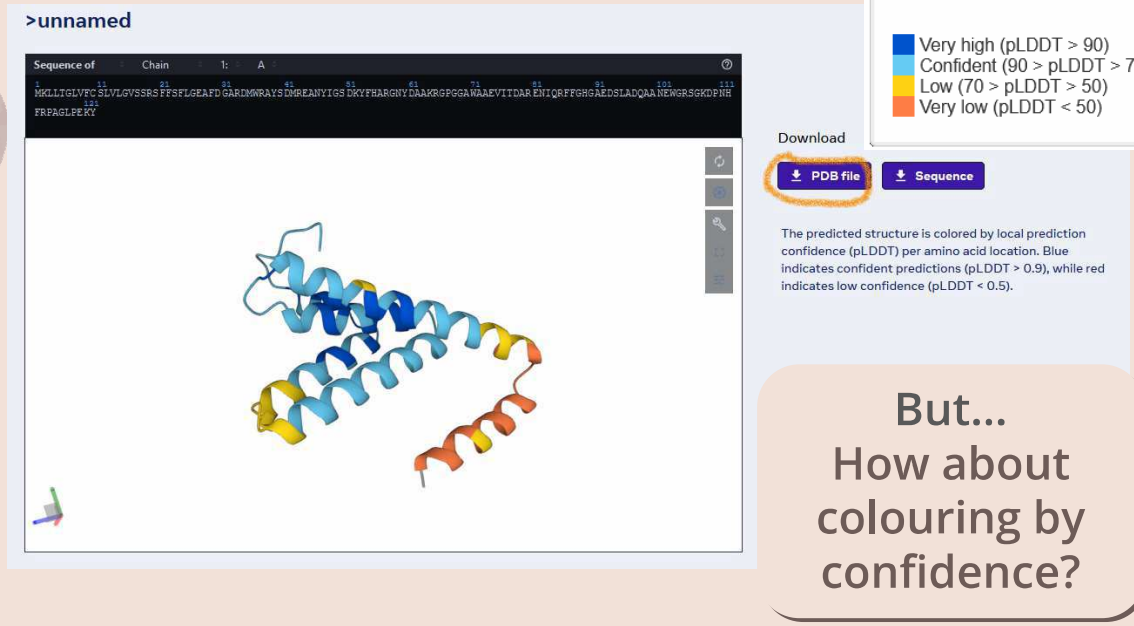


ESM prediction for urokinase-type plasminogen activator isoform 3

Prediction of folding (in NCBI + ESMFold)

1. Search for: "NCBI protein"
   NCBI.NLM.NIH.gov/protein/
   - There, search for, e.g.: Serum amyloid A-1
   - copy the sequence (FASTA link is the quickest way)
     MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYSDMRE
     ANYIGSDKYFHARGNYDAAKRGPGGAWAAEVITDARENIQRFFG
     HGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
2. Go to "ESM Fold sequence"
   esmatlas.com/resources?action=fold
   - paste your sequence (remove any newlines and ending spaces)
     and press Enter (or click the magnifying glass)

Saving results of ESMFold



But…
How about colouring by confidence?

Colouring ESMfold results by confidence

The download from ESMfold contains the pLDDT values of each atom as “B-factor” values in the pdb format

… but that’s not very useful as it is

Similarly for colouring results from the AlphaFold DB

Auxiliary tool:

`https://biomodel.uah.es/Jmol/fold_AI/`

- Apply colouring by pLDDT score of each alpha carbon
- Apply colouring by average pLDDT score of all atoms within each residue
- Save PNG image or PNGJ image reusable by Jmol
- Save to a new pdb file with pLDDT values converted to (temperature) B-factor values (for use in other programs)
- Extract pLDDT values to a text file

What if we “mutate” the sequence?

original sequence
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGS
LQKRGIVEQCCTSICSLYQLENYCN
modified (“mutated”)
MALWMRLLPLLALLALWGPDPAAAFVNQHLSGSHLVEALYLVSGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGS
LQKRGIVEQCCTSISSLYQLENYSN
L11E + A14D
MALWMRLLPLEALLDLWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGS
LQKRGIVEQCCTSICSLYQLENYCN

serum amyloid A1 [Homo sapiens] GenBank: AAA64799.1
MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYSDMREANYIGSDKYFHARGNYDAAKRGPGGVWAAEAISDARENIQRF
FGHGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDFWRFYSDMREANYIGSDKYFHARFNYDAAKRGPGGVWAAEAISDARENIQRFFGHG
AEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYSDMREANYIGSDKYFHARGNYDAAKRGPGGVWAAEPISDARENIQRFFGH
GAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
>NP_001306120.2 urokinase-type plasminogen activator isoform 3 [Homo sapiens]
MGRPCLPWNSATVLQQTYHAHRSDALQLGLGKHNYCRNPDNRRRPWCYVQVGLKLLVQECMVHDCADGKKPSSPPEELKFQCG
QKTLRPRFKIIGGEFTTIENQPWFAAIYRRHRGGSVTYVCGGSLISPCWVISATHCFIDYPKKEDYIVYLGRSRLNSNTQGEMKFEVEN
LILHKDYSADTLAHHNDIALLKIRSKEGRCAQPSRTIQTICLPSMYNDPQFGTSCEITGFGKENSTDYLYPEQLKMTVVKLISHRECQQ
PHYYGSEVTTKMLCAADPQWKTDSCQGDSGGPLVCSLQGRMTLTGIVSWGRGCALKDKPGVYTRVSHFLPWIRSHTKEENGLAL
>AAA52075.1 C-reactive protein [Homo sapiens]
MEKLLCFLVLTSLSHAFGQTDMSRKAFVFPKESDTSYVSLKAPLTKPLKAFTVCLHFYTELSSTRGYSIFSYATKRQDNEILIFWSKDIGY
SFTVGGSEILFEVPEVTVAPVHICTSWESASGIVEFWVDGKPRVRKSLKKGYTVGAEASIILGQEQDSFGGNFEGSQSLVGDIGNVNM
WDFVLSPDEINTIYLGGPFSPNVLNWRALKYEVQGEVFTKPQLWP
-

