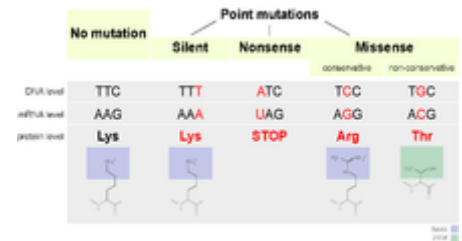


# Point accepted mutation

A **point accepted mutation** — also known as a **PAM** — is the replacement of a single amino acid in the primary structure of a protein with another single amino acid, which is accepted by the processes of natural selection. This definition does not include all point mutations in the DNA of an organism. In particular, silent mutations are not point accepted mutations, nor are mutations which are lethal or which are rejected by natural selection in other ways.

A PAM matrix is a matrix where each column and row represents one of the twenty standard amino acids. In bioinformatics, PAM matrices are regularly used as substitution matrices to score sequence alignments for proteins. Each entry in a PAM matrix indicates the likelihood of the amino acid of that row being replaced with the amino acid of that column through a series of one or more point accepted mutations during a specified evolutionary interval, rather than these two amino acids being aligned due to chance. Different PAM matrices correspond to different lengths of time in the evolution of the protein sequence.



An example of point mutations at an amino acid site coding for lysine. The missense mutations may be classed as point accepted mutations if the mutated protein is not rejected by natural selection.

## Contents

### Biological background

#### Terminology

#### Construction of PAM matrices

- Collection of data from phylogenetic trees
- Construction of the mutation matrix
  - Choice of the constant of proportionality
- Construction of the PAM<sub>n</sub> matrices

#### Properties of the PAM matrices

- Symmetry of the PAM matrices
- Relating the number of mutated amino acids and the number of mutations

#### An example - PAM250

#### Uses in bioinformatics

- Determining the time of divergence in phylogenetic trees
- Use in BLAST
  - Comparing PAM and BLOSUM

#### See also

#### References

#### External links

## Biological background

The genetic instructions of every replicating cell in a living organism are contained within its DNA.<sup>[1]</sup> Throughout the cell's lifetime, this information is transcribed and replicated by cellular mechanisms to produce proteins or to provide instructions for daughter cells during cell division, and the possibility exists that the DNA may be altered during these processes.<sup>[1][2]</sup> This is

known as a mutation. At the molecular level, there are regulatory systems that correct most — but not all — of these changes to the DNA before it is replicated.<sup>[2][3]</sup>

One of the possible mutations that occurs is the replacement of a single nucleotide, known as a point mutation. If a point mutation occurs within an expressed region of a gene, an exon, then this will change the codon specifying a particular amino acid in the protein produced by that gene.<sup>[2]</sup> Despite the redundancy in the genetic code, there is a possibility that this mutation will then change the amino acid that is produced during translation, and as a consequence the structure of the protein will be changed.

The functionality of a protein is highly dependent on its structure.<sup>[4]</sup> Changing a single amino acid in a protein may reduce its ability to carry out this function, or the mutation may even change the function that the protein carries out.<sup>[2]</sup> Changes like these may severely impact a crucial function in a cell, potentially causing the cell — and in extreme cases, the organism — to die.<sup>[5]</sup> Conversely, the change may allow the cell to continue functioning albeit differently, and the mutation can be passed on to the organism's offspring. If this change does not result in any significant physical disadvantage to the offspring, the possibility exists that this mutation will persist within the population. The possibility also exists that the change in function becomes advantageous. In either case, while being subjected to the processes of natural selection, the point mutation has been accepted into the genetic pool.

The 20 amino acids translated by the genetic code vary greatly by the physical and chemical properties of their side chains.<sup>[4]</sup> However, these amino acids can be categorised into groups with similar physicochemical properties.<sup>[4]</sup> Substituting an amino acid with another from the same category is more likely to have a smaller impact on the structure and function of a protein than replacement with an amino acid from a different category. Consequently, acceptance of point mutations depends heavily on the amino acid being replaced in the mutation, and the replacement amino acid. The PAM matrices are a mathematical tool that account for these varying rates of acceptance when evaluating the similarity of proteins during alignment.

