



Sep 08, 2020

# Prediction of ligand binding using FunFOLD2

In 1 collection

Chris Berndsen<sup>1</sup><sup>1</sup>James Madison University

In Development

This protocol is published without a DOI.

Chris Berndsen  
James Madison University

## ABSTRACT

Clues to functionality can be gleaned from comparing unknown or predicted structures with previously characterized structures of known function and characteristics. These scans can be biased against novel proteins or proteins with similar structures but distinct functions, but for initial guesses can be powerful. These methods align the structure and/or amino acid sequence to a database of structures with known ligand binding sites and look for structures with the similarity in amino acid composition, position, and overall 3-D similarity. The idea being that similar structures lead to similar functions. This protocol describes the use of the FunFOLD2 server which predicts ligand identity and binding site from the amino acid sequence.

## THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Roche, D. B., Buenavista, M. T., and McGuffin, L. J. (2013) The FunFOLD2 server for the prediction of protein-ligand interactions. *Nucleic Acids Res.* 41, W303–7.

## PROTOCOL CITATION

Chris Berndsen 2020. Prediction of ligand binding using FunFOLD2. **protocols.io**  
<https://protocols.io/view/prediction-of-ligand-binding-using-funfold2-bkrkv6n>

## MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Roche, D. B., Buenavista, M. T., and McGuffin, L. J. (2013) The FunFOLD2 server for the prediction of protein-ligand interactions. *Nucleic Acids Res.* 41, W303–7.

## COLLECTIONS ⓘ

 **Biochemistry I methods**

## LICENSE

— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

## CREATED

Sep 03, 2020

## LAST MODIFIED

Sep 08, 2020

## PROTOCOL INTEGER ID

41491

## PARENT PROTOCOLS

Part of collection

[Biochemistry I methods](#)

## MATERIALS TEXT

A protein sequence in one letter code

Internet connection

## Uploading sequence

- 1 Navigate to the FunFOLD2 server submission form.
- 2 Paste in your sequence in single letter amino acid code.

Input sequence of protein target (in single letter amino acid code) [Sample sequence](#)

```
MGEILAVDDYVGISFWLAAAIMLASTVFFVERSDDVPVKWKTSLTVAGLVTGVAFWHY
LYMRGVNIYAGETPTVFRYIDWLITVPLQIIEFYLIIAAVTAISSAVFWKLLIASLVM
LIGGFIGEAGLGDVVVWIVGMIAWLYIIEIFLGETAKANAGSGNAASQQAFTIKW
IVTVGMIAIPIGYAWGYFGDGLNEDALNIVYNLADLINKAAFGLAIAAAMKDKETST
SHA
```

E-mail address (optional) [Help](#)

Short name for protein target (optional) [Help](#)

References

Reset ▶

Predict ▶

- 3 Provide an email address and a short name for your protein. Record the short name as a note in this step.
- 4 Press Predict to submit your sequence.

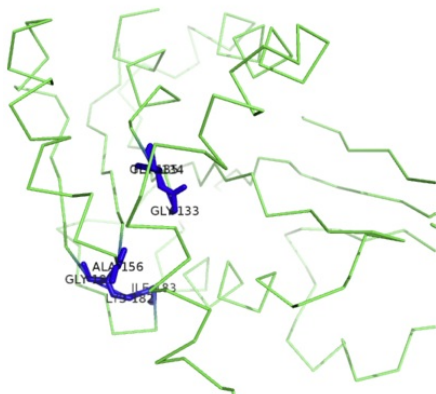


FunFOLD takes 24 to 72 hours to respond. You should record the results link as a note in this step in case you don't get the email.

## Analyzing data

- 5 When the results are ready there will be four areas of results.
- 6 At top is the predicted binding site in a homology model. The binding site is shown and labeled to be the amino acid numbers in the submitted sequence.  
  
Download the structure and view it in YASARA.

## PyMOL generated image of ligand binding residues prediction for test



[Click here to download PDB file of this model with the superposition of all identified ligands.](#)

### 6.1 Save your downloaded file as

```
FunFOLD2_[Group_name]_[sequence_name].pdb
```

Replace **[Group\_name]** with your name/group name without the brackets. Replace **[sequence\_name]** with the name of the sequence.

### 6.2 Indicate your OSF file location as a link within a note on this step.

- 7 Below this is information on the ligand identity and the binding site. The ligand identity is abbreviated as it would be in a crystal structure.

**Predicted ligand binding residues are shown as blue sticks in the image above.**

**Binding site: 133, 134, 135, 156, 181, 182, 183**

**Most likely ligand (Type): SAH**

**Centroid ligand (Type): SAH**

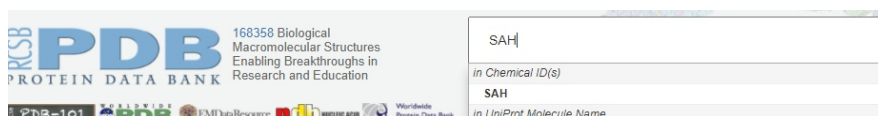
**All ligands clustered at site (Type-Number): SAH-6**



If there are multiple potential ligands, you will need to search each individually.

- 7.1 To identify the ligand, in a new browser window go to the RCSB ([www.rcsb.org](http://www.rcsb.org)) and in the search bar, look for your ligand name and wait for the search options to drop down (do not press enter, just input the code and wait).

