
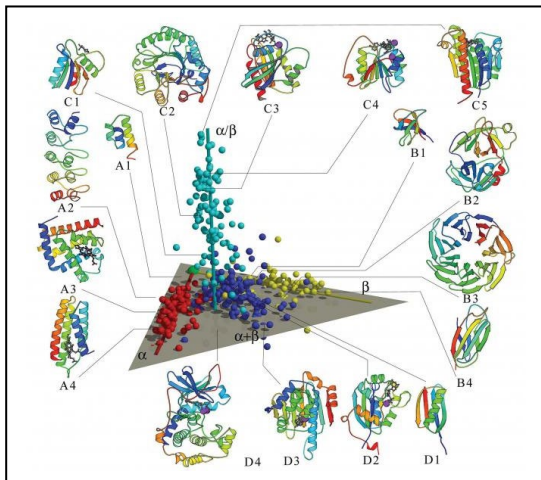
	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

BACKGROUND


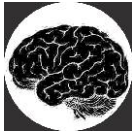
Predictive simulation technology has given rise to the ability to propose a possible tertiary structure of a protein based on its sequence similarity to a known protein whose structure has been experimentally determined. This kind of predictive technology, called homology modeling, is extremely useful as the amount of known protein sequences greatly outnumbers the proteins whose structure has been identified. Protein homology modeling allows structures to be estimated based on established data.



Protein homology modeling constructs a protein's tertiary structure based on similar sequences, on the basis that tertiary structure is better conserved than the amino acid sequence itself (Pitman and Menz, 2006). Throughout the course of evolution, in order for a protein to maintain its catalytic function or other purpose, it must maintain a similar structure that allows for it to continue to perform that function even if other mutations occur which alter the amino acid sequence. A protein which mutates in its active site to a point where it cannot serve its purpose is less likely to be passed down by the organism through the process of natural selection. Although many diseases involving protein

mutations do exist today, when observing the long-term conservation of protein structure, most of the mutations occur in areas outside of the active site. This does not have such a great effect on the overall structure of the protein; with this mutation the protein is still able to serve its purpose. These mutations lead to divergent evolution: different proteins stemming from a common ancestor, often carrying related domains. This is what accounts for the fact that many evolutionarily related proteins will have a more conserved tertiary structure yet a less conserved amino acid sequence. If an amino acid mutates to one of similar character (for example, a hydrophobic amino acid mutating to another hydrophobic amino acid, such as a Valine to Isoleucine mutation), it will not have a great effect on the proteins overall structure. This is the principle that allows protein homology modeling to predict a structure: sequences which are aligned may have accumulated mutations over time, but the overall fold structure will still likely be much more conserved than the sequence itself.

The PHYRE2 program (Protein Homology analogY Recognition Engine) was launched in 2011 as the sequel to the original PHYRE program and is one of the most popular protein prediction software used today, intaking nearly 1000 protein submissions each day. The PHYRE2 program will be explored in this module as a predictive tool for protein structure analysis. (Kelley, et all., 2015).

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

1. PURPOSE

The purpose of this procedure is to develop an understanding of how protein fold recognition software can be used to predict structural aspects of proteins based on their amino acid sequence alignment to better-studied sequences and structures.

2. SCOPE

This procedure applies to qualified skills center users.

3. RESPONSIBILITY

- 3.1. It is the responsibility of the user to understand and perform the procedure described in this document.
- 3.2. It is the responsibility of the user performing the procedure to fully document any deviations from the written procedure.
- 3.3. It is the responsibility of the user to become trained in the use of this application.

4. DEFINITIONS



- 4.1. Homology: Similarity due to shared ancestry between a pair of structures or genes in different taxa.
- 4.2. Domain: Distinct structural/functional unit in a protein. Typically responsible for a particular function or interaction, contributing to the overall role of a protein. Most full proteins are assembled of multiple domains that have their own independent fold and come together as a complex of multiple subunits.
- 4.3. Convergent Evolution: When two species not sharing a close common ancestor develop similar characteristics.
- 4.4. Divergent Evolution: When two species diverge from a common ancestor and develop different characteristics.
- 4.5. PHYRE2-Fold: Broad categories of protein tertiary structure topology.
- 4.6. PHYRE2-Family: A group of evolutionary related proteins.
- 4.7. PHYRE2-Superfamily: A group compiling multiple protein families, the families related more distantly than those within the family itself.

5. MATERIALS/EQUIPMENT

- 5.1. A computer to perform the procedure.
- 5.2. The PHYRE2 server.

