

BacRegDB: a database of bacterial regulatory elements with structural evidence

Vychyk P.¹, Duvalov E.², Digris A.², Skakun V.², Nikolaichik Y.¹

1) Department of Molecular Biology, Belarusian State University, Minsk, Belarus

2) Department of Systems Analysis and Computer Modelling,
Belarusian State University, Minsk, Belarus

Motivation and Aim

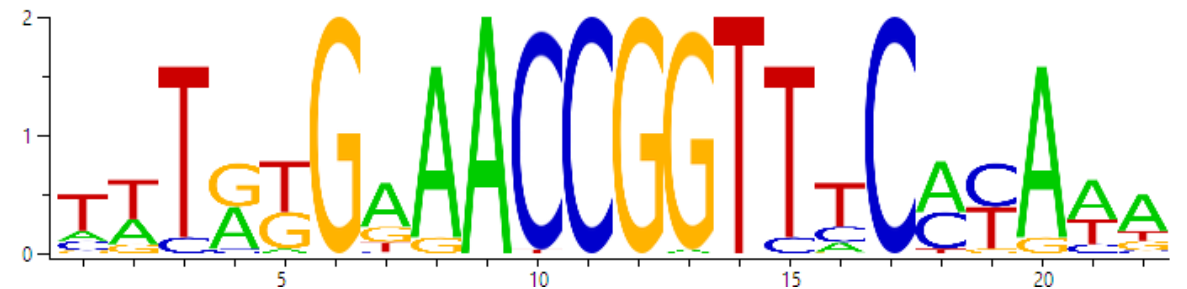
Most of the available bacterial genome annotations provide information about ORFs and their products. Adding regulatory elements to the annotation shows which **transcription factor (TF)** controls them and when the genes might be expressed, allows answering basic questions and facilitates constructing strains with desired properties.

RegulonDB (largest curated database for *Escherichia coli* regulatory sequences with experimental support) provides limited information for many TFs, therefore TFBS models useful for *in silico* binding site predictions can't be built with tools like MEME or HMMER.

Sequence	LeftPos	RightPos
aaaatattcaGGTGACCGGTTTCACaaatataaaa	2839350	2839364
catgattttgTGAAACCGGTTTCTTaattoogttt	2839426	2839440
nd	nd	nd
cagtcagaagTGAAACAGCTTCTGCacotgatgat	495157	495171
tttcagtttaaCGTTTCaggcaatgttt	348527	348533
tttcaggcaaTGTTTCAccgcggtttca	348515	348521
tgtttcaccgCGTTTCAAttgcaacaat	348505	348511
ccgcggtttcaTTGCAACaattatgaaa	348498	348504
tgcaacaattATGAAACaagactaaac	348487	348493
cagctcacagTGAAATCAGATGTGTacgaaatcac	855832	855846

Part of RegulonDB interface showing operators for [AscG](#) -> no usable motif

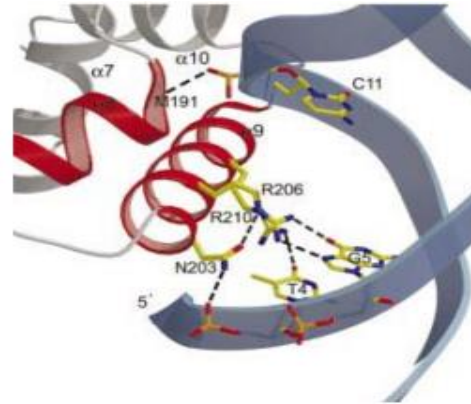
We aimed to combine results of [previously presented](#) Sigmoid pipeline for *in silico* regulatory motifs inference with experimental information from databases in new online database of calibrated HMM profiles representing richer TF DNA-binding motif in terms of search sensitivity.



AscG motif derived with Sigmoid *de novo* TFBS inference pipeline

Methods and Algorithms

We use a TF 3D structure based strict formal criterion – **CR-tag** – for applying regulatory site motifs created for TFs from a model organism to any bacterial genome.



