Module 1b

Bioinformatics for Biologists

How similar are two sequences?

1. How similar are two sequences?

Seq1: GANDALF **Seq2:** SCANDAL

Pairwise alignment

GANDALF	-GANDALF	G-ANDALF	--GANDALF	G--ANDALF
SCANDAL	SCANDAL-	SCANDAL-	SC-ANDAL-	-SCANDAL-

Which is the better alignment?

How can we evaluate which is the better alignment?

1. How similar are two sequences?

Seq1: GANDALF **Seq2:** SCANDAL

Pairwise alignment

GANDALF	-GANDALF	G-ANDALF	--GANDALF	G--ANDALF
SCANDAL	SCANDAL-	SCANDAL-	SC-ANDAL-	-SCANDAL-

Which is the better alignment?

We need a substitution matrix and a gap penalty to evaluate the alignments.

Substitution matrix

Gives the score or probability for an amino acid i to be substituted into amino acid j.

Amino acid frequencies for pairs of amino acids are estimated from a database or a large dataset.

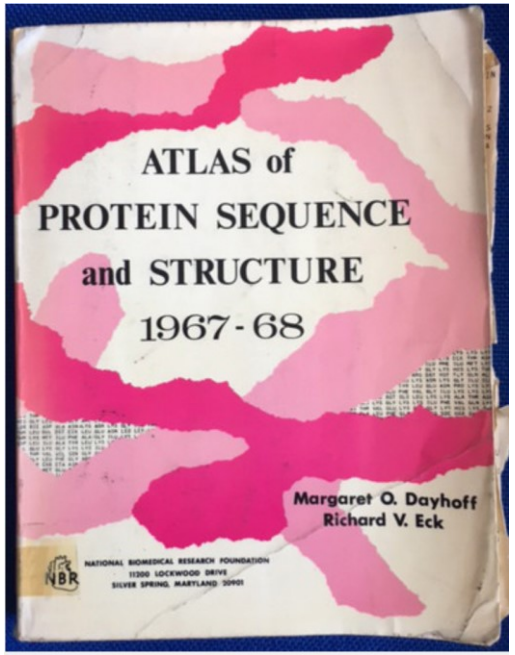
Margaret Dayhoff

Pioneer of bioinformatics

Atlas of protein sequence and structure

was a book with sequences

PAM matrices



key for the development of molecular biology, molecular evolution and bioinformatics.

PAM Step 1: **A**ccepted **P**oint **M**utations (APM or PAM) are defined not by the pairwise alignment but with respect to the common ancestor

Dayhoff et al. evaluated amino acid changes. They applied an evolutionary model to compare changes such as 1 versus 2 not to each other but to an inferred common ancestor at position 5.

PAM Step 2: Frequency of amino acids

TABLE 3.1 Normalized frequencies of amino acid. These values sum to 1. If the 20 amino acids were equally represented in proteins, these values would all be 0.05 (i.e., 5%); instead, amino acids vary in their frequency of occurrence.

Gly	0.089	Arg	0.041
Ala	0.087	Asn	0.040
Leu	0.085	Phe	0.040
Lys	0.081	Gln	0.038
Ser	0.070	Ile	0.037
Val	0.065	His	0.034
Thr	0.058	Cys	0.033
Pro	0.051	Tyr	0.030
Glu	0.050	Met	0.015
Asp	0.047	Trp	0.010