## Terminology

---

The term *accepted point mutation* was initially used to describe the mutation phenomenon. However, the acronym PAM was preferred over APM due to readability, and so the term *point accepted mutation* is used more regularly.<sup>[6]</sup> Because the value  $n$  in the  $PAM_n$  matrix represents the number of mutations per 100 amino acids, which can be likened to a percentage of mutations, the term *percentage accepted mutation* is sometimes used.

It is important to distinguish between point accepted mutations (PAMs), point accepted mutation matrices (PAM matrices) and the  $PAM_n$  matrix. The term 'point accepted mutation' refers to the mutation event itself. However, 'PAM matrix' refers to one of a family of matrices which contain scores representing the likelihood of two amino acids being aligned due to a series of mutation events, rather than due to random chance. The ' $PAM_n$  matrix' is the PAM matrix corresponding to a time frame long enough for  $n$  mutation events to occur per 100 amino acids.

## Construction of PAM matrices

---

PAM matrices were introduced by Margaret Dayhoff in 1978.<sup>[7]</sup> The calculation of these matrices were based on 1572 observed mutations in the phylogenetic trees of 71 families of closely related proteins. The proteins to be studied were selected on the basis of having high similarity with their predecessors. The protein alignments included were required to display at least 85% identity.<sup>[6][8]</sup> As a result, it is reasonable to assume that any aligned mismatches were the result of a single mutation event, rather than several at the same location.

Each PAM matrix has twenty rows and twenty columns — one representing each of the twenty amino acids translated by the genetic code. The value in each cell of a PAM matrix is related to the probability of a row amino acid before the mutation being aligned with a column amino acid afterwards.<sup>[6][7][8]</sup> From this definition, PAM matrices are an example of a substitution matrix.

## Collection of data from phylogenetic trees

For each branch in the phylogenetic trees of the protein families, the number of mismatches that were observed were recorded and a record kept of the two amino acids involved.<sup>[7]</sup> These counts were used as entries below the main diagonal of the matrix  $\mathbf{A}$ . Since the vast majority of protein samples come from organisms that are alive today (extant species), the 'direction' of a mutation cannot be determined. That is, the amino acid present before the mutation cannot be distinguished from the amino acid that replaced it after the mutation. Because of this, the matrix  $\mathbf{A}$  is assumed to be symmetric, and the entries of  $\mathbf{A}$  above the main diagonal are computed on this basis. The entries along the diagonal of  $\mathbf{A}$  do not correspond to mutations and can be left unfilled.

In addition to these counts, data on the mutability and the frequency of the amino acids was obtained.<sup>[6][7]</sup> The mutability of an amino acid is the ratio of the number of mutations it is involved in and the number of times it occurs in an alignment.<sup>[7]</sup> Mutability measures how likely an amino acid is to mutate acceptably. Asparagine, an amino acid with a small polar side chain, was found to be the most mutable of the amino acids.<sup>[7]</sup> Cysteine and tryptophan were found to be the least mutable amino acids.<sup>[7]</sup> The side chains for cysteine and tryptophan have less common structures: cysteine's side chain contains sulfur which participates in disulfide bonds with other cysteine molecules, and tryptophan's side chain is large and aromatic.<sup>[4]</sup> Since there are several small polar amino acids, these extremes suggest that amino acids are more likely to acceptably mutate if their physical and chemical properties are more common among alternative amino acids.<sup>[6][8]</sup>

## Construction of the mutation matrix

For the  $j$ th amino acid, the values  $m(j)$  and  $f(j)$  are its mutability and frequency. The frequencies of the amino acids are normalised so that they sum to 1. If total number of occurrences of the  $j$ th amino acid is  $n(j)$ , and  $N$  is the total number of all amino acids, then

$$f(j) = \frac{n(j)}{N}$$

Based on the definition of mutability as the ratio of mutations to occurrences of an amino acid

$$m(j) = \frac{\sum_{i=1, i \neq j}^{20} A(i, j)}{n(j)}$$

or

$$\frac{1}{Nf(j)} = \frac{1}{n(j)} = \frac{m(j)}{\sum_{i=1, i \neq j}^{20} A(i, j)}$$

