H3ABioNet

H3ABioNet

Pan African Bioinformatics Network for H3Africa

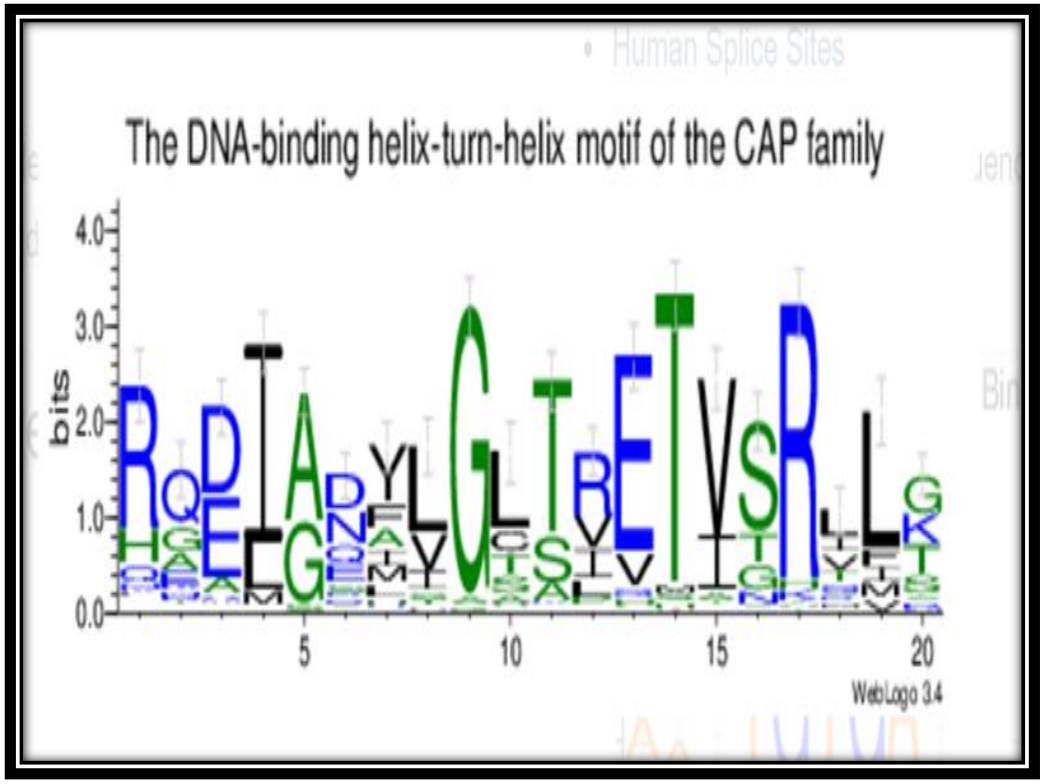
Introduction to Bioinformatics Online Course: IBT