6. PROCEDURE

In this procedure, the online protein fold recognition server PHYRE2 will be used to generate a 3D tertiary protein structure based on alignment with available sequences of

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

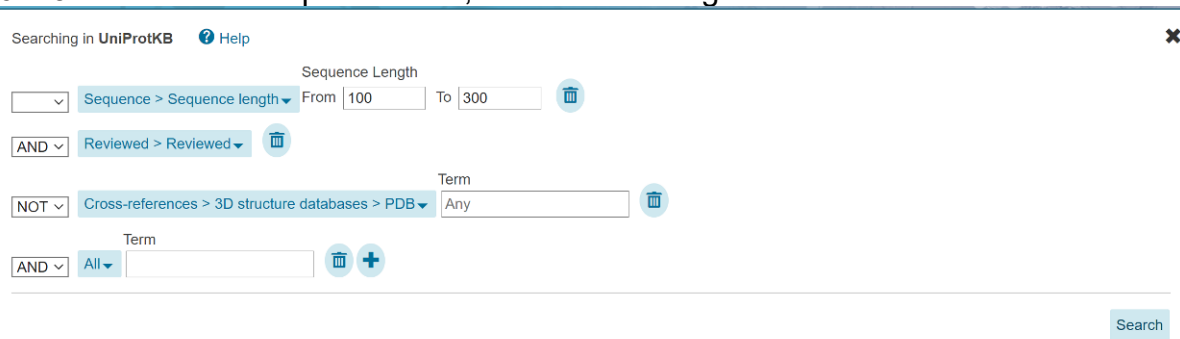
known structures. Other features of the PHYRE2 server will also be explored to show the many different functionalities of the protein fold recognition software.

6.1. Standard Mode: Predict Protein Structure

6.1.1. Find the amino acid sequence of a protein WITH UNKNOWN STRUCTURE to run for analysis using PHYRE2. To find proteins which have been studied and sequenced but lack an experimentally determined tertiary structure, first go to the UniProt database: <https://www.uniprot.org/>.

6.1.2. To conduct a search for proteins sequences which lack a structure, click the “Advanced” button next to the search bar.

6.1.3. For the search parameters, use the following:



Searching in UniProtKB Help

Sequence Length
 Sequence > Sequence length From 100 To 300

AND Reviewed > Reviewed

NOT Cross-references > 3D structure databases > PDB Term Any

AND All Term

Search

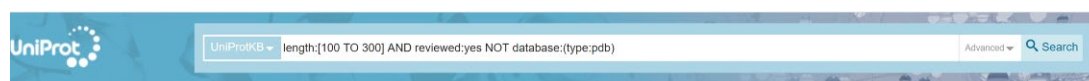
The sequence length parameters indicate proteins with a sequence of 100 to 300 amino acids will be shown.

The “AND” Reviewed indicates the proteins have been reviewed and studied.

The “NOT” Cross-referenced in 3D structure databases (PDB) indicates that a structure is not available for the protein on the PDB database, indicating that it is most likely not determined as the PDB database is one of the largest protein structure databases available.


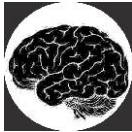
NOTE the “AND” and “NOT” options and make sure they are input so that the database is being searched for “AND Reviewed” and “NOT PDB”.

A shortcut for these parameters is the following: length:[100 TO 300] AND reviewed:yes NOT database:(type:pdb). These parameters can be directly typed into the UniProt search bar to yield proteins fitting these parameters.



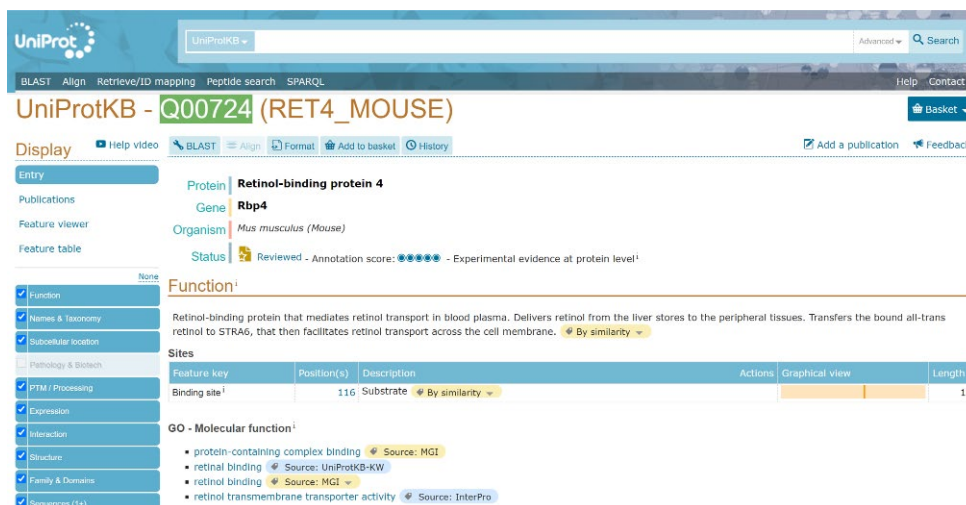
UniProtKB length:[100 TO 300] AND reviewed:yes NOT database:(type:pdb) Advanced Search

Once the parameters have been specified in the search bar, select “Search.”

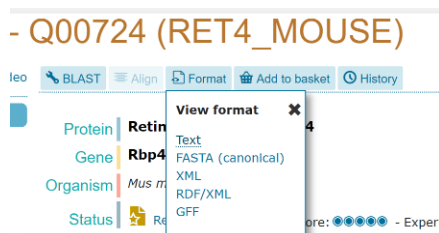
	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

6.1.4. Proteins matching the parameters will be displayed on the page. Browse through the protein options and select a protein of interest to use for PHYRE2. Select the protein by clicking its 6 character entry code (for example, Q00724.)