The mutation matrix  $\mathbf{M}$  is constructed so that the entry  $M(i, j)$  represents the probability of the  $j$ th amino acid mutating into the  $i$ th amino acid. The non-diagonal entries are computed by the equation<sup>[7]</sup>

$$M(i, j) = \lambda A(i, j) \frac{m(j)}{\sum_{i=1, i \neq j}^{20} A(i, j)} = \frac{\lambda A(i, j)}{Nf(j)}$$

where  $\lambda$  is a constant of proportionality. However, this equation does not compute the diagonal entries. Each column in the matrix  $\mathbf{M}$  lists each of the twenty possible outcomes for an amino acid — it can mutate into one of the 19 other amino acids, or remain unchanged. Since the non-diagonal entries listing the probabilities of each of the 19 mutations are known, and the sum of the probabilities of these twenty outcomes must be 1, this last probability can be calculated by

$$M(j, j) = 1 - \sum_{i=1, i \neq j}^{20} M(i, j)$$

which simplifies to<sup>[7]</sup>

$$M(j, j) = 1 - \lambda m(j)$$

### Calculation of the diagonal entries

$$M(j, j) = 1 - \sum_{i=1, i \neq j}^{20} M(i, j)$$

Substituting in the expression for the non-diagonal entries mutation matrix:

$$M(j, j) = 1 - \sum_{i=1, i \neq j}^{20} \frac{\lambda m(j) A(i, j)}{\sum_{i=1, i \neq j}^{20} A(i, j)}$$

$$M(j, j) = 1 - \frac{\sum_{i=1, i \neq j}^{20} \lambda m(j) A(i, j)}{\sum_{i=1, i \neq j}^{20} A(i, j)}$$

Since the values of  $\lambda$  and  $m(j)$  are constants that don't change with the value of  $i$

$$M(j, j) = 1 - \frac{\lambda m(j) \sum_{i=1, i \neq j}^{20} A(i, j)}{\sum_{i=1, i \neq j}^{20} A(i, j)}$$

And thus cancellation reveals that

$$M(j, j) = 1 - \lambda m(j)$$

A result of particular significance is that for the non-diagonal entries

$$f(j)M(i, j) = \frac{\lambda}{N} A(i, j) = \frac{\lambda}{N} A(j, i) = f(i)M(j, i)$$

Which means that for all entries in the mutation matrix

$$f(j)M(i, j) = f(i)M(j, i)$$

### Choice of the constant of proportionality

The probabilities contained in  $\mathbf{M}$  vary as some unknown function of the amount of time that a protein sequence is allowed to mutate for. Instead of attempting to determine this relationship, the values of  $\mathbf{M}$  are calculated for a short time frame, and the matrices for longer periods of time are calculated by assuming mutations follow a Markov chain model.<sup>[9][10]</sup> The base unit of

time for the PAM matrices is the time required for 1 mutation to occur per 100 amino acids, sometimes called 'a PAM unit' or 'a PAM' of time.<sup>[6]</sup> This is precisely the duration of mutation assumed by the PAM<sub>1</sub> matrix.

The constant  $\lambda$  is used to control the proportion of amino acids that are unchanged. By using only alignments of proteins that had at least 85% similarity, it could be reasonably assumed that the mutations observed were direct, without any intermediate states. This means that scaling down these counts by a common factor would provide an accurate estimate of the mutation counts had the similarity been closer to 100%. It also means that the number of mutations per 100 amino acids, the  $n$  in PAM <sub>$n$</sub>  is equal to the number of mutated amino acids per 100 amino acids.

