

Profile_1

Gegeben sei das folgende MSA von Erkennungssequenzen und die daraus abgeleitete, unvollständige Scoring-Matrix. Die Scores wurden mithilfe der log10-Funktion errechnet.

	1	2	3	4	
1234					
AATG	A	0.125	-0.18	0.125	?
ACCG	C	0	0	0	?
GGAA					
CTAC	G	-0.10	-0.10	-0.10	?
TTGT	T	-0.10	+0.20	-0.10	?

Wir müssen zunächst die Scores für Spalte vier errechnen.

Es gilt:

$$f(A, 4) = 0.2 \quad f(C, 4) = 0.2 \quad f(G, 4) = 0.4 \quad f(T, 4) = 0.2$$

Die mittleren Häufigkeiten sind:

$$f(A) = 6/20 = 0.3, \quad f(C) = 4/20 = 0.2, \quad f(G) = 5/20 = 0.25, \quad f(T) = 5/20 = 0.25$$

Damit folgt für den Score von S(ATTG):

$$S(A,4) = \log_{10}(0.2/0.3) = -0.18 \quad S(C,4) = (0.2/0.2) = 0, \quad S(G,4) = \log_{10}(0.4/0.25) = 0.20, \quad S(T,4) = \log_{10}(0.2/0.25) = -0.10$$

$$S(ATTG) = S(A,1) + S(T,2) + S(T,3) + S(G,4) = 0.125 + 0.20 - 0.10 + 0.20 = 0.425$$