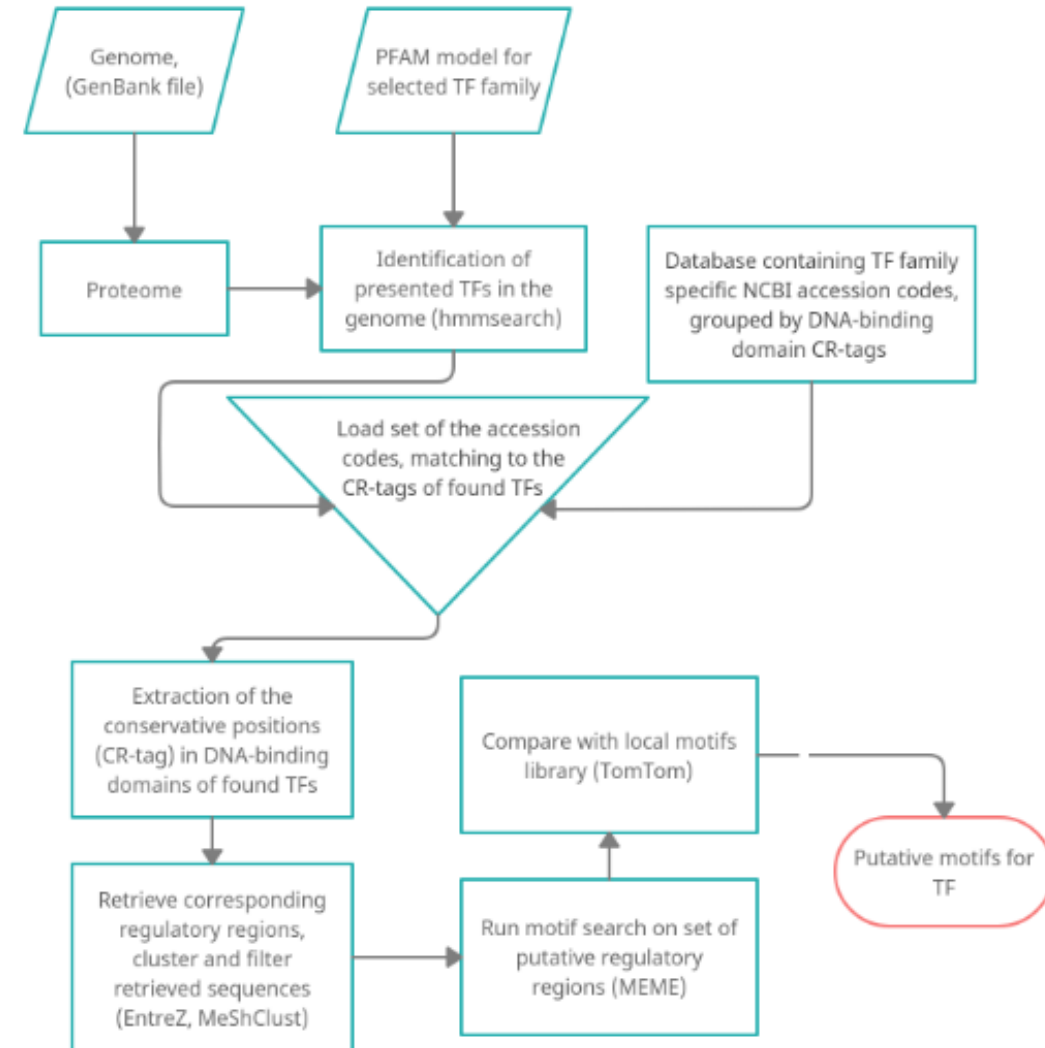
TraR TF-DNA complex
(CR-tag YNSRVKRE)
EMBO j. 21:4393-4401 (2002)

CR-tag is composed from the amino acid residues of a transcriptional regulator that specifically contacts the nitrogenous bases of the regulatory element in genomic DNA.

In silico motif inference pipeline implemented within Sigmoid exploits the idea of TFs with identical CR-tags recognizing identical (or highly similar) TFBSs.

Search for conserved patterns (with MEME or other tools) within the regulatory regions of genes encoding TFs with identical CR-tags usually gives correct regulatory motif common for these TFs.