6.1.5. After clicking the entry code, the screen below for the individual protein will be shown.



This page shows the gene name and organism the protein belongs to, along with known information about the function of the protein and its role in biological processes. To access the sequence of the protein, select the “Format” button which has the sequence available in several formats. A couple different formats will be used in this module; but for the standard prediction, simply use the text format.




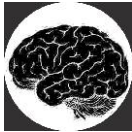
Select the “Text” format, underlined here.

6.1.6. Scroll to the very bottom of the text page and copy the amino acid sequence. Make sure to include JUST the sequence consisting of the one letter amino acid codes (shown below); no header, unlike one of the formats used later in this module.

```

SQ SEQUENCE 201 AA; 23206 MW; 8D187E293B37A75B CRC64;
MEWVHALVLL AALGGGSAER DCRVSSFRVK ENFDKARFSG LWYATAKKDP EGLFLQDNIT
AEFSVDEKGH MSATAGRRVR LLSNWEVCAD MVGTFDTED PAKFKMKYMG VASFLQRGND
DHWIIDTDYD TFALOYSCRL QNLDTGCADS YSFFVSRDPN GLSPETRRLV RQRQEEELCLE
RQYRWIEHNG YCQSRPSRNS L
//

```

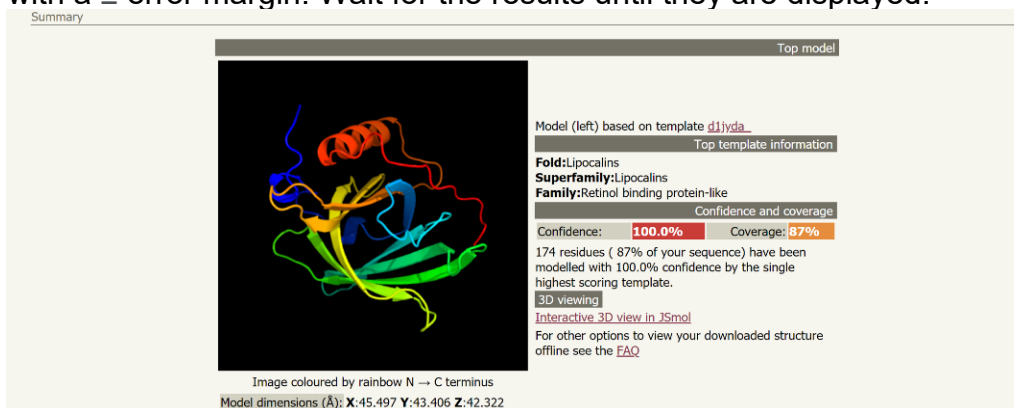
	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

6.1.7. Now that the amino acid sequence of the protein of interest has been copied, go to the PHYRE2 server to begin the homology search: <http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>.

6.1.8. On the top left corner, select the “Login” option. From here an account can be created. An account is not needed to run a standard search, but it is needed to run the expert search later. Put in an email address and a password will be sent to that email; from there you can login on the PHYRE2 server.

6.1.9. Run a standard search by pasting the amino acid sequence into the large input box shown beneath. Make sure to also include an email and have the modeling mode set to “normal.”

6.1.10. The search engine will give an estimate of how long it will take to run the results with a \pm error margin. Wait for the results until they are displayed.



Summary

Top model

Model (left) based on template [d1jyda_](#)

Top template information

Fold:Lipocalins
Superfamily:Lipocalins
Family:Retinol binding protein-like

Confidence and coverage

Confidence: **100.0%** Coverage: **87%**


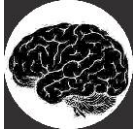
174 residues (87% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

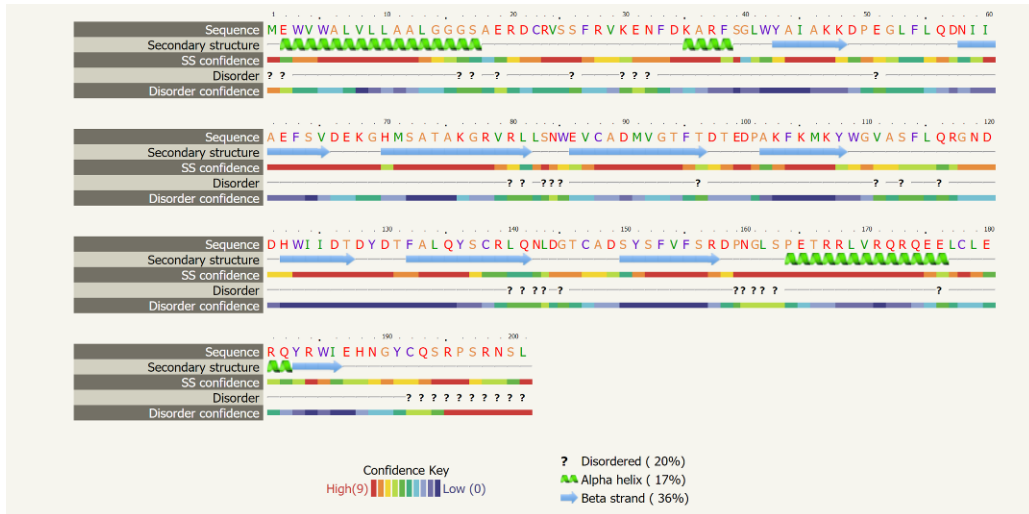
[3D viewing](#)
[Interactive 3D view in JSmol](#)
For other options to view your downloaded structure offline see the [FAQ](#)

