

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

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

Plan

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>

Structure Summary | Structure | Annotations | Experiment | Sequence | Genome | Ligands | Versions

Biological Assembly 1 ?

Display Files | **Download Files** | Data API

8PJA

FKBP51FK1 F67E/K58 (I, I+9) in complex with SAFit1

PDB DOI: <https://doi.org/10.2210/pdb8PJA/pdb>

Classification: ISOMERASE

Organism(s): Homo sapiens

Mutation(s): Yes

Deposited: 2023-06-23 Released: 2024-03-06

Deposition Author(s): Meyners, C., Charalampidou, A., Hausch, F.

Funding Organization(s): Hessian Ministry of Science, Higher Education and Art (HMWK)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.60 Å

R-Value Free: 0.210

R-Value Work: 0.186

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.219
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		0.8%

Ligand Structure Quality Assessment

Explore in 3D: Structure | Sequence

Annotations | Electron Density | Validation Report | Ligand Interaction (GY1)

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Biological assembly 1 assigned by authors.

Databases: PDB @ USA (RCSB)

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

- FASTA Sequence
- PDBx/mmCIF Format**
- PDBx/mmCIF Format (gz)
- BinaryCIF Format (gz)
- PDB Format**
- PDB Format (gz)
- PDBML/XML Format (gz)
- Structure Factors (CIF)
- Structure Factors (CIF - gz)
- Validation Full PDF
- Validation (XML - gz)
- Validation (CIF - gz)
- Biological Assembly 1 (CIF - gz)**
- Biological Assembly 1 (PDB - gz)

Databases: PDB @ Europe (PDBe)

<https://pdbe.org/8PJA>

PDBe > 8pja

FKBP51FK1 F67E/K58 (I, I+9) in complex with SAFit1

Source organism: *Homo sapiens*

Primary publication:

FKBP51FK1 F67E/K58 (I, I+9) in complex with SAFit1

Meyners C, Charalampidou A, Hausch F
ACS Cent Sci (2024)

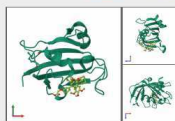
X-ray diffraction
1,6Å resolution

Released: 06 Mar 2024

DOI: 10.2210/pdb8pja/pdb

Model geom...

Fit model/da...



Quick links

8pja overview

- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

- View
- Downloads
- 3D Visualisation

Function and Biology

Details

Ligands and Environments

Reaction catalysed:

Peptidylproline (omega=180) =
peptidylproline (omega=0)

Biochemical function: not assigned

Biological process: not assigned

Cellular component: not assigned

Structure analysis

Details

Assembly composition: monomeric (preferred)

Assembly name: Peptidyl-prolyl cis-trans
isomerase FKBP5 (preferred)

1 bound ligand:



1 x GY1

1 modified residue:



2 x NLE

Experiments and Validation

Details

Downloads

Close

- Archive mmCIF file
- Updated mmCIF file
- PDB file
- PDB header
- PDB file (gz)
- PDBML
- PDBML (ATOM lines)
- PDBML (no atoms)
- Structure Factors
- Assembly composition XML
- Assembly 1 (mmCIF; gz)
- Assembly 1 (atom only; mmCIF)
- FASTA (Entry)
- SIFTS XML file with residue-level mappings
- Summary report (PDF)
- Full report (PDF)
- Percentile plot (PNG)
- Percentile plot (SVG)
- Validation data (XML)

Databases: PubChem

<https://pubchem.org> → <https://pubchem.ncbi.nlm.nih.gov>

Search for: NADPH → <https://pubchem.ncbi.nlm.nih.gov/compound/5884>

PubChem

About Docs Submit Contact

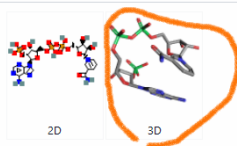
Search PubChem

COMPOUND SUMMARY

NADPH

PubChem CID 5884

Structure



Chemical Safety Laboratory Chemical Safety Summary (LCSS) Datasheet

Molecular Formula C₂₁H₃₀N₇O₁₇P₃

Synonyms NADPH
53-57-6
TPNH

Cite

Download

CONTENTS

Title and Summary

- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Pharmacology and Biochemistry
- 8 Use and Manufacturing
- 9 Safety and Hazards
- 10 Associated Disorders and Diseases
- 11 Literature
- 12 Patents
- 13 Interactions and Pathways

Databases: PubChem

DOWNLOAD

Download data used to display this page

JSON XML ASNT

2D Structure

SDF JSON XML

ASNT

3D Conformer

SDF JSON XML

ASNT

Server: CACTUS

<https://cactus.nci.nih.gov>

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

Chemical Identifier Resolver

Structure Identifier:

convert to:

by
name
other ID
SMILES

URL: <https://cactus.nci.nih.gov/chemical/structure/NADPH/file?format=sdf>

C21H30N7O17P3

APtclcactv03132411363D 0 0.00000 0.00000

78 82 0 0 0 0 0 0 0 0999 V2000
3.1351 1.8647 1.0617 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6.0365 4.5832 -0.8939 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.8600 -2.4036 1.1523 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-4.8092 -2.8822 -0.4979 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5.1441 -3.2195 1.4040 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

