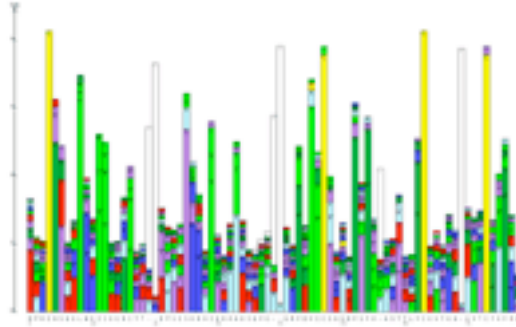


LOGOBAR



by Åsa Pérez-Bercoff, Johan Koch, Thomas R. Bürklin

LogoBar is a Java application to display protein sequence logos.

Main features of the program:

- the frequency of the residues at particular positions is represented with colored bars.
- amino acids can be placed into groups (e.g., to have similar amino acids in the same group), colors and groups can be selected according to the users own choice.
- bars can be sorted according to groups, so similar amino acids will be placed together in the bar graph.
- gaps are taken into account and also displayed in the logo as bars.

Detailed features

- Assign residues to particular groups. 21 groups are available (20 amino acids and 1 gap). Each group can be assigned to a color. For example, the small hydrophobic residues I, V, L, M can be assigned to a single group that is colored green. Several different default color assignments are provided. User can create their own amino acid groups and assign colors to these groups that can be saved to disk and loaded later.
- Bars in the graph can be sorted according to color groups. Note: gaps will always be treated separately.
- Gaps can be placed at the top or the bottom of the bars.
- Increase/decrease the vertical size of the bars (zooming in/out).
- Increase/decrease the font size for the characters placed into the bars.
- The lettering in the bars can be placed either at the top or bottom.
- Longer sequence motifs can be split into blocks, the size of the blocks can be specified.
- Optionally turn off small number error correction.
- Instead of bar graphs, traditional character-based logos can be created as an option.
- Underneath the bar graph, the residues at each position can be listed in decreasing order of frequency. A cut-off value can be given to determine at which frequency no further residues are displayed.
- An output file (output.txt) is created that contains a table with the frequency of all residues at each position.
- The bar graph logo can optionally display only the most conserved residue at each position.
- Exports graphics as EPS files for further processin.

Installation

Two downloadable versions are provided

- a) A compressed double-clickable application for Apple Mac Os X computers,
- b) Jar files for any computer system

Mac Os X

This is a double-clickable application that runs in Mac Os X. The software to run Java applications should be installed by default. Note: the "File Open" dialog behaves like in a Unix environment, not like MacOSX. To navigate to other partitions you have to go to the /volumes folder in the root of the system hard disk.

Jar files for any computer system

In order to run Java software on your computer JVM (Java Virtual Machine) has to be installed on the computer. This can be downloaded for free at Sun's web pages at <http://java.sun.com>.

Download the compressed file LogoBar.tgz to an appropriate directory on your computer, and extract it by typing the following in a terminal window:

```
tar -xzf LogoBar.tgz
```

Alternatively, download the two jar files ([LogoBar.jar](#) and [epsgraphics.jar](#)) to your computer, and place them in the same directory. Note: The Java EPS Graphics2D package (epsgraphics.jar) package is developed by Paul Mutton and is available under the GNU general public license.

Open terminal window, move to the directory with the LogoBar.jar file and type:

```
java -jar LogoBar.jar
```

Usage

- Under the File menu, selection "Open". A file dialog will appear. Note for Mac Os X: For navigation, the dialog behaves like a UNIX file dialog, not like a Mac Os X dialog. To change to another volume, you have to go to the root, and locate the "volumes" directory, then go the partition/volume that you like.

- The program expects Clustal multiple sequence alignment files. You can choose to just see the .aln files (default), or you can elect to see all files in the dialog.

- After reading of the file, an output file (output.txt) will be automatically created. NOTE: make sure the directory has the right permissions for allowing to write files, otherwise this will fail.

If the sequences contain any unrecognized character (hopefully rarely), i.e. X, J, O, U, Z, B, they will be grouped at present with the "gap" character. A warning dialog will be displayed, but the program will continue to run.