Image coloured by rainbow N → C terminus
Model dimensions (Å): X:45.497 Y:43.406 Z:42.322

An image of the protein will be shown constructed from sequence alignment comparison with the database-the highest matching protein structure template which the result is based on is indicated: “d1jyda_.” You can select this template to view the protein domain which has a high sequence similarity to the query protein and view the structure. The “Fold”, “Superfamily”, and “Family” (see DEFINITIONS) of the protein are all shown under top template information. Here, a mouse retinol binding protein was analyzed; the server recognized the family to be Retinol binding protein-like, and the structure also to belong to the larger Lipocalins superfamily. Lipocalins are a family of proteins which transport lipophilic molecules such as retinol.

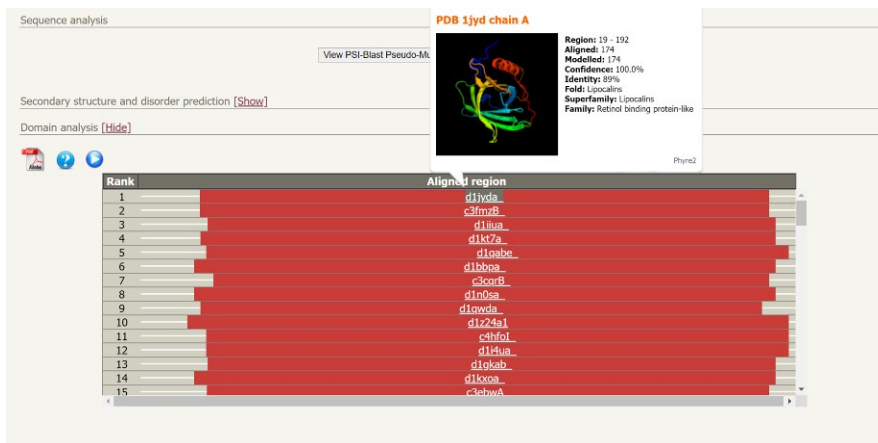
6.1.11. There are several more sequence analysis features that can be explored below the “Summary” result section. In the “Sequence analysis” section there is: “Secondary structure and disorder prediction”, “Domain analysis”, “Detailed template information”, and “Binding site prediction.” Go through each result by selecting the “show” feature, beginning with “Secondary structure and disorder prediction.”

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON 
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie




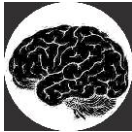
This tab gives a more detailed view of exactly which segments of the sequence correspond to which secondary structures. Here, the separate alpha helices, beta sheets, and disordered segments of Mouse RBP-4 are shown lining along their corresponding sequence segments.

6.1.12. The next tab, “Domain analysis”, gives a list of matching domains for the query protein.





The highest matching domains are listed. By hovering the cursor over the domain (here the cursor is over the top matching sequence for the query protein, d1jyda_) it will give statistics on this segment; by selecting the domain, the page will pull down to the next analysis section with more information about the specific domain.

6.1.13. “Detailed template information” gives more details about the proteins that were used as templates to determine the structure of the query protein, such as what exact

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

sequence it aligned with, the confidence in constructing the structure based off the sequence, the fold, protein family, etc.

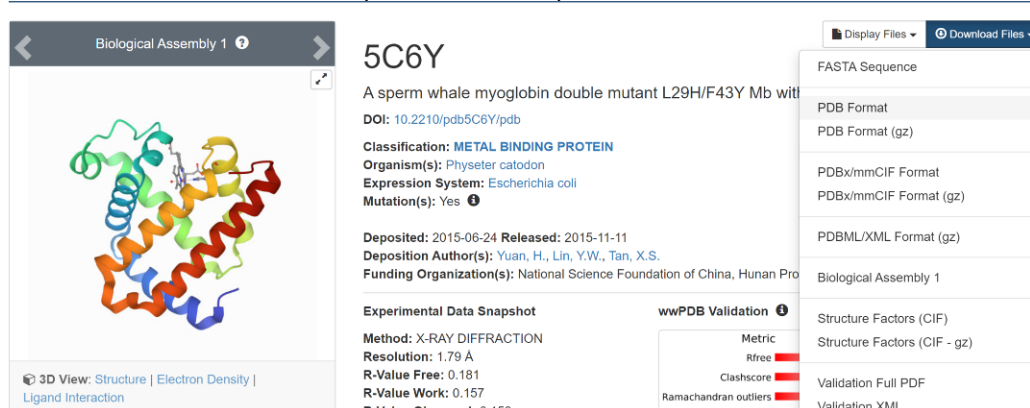
Detailed template information [Hide]

#	Template	Alignment Coverage	3D Model	Confidence	% I.d.	Template Information
1	d1hycd	Alignment		100.0	89	Fold: Lipocalins Superfamily: Lipocalins Family: Retinol binding protein-like Run Investigator
2	c1fmb6	Alignment		100.0	89	PDB header: transport protein Chain: B; PDB Molecule: retinol-binding protein 4; PDBTitle: crystal structure of retinol-binding protein 4 (rbp4) in2 complex with non-retinoid ligand Run Investigator

6.2. Expert Mode: One-to-One Threading

6.2.1. Expert mode offers several more functions for protein structure analysis. Select “Switch to Expert Mode” on the top left corner of the home page.