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

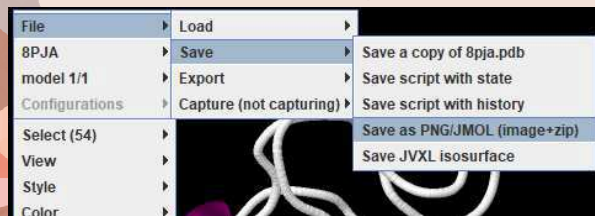
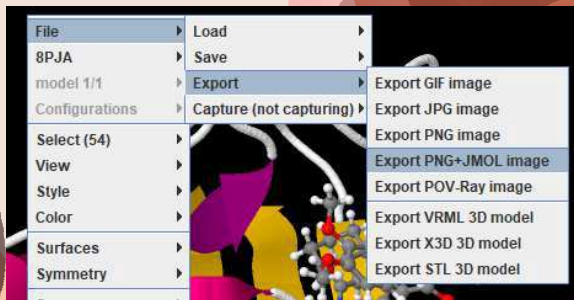
how to save to disk...

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/Aldehyde%20Dehydrogenase)

<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:

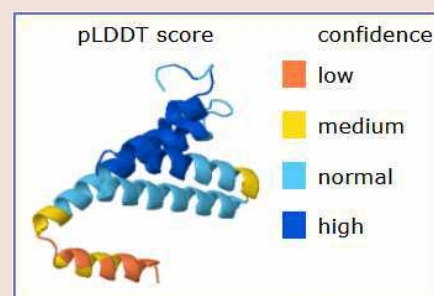
AlphaFold 2 (2020) → AlphaFold DataBase

AlphaFold 3 ← AlphaFold Server

ESM "Fold Sequence"

Proteopedia using ESMfold

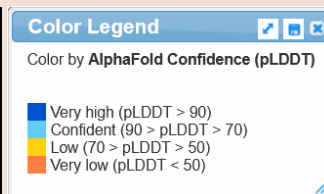
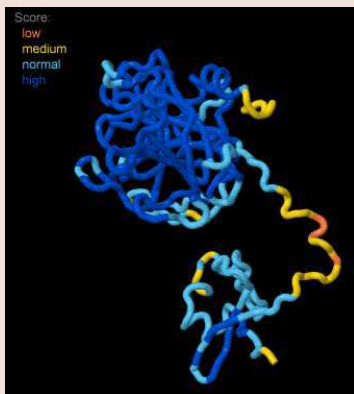
more about → AlphaFold



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/](https://ncbi.nlm.nih.gov/protein/)

- There, search for, e.g.: **Serum amyloid A-1**
- **copy the sequence** (FASTA link is the quickest way)

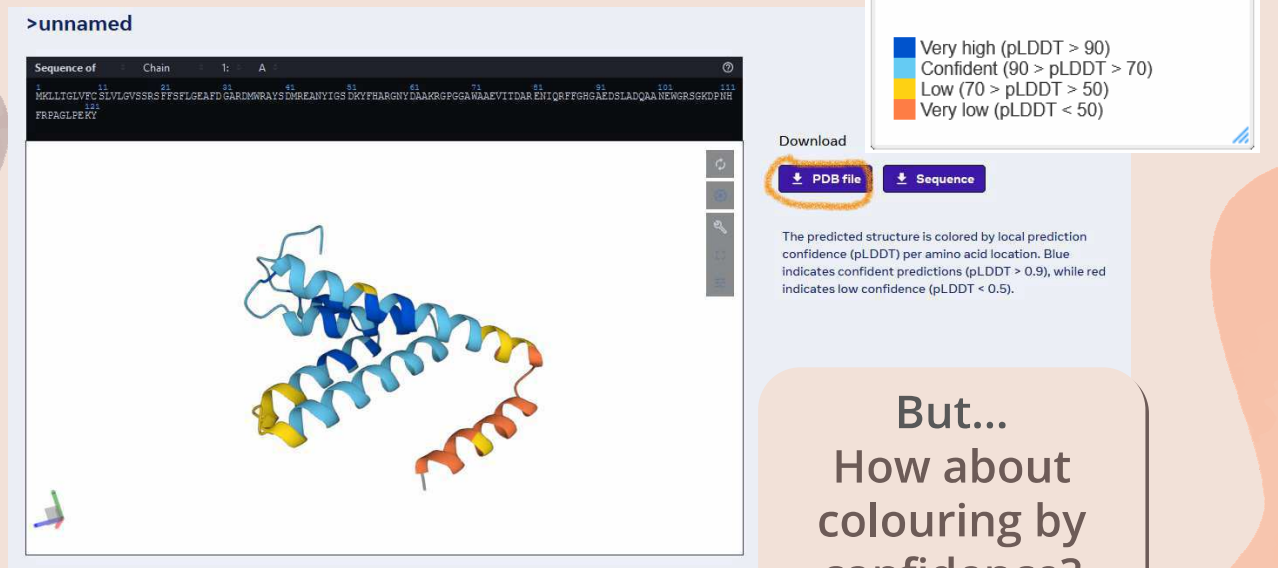
```
MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYSDMRE  
ANYIGSDKYFHARGNYDAAKRGPGGAWAAEVITDARENIQRFFG  
HGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
```

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



Color Legend

Color by **AlphaFold Confidence (pLDDT)**

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

Download

[PDB file](#) [Sequence](#)

The predicted structure is colored by local prediction confidence (pLDDT) per amino acid location. Blue indicates confident predictions (pLDDT > 0.9), while red indicates low confidence (pLDDT < 0.5).

