Practical Use of AlphaFold and AlphaFold-Multimer

Mitch Miller
Rice University, Dec. 7, 2022

Overview

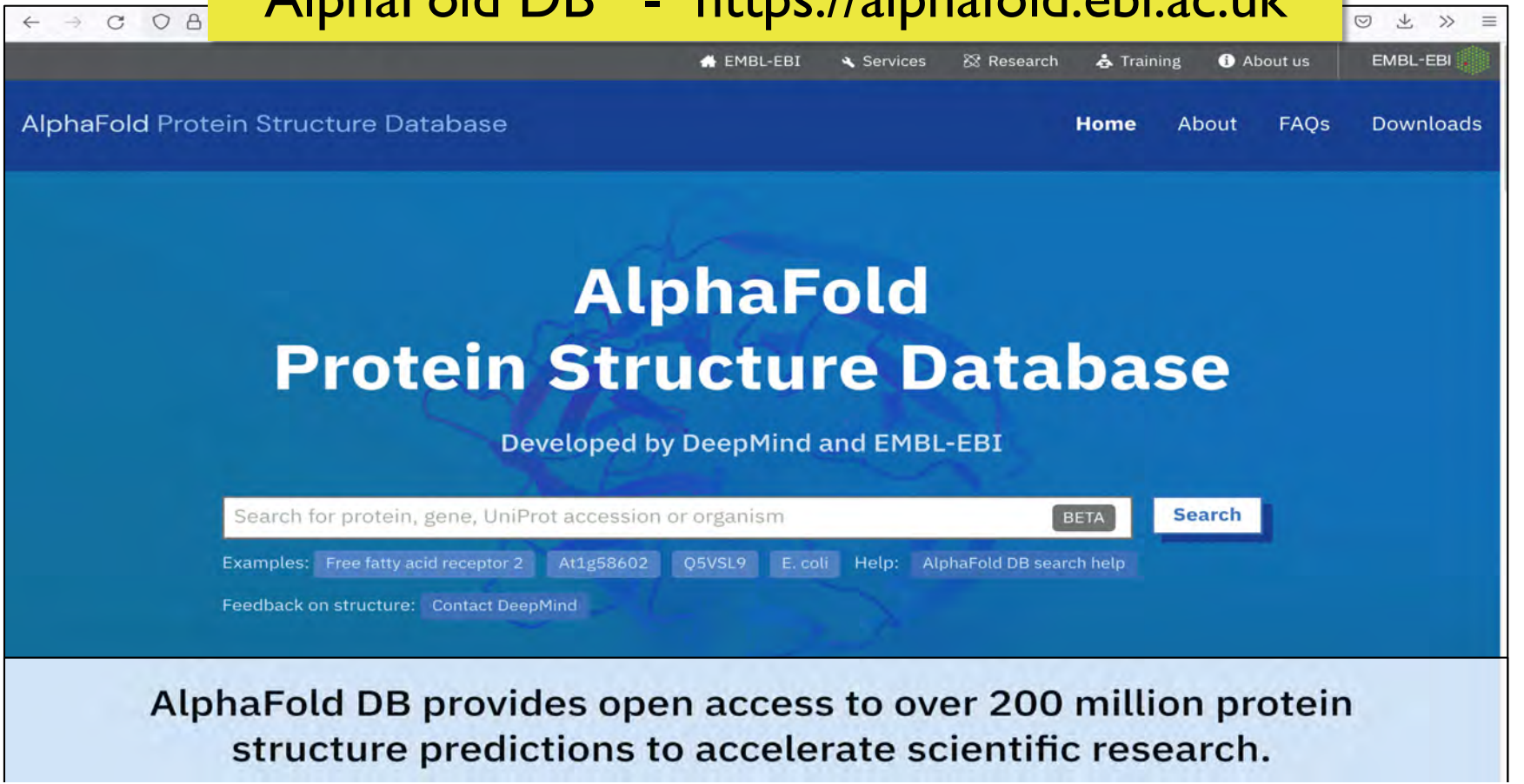
- Ways to obtain AlphaFold predictions
- How to assess AlphaFold predictions
- Example: running AlphaFold on Google Colab
- Some practical examples
  - Multimer predictions – finding interacting proteins
  - Ligands and Binding sites
  - Construct design
- Servers to help interpret the results

Ways to Obtain AlphaFold Predictions

- AlphaFoldDB
- AlphaFold via Google Colab
- Computer cluster
- Install locally

Highly accurate protein structure prediction with AlphaFold.
Jumper J, Evans R, Pritzel A, *et al.* 2021 *Nature* 596:583-589.

AlphaFold DB - https://alphafold.ebi.ac.uk



Tunyasuvunakool et al., (2021) Nature 596: 590–596

Tautomycetin biosynthetic PKS

AlphaFold structure prediction

Download **PDB file** **mmCIF file** **Predicted aligned error**

Information

Protein	Tautomycetin biosynthetic PKS
Gene	I547_1222
Source organism	Mycobacterium kansasii 824 go to search
UniProt	X7YHF0 go to UniProt
Experimental structures	None available in the PDB
Biological function	Not available. go to UniProt

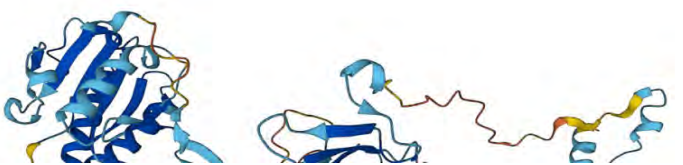
3D viewer

Model Confidence:

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

Sequence of AF-X7YHF0-F1 Chain 1: Tautomyce... A

```
MTATLARLHVHGQSPAWPALYPHARQVELPTYPFQHHRYWLTPRSGTDAAGLGLDQPQHPLLGAITTLADRDEVIATGRVVLGSQAWLAAHRVGDVVVLPATGFVDLVLGVGDYVGCPV
IDELVLHTPLVLAEHTPTDIQISVADADGAGRRSVNVHARTGTDHRAEAGWVLHASGTLRFGQTAAGEPPVPALPATAVDVQRFYQQLADCGLHYDPPFCSVRGIGLQPGDPDHIYAEV
ALPAGTDITGYGMHPALLDAALHPIAAALGAGADAEPAVLRLPFVFSGVTLYATAATRLQVWLTRTGDDTFTLYASDPAGAPVIRIDTVVVRVLPDMATLSAPTTAAPSVPGLWELAWP
```



AlphaFold's Assessment of Model Quality

3D viewer

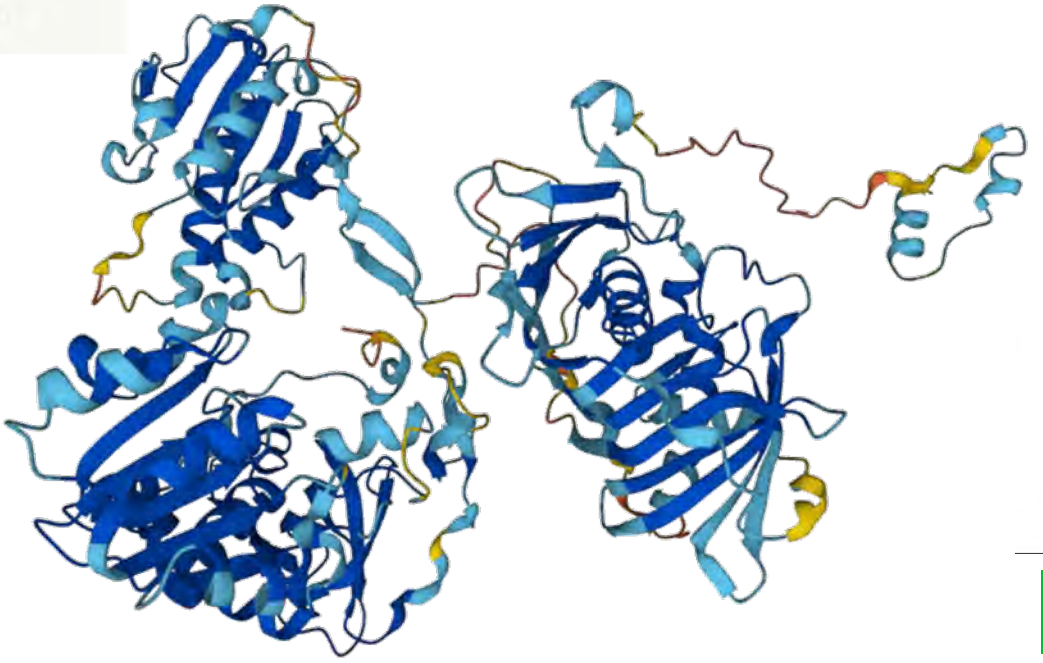
Model Confidence:

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

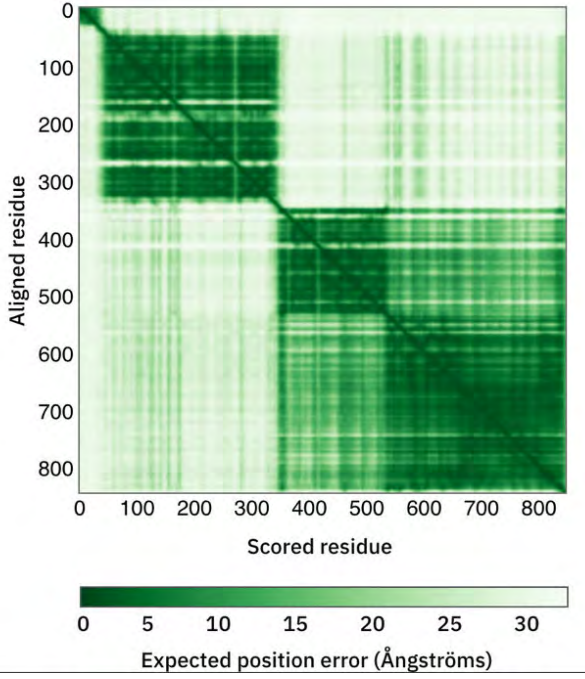
Sequence of AF-X7YHF0-F1 Chain 1: Tautomyce... A

```
MTATLARLHVHGQSPAWPALYPHARQVELPTYPFQHHRYWLTPRSGTDAAGLGLDQPQHPLLGAITTLADRDEVIATGRVVLGSQAWLAAHRVGDVVVLPATGFVDLVLGVGDYVGCPV
IDELVLHTPLVLAEHTPTDIQISVADADGAGRRSVNVHARTGTDHRAEAGWVLHASGTLRPGQTAAGEPPVPALPATAVDVQRFYQQLADCGLHYDPPFCSVRGIGLQPGDPDHIYAEV
ALPAGTDITGYGMHPALLDAALHPIAAALGAGADAEPAVLRLPFVFSGVTLYATAATRLQVWLTRTGDDTFTLYASDPAGAPVIRIDTVVVRVLPDMATLSAPTTAAPSVPGLWELAWP
```

predicted Local Distance Difference Test



PAE



Predicted Aligned Error

Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

New Results

Follow this preprint

Posted March 10, 2022

Protein complex prediction with AlphaFold-Multimer

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, Demis Hassabis

doi: https://doi.org/10.1101/2021.10.04.463034

nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-022-01488-1

Check for updates

OPEN

ColabFold: making protein folding accessible to all

Milot Mirdita[1,10] ✉, Konstantin Schütze[2], Yoshitaka Moriwaki[3,4], Lim Heo[5], Sergey Ovchinnikov[6,7,10] ✉ and Martin Steinegger[2,8,9,10] ✉

ColabFold offers accelerated prediction of protein structures and complexes by combining the fast homology search of MMseqs2 with AlphaFold2 or RoseTTAFold. ColabFold's 40–60-fold faster search and optimized model utilization enables prediction of close to 1,000 structures per day on a server with one graphics processing unit. Coupled with Google Colaboratory, ColabFold becomes a free and accessible platform for protein folding. ColabFold is open-source software available at https://github.com/sokrypton/ColabFold and its novel environmental databases are available at https://colabfold.mmseqs.com.

Predicting the three-dimensional (3D) structure of a protein from its sequence alone remains an unsolved problem. However, protein sizes of ~1,000 residues. For these, however, the MSA generation dominates the overall run time.

To enable researchers without these resources to use AlphaFold2, independent solutions based on Google Colaboratory were developed. Colaboratory is a proprietary version of Jupyter Notebook hosted by Google. It is accessible for free to logged-in users and includes access to powerful GPUs. Concurrently, Tunyasuvunakool et al.[9] developed an AlphaFold2 Jupyter Notebook for Google Colaboratory (referred to as AlphaFold-Colab), for which the input MSA is built by searching with HMMer against the UniProt Reference Clusters (UniRef90) and an eightfold-reduced environmental database. This results in less accurate predictions while still requiring long search times.

https://www.google.com/search?q=colabfold&client=firefox-b-1-d&ei=WU-OY

colabfold

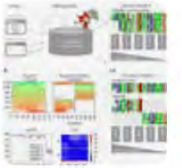
All News Images Videos Shopping More Tools

About 13,500 results (0.31 seconds)

https://www.nature.com › ... › brief communications

ColabFold: making protein folding accessible to all - Nature

by M Mirdita · 2022 · Cited by 805 — Here, we present **ColabFold**, a fast and easy-to-use software for the prediction of protein structures and homo- and heteromer complexes, for use ...



https://github.com › sokrypton › ColabFold

GitHub - sokrypton/ColabFold: Making Protein folding ...

LocalColabFold is an installer script designed to make **ColabFold** functionality available on local users' machines. It supports wide range of operating systems, ...

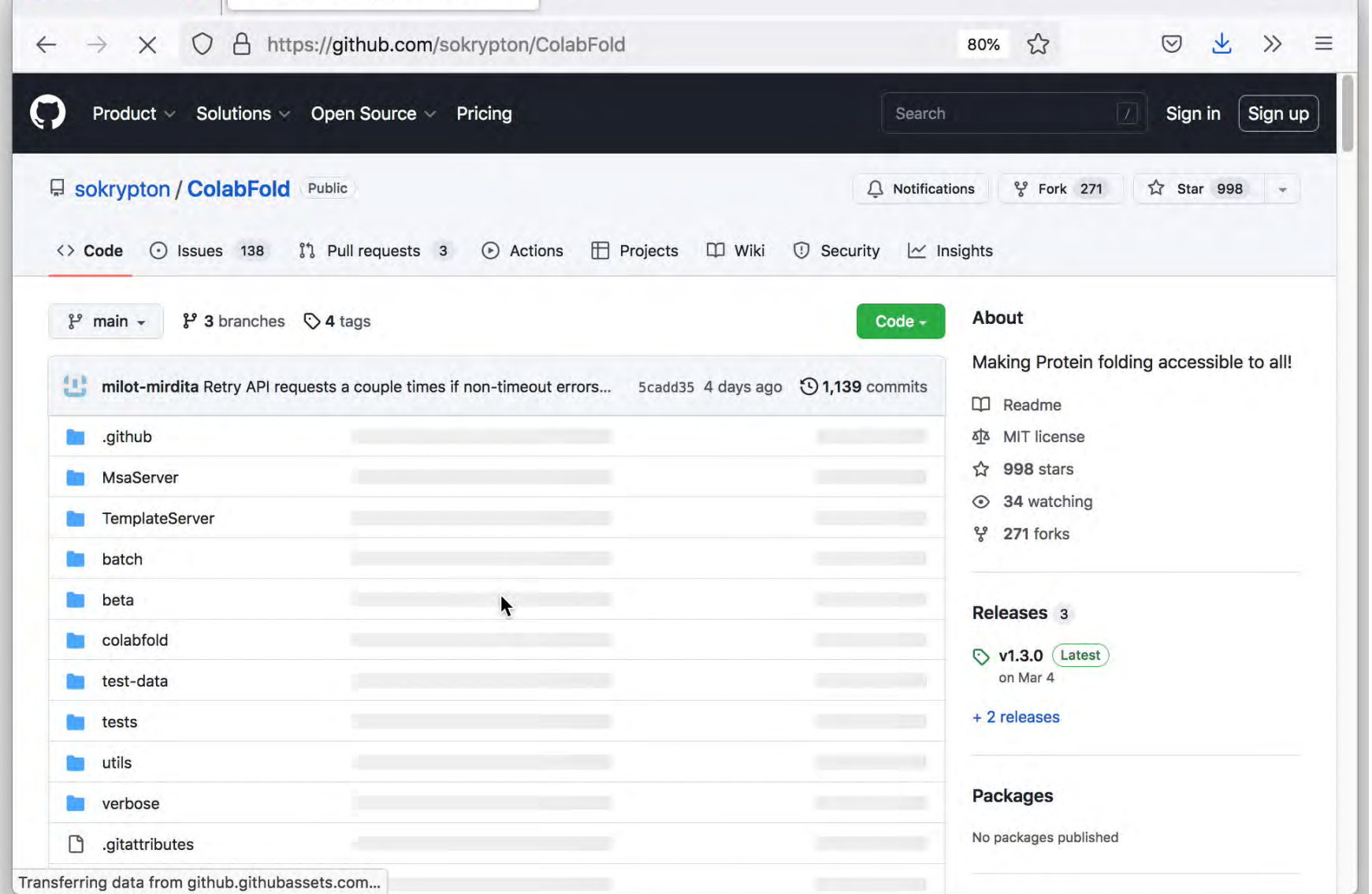
AlphaFold2.ipynb · Issues 137 · AlphaFold2_complexes.ipynb · README.md

https://colabfold.mmseqs.com

ColabFold

ColabFold databases are MMseqs2 expandable profile databases to generate diverse multiple

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj4...&url=https://github.com/sokrypton/ColabFold&usg=AOvVaw1vZfnkm31DstfuSrWUjcd3


https://github.com/sokrypton/ColabFold
80%
Product
Solutions
Open Source
Pricing
Search
Sign in
Sign up
sokrypton / ColabFold
Public
Notifications
Fork 271
Star 998
Code
Issues 138
Pull requests 3
Actions
Projects
Wiki
Security
Insights
main
3 branches
4 tags
Code
milot-mirdita Retry API requests a couple times if non-timeout errors...
5cadd35 4 days ago
1,139 commits
.github
MsaServer
TemplateServer
batch
beta
colabfold
test-data
tests
utils
verbose
.gitattributes
About
Making Protein folding accessible to all!
Readme
MIT license
998 stars
34 watching
271 forks
Releases 3
v1.3.0 Latest
on Mar 4
+ 2 releases
Packages
No packages published
Transferring data from github.githubassets.com...