The bar graph is displayed. You can increase/decrease the height of the bars with the zoom in/out commands under the "Options" menu.

In a floating window the user has a color legend, where one can see how amino acids are assigned to groups, and what color each group has. One can click on the color block and a color selection

dialog pops up that allows one to change the color. For additional control of residue grouping and color assignment, use the "Color settings" dialog".

Color settings dialog

To assign residues to groups, choose "Color settings". You can assign each residue to a group (group 1 to 21). Each group can then have its own color. Several default color schemes have been provided, adapted from textbooks and multiple sequence alignment programs (see below) and modified schemes by us: Mod. Cinema, Cysteine (cysteine in yellow, e.g. to highlight residues in extracellular molecules), helix to highlight proline and glycine in yellow as helix breakers.

Color schemes can also be saved to disk and loaded again.

Note: if a default color scheme is selected and applied (e.g., Cinema), the next time the "Color settings" dialog is opened again, the Pop-up menu has reset to the top choice (i.e. Cell), but the colors nevertheless are still those of the previous selection. Also note, how the colors in the boxes changes when you choose different default color schemes.

The graph style can be "Continuous" or "Block divided". Longer sequences motifs may be wrapped into blocks. The size of the blocks can be set in the Graph settings dialog.

Graph settings dialog

In the "Graph settings" dialog there are many options to customize the appearance of the logo, and what should be displayed.

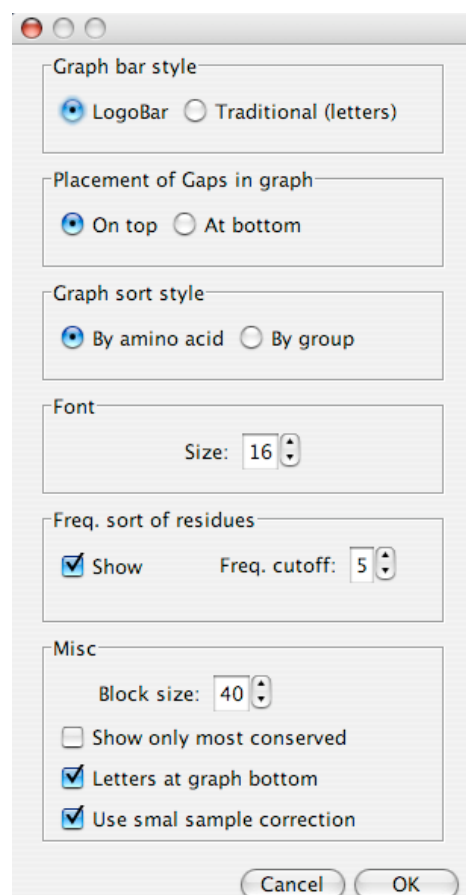
- Graph bar style: This can either be the bar graph (default) or if desired, a traditional, character based logo, where the height to the characters represents the entropy.

- Gaps placement: Gaps can either be placed on top of the amino acid bars (default), or at the bottom of the amino acid bars.

- Graph sort style: Bars of amino acids are sorted by individual frequency in a stack, or based on the group frequency. I.e. if E and D both belong to group 1 (acid residues, red) then they will be grouped together in the bar graph, even if their individual frequencies are very different. Otherwise, it is possible that - when D has a very low frequency - that other residues, e.g. N, will be placed between E and D.

- Font size: Allows setting the font size. Note: if the font size is lower than 10, the text entry window will shrink in size. If one increases then from 9 to higher (i.e. 10), the window unfortunately does not increase (Java bug), but the correct size will still be there, and will be displayed after closing and opening the dialog.

- Frequency sort of residues: Under the logo bar graph normally only a consensus is displayed. If this option is turned on, then it is possible to display more residues. On the right the cut-off



frequency is set, e.g., 5%. Then all residues at a particular position are stacked in decreasing order of frequency under the consensus until the cut-off value is reached. If the cut-off value is set to 0%, all residues that occur at a particular position are listed in this stack.

- Block size: When the graph is split into blocks (see above), one can set the size of the blocks here.
- Show only most conserved: The bar graph logo will only show the most conserved residue at each position.
- Letters at graph bottom. In the bar graph logo, within each bar, the position of the letter representing that amino acid can be set to be either at the top, or the bottom within each individual bar
- Use small sample correction: The small sample correction can be turned off, if desired.