L, S, R =6 codons; M, W =1 codon in the genetic code

PAM Step 3: amino acid substitutions

	A Ala	R Arg	N Asn	D Asp	C Cys	Q Gln	E Glu	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
A																				
R	30																			
N	109	17																		
D	154	0	532																	
C	33	10	0	0																
Q	93	120	50	76	0															
E	266	0	94	831	0	422														
G	579	10	156	162	10	30	112													
H	21	103	226	43	10	243	23	10												
I	66	30	36	13	17	8	35	0	3											
L	95	17	37	0	y	75	15	17	40	253										
K	57	477	322	85	0	147	104	60	23	43	39									
M	29	17	0	0	0	20	7	7	0	57	207	90								
F	20	7	7	0	0	0	0	17	20	90	167	0	17							
P	345	67	27	10	10	93	40	49	50	7	43	43	4	7						
S	772	137	432	98	117	47	86	450	26	20	32	168	20	40	269					
T	590	20	169	57	10	37	31	50	14	129	52	200	28	10	73	696				
W	0	27	3	0	0	0	0	0	3	0	13	0	0	10	0	17	0			
Y	20	3	36	0	30	0	10	0	40	13	23	10	0	260	0	22	23	6		
V	365	20	13	17	33	27	37	97	30	661	303	17	77	10	50	43	186	0	17	
	A Ala	R Arg	N Asn	D Asp	C Cys	Q Gln	E Glu	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val

PAM Step 3: amino acid substitutions

	A Ala	R Arg	N Asn	D Asp	C Cys	Q Gln	E Glu	G Gly
A								
R	30							
N	109	17						
D	154	0	532					
C	33	10	0	0				
Q	93	120	50	76	0			
E	266	0	94	831	0	422		
G	579	10	156	162	10	30	112	
H	21	103	226	43	10	243	23	10

Zooming in on the previous table, note that substitutions are very common (e.g. D → E, A → G) while others are rare (e.g. C → Q, C → E). The scoring system we use for pairwise alignments should reflect these trends.

PAM Step 4: Mutation probability matrix for the evolutionary distance of 1 PAM (= 1 per cent accepted mutations; 99 % sequence identity)

Replacement amino acid	Original amino acid: A Ala	R Arg	N Asn	D Asp	C Cys	Q Gln	E Glu	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
A	98.7	0.0	0.1	0.1	0.0	0.1	0.2	0.2	0.0	0.1	0.0	0.0	0.1	0.0	0.2	0.4	0.3	0.0	0.0	0.2
R	0.0	99.1	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0
N	0.0	0.0	98.2	0.4	0.0	0.0	0.1	0.1	0.2	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0
D	0.1	0.0	0.4	98.6	0.0	0.1	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Q	0.0	0.1	0.0	0.1	0.0	98.8	0.3	0.0	0.2	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
E	0.1	0.0	0.1	0.6	0.0	0.4	98.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	0.2	0.0	0.1	0.1	0.0	0.0	0.1	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1
H	0.0	0.1	0.2	0.0	0.0	0.2	0.0	0.0	99.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.7	0.1	0.0	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.3
L	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	99.5	0.0	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.2
K	0.0	0.4	0.3	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	99.3	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.0
M	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	98.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
F	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	99.5	0.0	0.0	0.0	0.0	0.3	0.0
P	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	99.3	0.1	0.0	0.0	0.0	0.0
S	0.3	0.1	0.3	0.1	0.1	0.0	0.1	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.2	98.4	0.4	0.1	0.0	0.0
T	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.3	98.7	0.0	0.0	0.1
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	99.5	0.0
V	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	99.0

This mutation probability matrix includes original amino acids (columns) and replacements (rows). The diagonals show that at a distance of 1 PAM most residues remain the same about 99% of the time (see shaded entries). Note how cysteine (C) and tryptophan (W) undergo few substitutions, and asparagine (N) many.

PAM Step 5: PAM0 and PAM∞

original amino acid

replacement amino acid

PAM0	A	R	N	D	C	Q	E	G
A	100	0	0	0	0	0	0	0
R	0	100	0	0	0	0	0	0
N	0	0	100	0	0	0	0	0
D	0	0	0	100	0	0	0	0
C	0	0	0	0	100	0	0	0
Q	0	0	0	0	0	100	0	0
E	0	0	0	0	0	0	100	0
G	0	0	0	0	0	0	0	100

original amino acid

replacement amino acid

PAM∞	A	R	N	D	C	Q	E	G
A	8.7	8.7	8.7	8.7	8.7	8.7	8.7	8.7
R	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.1
N	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
D	4.7	4.7	4.7	4.7	4.7	4.7	4.7	4.7
C	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3
Q	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8
E	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
G	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9

At the extreme of perfectly conserved proteins (PAM0) there are no amino acid replacements. At the extreme of completely diverged proteins (PAM∞) the matrix converges on the background frequencies of the amino acids.