README.md

Making Protein folding accessible to all via Google Colab!

Notebooks	monomers	complexes	mmseqs2	jackhmmer	temp
AlphaFold2_mmseqs2	Yes	Yes	Yes	No	Yes
AlphaFold2_batch	Yes	Yes	Yes	No	Yes
RoseTTAFold	Yes	No	Yes	No	No
AlphaFold2 (from Deepmind)	Yes	Yes	No	Yes	No
ESMFold	Yes	Maybe	No	No	No
BETA (in development) notebooks					
AlphaFold2_advanced	Yes	Yes	Yes	Yes	No
OmegaFold	Yes	Maybe	No	No	No

https://colab.research.google.com/github/sokrypton/ColabFold/blob 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help

Share M

+ Code + Text Copy to Drive Connect Editing

ColabFold: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.0, v1.1, v1.2, v1.3

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. *Nature Methods*, 2022

Input protein sequence(s), then hit `Runtime -> Run all`

query_sequence: "PIAQIHILEGRSDEQKETLIREVSEAISRSLDAPLTSVRVIITEMAKGHFGIGGELASK"

- Use **:** to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: "test"

use_amber: ☐

template_mode: none

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (PDB or mmCIF format, see notes below)

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share M

+ Code + Text Copy to Drive Connect Editing

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.0, v1.1, v1.2, v1.3

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. *Nature Methods*, 2022

Input protein sequence(s), then hit `Runtime -> Run all`

query_sequence: "FH PNVHFKTGEICLDILKNAWSPAWTLQSVCRAIIALMAHPEPDSPLNCDSGNLLRSGDVRGFNSMAQMYTRLAAMPKKGLE VLFQGP"

- Use **:** to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

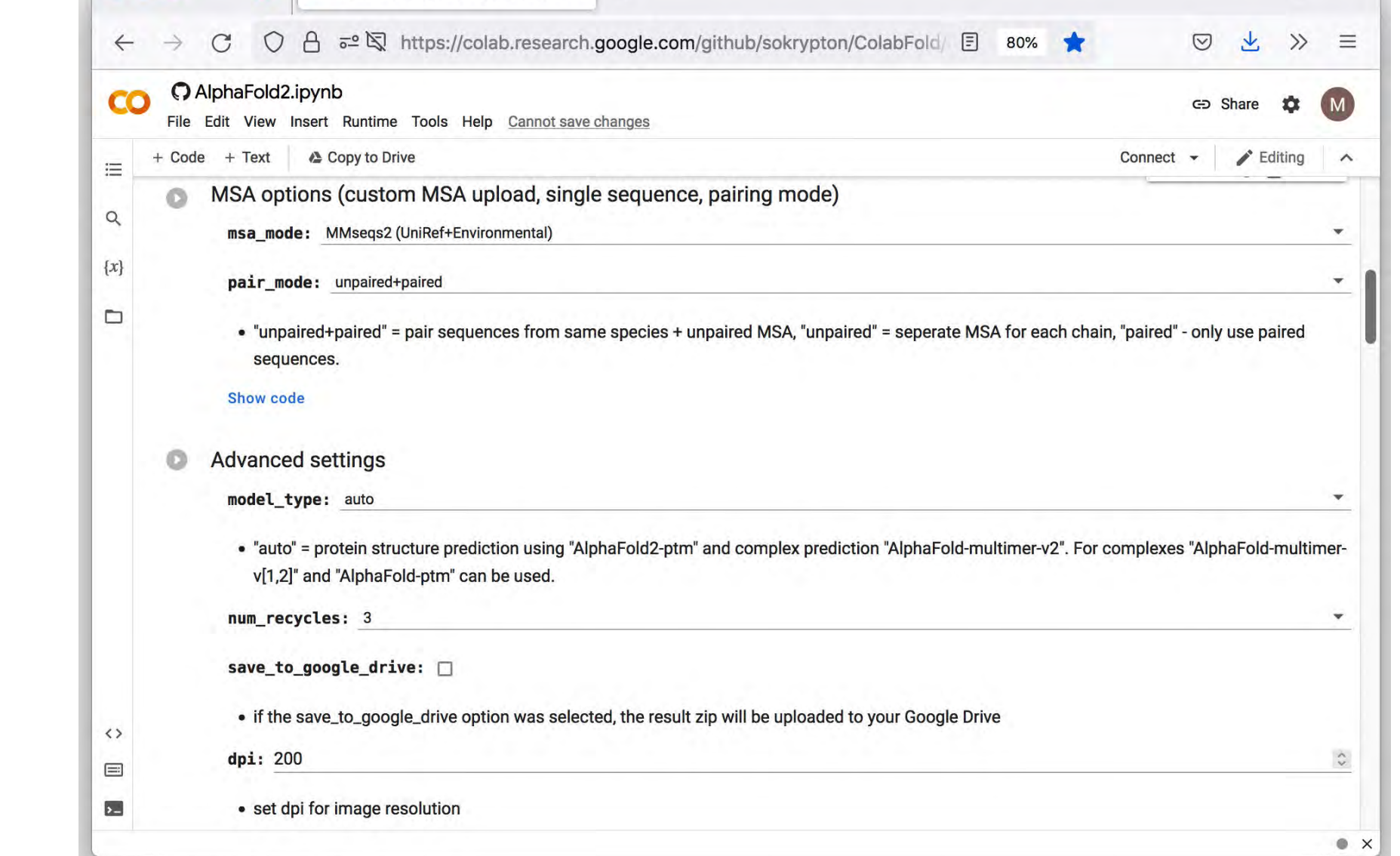
jobname: "PEX4"

use_amber: ☐

template_mode: none

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (PDB or mmCIF format, see notes below)

Show code


https://colab.research.google.com/github/sokrypton/ColabFold/
80%
AlphaFold2.ipynb
File Edit View Insert Runtime Tools Help Cannot save changes
Share
M
+ Code + Text Copy to Drive
Connect Editing
MSA options (custom MSA upload, single sequence, pairing mode)
msa_mode: MMseqs2 (UniRef+Environmental)
pair_mode: unpaired+paired
"unpaired+paired" = pair sequences from same species + unpaired MSA, "unpaired" = seperate MSA for each chain, "paired" - only use paired sequences.
Show code
Advanced settings
model_type: auto
"auto" = protein structure prediction using "AlphaFold2-ptm" and complex prediction "AlphaFold-multimer-v2". For complexes "AlphaFold-multimer-v[1,2]" and "AlphaFold-ptm" can be used.
num_recycles: 3
save_to_google_drive:
if the save_to_google_drive option was selected, the result zip will be uploaded to your Google Drive
dpi: 200
set dpi for image resolution

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share M

+ Code + Text Copy to Drive Connect Editing

Show code

Instructions

Quick start

1. Paste your protein sequence(s) in the input field.
2. Press "Runtime" -> "Run all".
3. The pipeline consists of 5 steps. The currently running step is indicated by a circle with a stop sign next to it.

Result zip file contents

1. PDB formatted structures sorted by avg. pLDDT and complexes are sorted by pTMscore. (unrelaxed and relaxed if `use_amber` is enabled).
2. Plots of the model quality.
3. Plots of the MSA coverage.
4. Parameter log file.
5. A3M formatted input MSA.
6. A `predicted_aligned_error_v1.json` using AlphaFold-DB's format and a `scores.json` for each model which contains an array (list of lists) for PAE, a list with the average pLDDT and the pTMscore.
7. BibTeX file with citations for all used tools and databases.

At the end of the job a download modal box will pop up with a `jobname.result.zip` file. Additionally, if the `save_to_google_drive` option was selected, the `jobname.result.zip` will be uploaded to your Google Drive.

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share M

+ Code + Text C

Connect Editing

Run all ⌘/Ctrl+F9
Run before ⌘/Ctrl+F8
Run the focused cell ⌘/Ctrl+Enter
Run selection ⌘/Ctrl+Shift+Enter
Run after ⌘/Ctrl+F10
Interrupt execution ⌘/Ctrl+M I
Restart runtime ⌘/Ctrl+M .
Restart and run all
Disconnect and delete runtime
Change runtime type
Manage sessions
View runtime logs

ColabFold: Alph

Easy to use protein stru d2 and Alphafold2-multimer. Sequence alignments/templates ch. For more details, see bottom of the notebook, checkout the Old versions: v1.0, v1.1, v1.2, v1.3

Mirdita M, Schütze K, M l. ColabFold: Making protein folding accessible to all. *Natur*

Input protein sequ l

query_sequer QSVCRAIIALMAHPEPDSPLNCDSGNLLRSGDVRGFNSMAQMYTRLAAMPKKGLE VLFQGP

- Use **:** to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: PEX4

use_amber: ☐

template_mode: none

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (PDB or mmCIF format, see notes below)

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share M

+ Code + Text Copy to Drive Connect Editing

ColabFold: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.0, v1.1, v1.2, v1.3

Mirdita M, Schütze K, Moriwaki Y
accessible to all. *Nature Method*

Warning: This notebook was not authored by Google.