Output

LogoBar produces two types of output.

EPS file of graphics

For further processing of the LogoBar bar graph, the screen image can be saved as an EPS file (encapsulated postscript). This can be printed on postscript printers, or it can be further edited in programs like Adobe Photoshop (© Adobe Inc.).

Output.txt file

When an input file is read in by LogoBar an output text file [output.txt] is created in the background. This output file stores the name of the input file, the alignment length, the number of sequences, and three tables consisting of the number, the frequency, and the height of the amino acids at every position in the alignment. Thus a complete data table with the frequency of each residue at each position is generated. A consensus sequence, created from the most frequently occurring residue at a position, will also be stored in this output text file. Should there be several residues with equal frequency, the one with the one letter abbreviation that come last in the alphabet is displayed.

Grouping and color schemes

Several default amino acid grouping and color schemes are provided.

Cinema

Multiple sequence alignment program (Parry-Smith, D.J., Payne, A.W.R, Michie, A.D. and Attwood, T.K. (1997) "CINEMA - A novel Colour INteractive Editor for Multiple Alignments." *Gene*, 211(2), GC45-56.

blue:	polar positive	H, K, R
red:	polar negative	D, E
green:	polar neutral	S, T, N,Q
white (>grey)	non-polar aliphatic	A, V, L, I, M

purple	non-polar aromatic	F, Y, W
brown		P, G
yellow		C

Lesk

Lesk A.M. (2002) Introduction to Bioinformatics, Oxford University Press, page 187

yellow:	small nonpolar	G, A, S, T
green:	hydrophobic	C, V, I, L, P, F, Y, M, W
magenta:	polar	N, Q, H
red:	negatively charged	D, E
blue:	positively charged	K, R

Clustal

Clustal_X 1.83 multiple sequence alignment program. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25, 4876-4882.

Cell

Alberts et al., Molecular Biology of the Cell textbook. Colors are assigned arbitrarily according to groups in the book.

WebLogo

Colors based on WebLogo (<http://weblogo.berkeley.edu/>), black => grey

Mod. Cinema

Modified Cinema

blue:	polar positive	H, K, R
red:	polar negative	D, E
green:	hydrophobic	V, L, I, M
dark green:	non-polar aromatic	F, Y, W
purple:	polar neutral	S, T, N, Q
light blue		A, P, G, C

Cysteine

Modified Cinema with cysteine yellow

Zinc finger

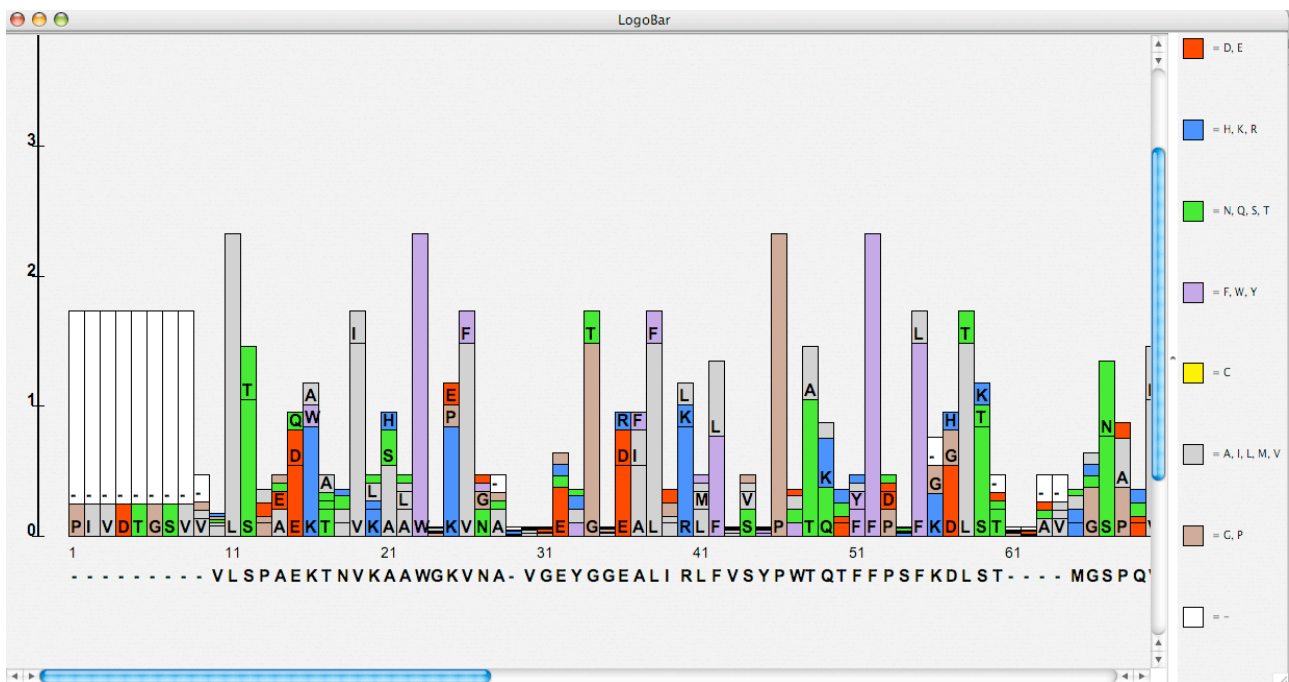
Modified Cinema with histidine and cysteine yellow