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

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

Similarly for colouring
results from the
AlphaFold DB

What if we “mutate” the sequence?

original sequence

MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI~~C~~SLYQLENY~~C~~N

modified (“mutated”)

MALWMRLLPLLALLALWGPDPAAAFVNQHLSGSHLVEALYLVSGERGFFYTPKTRREAEDLQVQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI~~S~~SLYQLENY~~S~~N

L11E + A14D

MALWMRLLPL~~E~~ALL~~D~~LWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI~~C~~SLYQLENY~~C~~N

serum amyloid A1 [Homo sapiens] GenBank: AAA64799.1

MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGDARMWRAYSMDREANYIGSDKYFHARGNYDAAKRGPGGVWAAEISDARENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY

MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGDAR~~D~~EWRE~~F~~YSMDREANYIGSDKYFHAR~~E~~NYDAAKRGPGGVWAAEISDARENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY

MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGDARMWRAYSMDREANYIGSDKYFHARGNYDAAKRGPGGVWAAE~~P~~ISDARENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY

>NP_001306120.2 urokinase-type plasminogen activator isoform 3 [Homo sapiens]

MGRPCLPWNSATVLQQTTHAHRSDALQLGLGKHNCRNPDRRRPWCYVQVGLKLLVQECMVHDCADGKKPSSPPEELKFQCGQKTLRPRFKIIGGEFTTIENQPWFAAIYRRHRGGSVTYVCGGSLISPCWVISATHCFIDYPKKEDYIVYLGRSRLNSNTQGEMKFEVENLILHKDYSADTLAHHNDIALLKIRSKEGRCAQPSRTIQTICLPSMYNDPQFGTSCFITGFGKENSTDYLYPEQLKMTVVKLISHRECQQPHYYGSEVTTKMLCAADPQWKTDSQCQDSSGGPLVCSLQGRMTLTGIVSWGRGCALKDKPGVYTRVSHFLPWIRSHTEENGLAL

>AAA52075.1 C-reactive protein [Homo sapiens]

MEKLLCFLVLTSLSHAFGQTDMSRKAFVFPKESDTSYVSLKAPLTKPLKAFVCLHVFYTELSTSRGYSIFSYATKRQDNEILIFWSDIGYSFTVGGSEILFEVPEVTVAPVHICTSWESASGIVEFWVDGKPRVRKSLKKGTVGAEASIIILGQEQDSFGGNGFEQSLSLVDIGNVNMDWDFLSPDEINTIYLGPPFSPNVLNWRALKYEVQGEVFTKPLWP

-

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)

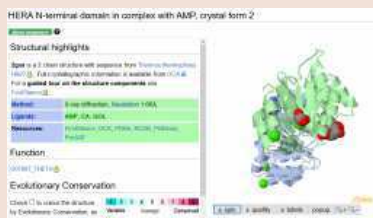
> Save page

The screenshot shows the Proteopedia website interface. At the top, the user is logged in as "Student". A red circle highlights the "Student" link, with a red number "1" next to it. Below the login bar, a message says "Login successful" and "You are now logged in to Proteopedia as 'Student'". A navigation menu is visible on the left, with a red circle around the "Go" button in the search section, and a red number "2" next to it. A red number "3" is placed below the "Go" button. A green callout box on the right contains the text "Sandbox_RSBMB_01", "Sandbox_RSBMB_02", etc. (different for each person). A red arrow points from this box to the "Go" button. Another red callout box contains the text "type the name of the page and click 'Go'", with a red arrow pointing to the search input field.

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 different 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"

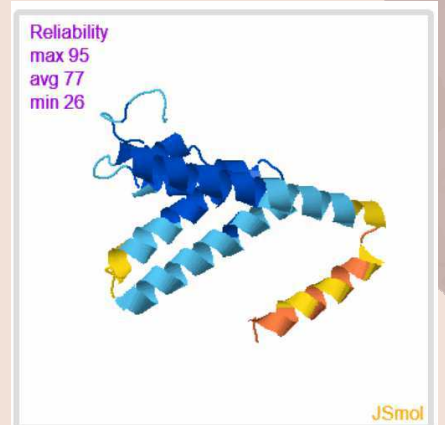
(go down) > SAT (Scene Authoring Tools) [[show](#)]

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

Type a name (Serum amyloid A-1) and paste the sequence

```
MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGMWRAYS  
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/>

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), infographics & images by [Freepik](#).