This notebook is being loaded from **GitHub**. It may request access to your data stored with Google, or read data and credentials from other sessions. Please review the source code before executing this notebook.

Cancel Run anyway

Input protein sequence(s)

query_sequence: "M LQPPQVRFLTKIFH PNVHFKTGEICLD"

- Use : to specify inte For example **PI...SK:PI...SK** for a homodimer

jobname: "PEX4"

use_amber: ☐

template_mode: none

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (PDB or mmCIF format, see notes below)

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share M

+ Code + Text Copy to Drive RAM Disk Editing

Don't forget to hit Runtime -> Run all after updating the form.

Show code

Install dependencies

[4] Show code

Run Prediction

Show code

```
/usr/local/lib/python3.8/dist-packages/Bio/Data/SCOPData.py:18: BiopythonDeprecationWarning: The 'Bio.Data.SCOPData' module will
  warnings.warn(
Downloading alphafold2 weights to .: 100%|██████████| 3.47G/3.47G [01:48<00:00, 34.3MB/s]
2022-12-06 03:39:21,053 Found 5 citations for tools or databases
2022-12-06 03:39:21,073 Unable to initialize backend 'tpu_driver': NOT_FOUND: Unable to find driver in registry given worker:
2022-12-06 03:39:21,819 Unable to initialize backend 'rocm': NOT_FOUND: Could not find registered platform with name: "rocm". Ava
2022-12-06 03:39:21,820 Unable to initialize backend 'tpu': module 'jaxlib.xla_extension' has no attribute 'get_tpu_client'
2022-12-06 03:39:21,821 Unable to initialize backend 'plugin': xla_extension has no attributes named get_plugin_device_client. C
2022-12-06 03:39:27,786 Query 1/1: PEX4_71d3d (length 165)
PENDING:   0%|          | 0/150 [elapsed: 00:00 remaining: ?]2022-12-06 03:39:28,608 Sleeping for 9s. Reason: PENDING
RUNNING:   6%|▌         | 9/150 [elapsed: 00:10 remaining: 02:46]2022-12-06 03:39:38,425 Sleeping for 10s. Reason: RUNNING
```

Display 3D structure

rank_num: 1

Executing (2m 10s) Cell > run() > get_msa_and_templates() > run_mmseqs2()

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

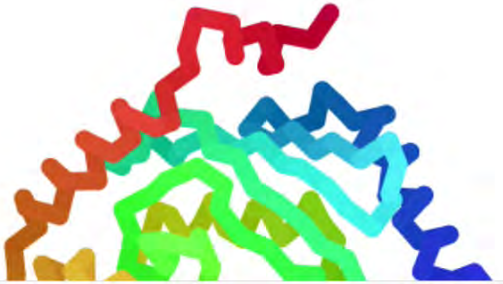
Share M

+ Code + Text Copy to Drive RAM Disk Editing

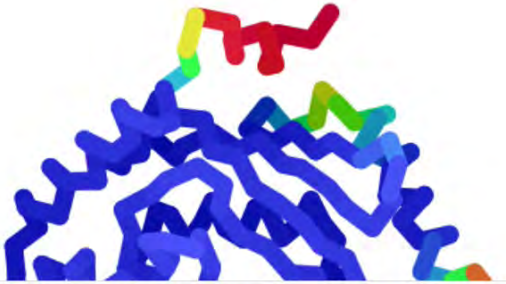
Show code

```
/usr/local/lib/python3.8/dist-packages/Bio/Data/SCOPData.py:18: BiopythonDeprecationWarning: The 'Bio.Data.SCOPData' module will
  warnings.warn(
Downloading alphafold2 weights to .: 100%|██████████| 3.47G/3.47G [01:48<00:00, 34.3MB/s]
2022-12-06 03:39:21,053 Found 5 citations for tools or databases
2022-12-06 03:39:21,073 Unable to initialize backend 'tpu_driver': NOT_FOUND: Unable to find driver in registry given worker:
2022-12-06 03:39:21,819 Unable to initialize backend 'rocm': NOT_FOUND: Could not find registered platform with name: "rocm". Ava
2022-12-06 03:39:21,820 Unable to initialize backend 'tpu': module 'jaxlib.xla_extension' has no attribute 'get_tpu_client'
2022-12-06 03:39:21,821 Unable to initialize backend 'plugin': xla_extension has no attributes named get_plugin_device_client. C
2022-12-06 03:39:27,786 Query 1/1: PEX4_71d3d (length 165)
PENDING:   0%|          | 0/150 [elapsed: 00:00 remaining: ?]2022-12-06 03:39:28,608 Sleeping for 9s. Reason: PENDING
RUNNING:   6%|▌         | 9/150 [elapsed: 00:10 remaining: 02:46]2022-12-06 03:39:38,425 Sleeping for 10s. Reason: RUNNING
RUNNING:  13%|█▎        | 19/150 [elapsed: 00:21 remaining: 02:26]2022-12-06 03:39:49,230 Sleeping for 5s. Reason: RUNNING
COMPLETE: 100%|██████████| 150/150 [elapsed: 00:29 remaining: 00:00]
2022-12-06 03:39:58,230 Running model_1
2022-12-06 03:41:32,133 model_1 took 88.5s (3 recycles) with pLDDT 90.2 and ptmscore 0.855
```

colored by N→C



colored by pLDDT



Executing (4m 16s) Cell > run() > predict_structure() > predict()

https://colab.research.google.com/github/sokrypton/ColabFold/ 80%

AlphaFold2.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Share M

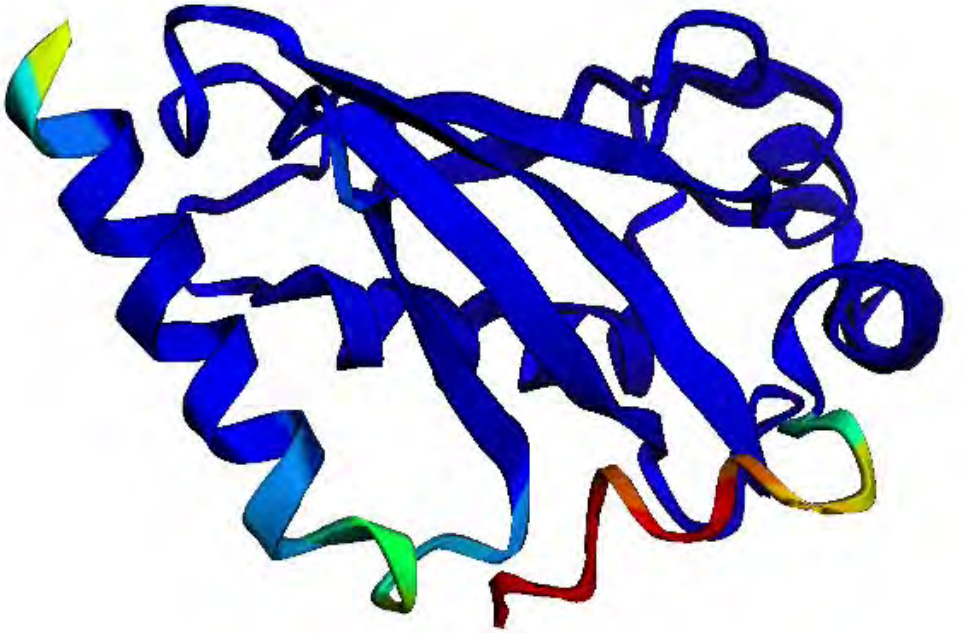
+ Code + Text Copy to Drive RAM Disk Editing

0s

show_sidechains:

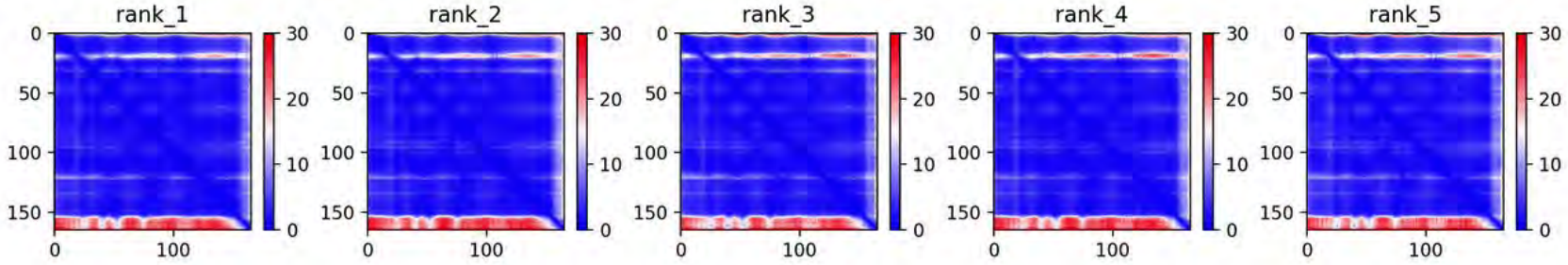
show_mainchains:

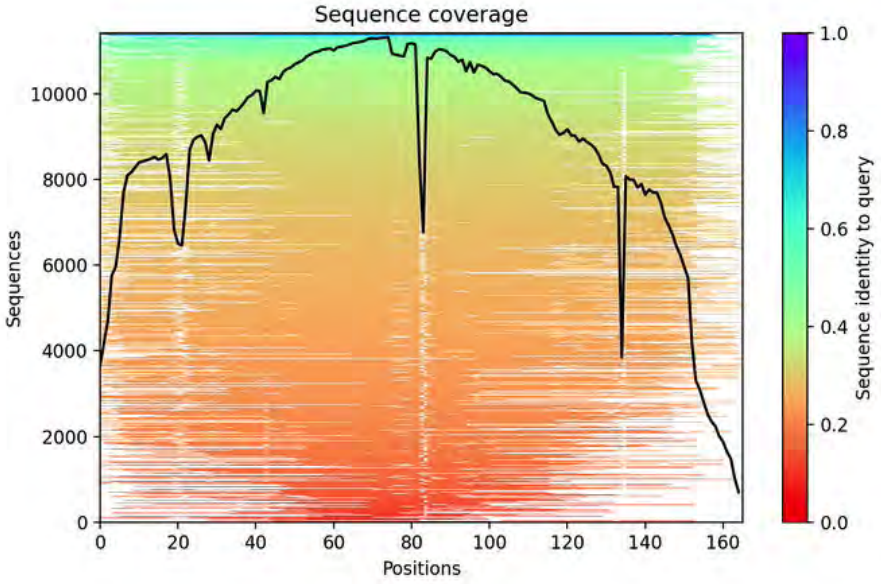
Show code

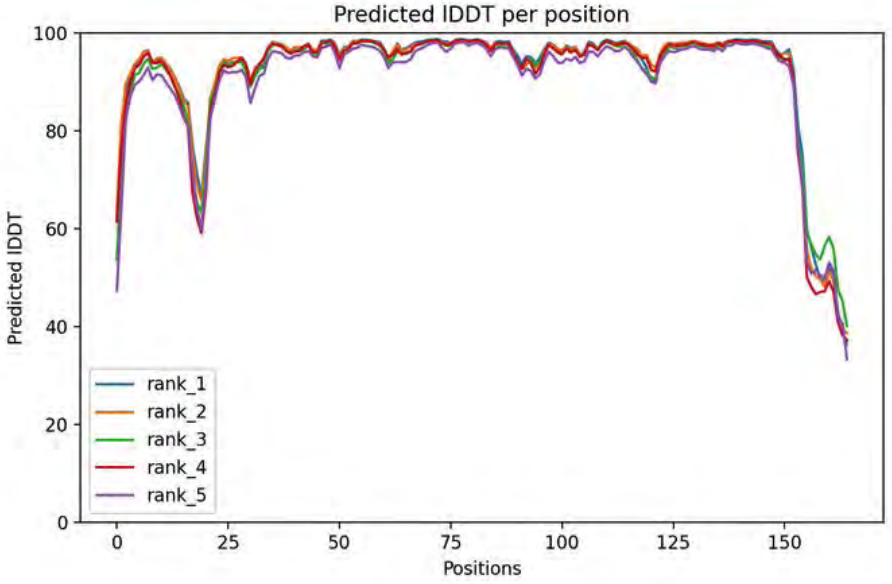


0s completed at 9:45 PM

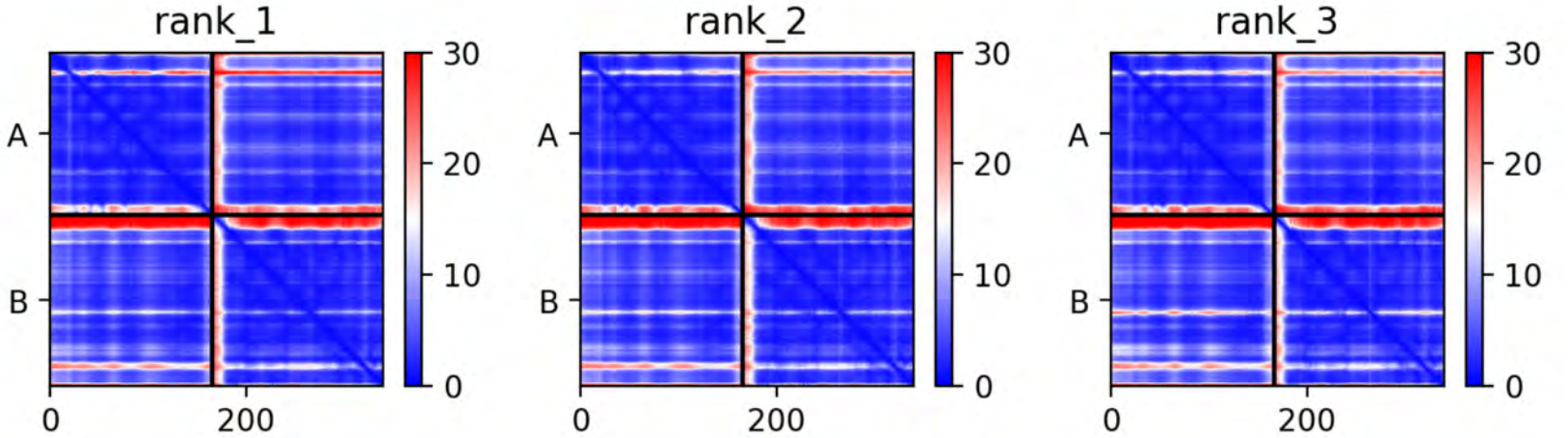
Plots for PEX4_71d3d

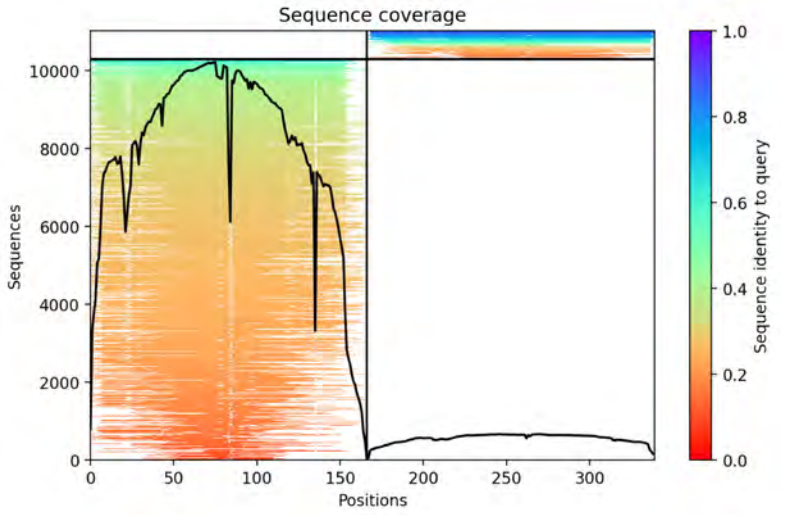


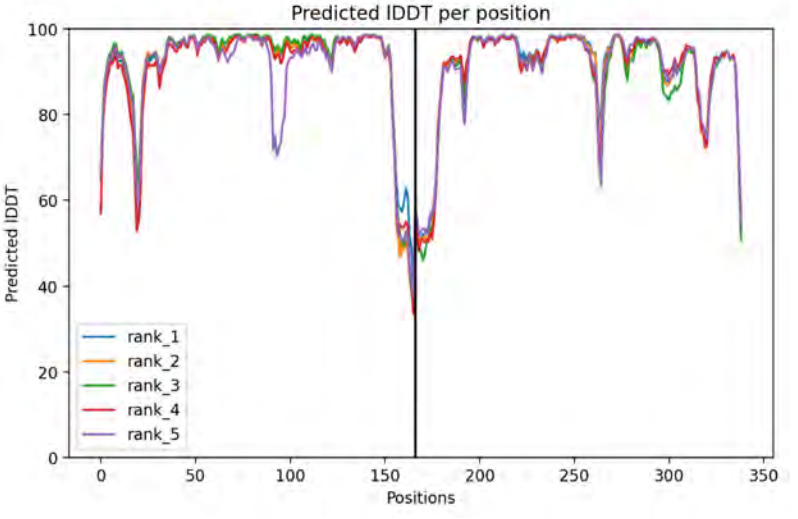