6.2.2. Begin with the first option, “One-to-one threading.” This feature will take both a sequence and a structure into account and attempt to create a structure based on both. This could be useful to users who wish to base their sequence prediction on a certain structure for reasons such as functional information about the protein or related proteins indicating the type of structure it may take on. Choose both a sequence (can be the same sequence used before) and a protein structure from the PDB database to upload: <https://www.rcsb.org/>. Search a protein and select its 4 character PDB code (for example, 5C6Y); then select “Download files”, and choose “PDB Format.” (shown below)



Biological Assembly 1

5C6Y

A sperm whale myoglobin double mutant L29H/F43Y Mb with

DOI: 10.2210/pdb5C6Y/pdb

Classification: METAL BINDING PROTEIN
 Organism(s): Physeter catodon
 Expression System: Escherichia coli
 Mutation(s): Yes

Deposited: 2015-06-24 Released: 2015-11-11
 Deposition Author(s): Yuan, H., Lin, Y.W., Tan, X.S.
 Funding Organization(s): National Science Foundation of China, Hunan Pro

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
 Resolution: 1.79 Å
 R-Value Free: 0.181
 R-Value Work: 0.157
 R-Value Observed: 0.158


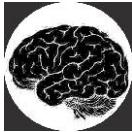
wwPDB Validation

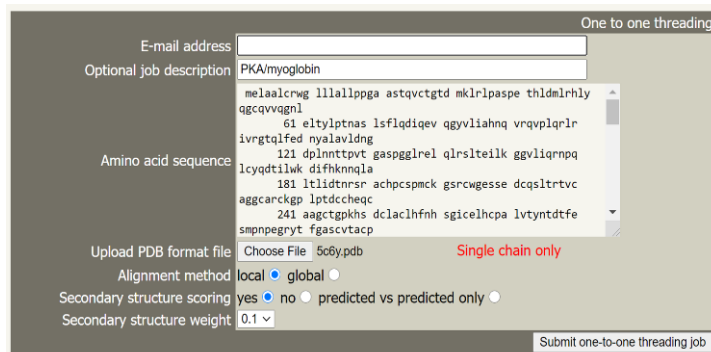
Metric: Rfree, Clashscore, Ramachandran outliers

Download Files

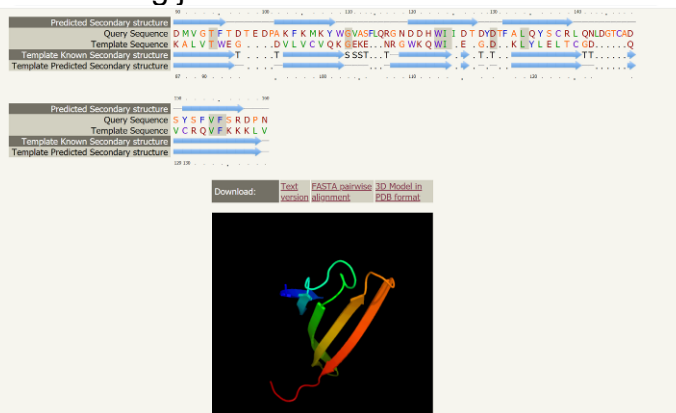
- FASTA Sequence
- PDB Format
- PDB Format (gz)
- PDBx/mmCIF Format
- PDBx/mmCIF Format (gz)
- PDBML/XML Format (gz)
- Biological Assembly 1
- Structure Factors (CIF)
- Structure Factors (CIF - gz)
- Validation Full PDF
- Validation XML

6.2.3. Once an amino acid sequence and a protein structure (PDB file) have been chosen, they can be input into the one-to-one threading search as shown below.


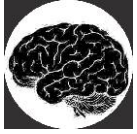
	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie



Make sure “alignment method” is set to local and “secondary structure scoring” is set to yes so the results will show analysis of the secondary structures present in the predicted structure. The “secondary structure weight” refers to how much of the predicted structure will be based on the uploaded PDB structure; values of 0.1 to 0.9 are available as choices of how much influence the user wishes the PDB structure to have on the prediction of the input structure. Choose a middle secondary structure weight value (i.e. 0.5) so the secondary structure has an intermediate amount of influence on the predicted structure of the sequence. Click “Submit one-to-one threading job.”



