

11 Exercises – Sequence profiles

1. PPM (Position probability matrix)

PWM (position weight matrix) is a popular method to find sequenced patterns. It can be generated from PPM (position probability matrix) and PFM (position frequency matrix).

Seq1 CAA

Seq2 CAG

Seq3 GAC

Seq4 ATT

(a) Create a PFM from Seq1, Seq2, Seq3, and Seq4.

	1	2	3
A			
G			
C			
T			

(b) Create a PPM from Seq1, Seq2, Seq3, and Seq4.

	1	2	3
A			
G			
C			
T			

2. Sequence profile

A sequence profile is similar to PWM, but it uses a scoring scheme. Use the following definitions to calculate the profile values.

$$Prof_{ra} : \frac{1}{m_r} \sum_{b \in M} R_{ba} F_{rb}$$

F_{rb} : The number of occurrences of b at position r

R_{ba} : Pairwise score between b and a

m_r : The number of residues without gaps at position r

Scoring matrix:

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

MSA

Seq1 GT
Seq2 -G
Seq3 CA

(a) Calculate the profile values of position 1.

A1:

G1:

C1:

T1:

(b) Calculate the profile values of position 2.

A2:

G2:

C2:

T2:

(c) Make a profile matrix.

	1	2
A		
G		
C		
T		

3. Profile search

A sequence profile can take gap penalties into account. Calculate the score of the alignment between the DNA profile below and a DNA segment.

A DNA profile of length 4

	A	G	C	T	Gap
1	5	-5	-2	-1	10
2	-2	3	4	-7	10
3	1	2	1	-1	5
4	-3	3	-2	7	10

- P1, P2, P3, P4: profile blocks at positions 1 - 4
- Gap penalty (for segments): 4

(a) Profile search on segment D1

Profile:	P1	P2	P3	P4
D1:	A	C	G	T

(b) Profile search on segment D2

Profile:	P1	P2	-	P3	P4
D2:	A	C	C	G	T

(c) Profile search on segment D3

Profile:	P1	P2	P3	P4
D3:	A	-	G	T

(d) Profile search on segment D4

Profile:	P1	P2	-	P3	P4
D4:	-	A	C	G	T