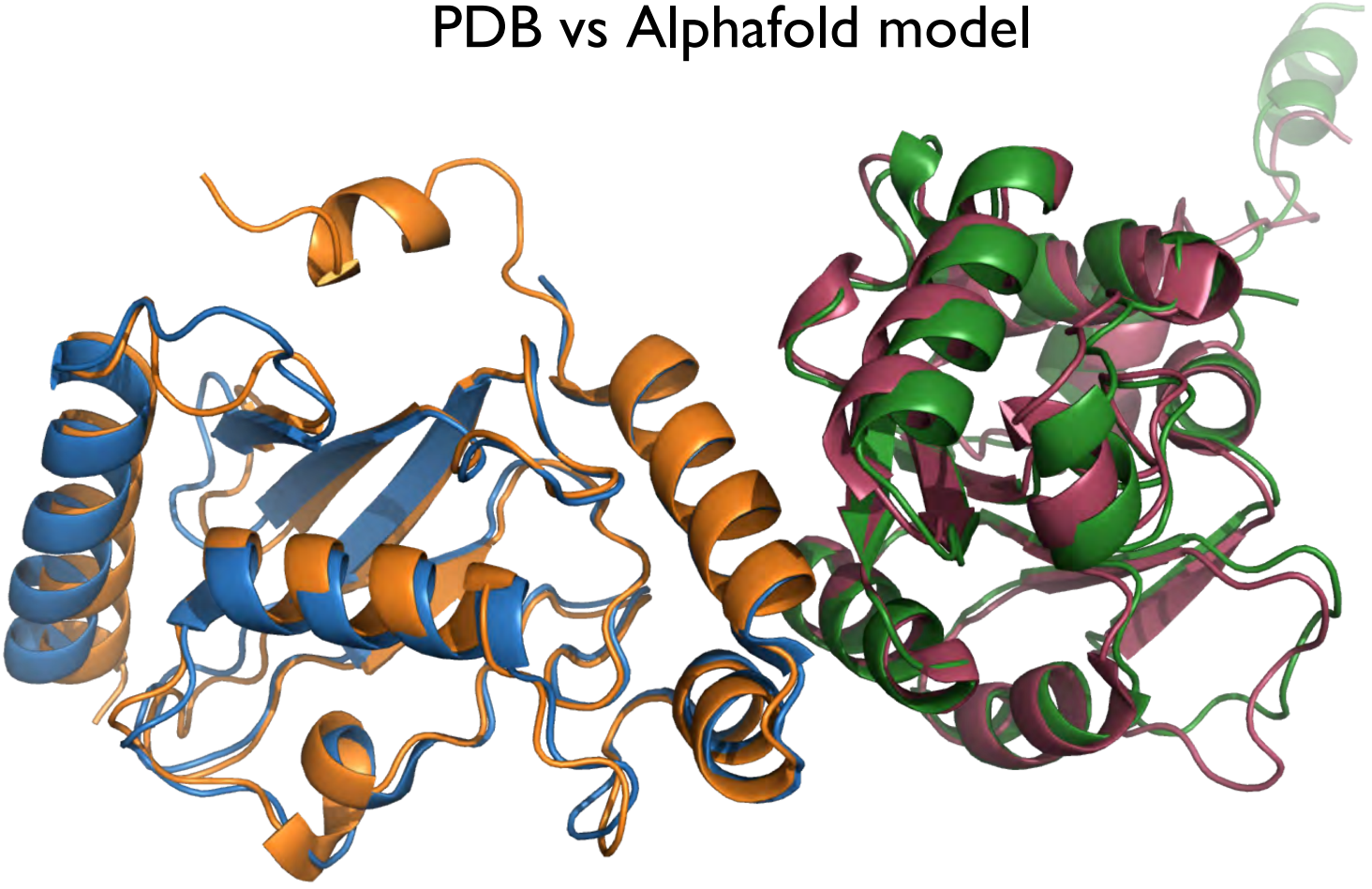
PEX4:PEX22 heterodimer







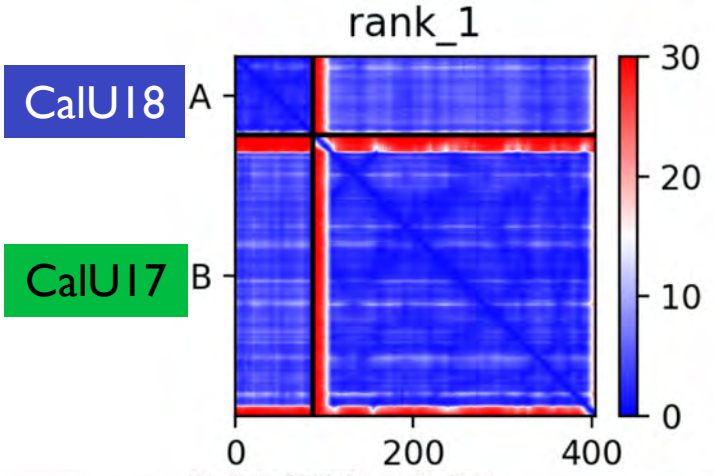
PEX4:PEX22 heterodimer
PDB vs Alphafold model

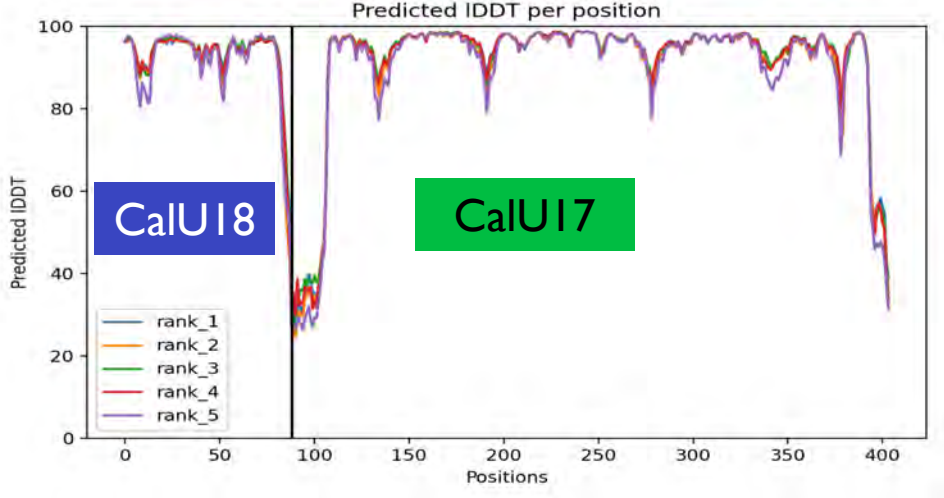


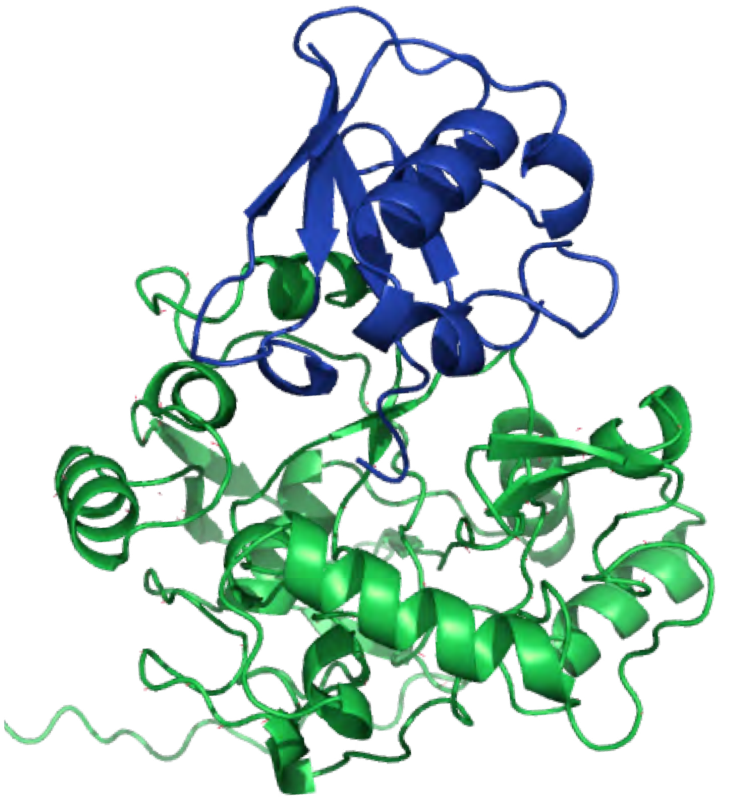
PDB 6xod – Traver *et al.*, (2022) *Front Cell Dev Biol* 10: 838923