The results will indicate what sequence segments correspond to which secondary structures. This is a threading job of the Mouse RBP-4 done against a Human RBP-2; the homologous similarity between the two proteins yielded a stable structure, with a 98% confidence value. A one-to-one threading job of Mouse RBP-4 done against a functionally dissimilar protein, keratin, yielded a confidence value of 0% with an essentially impossible linear structure. Forming a structure of the globular protein RBP-4 was nearly impossible using a 0.5 secondary weight structure of the fibrous keratin because the amino acid sequence of RBP-4 contains amino acid side chains which do not have the properties to take an alpha helical, fibrous form.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

6.3. Expert Mode: Batch Processing

6.3.1. The next option is “Batch processing.” This can be used to predict the topology of up to 100 proteins at once. For this option, submit a file containing the FASTA format of multiple proteins; to find the FASTA format versions of protein sequences on the UniProt database, follow the same steps as before (6.1.1. to 6.1.5.) except select “FASTA” instead of “Text” when choosing a protein format.



6.3.2. Copy the ENTIRE page of text and paste it into a separate note/text file. Do this for multiple different protein sequences but make sure the entire page of text is highlighted and copy/pasted, as the FASTA header will serve to separate the proteins in the single file.

```
>sp|Q00724|RET4_MOUSE Retinol-binding protein 4 OS=Mus musculus OX=10090 GN=Rbp4 PE=1 SV=2
MEIWWALVLLAALGGGSAERDCRVSSFRVKNFKARFSGLWYATAKKDPEGLFLQNNI
AEFSVDEKGHMSATAKGRVRLLSNWEVCADMVGTFFDTEPAKFKMYGVAASFLQKND
DHWZIDTDYDTFALQYSCLQLQLDGTCAOSYSFVSRDPNGLSPETRLRVQRQEECLCE
RQYRWIEHNGYQCSRPSRNSI
```

6.3.3. Once a file with multiple different protein sequences has been created, upload the file and select “Submit batch job.” The confirmation page below will be shown:

Batch Submission

User	michaeljguzie@gmail.com
Your maximum allowed sequences	100
Number submitted	3



Your sequences are listed below. Please check that they have been parsed correctly

Sequences must be between 30 and 6000 residues long and only contain the 20 canonical amino acids (no gaps, no wild cards, no nucleic acids)

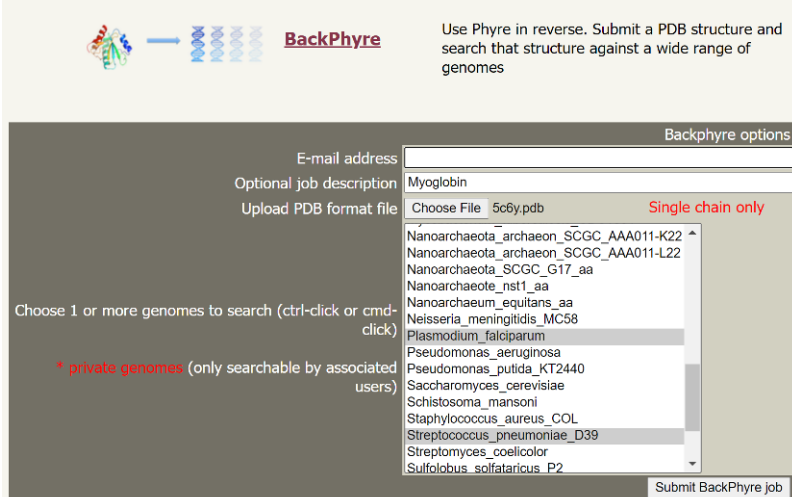
Seq. No.	Sequence descriptor	Sequence
1	sp Q00724 RET4_MOUSE Retinol-binding protein 4 OS=Mus musculus OX=10090 GN=Rbp4 PE=1 SV=2	MEIWWALVLLAALGGGSAERDCRVSSFRVKNFKARFSGLWYATAKKDPEGLFLQNNI AEFSVDEKGHMSATAKGRVRLLSNWEVCADMVGTFFDTEPAKFKMYGVAASFLQKND DHWZIDTDYDTFALQYSCLQLQLDGTCAOSYSFVSRDPNGLSPETRLRVQRQEECLCE RQYRWIEHNGYQCSRPSRNSI
2	sp P00594 PA21B_HORSE Phospholipase A2 (Fragment) OS=Equus caballus OX=9796 GN=PLA2G1B PE=1 SV=1	ENIEISPRVAVVQPSMIOCTIPNSKYLLENVDYGCYCGGGSTPVEILDACCQYHRCYQAKELSSCRFLVDMPTYESYKPSGTEVYTCSDKN ACEAFZCQDRNALDCSPYWFENILDSRGLCA
3	sp P31582 RAF2A_ARATH Ras-related protein RABF2a OS=Arabidopsis thaliana OX=3702 GN=RABF2A PE=1 SV=1	MSGGKNNKMLVLDGAGAKSSLLSFKDKQRFREFQSTIGAAFTSQTLVNDATWKEEWDTAGQERHSLAPHYHGAAGADYFQETNQS FERAKWVQELQKQNPMMYMALAGKAKLLDARKVSAEAEIYQENLFFMETSAKTATNWDIYELAKRLPQVPAENFTQRYLPRGPGAT AVSSSCCA

Make sure that the system has recognized all the different protein sequences in the file as separate sequences, which should be the case if the entire FASTA header was pasted in with the sequence. If all the sequences are identified separately, hit “Confirm batch submission.” The results will yield predicted structures of each protein.