Select in Chemical IDs to search the ligands database.



7.2 The search should take you to an information page on the ligand lots of information and the name.

**SAH**  
S-ADENOSYL-L-HOMOCYSTEINE

Find entries where:  
SAH is present as a standalone ligand  
SAH as a non-polymer is covalently linked to polymer or other heterogen groups  
SAH is present in a polymer sequence

Find related ligands:  
Similar Ligands (Stereo-Specific)  
Similar Ligands (incl' Stereoisomers)  
Similar Ligands

View summary at Ligand Expo

Chemical Component Summary	
Name	S-ADENOSYL-L-HOMOCYSTEINE
Identifiers	(2S)-2-amino-4-[[[(2S,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-dihydroxy-oxolan-2-yl]methylsulfanyl]butanoic acid
Formula	C <sub>14</sub> H <sub>20</sub> N <sub>6</sub> O <sub>5</sub> S
Molecular Weight	384.41
Type	L-PEPTIDE LINKING
Isomeric SMILES	<chem>N[C@@H](CCSC[C@H]1O[C@H](C[C@H](O)[C@@H]1O)n2cnc3c(N)ncnc23C(=O)O</chem>
InChI	InChI=1S/C14H20N6O5S/c15-6/(14/23/24)/1-2-26-3-7-9/2110/2213/25-7/20-5-19-8-11/16/17-4-18-12/8/20/h4-

Chemical Details	
Formal Charge	0
Atom Count	46
Chiral Atom Count	5
Bond Count	48
Aromatic Bond Count	9

7.3 Record the ligand binding residues and other statistics as well as the ligand identity. Add rows as needed.

Ligand binding residue(s)	
Ligand code(s)	
Ligand identity(ies)	

8 Below the ligand information is a JSmol window for viewing the ligand and binding site in the browser. This is convenient for when you do not have a molecular visualization software available.

#### JSmol view of ligand binding residues prediction for test

**Predicted binding residues - display options**

- ☒ High detail (unchecked for faster rendering on phones/tablets)
- ☒ Show residue labels
- ☒ Wireframe on
- Spacefill options: ☒ off ☐ 25% ☐ 100%
- Zoom options: ☐ 100% ☒ 200% ☐ 300% ☐ 400%

**Predicted ligands - display options**

- ☒ Wireframe on
- Spacefill options: ☒ off ☐ 25% ☐ 100%
- ☐ Spin model

**Basic mouse controls:**

Zoom: SHIFT + click and hold the left mouse button and move up or down.  
Rotate: click and hold the left mouse button and drag.  
Translate: CTRL + click and hold the right mouse button and drag.

9 The last section contains the quality of the fit.

### FunFOLDQA scores:

**BDTalign = 0.33695015486337826**

**Identity = 0.19236092**

**Rescaled BLOSUM62 = 0.2044577**

**Equivalent Residue Ligand Distance = 0.31665778**

**Model Quality = 0.86413**

**Predicted BDT score = 0.5674837**

**Predicted MCC score = 0.4127715**

#### 9.1 Explanation of scores:

<b>BDTalign</b>	Comparison of the structural match in 3-D between known binding site and modeled binding site. 1 indicates a perfect prediction, while 0 is associated with a random and unreliable prediction.
<b>Identity</b>	Number of amino acids in the binding site that are equivalent to the known binding site. 1 is a perfect match, 0 is imperfect match
<b>Rescaled BLOSUM62</b>	An additional measure of binding site sequence match using the BLOSUM62 algorithm. 1 is a perfect match, 0 is an imperfect match.
<b>Equivalent Residue Ligand Distance</b>	For the amino acids that are equivalent, this measure looks at how well the distances match. 1 is a perfect match, 0 is an imperfect match.
<b>Model Quality</b>	Score based on the ModFOLDclust2 algorithm to give a sense of the overall match of the model. 1 is a perfect match, 0 is an imperfect match.
<b>Predicted BDT score</b>	Binding site distance test indicates how far a predicted binding site residue is from the observed binding residue location. 1 indicates a perfect prediction, while 0 is associated with a random and unreliable prediction.

### Predicted MCC score

Matthews Correlation Coefficient score is a statistic for looking at accuracy of predicted vs. observed binding site amino acids. 1 indicates a perfect prediction, while 0 is associated with a random and unreliable prediction.

Information from Roche, D. B., Buenavista, M. T., and McGuffin, L. J. (2012) FunFOLDQA: a quality assessment tool for protein-ligand binding site residue predictions. *PLoS One* 7, e38219.

## 9.2 Record the scores of the prediction below:

<b>BDTalign</b>	
<b>Identity</b>	
<b>Rescaled BLOSUM62</b>	
<b>Equivalent Residue Ligand Distance</b>	
<b>Model Quality</b>	
<b>Predicted BDT score</b>	
<b>Predicted MCC score</b>	

- 10 If you have homology models from other methods (Phyre2, SWISS-Model, etc.) and refined any of those models, it is worth aligning the model with the ligand bound from [go to step #6](#) to all of these models to compare the predicted binding site.

Different programs may predict different conformations for the binding site and by comparing the different models, the reliability of the prediction can be further assessed. If there are distinct binding site shapes, then the prediction may be less reliable.