To find the mutation matrix for the PAM<sub>1</sub> matrix, the requirement that 99% of the amino acids in a sequence are conserved is imposed. The quantity  $n(j)M(j, j)$  is equal to the number of conserved amino acid  $j$  units, and so the total number of conserved amino acids is

$$\sum_{j=1}^{20} n(j)M(j, j) = \sum_{j=1}^{20} n(j) - \lambda \sum_{j=1}^{20} n(j)m(j) = N - N\lambda \sum_{j=1}^{20} f(j)m(j)$$

The value of  $\lambda$  needed to be pick to produce 99% identity after mutation is then given by the equation

$$0.99 = 1 - \lambda \sum_{j=1}^{20} f(j)m(j)$$

This  $\lambda$  value can then be used in the mutation matrix for the PAM<sub>1</sub> matrix.

## Construction of the PAM <sub>$n$</sub> matrices

The Markov chain model of protein mutation relates the mutation matrix for PAM <sub>$n$</sub> ,  $M_n$ , to the mutation matrix for the PAM<sub>1</sub> matrix,  $M_1$  by the simple relationship

$$M_n = M_1^n$$

The PAM <sub>$n$</sub>  matrix is constructed from the ratio of the probability of point accepted mutations replacing the  $j$ th amino acid with the  $i$ th amino acid, to the probability of these amino acids being aligned by chance. The entries of the PAM <sub>$n$</sub>  matrix are given by the equation<sup>[11][12]</sup>

$$\text{PAM}_n(i, j) = \log \frac{f(j)M_n(i, j)}{f(i)f(j)} = \log \frac{f(j)M^n(i, j)}{f(i)f(j)} = \log \frac{M^n(i, j)}{f(i)}$$

Note that in Gusfield's book, the entries  $M(i, j)$  and  $\text{PAM}_n(i, j)$  are related to the probability of the  $i$ th amino acid mutating into the  $j$ th amino acid.<sup>[11]</sup> This is the origin of the different equation for the entries of the PAM matrices.

When using the PAM <sub>$n$</sub>  matrix to score an alignment of two proteins, the following assumption is made:

*If these two proteins are related, the evolutionary interval separating them is the time taken for  $n$  point accepted mutations to occur per 100 amino acids.*

When the alignment of the  $i$ th and  $j$ th amino acids is considered, the score indicates the relative likelihoods of the alignment due to the proteins being related or due to random chance.

- If the proteins are related, a series of point accepted mutations must have occurred to mutate the original amino acid into its replacement. Suppose the  $j$ th amino acid is the original. Based on the abundance of amino acids in

proteins, the probability of the  $j$ th amino acid being the original is  $f(j)$ . Given any particular unit of this amino acid, the probability of being replaced by the  $i$ th amino acid in the assumed time interval is  $M_n(i, j)$ . Thus, the probability of the alignment is  $f(j)M_n(i, j)$ , the numerator within the logarithm.

- If the proteins are not related, the events that the two aligned amino acids are the  $i$ th and  $j$ th amino acids must be independent. The probabilities of these events are  $f(i)$  and  $f(j)$ , which means the probability of the alignment is  $f(i)f(j)$ , the denominator of the logarithm.
- Thus, the logarithm in the equation results in a positive entry if the alignment is more likely due to point accepted mutations, and a negative entry if the alignment is more likely due to chance.

## Properties of the PAM matrices

---

### Symmetry of the PAM matrices

While the mutation probability matrix  $M$  is not symmetric, each of the PAM matrices are.<sup>[6][7]</sup> This somewhat surprising property is a result of the relationship that was noted for the mutation probability matrix:

$$f(j)M(i, j) = f(i)M(j, i)$$

In fact, this relationship holds for all positive integer powers of the matrix  $M$ :

$$f(j)M^n(i, j) = f(i)M^n(j, i)$$

### Generalisation of property to positive integer matrix powers

This generalisation can be proven using mathematical induction.