6.4. Expert Mode: BackPhyre

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

6.4.1. The next option is “BackPhyre”-this is the reverse process. Upload a PDB structure and the protein structure will be ran through the database to find possible sequences which best fit this structure. Upload a PDB file in the same format as in 6.2.2. There are several different available genomes from a variety of different organisms; select at least two genomes to compare this structure against and hit “Submit BackPhyre Job.”



BackPhyre

Use Phyre in reverse. Submit a PDB structure and search that structure against a wide range of genomes

Backphyre options

E-mail address

Optional job description

Upload PDB format file 5c6y.pdb Single chain only

Choose 1 or more genomes to search (ctrl-click or cmd-click)

* private genomes (only searchable by associated users)

Nanoarchaeota_archaeon_SCGC_AAA011-K22

Nanoarchaeota_archaeon_SCGC_AAA011-L22

Nanoarchaeota_SCGC_G17_aa

Nanoarchaeote_nst1_aa

Nanoarchaeum_equitans_aa

Neisseria_meningitidis_MC58

Plasmodium_falciparum

Pseudomonas_aeruginosa

Pseudomonas_putida_KT2440

Saccharomyces_cerevisiae

Schistosoma_mansonii

Staphylococcus_aureus_COL

Streptococcus_pneumoniae_D39

Streptomyces_coelicolor

Sulfolobus_solfataricus_P2

This BackPhyre contains a PDB file and 2 genomes selected to run the myoglobin structure against; including the bacteria which causes pneumonia, *streptococcus pneumoniae*.

6.4.2. The results will be emailed and show the genomes which hit a match to the submitted structure and how much of a matching sequence was present.


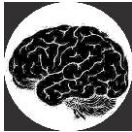
6.5. Expert Mode: GenSearch

6.5.1. Lastly, “GenSearch” can be used to run a sequence of interest against a variety of genome databases. Paste an amino acid sequence into the input box.

6.5.2. Choose at least 2 genomes to run the sequence against from the available organisms.

6.5.3. Select “Submit GenSearch”.

6.5.4. The results will be emailed and show the genomes containing a protein which hit a match to the submitted amino acid sequence and how much of the matching sequence was present.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

Rank	Description	Organism	Alignment Coverage	Confidence	% i.d.
1	gij55743122 ref NP_006735.2 retinol-binding protein 4, plasma precursor [Homo sapiens]	<i>Homo sapiens</i>	1-201 	100.0	86
2	gij28574799 ref NP_787960.1 CG33126-PA [Drosophila melanogaster]	<i>Drosophila melanogaster</i>	1-192 	100.0	13
3	gij24653274 ref NP_523727.2 CG4604-PA [Drosophila melanogaster]	<i>Drosophila melanogaster</i>	3-195 	100.0	17
4	gij24580854 ref NP_722702.1 CG31659-PA [Drosophila melanogaster]	<i>Drosophila melanogaster</i>	1-192 	100.0	13
5	gij24641392 ref NP_727559.1 CG4139-PA, isoform A [Drosophila melanogaster]	<i>Drosophila melanogaster</i>	4-194 	100.0	15
6	gij24641394 ref NP_572748.2 CG4139-PB, isoform B [Drosophila melanogaster]	<i>Drosophila melanogaster</i>	4-194 	100.0	15
7	gij4502163 ref NP_001638.1 apolipoprotein D precursor [Homo sapiens]	<i>Homo sapiens</i>	1-194 	100.0	22
8	gij38455402 ref NP_005555.2 lipocalin 2 (oncoogene 24p3) [Homo sapiens]	<i>Homo sapiens</i>	1-192 	100.0	16
9	gij89886464 ref NP_848631.2 lipocalin 12 [Homo sapiens]	<i>Homo sapiens</i>	1-197 	100.0	15
10	gij32171249 ref NP_000945.3 prostaglandin H2 D-isomerase [Homo sapiens]	<i>Homo sapiens</i>	1-195 	100.0	16
11	gij4557393 ref NP_000597.1 complement component 8, gamma polypeptide [Homo sapiens]	<i>Homo sapiens</i>	3-201 	100.0	21
12	gij42714611 ref NP_976222.1 MSFL2541 [Homo sapiens]	<i>Homo sapiens</i>	1-197 	100.0	17
13	gij4502067 ref NP_001624.1 alpha-1-microglobulin/bikunin precursor [-Homo sapiens]	<i>Homo sapiens</i>	1-197 	100.0	19
14	gij39573716 ref NP_945184.1 lipocalin 6 [Homo sapiens]	<i>Homo sapiens</i>	1-181 	99.9	15


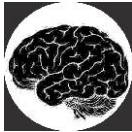
These are the results for mouse RBP-4 compared to 6 different genomes, including *Homo sapiens*. The results show what proteins from what organism hit a match, and what sequence alignment is present. *Homo sapiens* and *Drosophila melanogaster* show the highest number of matching proteins.