Multiple Sequence Alignment

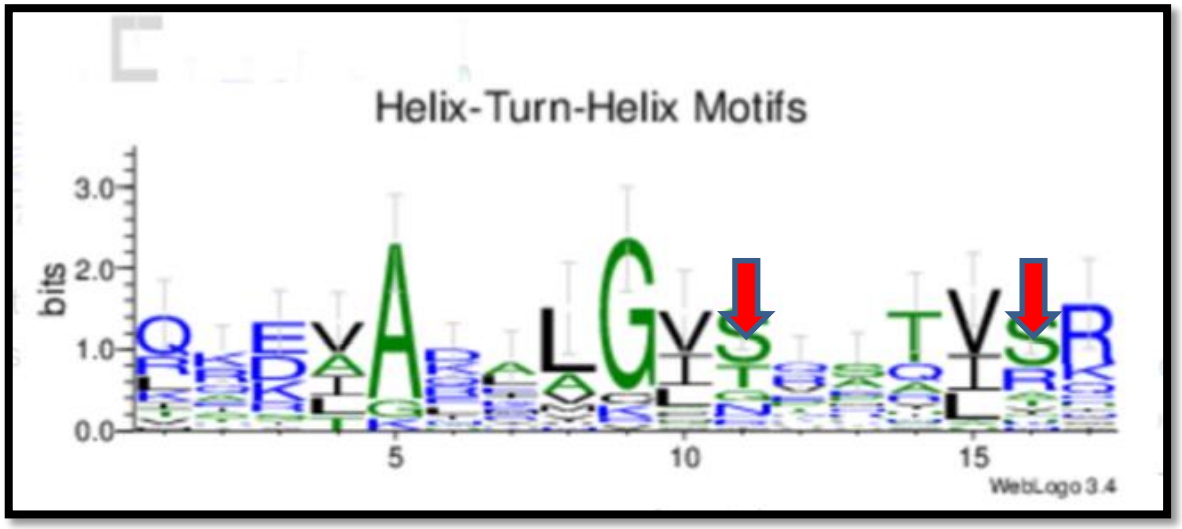
Building Multiple Sequence Alignment

Lec5: Interpreting your MSA Using Logos

Using Logos



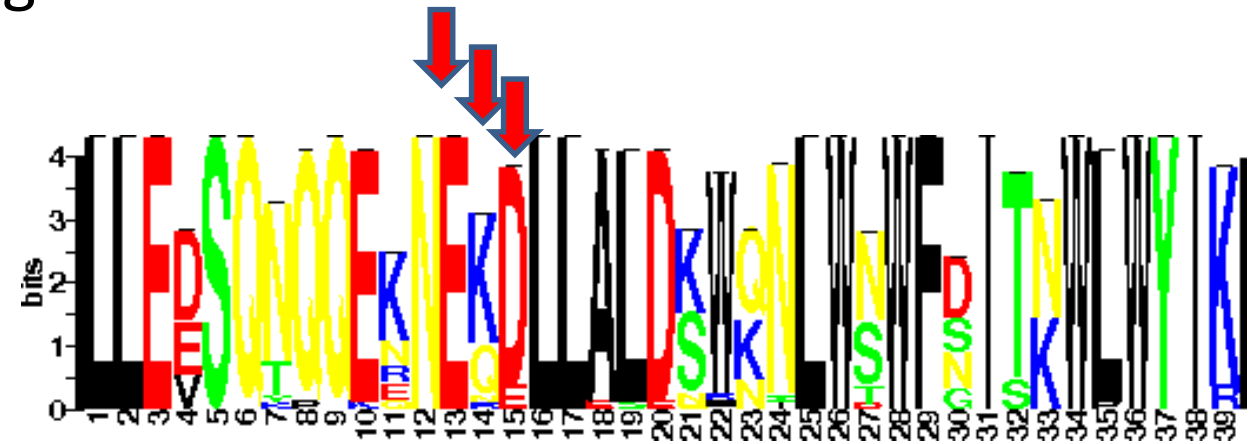
- Logos are a **terrific way** to generate **high-impact pictures** from MSA
- logo Figure is a representation of the alignment.
- Notice how the **conserved amino acids** (e.g **cysteines**) stick out, **indicating** regions of **potential biological importance**.



Claverie J, Notredame C (2007). Bioinformatics for Dummies (2nd Edn). *Wiley publishing, Inc*. 436 pp.

When looking at a sequence logo, you can consider the following elements:

- **Each position** corresponds to a **column** in **the multiple alignment.**
- The **total height** of a logo position depends on the **degree of conservation** in the corresponding multiple alignment column.

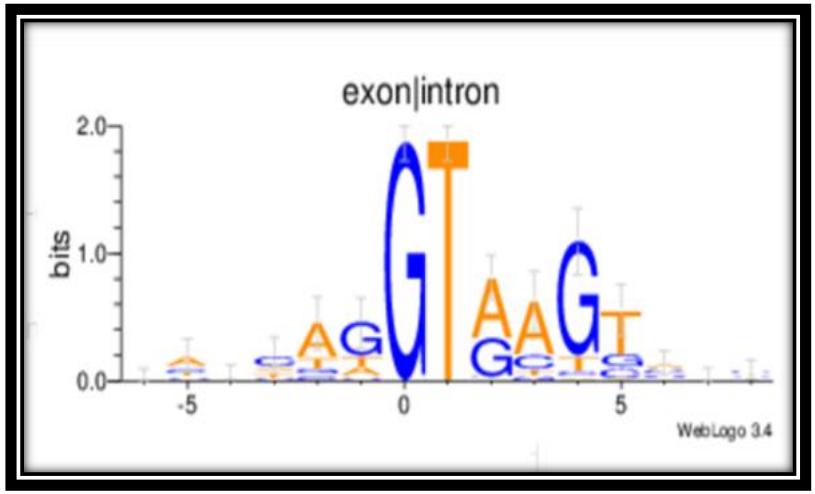


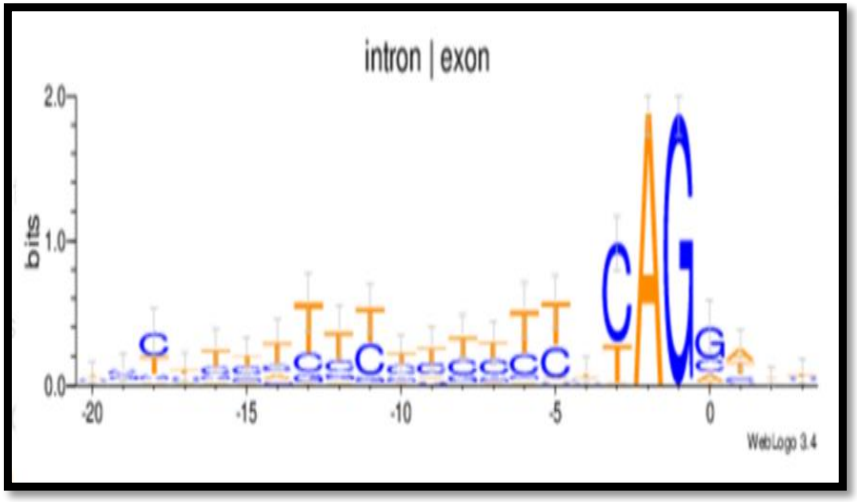
Claverie J, Notredame C (2007). Bioinformatics for Dummies (2nd Edn). *Wiley publishing, Inc*. 436 pp.