Helix breaker

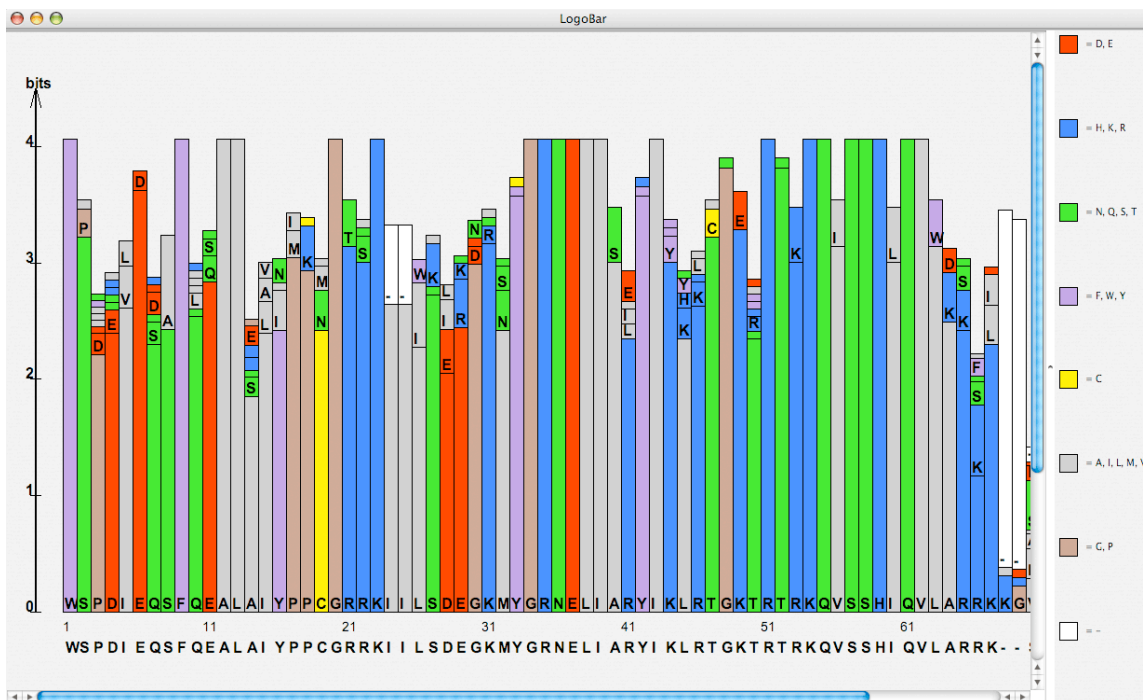
Modified Cinema with proline and glycine yellow

