

Tutorial: using SIFTED to predict the specificity of TALE proteins

1. Getting started

The SIFTED suite is implemented as a set of online tools within Galaxy, an online bioinformatics platform. If you are already familiar with Galaxy, you should find SIFTED very intuitive. Even if you are not, this tutorial will guide you through all the necessary steps. If you would like to learn more about Galaxy before using SIFTED, you can find an excellent tutorial here:

<https://usegalaxy.org/u/aun1/p/galaxy101>

2. Formatting data

The first step is to prepare a description of the protein(s) whose specificity you would like to predict. For simplicity, SIFTED's input is the target DNA sequence that would be predicted by the TALE code, using the following rules:

NI = A
HD = C
NN = G
NG = T

For example, if you would like to predict the specificity of a TALE protein with the following sequence of RVDs:

HD-NN-NN-NG-NN-NI-HD-HD-NN-NG-NN-HD

You would enter the corresponding DNA binding site (including the 5' T):

TCGGTGACCGTGC

3. Loading data into Galaxy

To start using SIFTED, simply click on "SIFTED Suite" at the top left of the screen, which will open the tools panel. Click on "Upload File" and either select the FASTA file containing the sequence you want to use or paste it in the "URL/Text" box. (A FASTA file is just a text file with a header line that starts with ">" and the DNA sequence starting in the next line). In addition, you should select "fasta" in the File Format box. Then simply enter a name(s) for your protein along with the DNA binding site, as described in step 2. Finally, click the 'Execute' button at the bottom of the page. **Note that you can input multiple proteins this way, along as each has its own separate header line.**

Upload File (version 1.1.4)

File Format:

fasta

Which format? See help below

File:

Choose File No file chosen

TIP: Due to browser limitations, uploading files large method (below) or FTP (if enabled by the site administrator)

URL/Text:

```
>TALE1
TCGGTGACCGTGC
```

4. Specificity prediction

To predict the specificity of the proteins that were entered in step 3, select the "Predict TALE PWM" option from the left menu.

Tools

search tools

SIFTED Suite

[Upload File from your computer](#)

[Generate candidate TALEs for targeting a given sequence.](#)

[Predict TALE target sequences](#)

[Predict TALE PWM](#)

[Predict genomic targets](#)

Predict TALE PWM (version 0.03)

Select a FASTA file that describes the TALEs to be used.:

1: Pasted Entry



Execute



This module of SIFTED takes a FASTA file containing TALE target sequences. An example input file would look like this (without blank lines):




```
> TALE1
```

Make sure the selected FASTA file is the one that contains the sequences that you entered on step 2. Your output should appear on the right, as shown in the figure below. Click the highlighted button to download the output file. The predicted specificity will be described in terms of a standard position frequency matrix (or position weight matrix). For more information about this format, see the following article:




http://en.wikipedia.org/wiki/Position_weight_matrix

History  

Unnamed history
434 bytes  

3: Predicted TALE PWMs   

5 lines
format: **text**, database: ?

```
>TALE1
0.0271 0.4359 0.1808 0.3268 0.1301
0.0457 0.5501 0.0545 0.0389 0.0644
0.0578 0.0050 0.7212 0.6212 0.0445
0.8694 0.0090 0.0435 0.0132 0.7611
```