- Very **conserved alignment** columns give you **high logo positions.**
- **Positions** that contain a **very heterogeneous** mixture of symbols yield **low logo positions**.
- The **size** of each letter in a logo position depends on **how frequent** this letter is in the column.
- The **top letter** is always the **most frequent** in the column.

Claverie J, Notredame C (2007). Bioinformatics for Dummies (2nd Edn). *Wiley publishing, Inc*. 436 pp.

- **Logos make sense** only if you have a **nice block** with a few **highly conserved** positions **surrounded by** highly **degenerated positions**.
- There is a **handy utility** on the Web that identifies blocks within your multiple alignments and turns each of them into a logo.
- **blocks.fhcrc.org./blocks/process_blocks.html**



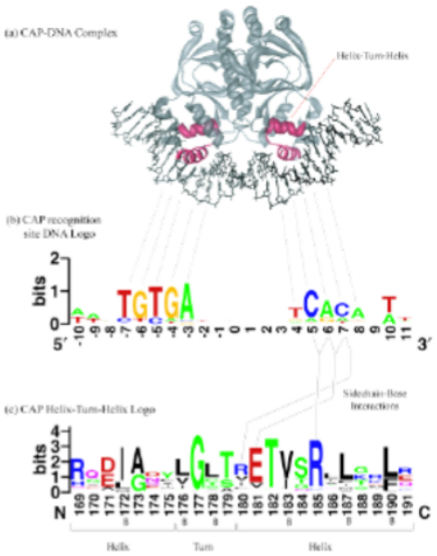


Claverie J, Notredame C (2007). Bioinformatics for Dummies (2nd Edn). *Wiley publishing, Inc*. 436 pp.

WebLogo 3 | home | create | examples | manual

Introduction

WebLogo is a web-based application designed to make the generation of sequence logos easy and painless. WebLogo has been featured in over 4000 scientific publications.



A sequence logo is a graphical representation of an amino acid or nucleic acid multiple sequence alignment. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence.

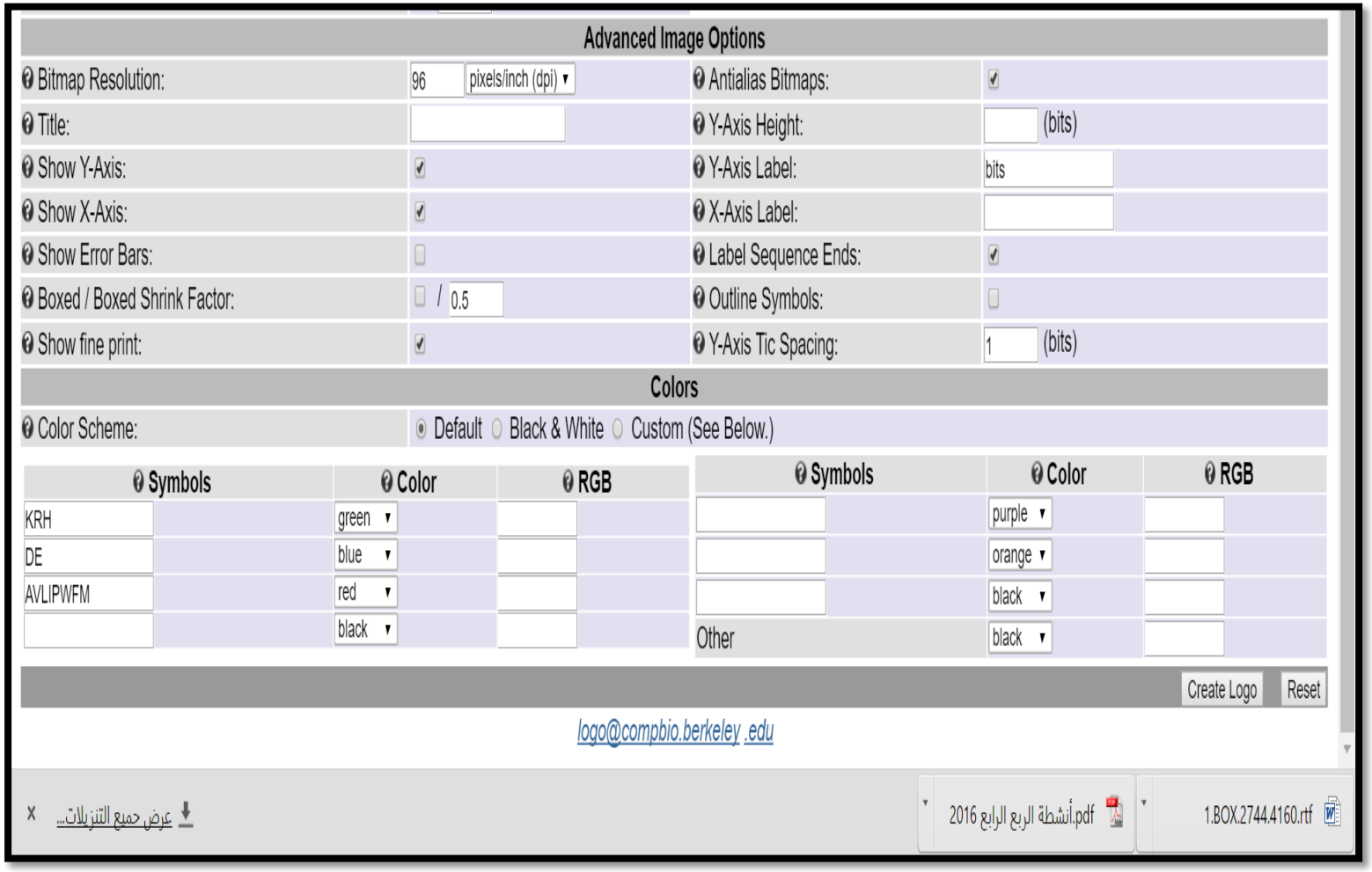
WebLogo is a web-based application designed to make the generation of sequence logos easy and painless. WebLogo has featured in over 4000 scientific publications.

