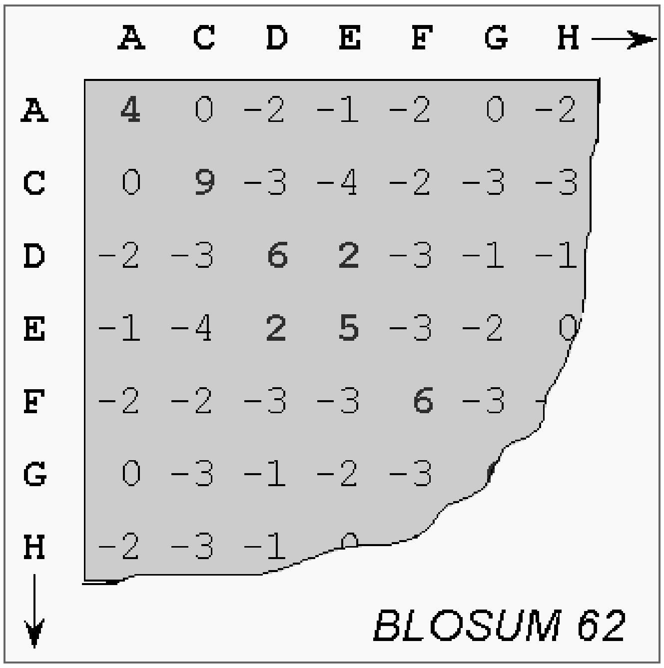
Scoring Systems

back to Education

Log-odds Substitution Matrices for Scoring Amino-Acid Alignments



	A	C	D	E	F	G	H
A	4	0	-2	-1	-2	0	-2
C	0	9	-3	-4	-2	-3	-3
D	-2	-3	6	2	-3	-1	-1
E	-1	-4	2	5	-3	-2	0
F	-2	-2	-3	-3	6	-3	
G	0	-3	-1	-2	-3		
H	-2	-3	-1	0			

BLOSUM62 Substitution Scoring Matrix. The BLOSUM 62 matrix shown here is a 20 x 20 matrix of which a section is shown here in which every possible identity and substitution is assigned a score based on the observed frequencies of such occurences in alignments of related proteins. Identities are assigned the most positive scores. Frequently observed substitutions also receive positive scores and seldom observed substitutions are given negative scores.

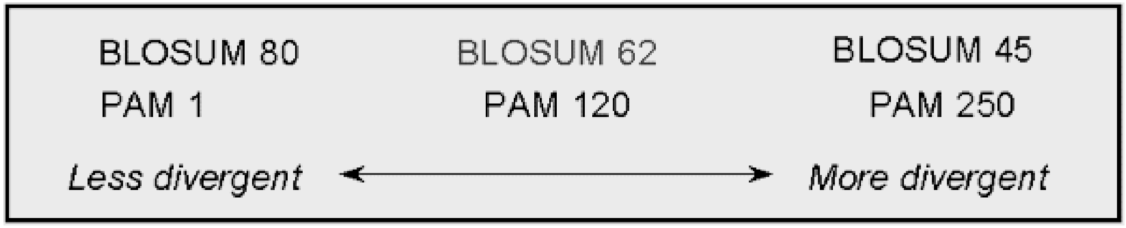
The PAM family

·PAM matrices are based on global alignments of closely related proteins.
·The PAM1 is the matrix calculated from comparisons of sequences with no more than 1% divergence.
·Other PAM matrices are extrapolated from PAM1.

The BLOSUM family

·BLOSUM matrices are based on local alignments.
·BLOSUM 62 is a matrix calculated from comparisons of sequences with no less than 62% divergence.
·All BLOSUM matrices are based on observed alignments; they are not extrapolated from comparisons of closely related proteins.
·BLOSUM 62 is the default matrix in BLAST 2.0. Though it is tailored for comparisons of moderately distant proteins, it performs well in detecting closer

relationships. A search for distant relatives may be more sensitive with a different matrix.



The relationship between BLOSUM and PAM substitution matrices. BLOSUM matrices with higher numbers and PAM matrices with low numbers are both designed for comparisons of closely related sequences. BLOSUM matrices with low numbers and PAM matrices with high numbers are designed for comparisons of distantly related proteins. If distant relatives of the query sequence are specifically being sought, the matrix can be tailored to that type of search.

Revised February 24, 2000