

Übung 6

6.1) $Pr = \frac{s+1}{n+1} = \frac{8}{9} = 0,88$

6.2) Motifs

	¹ G	² A	³ T	⁴ A	⁵ C
	G	A	T	C	A
	A	T	T	G	A
	G	C	T	C	G

		¹	²	³	⁴	⁵
COUNT (Motifs)	A:	1	2	0	1	2
	C:	0	1	0	2	1
	G:	3	0	0	1	1
	T:	0	1	1	0	0

PROFILE (Motifs)	A:	1/4	2/4	0	1/4	2/4
	C:	0	1/4	0	2/4	1/4
	G:	3/4	0	0	1/4	1/4
	T:	0	1/4	1/4	0	0



		¹	²	³	⁴	⁵
PSEUDOCOUNTS (Motifs)	A:	2	3	1	2	3
	C:	1	2	1	3	2
	G:	4	1	1	2	2
	T:	1	2	5	1	1

PROFILE (Motifs)	A:	2/8	3/8	1/8	2/8	3/8
	C:	1/8	2/8	1/8	3/8	2/8
	G:	4/8	1/8	1/8	2/8	2/8
	T:	1/8	2/8	5/8	1/8	1/8

- ↪ Pseudocounts: Wahrscheinlichkeit aller Ereignisse wird verändert
- ↪ v.a. wichtig f. Ereignisse mit $Pr = 0$
- ↪ in Exp. mit kleinen n bedeutet $Pr = 0$ nicht zwingend, dass das Ereignis nie auftritt, sondern seltener ↪ Pseudocounts ↪ Ereignis wird einbezogen

6.3)

	A:	1/2	1/4	0	1
Profile:	C:	1/2	1/4	0	0
	T:	0	1/4	1/2	0
	G:	0	1/4	1/2	0

$Pr(\text{ACTA} | \text{Profile}) = 1/2 \cdot 1/4 \cdot 1/2 \cdot 1 = 1/16$

$Pr(\text{ACTG} | \text{Profile}) = 1/2 \cdot 1/4 \cdot 1/2 \cdot 0 = 0$

↪ wäre bei Pseudocounts $\neq 0$

$Pr_1(\text{CCTAA} | \text{Profile}) = 1/2 \cdot 1/4 \cdot 1/2 \cdot 1 = 1/16$

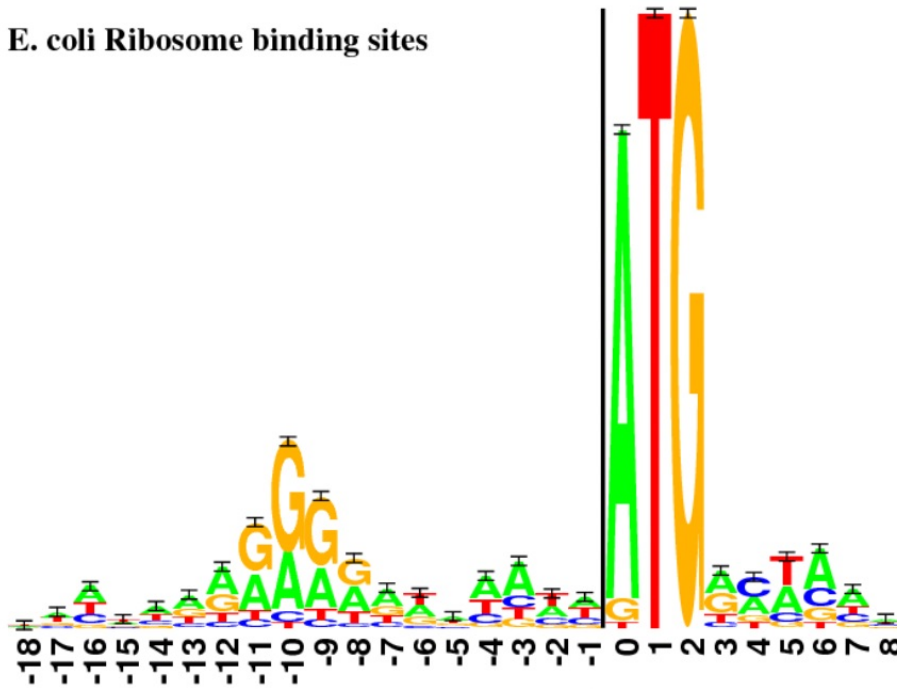
$Pr_2(\text{CCTAA} | \text{Profile}) = 1/2 \cdot 1/4 \cdot 0 \cdot 1 = 0$

↪ zwei Möglichkeiten, da String länger als Motifs

↪ wäre bei Pseudocounts $\neq 0$

6.4)

E. coli Ribosome binding sites



- Abb. ist ein Sequenzlogo
 - x-Achse: Position auf DNA-Strang
 - y-Achse: Häufigkeit des Nukleotids
 - Vergleich Häufigkeit d. Nukleotide auf DNA-Strang verschiedener Sequenzen (E. coli Ribosomenbindestelle)
- große Buchstaben kommen häufiger an dieser Position vor als kleine Buchstaben
- besonders deutlich: Pos. 1-3 => dort fast ausschließlich "ATG"
- ↗ konserviertes Motiv => wichtige Funktion!
 - ↗ ergibt Sinn, denn ATG = Startcodon ↗ hier oder davor müssen Ribosomen binden, um Translationsstart zu ermöglichen