- Create your own logos
- View example sequence logos and input data.
- Read the release notes for latest changes and updates.
- Read the User's Manual
- WebLogo source code
- WebLogo discussion group

References

Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190, (2004) [Full Text]

Schneider TD, Stephens RM. 1990. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res. 18*:6097-6100



Advanced Image Options

Bitmap Resolution:	96 pixels/inch (dpi)	Antialias Bitmaps:	☑
Title:		Y-Axis Height:	(bits)
Show Y-Axis:	☑	Y-Axis Label:	bits
Show X-Axis:	☑	X-Axis Label:	
Show Error Bars:	☐	Label Sequence Ends:	☑
Boxed / Boxed Shrink Factor:	☐ / 0.5	Outline Symbols:	☐
Show fine print:	☑	Y-Axis Tic Spacing:	1 (bits)

Colors

Color Scheme: ◉ Default ○ Black & White ○ Custom (See Below.)

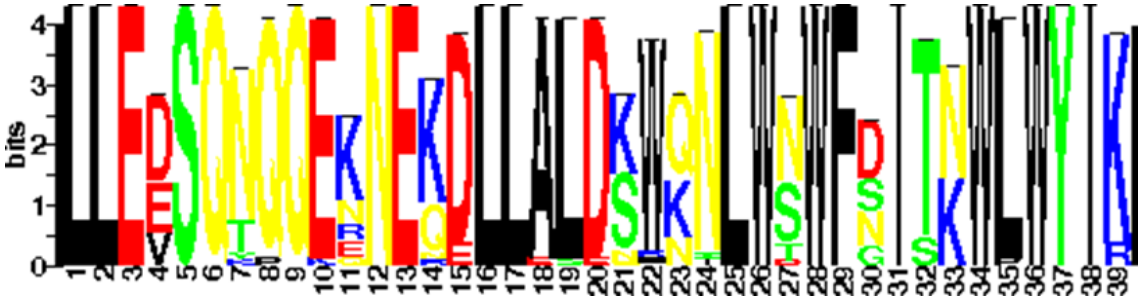
Symbols	Color	RGB	Symbols	Color	RGB
KRH	green			purple	
DE	blue			orange	
AVLIPWFM	red			black	
	black		Other	black	

Create Logo Reset

logo@compbio.berkeley.edu

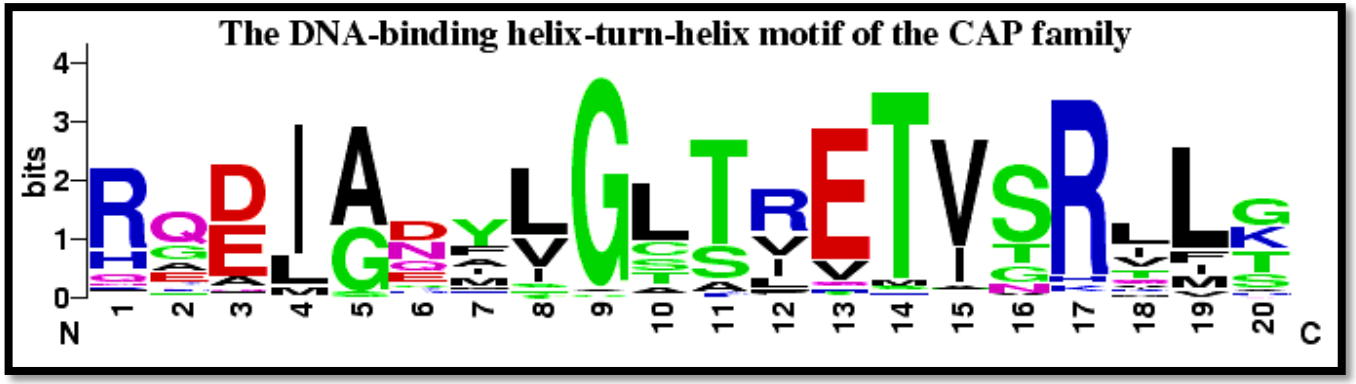
عرض جميع التنزيلات... 2016 أنشطة الربع الرابع.pdf 1.BOX.2744.4160.rtf