More details in BIATA 2020 [talk](#)



Sigmoid de novo motif inference pipeline

Results

The screenshot shows the BacRegDB website interface. At the top, there is a green navigation bar with the BacRegDB logo and menu items: Home, TF Families, About, Contacts, and Links. A search bar is located on the right side of the navigation bar. Below the navigation bar, the main content area is titled "Transcription factors profiles" and includes a sub-header "Browse BacRegDB by transcription factor and families." and a note: "Click on each family to expand it and see its associated information and links to member TF pages." The main content area displays a grid of transcription factor families with their names and profile counts:

LacI (profiles count: 22)	AraC (profiles count: 44)	TetR (profiles count: 5)	CitT (profiles count: 5)	XRE (profiles count: 1)	ROK (profiles count: 2)	RHH (profiles count: 1)
MerR (profiles count: 1)	LysR (profiles count: 9)	LexA (profiles count: 2)	IcIR (profiles count: 2)	HxIR (profiles count: 1)	DeoR (profiles count: 1)	RpiR (profiles count: 3)
GntR (profiles count: 16)	GerE (profiles count: 52)	OmpR (profiles count: 57)	Fis (profiles count: 21)	Rrf2 (profiles count: 2)	MarR (profiles count: 16)	

At the bottom of the page, there is a green footer section with the following text:

Addresses:
Belarusian State University,
4 Nezavisimosti Avenue, Minsk, 220030, Belarus.

©BacRegDB - 2022. Belarusian State University.

<https://BacRegDB.bsu.by>

Using Sigmoid *de novo* pipeline and experimental evidence for TF binding sites, the online database of curated motifs was created. Currently it includes records for 263 transcriptional regulators that have all undergone human curation with verification of experimental evidence and threshold scores determination for HMM models.

Example TFBS motif entry in BacRegDB

GerE family

- AETKSYRS_RamA
- AKTSSHGN_RcsA
- ATTKTHRN_MalT
- EANNFHGN_LasR
- EATKVHKS_NarP_Vibrionales
- EGTKVHKN_NarP_Shewanellaceae
- EHTRDYKT_CinR
- EKTKNHSN_DegU
- EKTKNYSR_DosR
- EQTKVHRN_NarP
- ERTKFHAN_CviR
- ERTKFHKQ_PcoR
- ERTKFHSN_AurR
- ERTKSRTS_YdfI_Bacillales
- ERTRFFTN_SinR
- ERTTFHTN_LuxR
- ESTKVHKKH_NarP
- EVTKFHNN_CarR

Profile logo:

Description:
RamA is a global transcriptional regulator in *Corynebacterium glutamicum* that controls genes for enzymes of the central metabolism (Auchter et al., 2011). It controls a rather large regulon of over 200 genes (many indirectly). This profile is built with sites found by meme in promoter fragments of the genes differentially expressed in ramA mutants and bound by RamA in EMSA experiments (Auchter et al., 2011). It should be noted that meme didn't find this motif in all fragments analysed. Experimental sites were supplemented by the ones found by Sigmoid in regulatory regions of ramA genes coding for a protein with the same CR tag. The final profile was calibrated to find the most reliably determined RamA targets. However, it misses some of the known targets and finds some of the unknown ones.

Reference	DOI	Evidence
1. Auchter, M., Cramer, A., Hfjser, A., Rfjckert, C., Emer, D., Schwarz, P., Arndt, A., Lange, C., Kalinowski, J., Wendisch, V. F., & Eikmanns, B. J. (2011). RamA and RamB are global transcriptional regulators in <i>Corynebacterium glutamicum</i> and control genes for enzymes of the central metabolism. <i>Journal of Biotechnology</i> , 154(2B5"3), 126B5"139. https://doi.org/10.1016/j.jbiotec.2010.07.001	10.1016/j.jbiotec.2010.07.001	[BPP]

BacRegDB delivers comprehensive information about regulatory motif of interest. One can:

- access general description of the TF for which the binding site model was built;
- view experimental evidence references;
- download sequences used for profile construction or HMM/PWM motif models.

Conclusions

- CR-tag is a fingerprint uniquely matching transcription factors with their operators.
- Every regulatory motif record in [BacRegDB](#) is associated with a CR-tag and, therefore, can be correctly used to annotate matching elements in any genomes encoding a transcriptional regulator with the identical CR-tag.
- While [BacRegDB](#) is being developed, its full contents and regulatory sequence analysis tools can be accessed via our [Sigmoid](#) desktop application.

You are welcome to try BacRegDB yourself!



BacRegDB

<https://BacRegDB.bsu.by>



Sigmoid

<https://github.com/nikolaichik/Sigmoid>