

# About LogoMat-M

LogoMat-M is a tool for visualization of *profile HMMs*. It creates so-called *HMM Logos*. HMM Logos are closely related to *sequence logos* invented by T. D. Schneider and R. M. Stephens.

## Overview

Sequence logos display the information content and the distribution of nucleotides or proteins from multiple alignments (see [LogoMat](#) if you wish to generate sequence logos). Profile HMMs on the other hand are a probabilistic model frequently used in computational biology to represent sequence families. They share a lot with multiple alignments or position specific scoring matrices. This is especially that every position of the sequence family is represented by a different distribution over the alphabet. The main difference between profile HMMs and other methods is the modeling of insertions and deletions. Every position in an HMM consists of a Match, Insert and Delete state, each connected to the next position's states through transitions of altering probability. For more information about profile HMMs, try HMM Logos try to provide the researcher with a quick overview of the features of a profile HMM while conserving as much information as possible.

## How to read HMM Logos

An HMM Logo looks quite a bit like a sequence logo. You'll see stacks of letters of different height. The height of the stack represents the *relative entropy* of the distribution of the emission probabilities within some state relative to the *background distribution* given for the whole profile. The relative size of a letter then expresses its emission probability from a state's distribution. In other words: The larger the letter, the more information it provides about the respective position in the protein family. Letters are sorted in descending order depending on their probability.

If you take a closer look, you'll realize that the width of the columns differ. This is to express the *relative contribution* of the position to the overall protein family. The different transition probabilities lead to varying probabilities of visiting different insert or match states at all. If a state is bypassed somehow, it will not emit. If a state has a self loop on the other hand, it can emit several letters. The width of the columns is a representation of both aspects, hitting probability as well as expected length of emitted sequence. Note that in HMMer HMMs, match states always emit exactly 1 letter.

Apart from the white columns, there are pink ones as well. These will most likely not contain any letters: These are the insert states, which normally have an emission distribution very close or equal to the background. Thus, the information content is usually so small that no letters can be seen. Insert states regularly have a probability less than 1 of reaching them. In HMM

Logos, they are still shown with a minimum width of 1 pixel to structurise the plot. Width are calculated relative to the overall width of the picture, so the wider the picture, the more differences can be seen.

Some insert states have a large expected length of emitted sequence, i.e. a strong self loop. To distinguish between the hitting probability and the expected contribution, two shades of pink are used. The darker color on the left corresponds to the hitting probability.

The amino-acid residue colours in logomat-m are reflect their respective biological properties:



## Options

### Descriptions

You can provide your own titles for the logo as well as the x- and y-axis. By default, the name of the family as given by the uploaded file is used as the title. If you don't want this, deselect "*use profile name from file*".

### Image

By default, the width of the image is calculated as 1/12 of the entered height for

every position in the sequence. If you want a fixed width, deselect "*calculate dynamically*".

To plot the image in grey for use in non-colour publications, select "*greyscale*".

## Split Logo

As long as "*show all positions at once*" is selected, the logo reaches over all positions in the profile. You can deselect "*show all positions at once*" and enter your own start and end positions. This will show a logo only displaying the positions between "*start from position*" (inclusive) and "*end with position*" (exclusive). If you want to browse through a profile, enter a number of positions that should be displayed per page. This will show a page with 2 buttons "*previous*" and "*next*" above the logo.

## Upload a file in HMMer-format

Currently, LogoMat-M reads files in plain HMMER2.0 format only.

## Literature

- [Sean R. Eddy. HMMER User's Guide: Biological sequence analysis using profile hidden Markov models, version 2.2. Washington University School of Medicine, 2001](#)

### **HMM Logos for visualization of protein families.**

*Schuster-Böckler B, Schultz J, Rahmann S*

BMC Bioinformatics. 2004;5;7. PMID: [14736340](#) DOI: [10.1186/1471-2105-5-7](#)

### **Sequence logos: a new way to display consensus sequences.**

*Schneider TD, Stephens RM*

Nucleic Acids Res. 1990;18;6097-100. PMID: [2172928](#)

### **Hidden Markov models in computational biology. Applications to protein modeling.**

*Krogh A, Brown M, Mian IS, Sjölander K, Haussler D*

J Mol Biol. 1994;235;1501-31. PMID: [8107089](#) DOI: [10.1006/jmbi.1994.1104](#)

- R. Durbin, S. Eddy, A. Krogh and G. Mitchison. Biological Sequence Analysis. Cambridge University Press, 1998