Examples

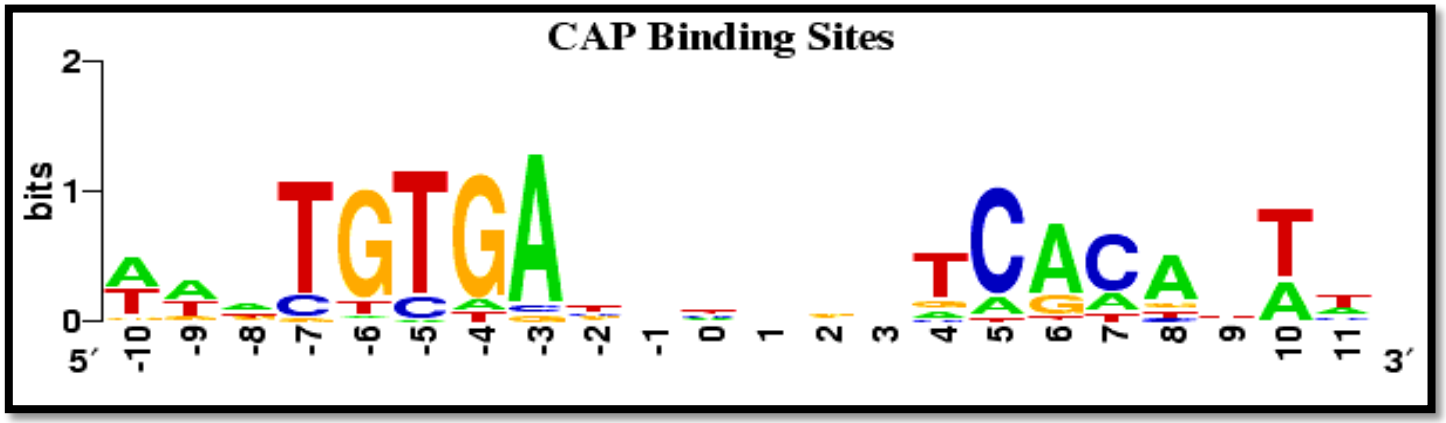


Catobolite Activator Protein (CAP)

The helix-turn-helix motif from the CAP family of homodimeric DNA binding proteins. CAP (Catabolite Activator Protein, also known as CRP for cAMP Receptor Protein) is a transcription promoter that binds at more than 100 sites within the *E. coli* genome. Residues 1-7 form the first helix, 8-11 the turn and 12-20 form the DNA recognition helix. The glycine at position 9 appears to be critical in forming the turn. Positions 4, 8, 10, 15 and 19 are partially or completely buried, and therefore tend to be populated by hydrophobic amino acids, which are colored black. Positions 11-14, 17 and 20 interact directly with bases in the major groove and are critical to the sequence specific binding of the protein. The data for this logo consists of 100 sequences from the full Pfam alignment of this family (Accession number PF00325). A few sequences with rare insertions were removed for convenience.



The two DNA recognition helixes of the CAP dimer insert themselves into into cThe two DNA recognition helixes of the CAP homodimer insert themselves into consecutive turns of the major groove. Several consequences can be observed in this CAP binding site logo. The logo is approximately palindromic, which provides two very similar recognition sites, one for each subunit of the dimer. However, the binding site is not perfectly symmetric, possible due to the inherent asymmetry of the operon promoter region. The displacement of the two parts is 11 base pairs, or approximately one full turn of the DNA helix. Additional interactions between the protein and the first and last two bases occur within the DNA minor groove, where it is difficult for the protein to distinguish A from T, or G from C\cite{Seeman76}. The data for this logo consists of 59 binding sites determined by DNA footprinting. *Robison, K., McGuire, A. M., Church, G. M. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K12 genome. Journal of Molecular Biology (1998) 284, 241-254.*