Suppose that for a matrix  $M$

$$f(j)M(i, j) = f(i)M(j, i)$$

And that for a positive integer  $k$

$$f(j)M^k(i, j) = f(i)M^k(j, i)$$

By expansion of the matrix product  $M^{k+1} = M^k \cdot M$ ,

$$f(j)M^{k+1}(i, j) = f(j) \sum_{n=0}^N M^k(i, n)M(n, j)$$

$$f(j)M^{k+1}(i, j) = \sum_{n=0}^N M^k(i, n)(f(j)M(n, j))$$

Using the property we have assumed of the matrix  $M$

$$f(j)M^{k+1}(i, j) = \sum_{n=0}^N M^k(i, n)(f(n)M(j, n))$$

$$f(j)M^{k+1}(i, j) = \sum_{n=0}^N (f(n)M^k(i, n))M(j, n)$$

And using the property for the matrix  $M^k$

$$f(j)M^{k+1}(i, j) = \sum_{n=0}^N (f(i)M^k(n, i))M(j, n)$$

$$f(j)M^{k+1}(i, j) = f(i) \sum_{n=0}^N M(j, n)M^k(n, i)$$

$$f(j)M^{k+1}(i, j) = f(i)M^{k+1}(j, i)$$

In this case, it is only known at first that the result holds for  $k = 1$ . However, the above argument shows that the property also holds for  $k = 2$ . This new knowledge then shows that the property also holds for  $k$  and this repeats to show that the property holds for all positive integers  $k$ .

As a result, the entries of the PAM<sub>n</sub> matrix are symmetric, since

$$\text{PAM}_n(i, j) = \log \frac{f(j)M^n(i, j)}{f(j)f(i)} = \log \frac{f(i)M^n(j, i)}{f(i)f(j)} = \text{PAM}_n(j, i)$$

## Relating the number of mutated amino acids and the number of mutations

The value  $n$  represents the number of mutations that occur per 100 amino acids, however this value is rarely accessible and often estimated. However, when comparing two proteins it is easy to calculate  $m$  instead, which is the number of mutated amino acids per 100 amino acids. Despite the random nature of mutation, these values can be approximately related by<sup>[13]</sup>

$$\frac{m}{100} = 1 - e^{-\frac{n}{100}}$$

### Derivation of relationship between $m$ and $n$

Mutations in the primary structure of a protein can occur anywhere along the sequence. If it is assumed the distribution of the mutations among amino acid positions is uniform, the problem is analogous to a distribution of "balls into bins", a common problem in combinatorics. In a case where  $K$  balls (i.e. mutations) are distributed amongst  $N$  bins (amino acid positions), the number of bins containing at least one ball,  $M$  has a distribution with a mean given by<sup>[14]</sup>

$$E(M) = N - N(1 - \frac{1}{N})^K$$

$$\frac{E(M)}{N} = 1 - (1 - \frac{1}{N})^K$$

If the rate of mutation is  $n$  mutations per 100 amino acids, then

$$\frac{n}{100} = \frac{K}{N}$$

And if there are  $m$  mutated amino acids per 100 amino acids, then it is approximately equal to

$$\frac{m}{100} = \frac{E(M)}{N}$$

Now  $m$  and  $n$  can be related by

$$\frac{m}{100} = 1 - \left(1 - \frac{1}{N}\right)^{\frac{nN}{100}}$$

For large values of  $N$ , an assumption that can be reasonably made for typical proteins, this expression is approximately equal to

$$\frac{m}{100} = 1 - e^{-\frac{n}{100}}$$

The validity of these estimates can be verified by counting the number of amino acids that remain unchanged under the action of the matrix  $M$ . The total number of unchanged amino acids for the time interval of the PAM<sub>n</sub> matrix is

$$\sum_{j=1}^{20} n(j)M^n(j, j)$$

and so the proportion of unchanged amino acids is

$$\frac{\sum_{j=1}^{20} n(j)M^n(j, j)}{N} = \sum_{j=1}^{20} f(j)M^n(j, j) = 1 - \frac{m}{100}$$

## An example - PAM250

---

A PAM250 is a commonly used scoring matrix for sequence comparison. Only the lower half of the matrix needs to be computed, since by their construction, PAM matrices are required to be symmetric. Each of the 20 amino acid are shown down the top and side of the matrix, with 3 additional ambiguous amino acids. The amino acids are most commonly shown listed alphabetically, or listed in groups. These groups are the characteristics shared among the amino acids.<sup>[7]</sup>

## Uses in bioinformatics

---

### Determining the time of divergence in phylogenetic trees

The molecular clock hypothesis predicts that the rate of amino acid substitution in a particular protein will be approximately constant over time, though this rate may vary between protein families.<sup>[13]</sup> This suggests that the number of mutations per amino acid in a protein increases approximately linearly with time.