7. TROUBLE SHOOTING

7.1. If trying to use the sequence finder function rather than inputting your own sequence, the search engine may seem perpetually stuck in the searching que. The sequence search feature is in the beta-testing phase so it may take awhile to load; it will be easiest to search the sequence of the protein on the UniProt database or another database and past it directly in the search box.

7.2. The “Binding site prediction” tab may be unavailable due to server load. An alternative ligand binding site is offered: <http://www.sbg.bio.ic.ac.uk/3dligandsite/>. For more on ligand-protein binding models, see the *In-silico* Drug Analysis 1 Skills Center Standard Operating Procedure on SwissDock.

7.3. For one-to-One threading, only one chain can be processed at a time, so if you upload a PDB file with multiple chains you will be prompted to select only one of the chains. Whichever chain is selected will be the chain which is used as a guiding structural template to make a structure prediction.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

Your PDB file contains multiple (2) chains

One to One Threading can only process 1 chain at a time. Please select the chain you wish to use.

Chain:	C TAGRHGDSVRNSKIEISELNRIQLRSEIDNVKKQISNLQQSISDAEQRGENALKDAKNKLNLDLEDALQQAKEDLARL
<input checked="" type="radio"/> Length:	LRDYQELMNTKLALDLEIATYRTLLEGE
	107
Chain:	X TEIDNNEIQISSYKSEITELRRNVQALEIEIQSALKQSLASLAETEGRYAVQLSQIQAQISALEEQQIRAETEC
<input type="radio"/> Length:	QNTHEYQLLDIKIRLENEIQYRSLLEGE
	108

7.4. For batch-processing, if the server is only recognizing one protein rather than the whole batch submitted, make sure the entire FASTA header has been copied and pasted into the batch text file and not just the amino acid sequence itself.

8. REFERENCES

Chen, Denise. (2009, December). *Structural genomics: Exploring the 3d protein landscape*. Retrieved from <http://www.bcr.org/content/structural-genomics-exploring-3d-protein-landscape>.

Pitman, M. R., & Menz, I. (2006). *Homology modeling*. Retrieved from <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/homology-modeling>.


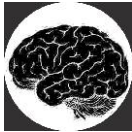
Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015, June). *The Phyre2 web portal for protein modeling, prediction and analysis*. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5298202/#:~:text=Phyre2%20is%20a%20suite%20of,protein%20structure%2C%20function%20and%20mutations.&text=This%20protocol%20will%20guide%20a,domain%20composition%20and%20model%20quality>.

9. MODULE MASTERY TASK

Use all the PHYRE2 protein fold recognition programs to attain structural information about an amino acid sequence.

9.1. Standard Mode: Predict Protein Structure

9.1.1. Using the set of one-letter abbreviations for amino acids, create a random amino acid sequence to submit for structural analysis. Make sure the sequence is at least 100 amino acids long. Submit the sequence to the standard mode PHYRE2 engine

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Protein Homology Modeling	Effective Date: 03/05/2021 Checked by A. Siclair	Revision # 1.0 M. Guzie

and record the results of what structure is recorded from your random sequence, and what secondary structures correspond to what regions of the sequence. Also make note of what domain is the highest match to the random sequence and how well it matches; what general function does this domain serve?

9.1.2. Select a protein with an unidentified structure (using the parameters described in section 6.1.) that may have a therapeutic implication (for example, a RAS-related protein as RAS mutations can have implications with cancerous cell division.) Run a prediction for this sequence. What can PHYRE2 offer as a therapeutic tool?

9.2. Expert Mode: One-to-One Threading

9.2.1. Based on literature, choose two proteins with very different functions. Upload the PDB structure of one, and the amino acid sequence of another (with unknown structure.) First run the job using a secondary weight of 0.1, then 0.5, then 0.9. Note with the increasing influence of secondary structure how the protein's morphology changed. For the job with 0.1 secondary structure weight, note if the structure still leans more towards your uploaded PDB structure or if it takes on a different morphology which may be more characteristic of your amino acid sequence.

9.3. Expert Mode: Batch Processing

9.3.1. Choose at least 5 different protein sequences and submit them for structural analysis to batch processing. Generally compare the resulting structure of each protein to its amino acid sequence.

9.4. Expert Mode: BackPhyre

9.4.1. Use literature to choose a protein which seems to be relatively conserved amongst living organisms (some form present in many species today; for example, myoglobin.) Submit the protein structure to BackPhyre as a PDB file for search against several genomes; select at least 5 different organism's genomes. Make note of several genes which match the protein structure you submitted and what this means about these gene sequences. Make a short descriptive sentence for each one on what you can generalize about their structure/commonalities based on the results.

9.5. Expert Mode: GenSearch

9.5.1. Choose an amino acid sequence to compare for sequence similarity against at least 5 genomes. Make a note of at least one match from each species (if there is one) and how well they match to your amino acid sequence. Indicate what the amount of matches indicates about your query amino acid sequence. What beneficial uses can you think of that come with the GenSearch tool?