PAM250 matrix: for proteins that share ~20% identity

Replacement amino acid \ Original amino acid	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	13	6	9	9	5	8	9	12	6	8	6	7	7	4	11	11	11	2	4	9
R	3	17	4	3	2	5	3	2	6	3	2	9	4	1	4	4	3	7	2	2
N	4	4	6	7	2	5	6	4	6	3	2	5	3	2	4	5	4	2	3	3
D	5	4	8	11	1	7	10	5	6	3	2	5	3	1	4	5	5	1	2	3
C	2	1	1	1	52	1	1	2	2	2	1	1	1	1	2	3	2	1	4	2
Q	3	5	5	6	1	10	7	3	7	2	3	5	3	1	4	3	3	1	2	3
E	5	4	7	11	1	9	12	5	6	3	2	5	3	1	4	5	5	1	2	3
G	12	5	10	10	4	7	9	27	5	5	4	6	5	3	8	11	9	2	3	7
H	2	5	5	4	2	7	4	2	15	2	2	3	2	2	3	3	2	2	3	2
I	3	2	2	2	2	2	2	2	2	10	6	2	6	5	2	3	4	1	3	9
L	6	4	4	3	2	6	4	3	5	15	34	4	20	13	5	4	6	6	7	13
K	6	18	10	8	2	10	8	5	8	5	4	24	9	2	6	8	8	4	3	5
M	1	1	1	1	0	1	1	1	1	2	3	2	6	2	1	1	1	1	1	2
F	2	1	2	1	1	1	1	1	3	5	6	1	4	32	1	2	2	4	20	3
P	7	5	5	4	3	5	4	5	5	3	3	4	3	2	20	6	5	1	2	4
S	9	6	8	7	7	6	7	9	6	5	4	7	5	3	9	10	9	4	4	6
T	8	5	6	6	4	5	5	6	4	6	4	6	5	3	6	8	11	2	3	6
W	0	2	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	55	1	0
Y	1	1	2	1	3	1	1	1	3	2	2	1	2	15	1	2	2	3	31	2
V	7	4	4	4	4	4	4	5	4	15	10	4	10	5	5	5	7	2	4	17

PAM1

Replacement amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	
A	98.7	0.0	0.1	0.1	0.0	0
R	0.0	99.1	0.0	0.0	0.0	0
N	0.0	0.0	98.2	0.4	0.0	0
D	0.1	0.0	0.4	98.6	0.0	0
C	0.0	0.0	0.0	0.0	99.7	0
Q	0.0	0.1	0.0	0.1	0.0	9
E	0.1	0.0	0.1	0.6	0.0	0
G	0.2	0.0	0.1	0.1	0.0	0
H	0.0	0.1	0.2	0.0	0.0	0
I	0.0	0.0	0.0	0.0	0.0	0
L	0.0	0.0	0.0	0.0	0.0	0
K	0.0	0.4	0.3	0.1	0.0	0
M	0.0	0.0	0.0	0.0	0.0	0
F	0.0	0.0	0.0	0.0	0.0	0
P	0.1	0.1	0.0	0.0	0.0	0
S	0.3	0.1	0.3	0.1	0.1	0
T	0.2	0.0	0.1	0.0	0.0	0
W	0.0	0.0	0.0	0.0	0.0	0
Y	0.0	0.0	0.0	0.0	0.0	0
V	0.1	0.0	0.0	0.0	0.0	0

Compare this to a PAM1 matrix, and note the diagonal still has high scores but much information content is lost.

PAM Step 6: from a mutation probability matrix to a relatedness odds matrix

$$R_{ij} = \frac{M_{ij}}{f_i}$$

A relatedness odds matrix reports the probability that amino acid *j* will change to *i* in a homologous sequence.

The numerator models the observed change. The denominator *fi* is the probability of amino acid residue *i* occurring in the second sequence by chance.