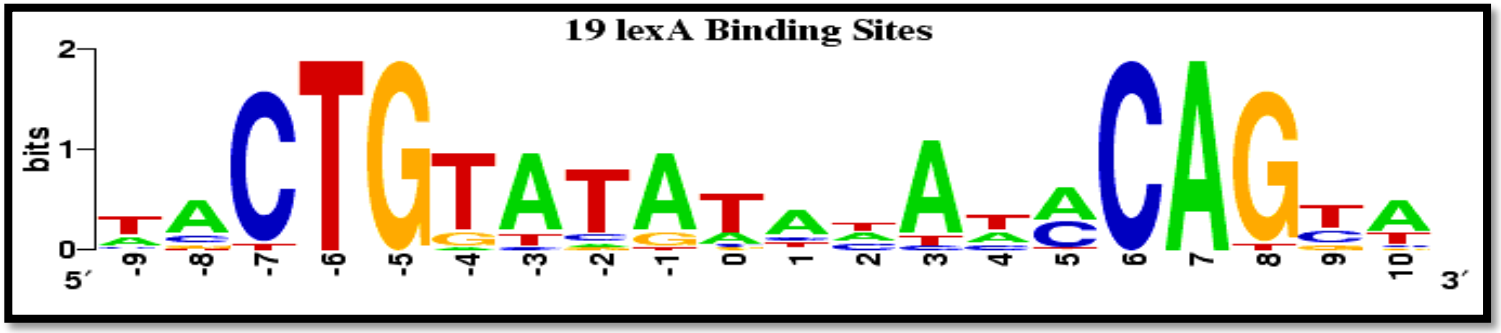


E. coli Transcription Factor Binding Sites

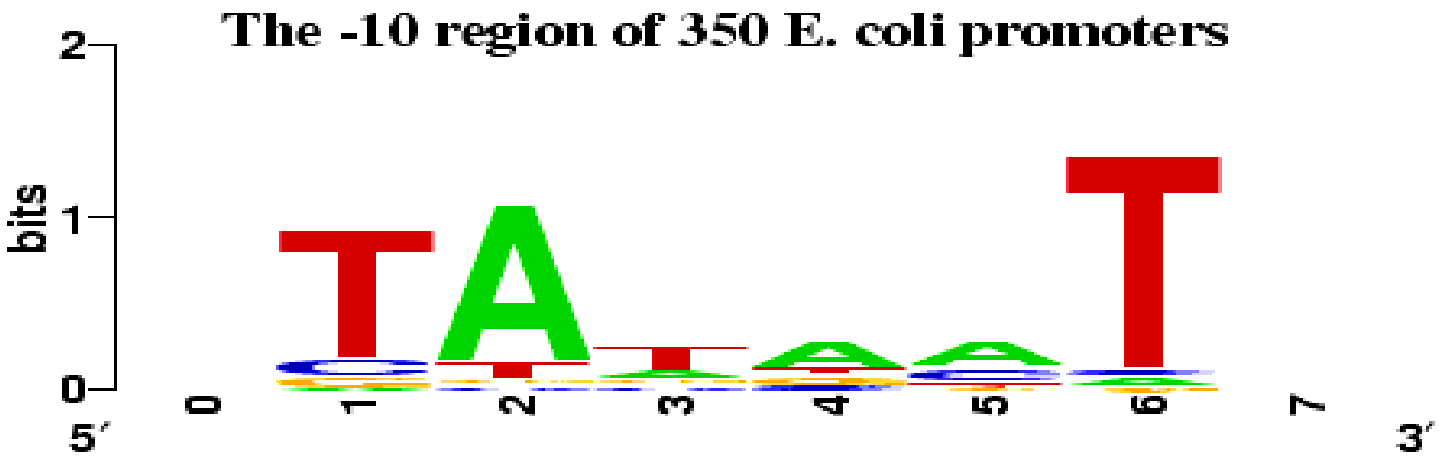
The following logos (along with the CAP logo above) display a selection of E. coli transcription factor binding sites determined by DNA footprinting. This data has been collated in the DPInteract database and has been used to search for additional binding sites within the E. coli genome.

- **LexA repressor is closely related to CAP, and has similar DNA protein interactions.**
- **H-NS: Histone like, nucleoid-associated DNA-binding protein.**
- **DNA biosynthesis initiation binding protein.**
- **Arginine Repressor.**

Robison, K., McGuire, A. M., Church, G. M. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K12 genome. Journal of Molecular Biology (1998) 284, 241-254.



E. coli Promoters (Transcription Start Signals)

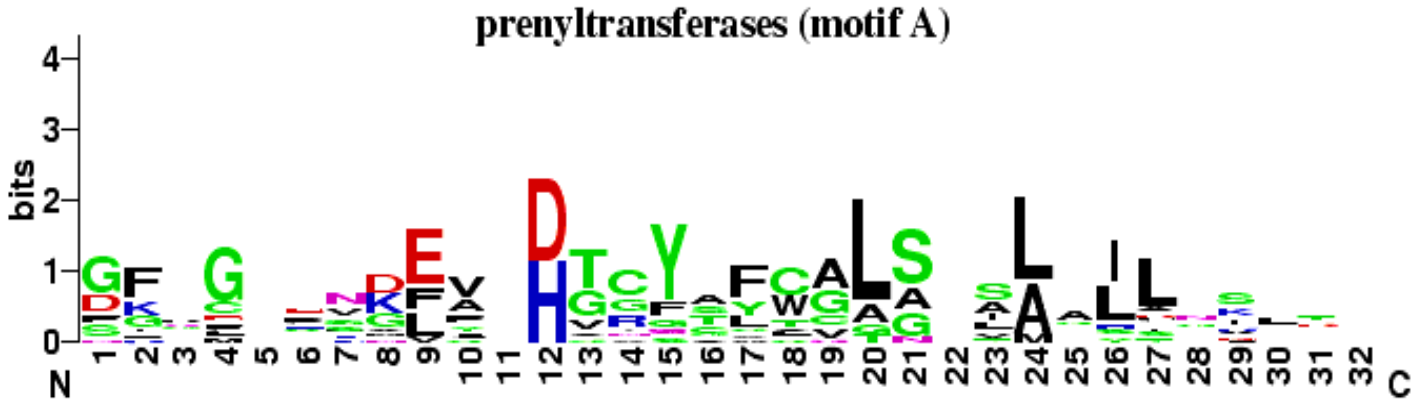


In prokaryotes the DNA sequence just upstream of the transcription start point contains two important conserved regions. The first such region is centered at around 35bp upstream and is involved in the initial recognition of the gene by RNA polymerase. The second region, sometimes referred to as the Pribnow box, is centered at about 10bp upstream. The typical separation between the -35 and -10 sites is 15-18 bp. See baseflip: Strong Minor Groove Base Conservation in Sequence Logos implies DNA Distortion or Base Flipping during Replication and Transcription Initiation for more information.

Globins

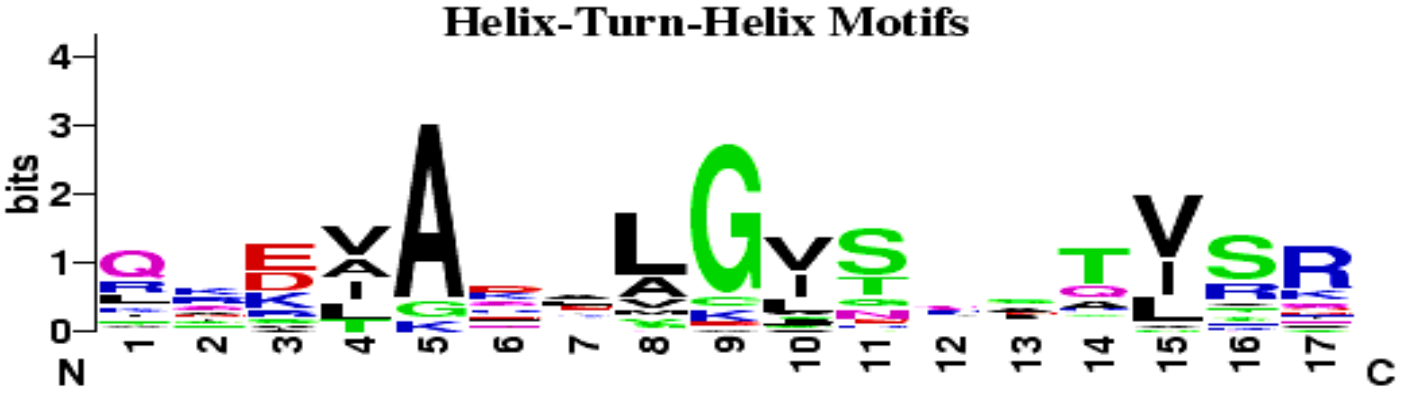
The end of the B helix through the beginning of the D helix of 34 globins.

Prenyltransferases (motif A)



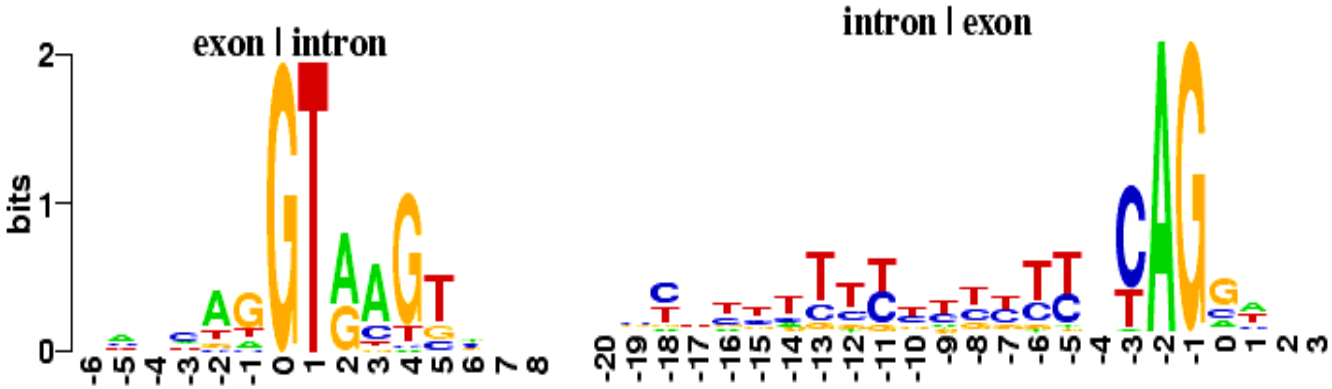
Here is an alignment found by the gibbs sampling system. Both the identified site and some context are shown. Note that spaces are significant, so that the spaces included below (to aid identification of the site) will end up being considered amino acid positions.

HTH Proteins



Helix-Turn-Helix DNA binding motifs found by the gibbs sampling system. Compared to the CAP HTH logo there is much less sequence conservation within the DNA binding helix (11-17), as might be expected for a diverse sample of proteins.

Human Splice Sites



These logos show a small sample of Human intron-exon splice boundaries. Sequences of experimentally confirmed genes were extracted from EID: the Exon-Intron database. Additional discussion of the features in this logo can be found in the paper Features of spliceosome evolution...

