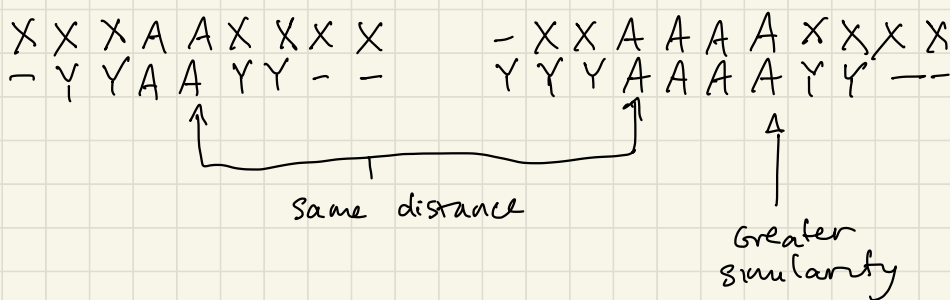


## Sequence Similarity

- Rather than aligning sequences so as to minimize the edit distance between them, we can align to maximize the similarity between them
- Aligning using similarity or distance typically produces same results when applied to global alignment problem
- Similarity is better criterion when concerned with local alignment

Consider these two alignments



$$w(a_i, b_i) = \begin{cases} 0 & \text{if } a_i = b_i \\ 1 & \text{otherwise} \end{cases}$$

$$s(a_i, b_i) = \begin{cases} 1 & \text{if } a_i = b_i \\ 0 & \text{otherwise} \end{cases}$$

# Global alignment w/ similarity scoring

$$a = A A C$$

$$b = A C A A$$

$$s(a_i, b_i) = \begin{cases} 1 & \text{if } a_i = b_i \\ -1 & \text{otherwise} \end{cases}$$

updating rules

		A	C	A	A
A	0	-1	-2	-3	-4
A	-1	1	0	-1	-2
A	-2	0	0	1	0
C	-3	-1	1	0	0

$$M_{0,0} = 0$$

$$M_{i,j} = \max \begin{cases} M_{i-1,j-1} + s(a_i, b_j) \\ M_{i-1,j} + s(a_i, -) \\ M_{i,j-1} + s(-, b_j) \end{cases}$$

1 A C A A  
A - A C

## Local Alignment Problem

- Smith-Waterman Algorithm

		A	C	A	A
A	0	0	0	0	0
A	0	1	0	1	1
A	0	1	0	1	2
C	0	0	2	1	1

$$H_{0,0} = 0$$

$$H_{i,0} = 0$$

$$H_{0,j} = 0$$

$$H_{i,j} = \max \begin{cases} 0 \\ H_{i-1,j-1} + s(a_i, b_j) \\ H_{i-1,j} + s(a_i, -) \\ H_{i,j-1} + s(-, b_j) \end{cases}$$

A: A A  
A A

B: A C  
A C

# Local Alignment Using Substitution Matrix

a = V E P P L S Q E

b = E L P L C

Gap penalty = -3

BLOSUM62 as scoring matrix

BLOSUM 62 scoring matrix

(positive values are shaded)

A	4																		
R	-1	5																	
N	-2	0	6																
D	-2	-2	1	6															
C	0	-3	-3	-3	9														
Q	-1	1	0	0	-3	5													
E	-1	0	0	2	-4	2	5												
G	0	-2	0	-1	-3	-2	-2	6											
H	-2	0	1	-1	-3	0	0	-2	8										
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4									
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4								
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5							
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5						
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6					
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7				
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4			
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5		
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	0	-3	-1	4
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

The values for amino acid substitutions were obtained from Henikoff S & Henikoff JG (1992) Amino acid substitutions matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89**: 10915-10919.

		+	V	E	P	P	L	S	Q	E
+	0	0	0	0	0	0	0	0	0	0
E	0	0	5	2	0	0	0	2	5	
L	0	1	2	2	0	4	1	0	2	
P	0	0	0	9	9	6	3	0	0	
L	0	1	0	6	6	13	10	7	4	
C	0	0	0	3	3	10	12	9	6	

E P P L

E L P L

alignment score = 13

# Global alignment using Substn Matrix

a = V E P P L S Q E

b = E L P L C

Gap Penalty = -3

BLOSUM62

	+	V	E	P	P	L	S	Q	E
+	0	-3	-6	-9	-12	-15	-18	-21	-24
E	-3	2	2	-1	-4	-7	-10	-13	-16
L	-6	-2	-1	-1	-4	0	-3	-6	-9
P	-9	-5	-3	6	6	3	0	-3	-6
L	-12	-8	-6	3	3	10	7	4	1
C	-15	-11	-9	0	0	7	9	6	-3

V E P P L S Q E  
- E L P L C - -

A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

BLOSUM 62 scoring matrix

(positive values are shaded)

BLOSUM 62 scoring matrix

(positive values are shaded)

The values for amino acid substitutions were obtained from Henikoff S & Henikoff JG (1992) Amino acid substitutions matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89**: 10915-10919.