Motif finder

Author: Hansi Thewarapperuma
Date: 10/03/2022

Importance

Transcription factor binding motifs and promoter elements are both crucial components in the regulation of gene expression. Transcription factors (TFs) are proteins that bind to specific short DNA sequences, known as transcription factor binding motifs or transcription factor binding sites (TFBS). When transcription factors bind to their respective motifs, they can either activate or repress the transcription of adjacent genes. The binding of transcription factors to these motifs is a key step in the regulation of gene expression. The presence of specific motifs in the regulatory regions of genes helps determine when and to what extent a gene will be transcribed.

Promoter elements are specific DNA sequences located in the promoter region of a gene. The promoter is a region upstream of the transcription start site where RNA polymerase and other transcription factors bind to initiate transcription. Promoter elements include various sequences that are recognized by transcriptional machinery. They are essential for the initiation of transcription by providing binding sites for RNA polymerase and other transcription factors. Common promoter elements include the TATA box, CCAAT box, and GC box.

"Motif Finder – Your Ultimate Solution for Sequence Analysis!"
Uncover the mysteries within your sequences effortlessly with Motif Finder. Identify sequence types, calculate AT/GC content, detect TATA and CCAAT boxes, search for desired patterns, and assess the presence of promoter regions. Save time, enhance accuracy, and elevate your sequence analysis game. Researchers, bioinformaticians, and students, Motif Finder is tailored for you! Try it now and unlock the potential of your sequences.

-Hansi -

This program includes classes and methods to find sequence motifs, such as transcription factor binding motifs, promoter elements etc.

- A method to return the sequence type.
- A method to search for a TATA box motif in a given DNA sequence.
- A method to search for a CCAAT box motif in a given DNA sequence.
- Methods to calculate AT content in DNA and RNA sequences.
- Methods to calculate GC content in DNA and RNA sequences.
- A method to search for the above elements (TATA, CCAAT, AT content, and GC content) when a FASTA file containing multiple sequences are given as the input. You can use another method parameter to prompt the user for the element type that the method needs to search.

- A method to search for a user given sequence motif pattern in a file containing multiple FASTA sequences. The user must provide the sequence pattern as a regular expression, which should be taken as a parameter for the method.
- A method to predict promoter sequences. The user can input a DNA sequence, and based on the presence of TATA, CCAAT boxes and based on the GC content, this method should predict whether the sequence contains a promoter region or not.