Examples

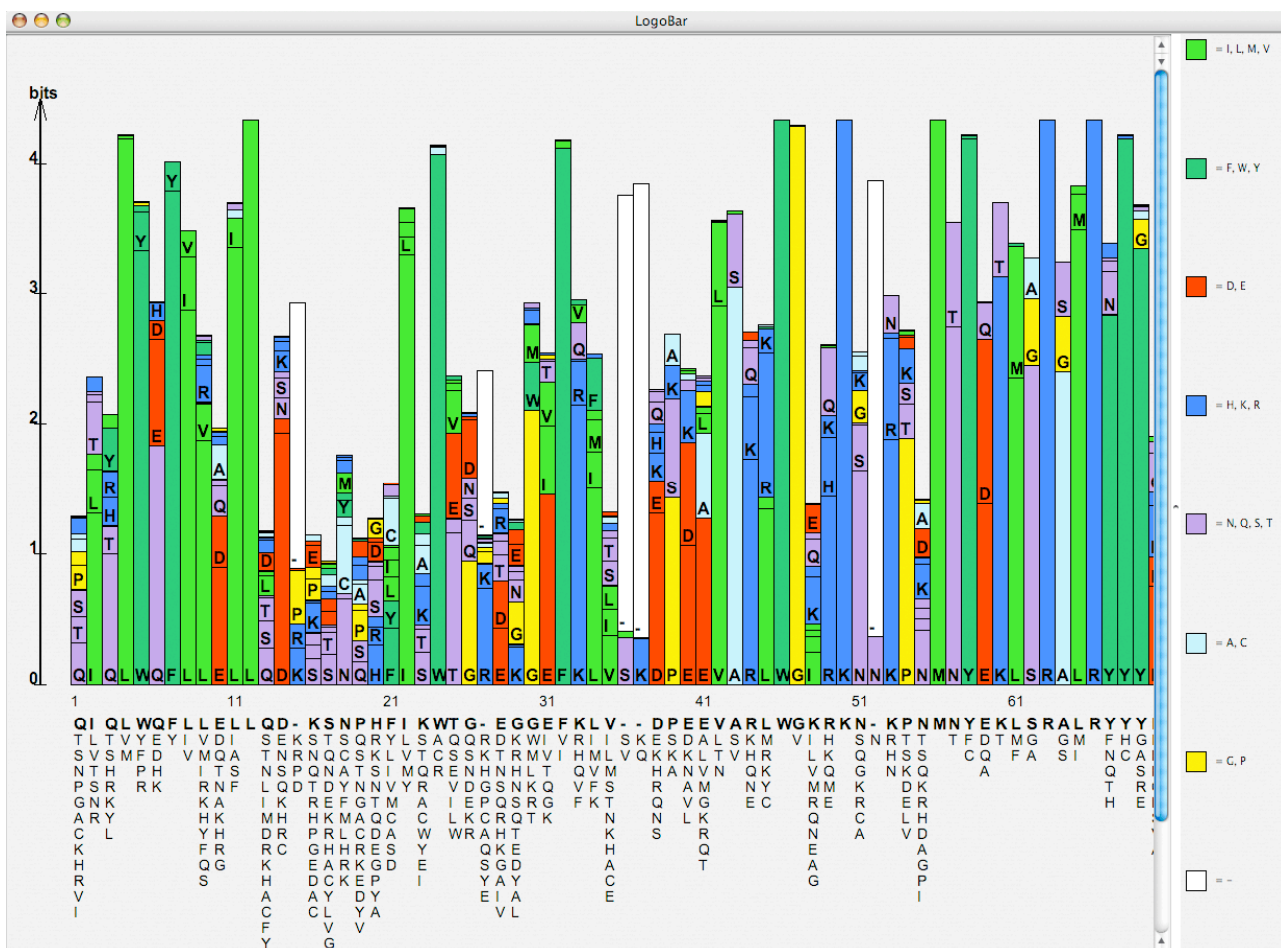
Screen shot of a LogoBar of 7 globin sequences, using Cinema coloring scheme:



Sample of 44 TEA domain sequences (retrieved from PFAM, and edited to remove incomplete sequences). Screen shot of a LogoBar, using Cinema coloring scheme:



Logo of 280 ETS domain sequences (retrieved from PFAM, edited to remove incomplete sequences). LogoBar screen shot, using custom coloring scheme, amino acids listed at bottom in decreasing order of frequency:



Comparison of ETS domain Logos created with WebLogo and LogoBar with high Y-axis, and similar color schemes:

