

Sequence alignment

SEQ1 :

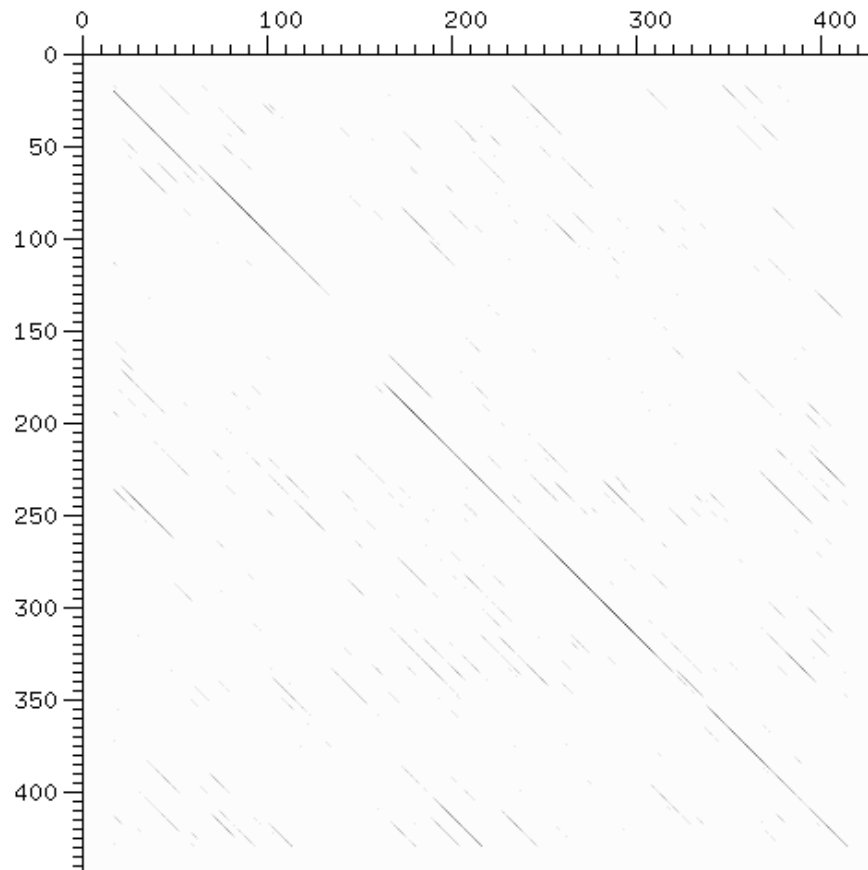
VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKK
VADALTNAVAHVDDPNALNSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHA
SLDKFLASVSTVLTISKYR

SEQ2 :

VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKVK
AHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLHVDPENFRLLGNVLCVLAHFFGKEF
TPPVQAAYQKVVAGVANALAHKYH

Are these two proteins alike?

Dotplots: visual sequence comparison



1. Place two sequences along axes of plot
 2. Place dot at grid points where two sequences have identical residues
 3. Diagonals correspond to conserved regions
-

Alignment scores: match vs. mismatch

Simple scoring scheme (too simple in fact...):

Matching amino acids: 5

Mismatch: 0

Scoring example:

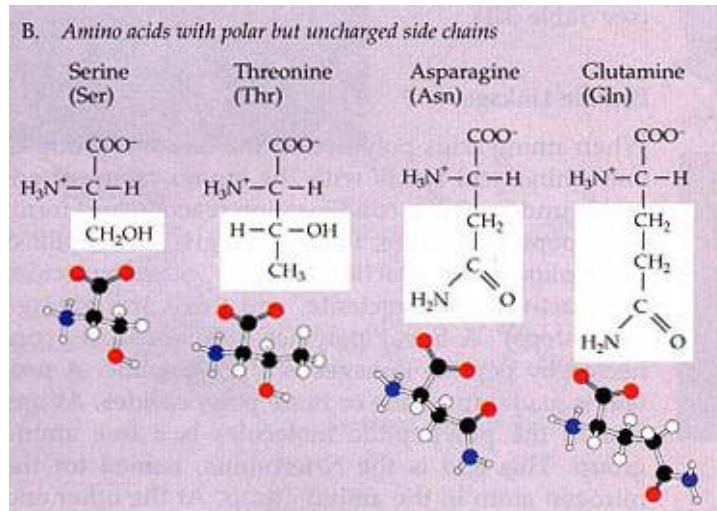
K A W S A D V

: : : : :

K D W S A E V

5+0+5+5+5+0+5 = 25

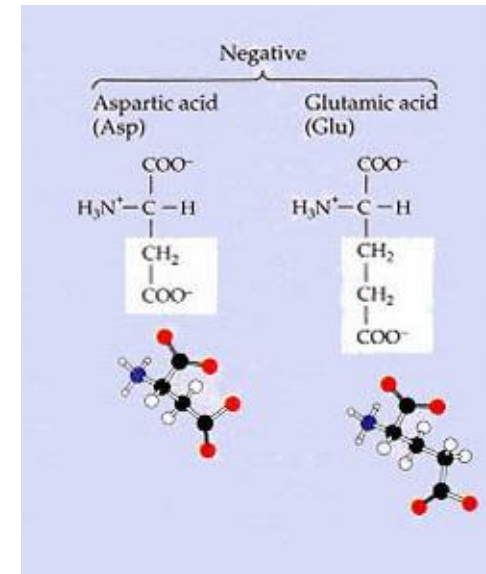
Amino acid properties



Serine (S) and Threonine (T) have similar physicochemical properties

=> Substitution of S/T or E/D occurs relatively often during evolution

=> Substitution of S/T or E/D should result in scores that are only moderately lower than identities



Aspartic acid (D) and Glutamic acid (E) have similar properties

Protein substitution matrices

A	5																			
R	-2	7																		
N	-1	-1	7																	
D	-2	-2	2	8																
C	-1	-4	-2	-4	13															
Q	-1	1	0	0	-3	7														
E	-1	0	0	2	-3	2	6													
G	0	-3	0	-1	-3	-2	-3	8												
H	-2	0	1	-1	-3	1	0	-2	10											
I	-1	-4	-3	-4	-2	-3	-4	-4	-4	5										
L	-2	-3	-4	-4	-2	-2	-3	-4	-3	2	5									
K	-1	3	0	-1	-3	2	1	-2	0	-3	-3	6								
M	-1	-2	-2	-4	-2	0	-2	-3	-1	2	3	-2	7							
F	-3	-3	-4	-5	-2	-4	-3	-4	-1	0	1	-4	0	8						
P	-1	-3	-2	-1	-4	-1	-1	-2	-2	-3	-4	-1	-3	-4	10					
S	1	-1	1	0	-1	0	-1	0	-1	-3	-3	0	-2	-3	-1	5				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	2	5			
W	-3	-3	-4	-5	-5	-1	-3	-3	-3	-3	-2	-3	-1	1	-4	-4	-3	15		
Y	-2	-1	-2	-3	-3	-1	-2	-3	2	-1	-1	-2	0	4	-3	-2	-2	2	8	
V	0	-3	-3	-4	-1	-3	-3	-4	-4	4	1	-3	1	-1	-3	-2	0	-3	-1	5
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

BLOSUM50 matrix:

- Positive scores on diagonal (identities)
- Similar residues get higher (positive) scores
- Dissimilar residues get smaller (negative) scores

When Blast works!

