

# Scoring Matrix

- 1. Scoring system is a set of values for qualifying the set of one residue being substituted by another in an alignment**
- 2. It is also known as substitution matrix**
- 3. Scoring matrix of nucleotide is relatively simple**
- 4. Scoring matrices for amino acids are more complicated because scoring has to reflect the physiochemical properties of amino acid residues.**

# Scoring Matrices for Aligning DNA Sequences

	A	T	C	G
A	1	0	0	0
T	0	1	0	0
C	0	0	1	0
G	0	0	0	1

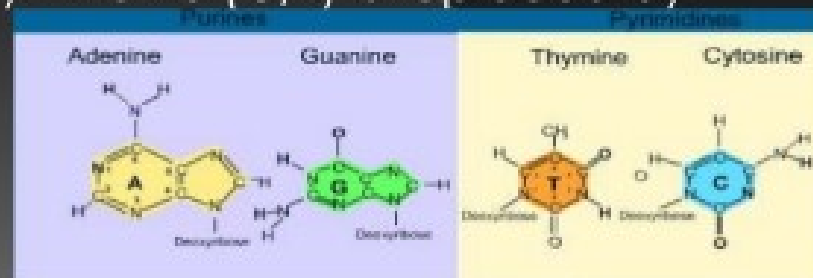
Identity matrix

	A	T	C	G
A	1	-5	-5	-1
T	-	1	-1	-5
C	5	-1	1	-5
G	5	-5	-5	1

Transition-Transversion matrix

**Transition** --- substitutions in which a purine (A/G) is replaced by another purine (A/G) or a pyrimidine (C/T) is replaced by another pyrimidine (C/T).

**Transversions** ---  
(A/G)  $\leftrightarrow$  (C/T)



# Scoring a sequence alignment

- Match score: +1
- Mismatch score: +0
- Gap penalty: -1

● ACGTCTGAT**A**CGCCGTAT**A**GTCTATCT  
          | | | | | | | | | | | | | | | |  
----CTGAT**T**CGC---AT**C**GTCTATCT

- Matches:  $18 \times (+1)$
- Mismatches:  $2 \times 0$
- Gaps:  $7 \times (-1)$

Score = +11

# Amino Acid Substitution matrices

**PAM: Point accepted mutation based on global alignment**

**PAM matrix were derived based on evolutionary divergence between sequences of protein structure.**

**Construction of PAM 1 matrix involves alignment of full length sequence and subsequent Construction of phylogenic trees using parsimony principle.**

**A PAM is defined as 1% aminoacid change or one mutation per 100 residues.**

**The increasing PAM numbers correlate with increasing PAM units and thus evolutionary distances of protein sequences.**

## **Limitations of PAM Matrices**

- 1. Constructed based on the phylogenetic relationships prior to scoring mutations.**
- 2. Difficulty of determining ancestral relationships among sequences.**
- 3. Based on a small set of closely related proteins**

# BLOSUM

**Blosum: Block substitutions based on local alignments**

- 1. It is derived on the basis of direct observation for every possible aminoacid substitution in multiple sequence alignment**
- 2. Sequence pattern is also called as block.**
- 3. Ungapped alignments are less than 60 aminoacids in length.**
- 4. Blosum matrix are actual % values of sequence selected for construction of matrix.**

# Comparison

<b>PAM</b>	<b>BLOSUM</b>
<b>Based on mutational model of evolution</b>	<b>Based on the multiple alignment of blocks</b>
<b>Based on sequences of 85% similarity</b>	<b>Good to be used to compare distant sequences</b>
<b>Designed to track the evolutionary origins</b>	<b>Designed to find proteins conserved domains</b>