

PSSM's with pseudocounts

		w=10									
Local multiple alignment ungapped	k=11 sequences w sites	G	T	C	A	G	G	T	G	T	G
		A	T	C	A	A	G	T	G	C	A
		A	C	C	A	C	C	T	G	C	A
		A	T	C	A	A	G	T	G	C	T
		G	G	C	A	G	G	T	G	T	C
		A	G	C	A	G	G	T	G	T	G
		G	G	C	A	G	T	T	G	A	A
		G	G	C	A	A	C	T	G	G	A
		G	T	C	A	G	G	T	G	G	C
		G	A	C	A	G	A	T	G	G	G
		G	T	C	A	A	G	T	G	G	G
Pseudocounts b=1	Σ =4	A	A	A	A	A	A	A	A	A	A
		G	G	G	G	G	G	G	G	G	G
		C	C	C	C	C	C	C	C	C	C
		T	T	T	T	T	T	T	T	T	T

Nucleotide counts

$c[x, i] + b$:	A	5	2	1	12	5	2	1	1	2	5
Number of copies of nucleotide x in column i.	G	8	5	1	1	7	8	1	12	5	5
	C	1	2	12	1	2	3	1	1	4	3
	T	1	6	1	1	1	2	12	1	4	2

Frequency matrix

$q[x, i] = \frac{c[x, i] + b}{k + \Sigma b}$	A	0.33	0.13	0.07	0.80	0.33	0.13	0.07	0.07	0.13	0.33
	G	0.53	0.33	0.07	0.07	0.47	0.53	0.07	0.80	0.33	0.33
	C	0.07	0.13	0.80	0.07	0.13	0.20	0.07	0.07	0.27	0.20
	T	0.07	0.40	0.07	0.07	0.07	0.13	0.80	0.07	0.27	0.13

$q[x, i]$:

Frequency of nucleotide x in column i, corrected with pseudocount b.

Likelihood ratio

$P[x, i] = \frac{q[x, i]}{p[x]}$	A	1.1	0.5	0.2	2.7	1.1	0.5	0.2	0.2	0.5	1.1
	G	2.6	1.6	0.3	0.3	2.3	2.6	0.3	3.9	1.6	1.6
	C	0.3	0.7	3.9	0.3	0.7	1.0	0.3	0.3	1.3	1.0
	T	0.2	1.4	0.2	0.2	0.2	0.5	2.7	0.2	0.9	0.5

Background frequency

A	0.295
G	0.205
C	0.205
T	0.295

PSSM: Log-odds scoring matrix

$S[x, i] = \log_2 P[x, i]$	A	0.18	-1.15	-2.15	1.44	0.18	-1.15	-2.15	-2.15	-1.15	0.18
	G	1.38	0.70	-1.62	-1.62	1.19	1.38	-1.62	1.96	0.70	0.70
	C	-1.62	-0.62	1.96	-1.62	-0.62	-0.04	-1.62	-1.62	0.38	-0.04
	T	-2.15	0.44	-2.15	-2.15	-2.15	-1.15	1.44	-2.15	-0.15	-1.15

Log-odds score of observing nucleotide x in column i.

Scoring a new sequence

PSSM: Log-odds scoring matrix

$$S[x, i] = \log_2 P[x, i]$$

A	0.18	-1.15	-2.15	1.44	0.18	-1.15	-2.15	-2.15	-1.15	0.18
G	1.38	0.70	-1.62	-1.62	1.19	1.38	-1.62	1.96	0.70	0.70
C	-1.62	-0.62	1.96	-1.62	-0.62	-0.04	-1.62	-1.62	0.38	-0.04
T	-2.15	0.44	-2.15	-2.15	-2.15	-1.15	1.44	-2.15	-0.15	-1.15