- **Exon-Intron (Donor) Sites**
- **Edit Logo Intron-Exon (Acceptor) Sites**

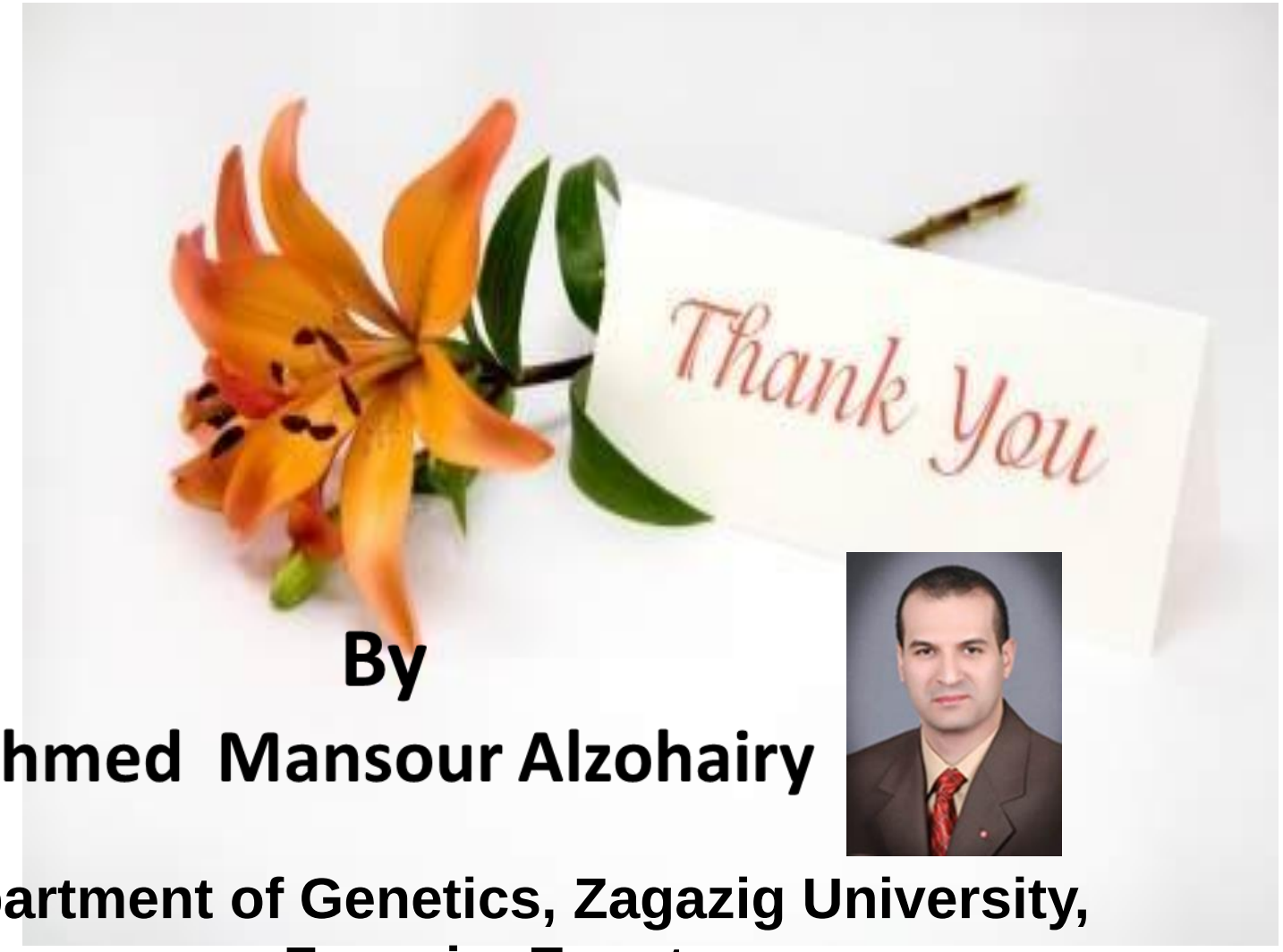
Practical

'to try in your own time”

- **Interpret the conserved amino acids in your alignments using logos**

Interpret the conserved amino acids in your alignments (e.g HSF1) using logos


BTU
BIOINFORMATICS
TRAINNING UNIT
GENOMICS
1st Module What Bioinformatics Can Do for You
2nd Module Manipulation of Biological Sequences
3rd Module Working with single DNA sequence
4th Module How to Build a Multiple Sequence Alignment?
5th Module Inferring Phylogenetics analysis using Jalview
6th Module Advanced Molecular Concepts
7th Module Inferring Protein Sequence (structure & Function)
8th Module RNAanalysis and Function
9th Module Editing and Publishing Alignments in your Manuscript
10th Module Building and Publishing Phylogenetic Trees
11th Module Working with Protein 3-D Structures
12th Module Advanced Bioinformatics Using R
Bioinformatics Workshops Series
EST by, Dr. Ahmed Mansour Alzohairy
Call us: 01026060324 / 01000727270
Facebook: https://www.facebook.com/Bioinformaticsunit



By

Ahmed Mansour Alzohairy

Department of Genetics, Zagazig University, Zagazig, Egypt