A positive value indicates a replacement happens more often than expected by chance. A negative value indicates the replacement is not favored.

PAM Step 7: log odds scoring matrix

$$s_{ij} = 10 \times \log_{10}\left(\frac{M_{ij}}{f_i}\right)$$

A log odds matrix is the logarithmic form of the relatedness odds matrix. s_{ij} is the score for aligning any two residues in a pairwise alignment. (There is also a score for aligning a residue with itself.) M_{ij} is of the observed frequency of substitutions for each pair of amino acids. These values ("target frequencies") are derived from a mutation probability matrix.

Example of a score for aligning cysteine and leucine using the values in a PAM250 scoring matrix:

$$s_{(\text{cysteine, leucine})} = 10 \times \log_{10}\left(\frac{0.02}{0.085}\right) = -6.3$$

Substitution matrix

Amino acid frequencies for pairs of amino acids are estimated from a database or a dataset.

PAM (Point Accepted Mutations)

Score of mutation $i \leftrightarrow j$ =

$$10 \times \log_{10} \frac{observed\ i \leftrightarrow j\ mutation\ rate}{mutation\ rate\ expected\ from\ amino\ acid\ frequencies}$$

1 PAM = 1 per cent accepted mutations; 99 % sequence identity

PAM30 – 75 % sequence identity

PAM80 – 50 % sequence identity

PAM110 – 60 % sequence identity

PAM200 – 25 % sequence identity

PAM250 – 20 % sequence identity

- This matrix was derived by Margaret Dayhoff in the 70s.
- The sequence data used to calculate the PAM distances consisted of less than 1600 mutations in 71 protein families. In each protein family, the proteins had at least 85% sequence identity, but the entire protein length was used (GLOBAL alignment) and sequences were analyzed in a phylogenetic context.

Substitution matrix

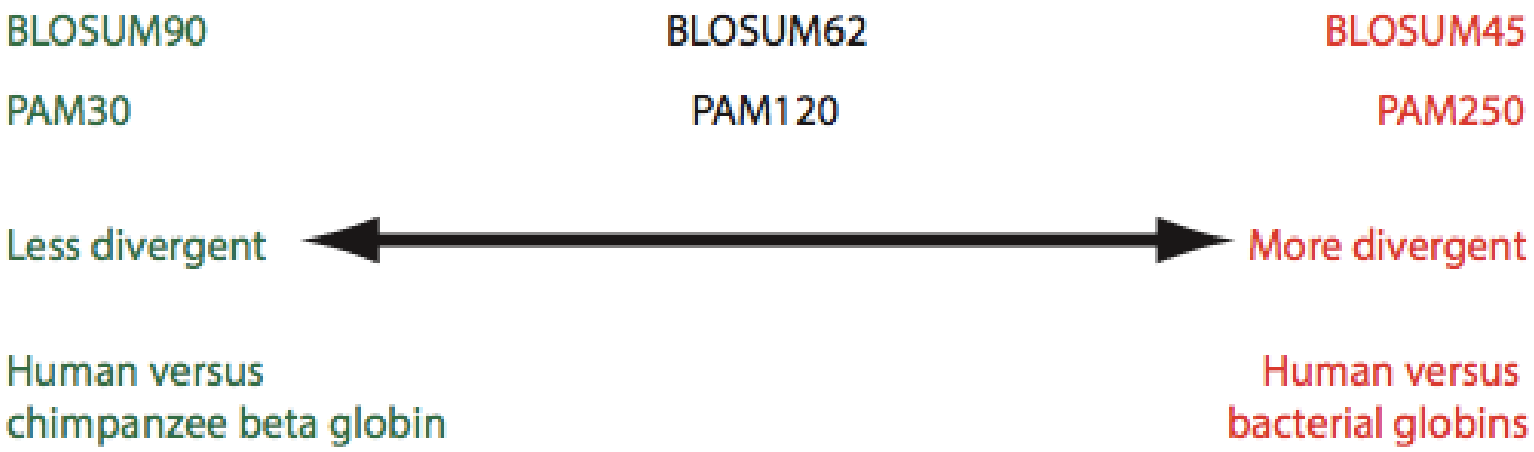
BLOSUM (BLOcks SUbstitution Matrix)

Goal: To generate matrices for less similar proteins.

