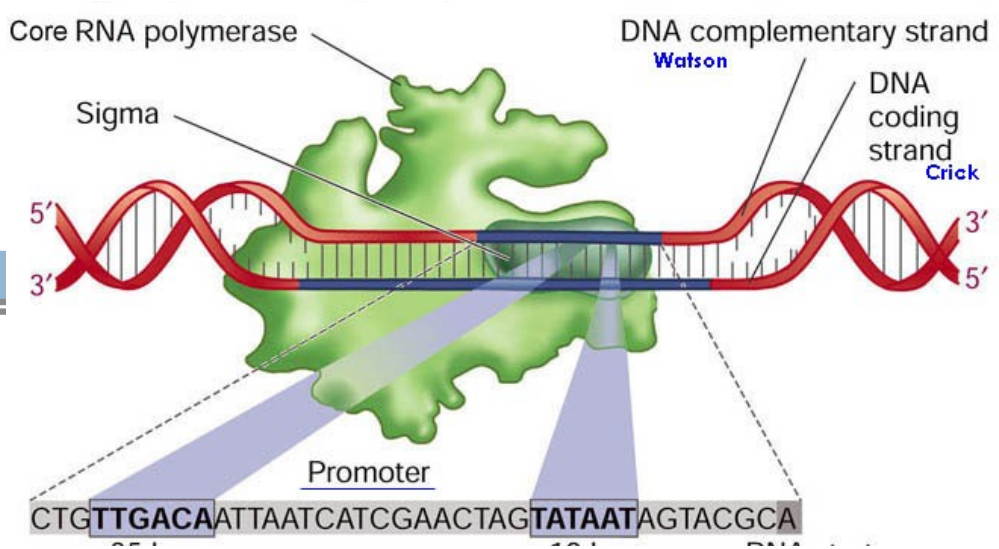
PROMOTOR MOTIF MINNING

Skryté markovské modely
(Some slides are courtesy of Mark Craven, U. of Wisconsin)

Motivation



these sequences are E. coli promoters

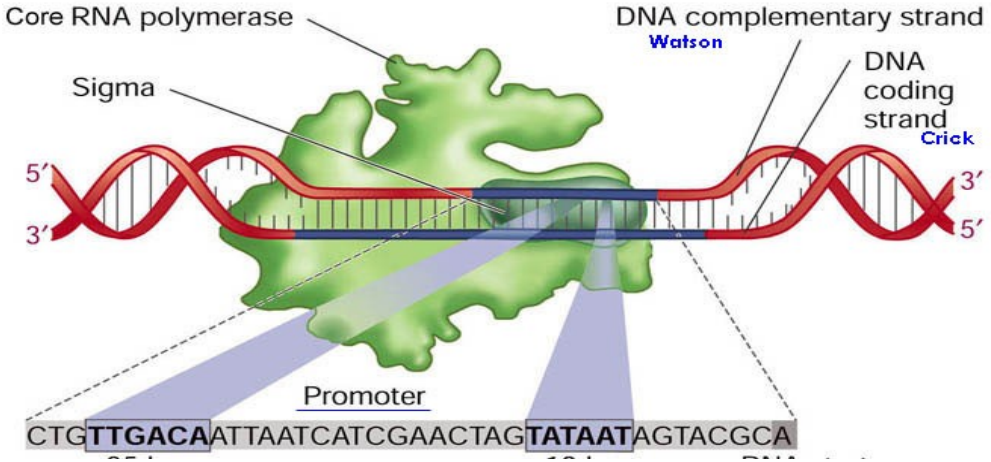
```
tctgaaatgagctgttgacaattaatcatcgaactagttaactagtacgcaagttca
accggaagaaaaccgtgacattttaacacgtttgttacaaggtaaaggcgacgccgc
aaattaaaatttttattgacttaggtcactaaatactttaaccaatataggcatagcg
ttgtcataatcgacttgtaaaccaaattgaaaagatttaggtttacaagtctacacc
catcctcgcaccagtcgacgacggtttacgctttacgtatagtggcgacaatttttt
tccagtataatttgttggcataattaagtacgacgagtaaaattacatacctgcccg
acagttatccactattcctgtggataaccatgtgtattagagttagaaaacacgagg
```

these sequences are not promoters

```
atagtctcagagtcttgacctactacgccagcattttggcggtgtaagctaaccatt
aactcaaggctgatacggcgagacttgcgagccttgtccttgcggtacacagcagcg
ttactgtgaacattattcgtctccgcgactacgatgagatgcctgagtgcttccgtt
tattctcaacaagattaaccgacagattcaatctcgtggatggacgttcaacattga
aacgagtcaatcagaccgctttgactctggtattactgtgaacattattcgtctccg
aagtgcttagcttcaaggtcacggatacgaccgaagcgagcctcgtcctcaatggcc
gaagaccacgcctcgccaccgagtagacccttagagagcatgtcagcctcgacaact
```

Assignment

- Download a local copy of BioProspector from http://motif.stanford.edu/distributions/bioprospector/
- Find binding motifs of SigA transcription factor for Bacillus subtilis (ref. Genome: http://www.ncbi.nlm.nih.gov/nuccore/AP012496.1?&feature=CDS)
- Compare found motifs with curated ones in http://dbtbs.hgc.jp/
- Deadline:
  next lesson
  award: 5 pt.



How?

- Linux: Make BioProspector.linux
- Windows: through http://cygwin.com/

Parameters:

- -i <seq_file> : promotor sequences (cases)
- -b <seq_file> : background sequences (controls), i.e. genetic background
- -W : 1st motif width
- -w=10 : 2nd motif width

Tasks

1. Learn on 7 promotor sequences only (bacil_red.fasta).
2. Learn on all the promotor sequences (bacil.fasta), but without refer. Genome.
3. Learn against only one reference gene (bacil_gene).
4. Learn with complete information
5. Compare the motifs found according to 1) - 4) in terms of dbtbs.hgc.jp/.
   Do they differ? If so, why?