

## Position Specific Scoring Matrices

### Local multiple alignment

$k$  sequences,  $w$  positions, no gaps

W	E	I	R	D
W	E	I	R	D
W	E	I	R	E
W	E	I	Q	H

$k$  rows,  $w$  columns

### Amino acid counts

$c[i, j]$  :

The number of copies of amino acid  $i$  in column  $j$ . For brevity, only residues that occur at least once are shown.

D	0	0	0	0	2
E	0	4	0	0	1
H	0	0	0	0	1
I	0	0	4	0	0
Q	0	0	0	1	0
R	0	0	0	3	0
W	4	0	0	0	0
$k$	4	4	4	4	4

$|\Sigma|$  rows,  $w$  columns

### Frequency matrix

$$q[i, j] = \frac{c[i, j]}{k}$$

The frequency of amino acid  $i$  in column  $j$ .

D					0.50
E		1.00			0.25
H					0.25
I			1.00		
Q				0.25	
R				0.75	
W	1.00				
SUM	1.0	1.0	1.0	1.0	1.0

### Propensity matrix

$$P[i, j] = \frac{q[i, j]}{p[i]}$$

**Note:** this is a likelihood ratio

D					9.6
E		16.1			4.0
H					10.9
I			18.9		
Q				6.1	
R				14.7	
W	71.4				

### Background Frequency, $p[i]$

D 0.052  
E 0.062  
H 0.023  
I 0.053  
Q 0.041  
R 0.051  
W 0.014

### Log odds scoring matrix

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

This is a log likelihood ratio

D						3.3
E		4.0				2.0
H						3.4
I			4.2			
Q				2.6		
R				3.9		
W	6.2					

### Scoring a new sequence:

	W	I	W	E	I	R	H
6.2	6.2						
0.0							
21.7			6.2	4.0	4.2	3.9	3.4