- Used blocks of protein alignments (LOCAL alignment) without gaps.
- Does not consider the phylogenetic context.
- Different threshold for protein sequence identity was used to generate different matrices.
- BLOSUM80 – 80 % sequence id
- BLOSUM62 – 62 % sequence id
- BLOSUM45 – 45 % sequence id

PAM & BLOSUM

BLOSUM62 is the default matrix for BLAST



A higher PAM number, and a lower BLOSUM number, tends to correspond to a matrix tuned to more divergent proteins.

BLOSUM62

(A)

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

PAM120

(B)

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9																			
S	−1	3																		
T	−3	2	4																	
P	−3	1	−1	6																
A	−3	1	1	1	3															
G	−5	1	−1	−2	1	5														
N	−5	1	0	−2	0	0	4													
D	−7	0	−1	−2	0	0	2	5												
E	−7	−1	−2	−1	0	−1	1	3	5											
Q	−7	−2	−2	0	−1	−3	0	1	2	6										
H	−4	−2	−3	−1	−3	−4	2	0	−1	3	7									
R	−4	−1	−2	−1	−3	−4	−1	−3	−3	1	1	6								
K	−7	−1	−1	−2	−2	−3	1	−1	−1	0	−2	2	5							
M	−6	−2	−1	−3	−2	−4	−3	−4	−4	−1	−4	−1	0	8						
I	−3	−2	0	−3	−1	−4	−2	−3	−3	−3	−4	−2	−2	1	6					
L	−7	−4	−3	−3	−3	−5	−4	−5	−4	−2	−3	−4	−4	3	1	5				
V	−2	−2	0	−2	0	−2	−3	−3	−3	−3	−3	−3	−4	1	3	1	5			
F	−6	−3	−4	−5	−4	−5	−4	−7	−6	−6	−2	−4	−6	−1	0	0	−3	8		
Y	−1	−3	−3	−6	−4	−6	−2	−5	−4	−5	−1	−6	−6	−4	−2	−3	−3	4	8	
W	−8	−2	−6	−7	−7	−8	−5	−8	−8	−6	−5	1	−5	−7	−7	−5	−8	−1	−1	12

Fig. from Zvelebli and Baum, Understanding Bioinformatics, 1st Ed.

Gap penalty

- Sequences also evolve by insertion and deletion (indel event).

- How can we score an indel event?

1. Starting a gap in an alignment is more costly (commonly -10).
2. Once the gap is there, extending it is cheaper (commonly -1).

→ *What is the biological rationale for making gap openings more costly than gap extensions?*

1. How similar are two sequences?

Seq1: GANDALF **Seq2:** SCANDAL

Pairwise alignment

-GANDALF	G-ANDALF	--GANDALF	G--ANDALF
SCANDAL-	SCANDAL-	SC-ANDAL-	-SCANDAL-

Which is the better alignment?

We need a substitution matrix and a gap penalty to evaluate the alignments.

1. How similar are two sequences?

Why Courier New font? Every letter is the same size so they will always line up!

Seq1: GANDALF **Seq2:** SCANDAL

Pairwise alignment

-GANDALF	G-ANDALF	--GANDALF	G--ANDALF
SCANDAL-	SCANDAL-	SC-ANDAL-	-SCANDAL-

Which is the better alignment?

We need a substitution matrix and a gap penalty to evaluate the alignments.