Abigael Kosgei

CalU17 CalU18 heterodimer prediction

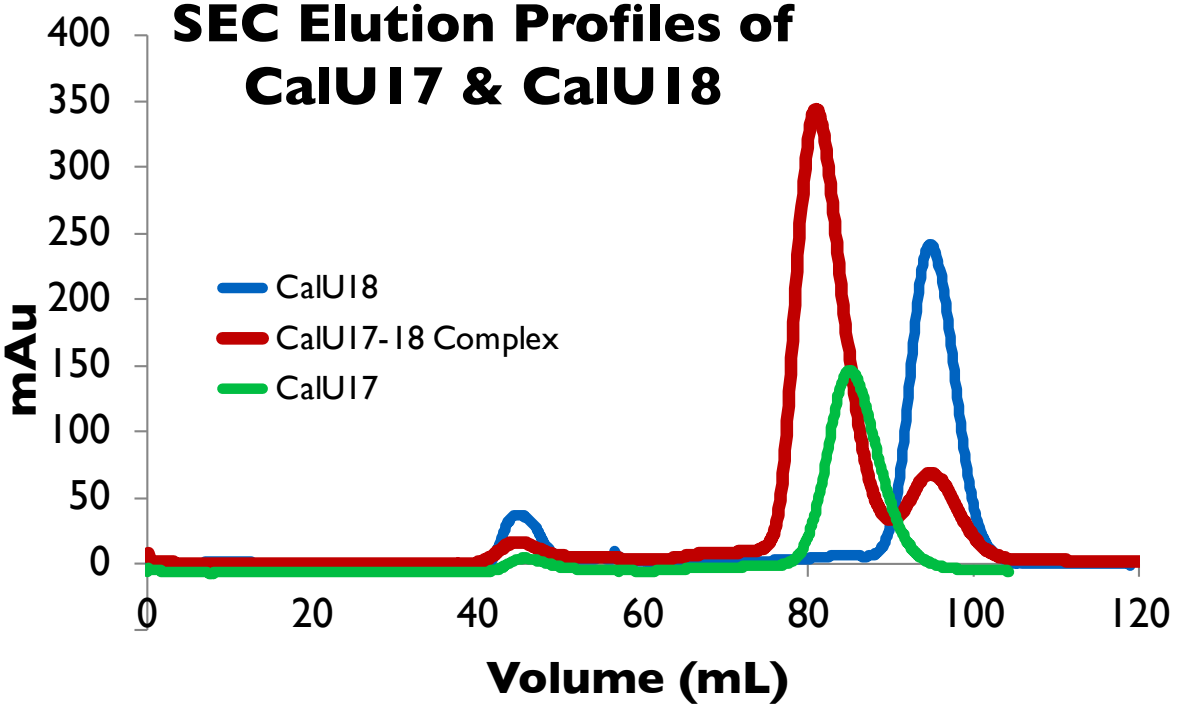


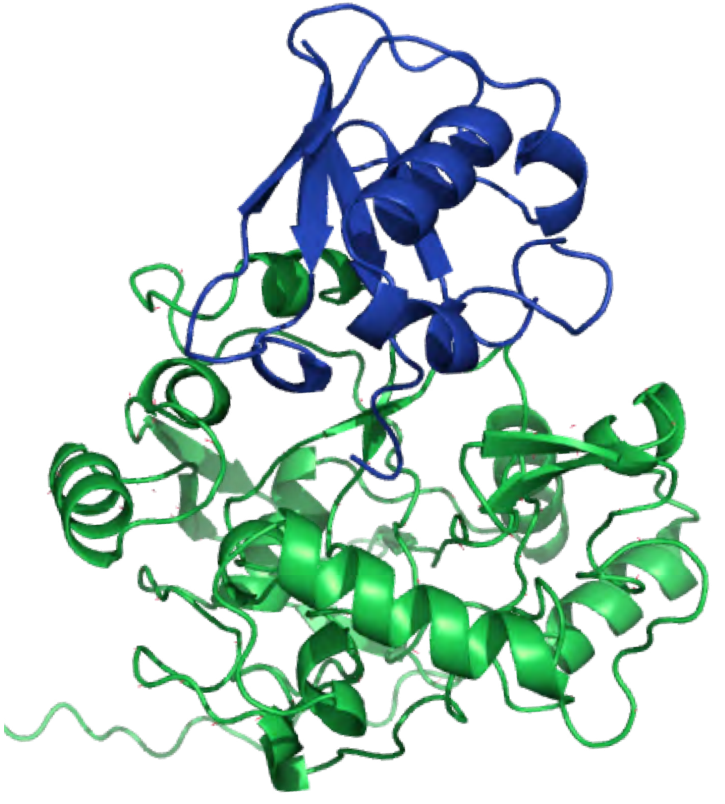


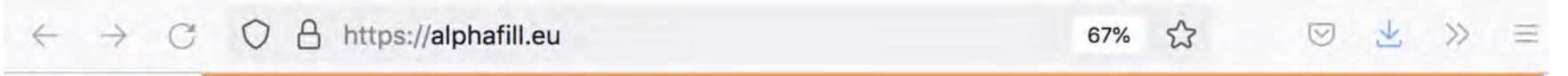


Abigael Kosgei

CalU17 CalU18 heterodimer prediction







αfill

NKI Research | Biochemistry | Perrakis group

Home Structures Compounds Model About Download

AlphaFill

AlphaFold models enriched with ligands and co-factors

Uniprot ID | Search for Uniprot ID | Browse... | No file selected. | Upload

AlphaFill

AlphaFill is an algorithm based on sequence and structure similarity that "transplants" missing ligands, cofactors and (metal) ions to the AlphaFold models. By adding the molecular context to these protein structures, the models can be more easily appreciated in terms of function and structural integrity. Consequently, the AlphaFill models can be helpful in designing downstream wet-lab experiments and/or computational studies.

AlphaFill databank

The AlphaFill databank contains all AlphaFold models including the transferred compounds, if present. For each entry, metadata is provided describing the origin of the ligands and indicating the quality based on global and local Root-Mean-Square deviation (RMSd) values.

Data for a single entry can be found at the corresponding entry page (search by UniProt ID above). Download details for the complete databank can be found at the download page.

For more AlphaFill details: https://www.nature.com/articles/s41592-022-01685-y

P0A8Z3

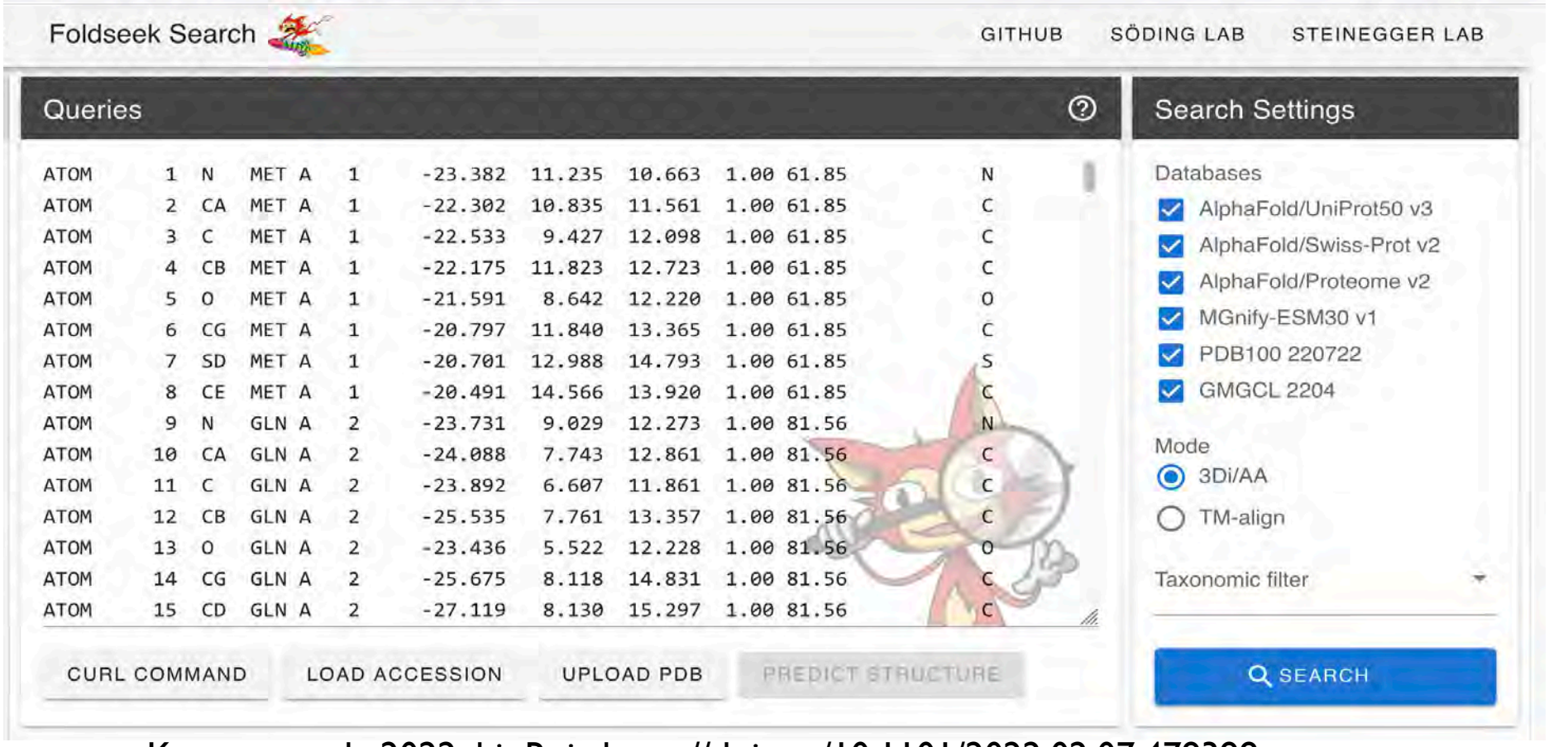
Acyl-CoA thioester hydrolase YbgC

Structure file	https://alphafill.eu/v1/aff/P0A8Z3-F1
Metadata	https://alphafill.eu/v1/aff/P0A8Z3-F1/json
Original AlphaFold model	https://alphafold.ebi.ac.uk/entry/P0A8Z3

