

About LogoMat-P

LogoMat-P is a tool for creation and visualization of *HMM alignments*. It creates *pairwise HMM Logos* by first aligning the HMMs with *prc*. Pairwise HMM Logos are closely related to HMM Logos.

Overview

HMM-Logos display the information content, the distribution of nucleotides or proteins and the relative contribution of each position from Hidden Markov Model files (see Logomat-M if you wish to generate HMM-Logos).

Pairwise HMM-Logos are an extension of HMM-Logos to visualise alignments of HMMs. This is possible by using a program called *prc*. It finds the paths through each HMM that are most similar to each other in terms of emission and transition probabilities.

How to read HMM Logos

A pairwise HMM Logo is very similar to an HMM-Logo. The pairwise HMM-Logo consists of 2 HMM-Logos, each representing one of the aligned families. The states that were paired by *prc* are framed in black and linked together by a line. Unaligned states are simply shaded in grey. The alignment might not span the entire HMM. Such a local alignment will show the start and end positions above or below the HMM-Logo. Let's give a short summary of the features of HMM-Logos now:

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 pixels to structure 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 title and x-axis descriptions for the logo. The two y-axis descriptions will appear vertically next to the HMM-Logos. By default, the name of the family as given by the HMM is used as the y-axis. If you don't want this, deselect "*use profile name*".

Image

By default, the width of the image is calculated as $25 * \text{length of the alignment}$. If you want a fixed width, deselect "*calculate dynamically*".

To plot the image in grey for use in non-colour publications, select "*greyscale*" (Not yet implemented).

Input HMMs

LogoMat-P reads files in plain HMMER2.0 format as well as multiple alignments in ClustalW, MSF and SELEX format. The multiple alignments will be used to create HMMs on-the-fly. Alternatively, you can choose to enter a Pfam ID that will be fetched from the current Pfam release.

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