1. How similar are two sequences?

Seq1: GANDALF

Seq2: SCANDAL

Pairwise alignment

```
-GANDALF    G-ANDALF    --GANDALF    G--ANDALF
SCANDAL-    SCANDAL-    SC-ANDAL-    -SCANDAL-
```

Substitution matrix

G ↔ S 4

G ↔ C 2

Same = 5

Gap penalty

open -5

extend -1

1. How similar are two sequences?

Seq1: GANDALF

Seq2: SCANDAL

Pairwise alignment

-GANDALF	G-ANDALF	--GANDALF	G--ANDALF
SCANDAL-	SCANDAL-	SC-ANDAL-	-SCANDAL-
*(-5)+2+(5*5)+(-5)=17*	*(4)+(-5)+(5*5)+(-5)=19*	*(-5)+(-1)+(-5)+(5*5)+(-5)=9*	*(-5)+(-5)+(-1)+(5*5)+(-5)=9*

Substitution matrix

G ↔ S 4

G ↔ C 2

Same = 5

Gap penalty

open -5

extend -1

With BLAST we need to find the sequences to align in a database…

And this will be the topic of the next module.

Download GenPept Graphics

phenylalanine-4-hydroxylase [Homo sapiens]

Sequence ID: ref|NP_000268.1| **Length:** 452 **Number of Matches:** 1

Range 1: 30 to 451 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
467 bits(1202)	7e-160	Compositional matrix adjust.	224/428(52%)	304/428(71%)	14/428(3%)

```
Query  78   DGNAVLNLLFSLRGTKPSSLSRAVKVFETFEAKIHHLETRPAQRPLAGSPHLEYFVRFEV  137
            + N  ++L+FSL+  +  +L++ +++FE  +  + H+E+RP++         E+F   +
Sbjct  30   NQNGAISLIFSLK-EEVGALAKVLRLFEENDVNLTHIESRPSR---LKKDEYEFFTHLDK  85

Query  138  PSGDLAALLSSVRRVSDDV--------RSAREDKVPWFPRKVSELDKCHHLVTKFDPDLD  189
             S  L AL + ++ +  D+         R  ++D VPWFPR + ELD+  + +  +  +LD
Sbjct  86   RS--LPALTNIIKILRHDIGATVHELSRDKKKDTVPWFPRTIQELDRFANQILSYGAELD  143

Query  190  LDHPGFSDQVYRQRRKLIAEIAFQYKHGEPIPHVEYTAEEIATWKEVYVTLKGLYATHAC  249
             DHPGF D VYR RRK  A+IA+ Y+HG+PIP VEY  EE  TW  V+ TLK LY THAC
Sbjct  144  ADHPGFKDPVYRARRKQFADIAYNYRHGQPIPRVEYMEEEKKTWGTVFKTLKSLYKTHAC  203

Query  250  REHLEGFQLLERYCGYREDSIPQLEDVSRFLKERTGFQLRPVAGLLSARDFLASLAFRVF  309
             E+   F LLE+YCG+ ED+IPQLEDVS+FL+  TGF+LRPVAGLLS+RDFL  LAFRVF
Sbjct  204  YEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCTGFRLRPVAGLLSSRDFLGGLAFRVF  263

Query  310  QCTQYIRHASSPMHSPEPDCCHELLGHVPMLADRTFAQFSQDIGLASLGASDEEIEKLST  369
             CTQYIRH S PM++PEPD CHELLGHVP+ +DR+FAQFSQ+IGLASLGA DE IEKL+T
Sbjct  264  HCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGAPDEYIEKLAT  323

Query  370  VYWFTVEFGLCKQNGELKAYGAGLLSSYGELLHSLSEEPEVRAFDPDTAAVQPYQDQTYQ  429
            +YWFTVEFGLCKQ   +KAYGAGLLSS+GEL + LSE+P++   + +  A+Q Y    +Q
Sbjct  324  IYWFTVEFGLCKQGDSIKAYGAGLLSSFGELQYCLSEKPKLLPLELEKTAIQNYTVTEFQ  383

Query  430  PVYFVSESFNDAKDKLRNYASRIQRPFSVKFDPYTLAIDVLDSPHTIQRSLEGVQDELHT  489
            P+Y+V+ESFNDAK+K+RN+A+ I RPFSV++DPYT  I+VLD+   ++   + +  E+
Sbjct  384  PLYYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQRIEVLDNTQQLKILADSINSEIGI  443

Query  490  LAHALSAI  497
             L  AL  I
Sbjct  444  LCSALQKI  451
```