25% identity | 30% identity | 40% identity | **50% identity** | 60% identity | 70% identity

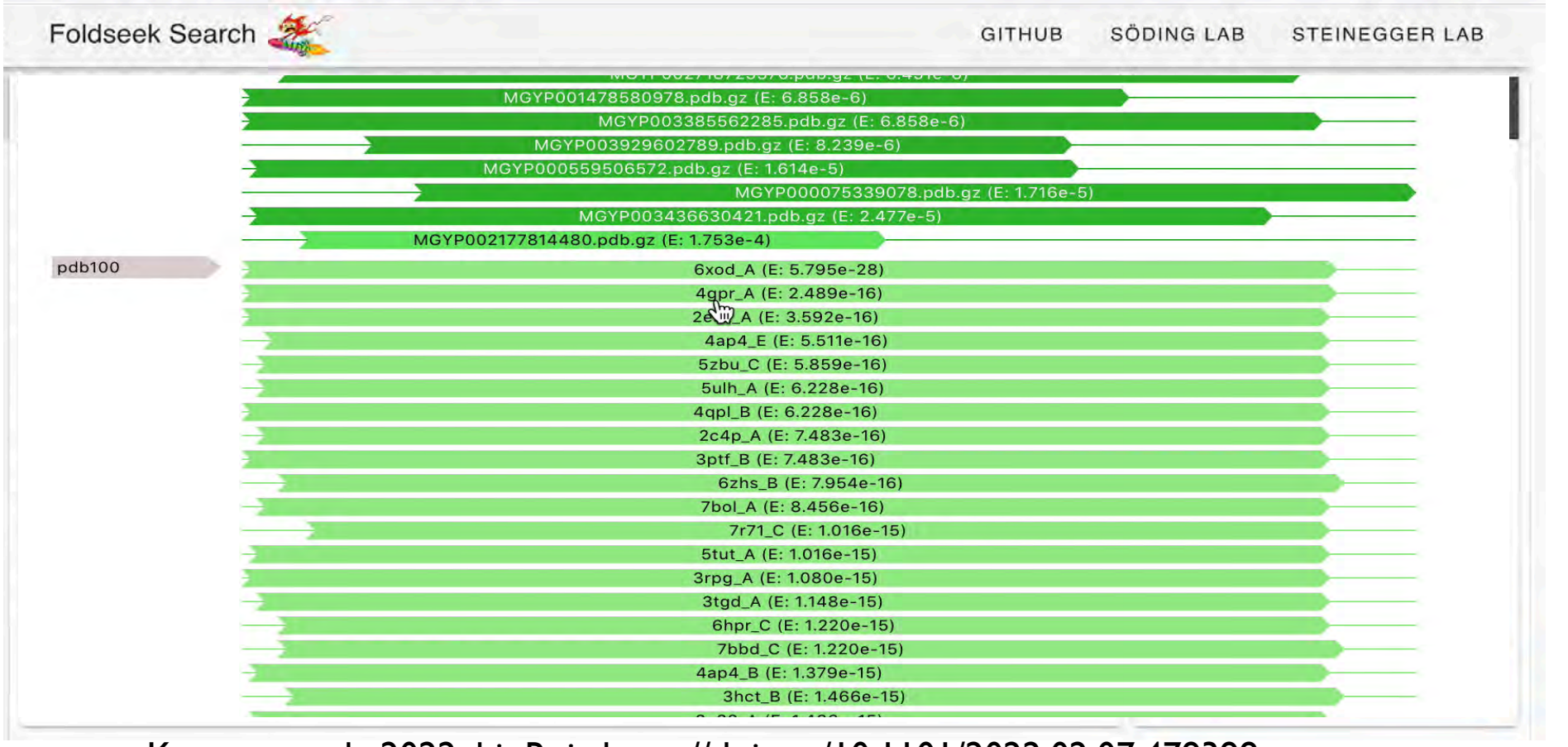
Compound	PDBID	Global RMSd	Asym	Local RMSd	TCS	Show
COA	5kl9.A	0.51	C	0.13	0.08	☑
			B	0.16	0.07	☑
HXC -> COA	5t06.A	0.29	D	0.14	0.07	☑
	5t06.C	0.84	E	0.18	0.04	☑

https://search.foldseek.com/search



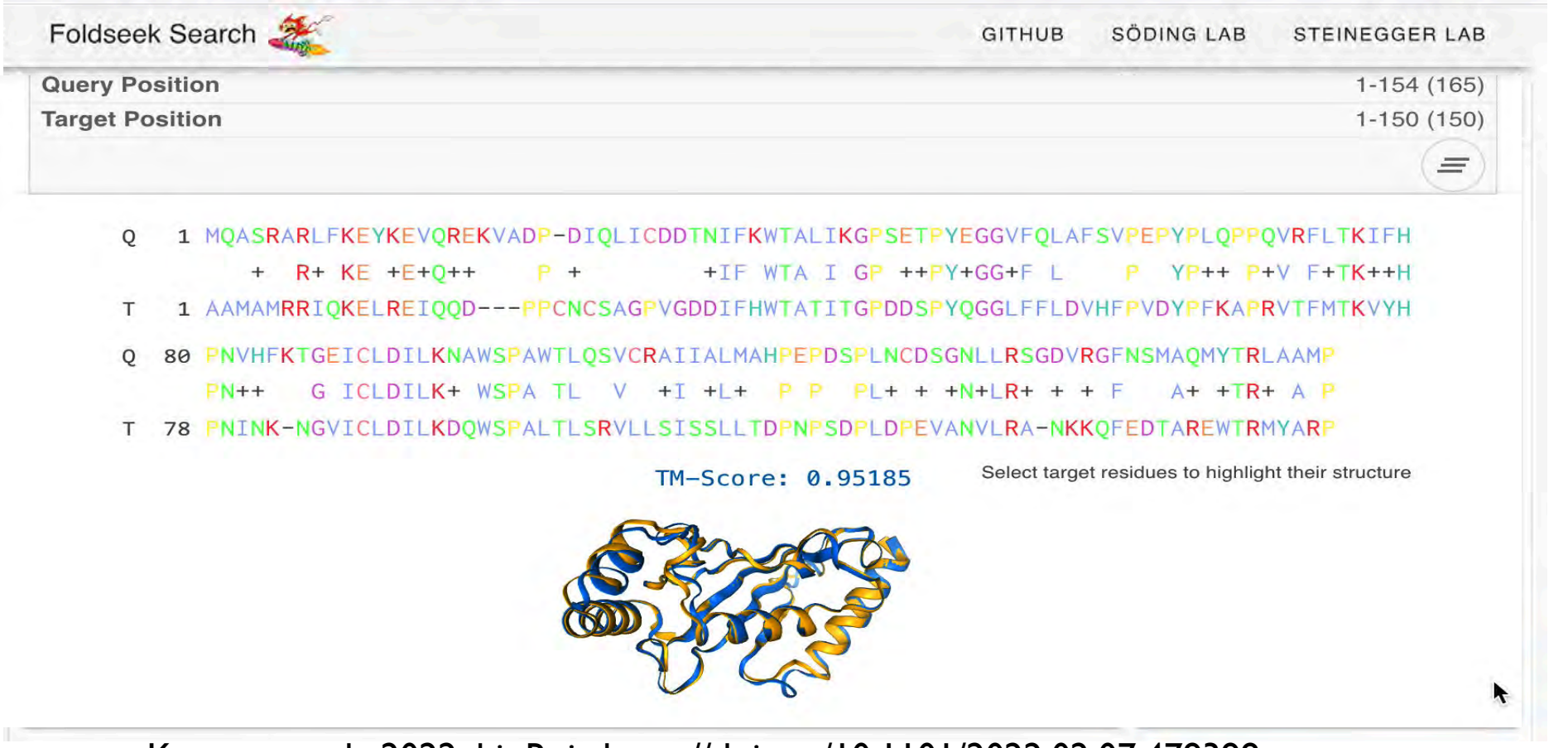
van Kempen et al., 2022. bioRxiv https://doi.org/10.1101/2022.02.07.479398

https://search.foldseek.com/search



van Kempen et al., 2022. bioRxiv https://doi.org/10.1101/2022.02.07.479398

https://search.foldseek.com/search



van Kempen et al., 2022. bioRxiv https://doi.org/10.1101/2022.02.07.479398

Conclusions

Exciting times

Easy to access structures

New creative ways of using structure predications

Predictions are great hypotheses

Acknowledgements

George Phillips

Alex Ditzel

Abigael Kosgei

Tom Pan

David Xu

BioXFEL

NSF

NIH

THE Welch FOUNDATION