5 Authoring Proteopedia

Engage your students into protein structure by telling a story,
either with examples already available
or writing your own pages

e.g.:

- Why is carbon monoxide so dangerous to breathe?
- How do Tamiflu and Relenza work as antiviral medication to and why do they sometimes fail?
- How do HIV drugs work to stop AIDS infection?
- How to design a human protein that can be expressed in bacteria?
- How does a repressor protein bind to a particular region of DNA?

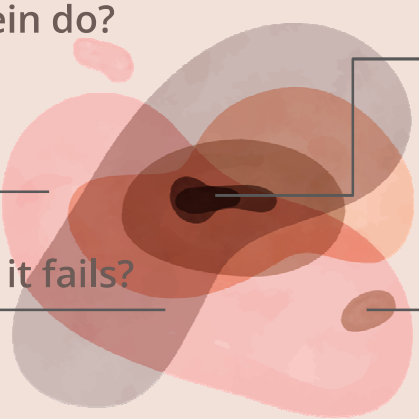
Credits: Jaime Prilusky, Joel Sussman

Advice for designing a case study

Pick an interesting protein

What does the protein do?
function

How does it happen?
structure



What happens when it fails?
disease

Are there any remedies?
drugs

(any newly created page in Proteopedia will have a template like these points)

Proteopedia Take Home Guidelines (pdf)

Credits: Jaime Prilusky, Joel Sussman

Your first page in Proteopedia

New pages are not empty, but filled with some template content:
Sections and text.
3D panel on the right.

You can use it ☺ or
delete everything and start from scratch ☹

Let's practice! Authoring Proteopedia

Go to proteopedia.org

> Log in Username: student Password:

To create a page:

Type the name of your page (see at the right)

Sandbox_RSBMB_00

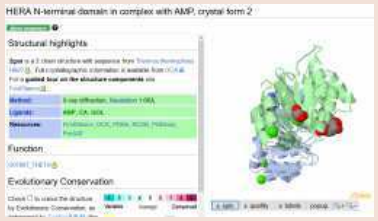
Press `Go´ and click on the red link (non-existing page)

> Save page

Credits: Jaime Prilusky

Flow of editing

View page



Edit page

Edit text and other content.

...

...

Show the SAT

...

...

Preview or save page

SAT (edit scenes)

Load molecule or load scene.

...

insert the green link

...

Save scene.

Hide the SAT.

Proteopedia page editing (simple)

If you know Wikipedia, text and content edition is similar.

Easy way to insert a 3D model:

- Click the green panels ╬ button (recommended) (toolbar above the textbox)
- Or click the **3D** button for an isolated Jmol.
- Write your **PDB code** between the quotes in the load parameter of the Structure tag.
- **Save** by clicking on the Save page button.
- You should see your molecule and text.
- Improve by editing options of the tag.

Proteopedia page editing (advanced)

For more control on the model style, create some scenes:

While editing, click on the link to **show** the *Scene authoring tools*.

Use the **load molecule** tab and enter your PDB code.

Customize your scene using the dfferent tabs, menus, buttons

Select a part of the molecule and
apply style of representation, color, labels... as you wish.

Use the **save scene** tab to save your scene.

From the *Wikitext* box, click the button that inserts the code,

or select and copy the **scene tag**, then paste into the main text area for the page (above the *SAT*).

Edit the *TextToBeDisplayed* portion to your custom text.

Hide the *Scene Authoring Tools* and **save** the page.

Try the green link that has been created in the page.

Prediction of folding in Proteopedia

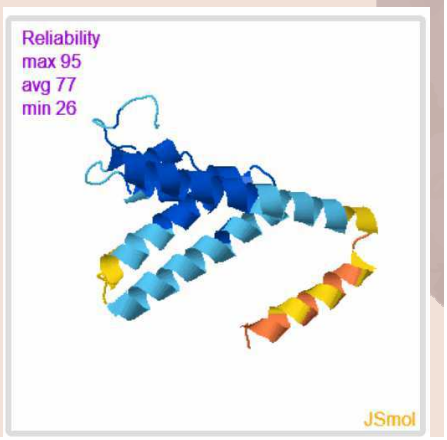
(up) > "Edit this page"

(go down) > SAT (Scene Authoring Tools) [show]

- "load molecule" tab
- "load structure after predicting from sequence"

Type a name (Serum amyloid A-1) and paste the sequence
MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYS
DMREANYIGSDKYFHARGNYDAAKRGPGGAWAAEVITD
ARENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPA
GLPEKY

- Click on "predict"
- ...
- "colors" tab > ESMfold
- 3D panel > right-click
  > File > Save > as PNG/Jmol



pLDDT score:

low, < 50	medium, 50-70	normal, 70-90	high, ≥ 90

Hope you are now enthusiastic about trying all this

Thank you!

Stay in contact:
angel.herraez@uah.es
https://biomodel.uah.es/en/