Determining the time at which two proteins diverged is an important task in phylogenetics. Fossil records are often used to establish the position of events on the timeline of the Earth's evolutionary history, but the application of this source is limited. However, if the rate at which the molecular clock of protein family ticks — that is, the rate at which the number of mutations per amino acid increases — is known, then knowing this number of mutations would allow the date of divergence to be found.

Suppose the date of divergence for two related proteins, taken from organisms living today, is sought. The two proteins have both been accumulating accepted mutations since the date of divergence, and so the total number of mutations per amino acid separating them is approximately twice that which separates them from their common ancestor. If a range of PAM matrices are used to align two proteins that are known to be related, then the value of  $n$  in the PAM<sub>n</sub> matrix which results in the best score is most likely to correspond to the mutations per amino acid separating the two proteins. Halving this value and dividing by the rate at which accepted mutations accumulate in the protein family provides an estimate of the time of divergence of these two proteins from their common ancestor. That is, the time of divergence in myr is<sup>[13]</sup>

$$T = \frac{K}{2r}$$

Where  $K$  is the number of mutations per amino acid, and  $r$  is the rate of accepted mutation accumulation in mutations per amino acid site per million years.

## Use in BLAST

PAM matrices are also used as a scoring matrix when comparing DNA sequences or protein sequences to judge the quality of the alignment. This form of scoring system is utilized by a wide range of alignment software including BLAST.<sup>[15]</sup>

## Comparing PAM and BLOSUM

In addition to PAM matrices, another more recently developed scoring matrix can be used. This is known as a BLOSUM. The two result in the same scoring outcome, but use differing methodologies. BLOSUM directly look at mutations in motifs of related sequences while PAM's extrapolate evolutionary information based on closely related sequences.<sup>[16]</sup>

Since both PAM and BLOSUM are different methods for showing the same scoring information, the two can be compared but due to the very different method of obtaining this score, a PAM100 does not equal a BLOSUM100.<sup>[17]</sup>

## See also

- Point mutation
- Sequence alignment
- Margaret Dayhoff
- Molecular clock
- BLOSUM
- BLAST

## References

1. Campbell NA, Reece JB, Meyers N; Urry LA; Cain ML; Wasserman SA; Minorsky PV; Jackson RB (2009). "The

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	J	Z	X
A	2	-2	0	0	-2	0	0	1	-1	-2	-1	-1	-3	1	1	1	1	4	1	0	0	-1	0	-1
R	-2	6	0	-1	-4	1	-1	-3	-2	-2	-3	3	0	-4	0	0	-1	-2	-4	-2	-1	-3	0	-1
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-3	0	1	0	-4	-2	-2	-3	1	-1	-1
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	-3	-3	-3
C	-2	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	-5	-1	-1
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	-2	3	-3
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	3	-1	1	-1
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	0	1	0	-7	-5	-1	0	-4	0	-5
H	-1	2	2	1	-1	3	1	-2	6	-2	0	-2	-2	0	-1	-1	0	-2	1	-2	2	-1	-2	-1
I	-1	-2	-2	-2	-2	-2	-3	-2	5	2	2	2	1	-2	-1	0	-5	-1	4	-2	3	-2	-5	-1
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	6	3	4	2	-3	-3	-2	-1	2	3	5	-1	-1	-1	-1
K	-1	3	1	0	-5	1	0	-2	0	-2	3	5	0	-5	-1	0	0	-3	-4	-2	1	-3	0	-1
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6	0	-2	-2	-1	-4	-2	-2	3	-2	-1	-1
F	-3	-4	-3	-6	-4	-5	-5	-5	-2	1	2	5	0	9	-5	-3	-3	0	7	1	4	-2	-5	-1
P	1	0	0	-1	-1	0	-1	0	0	-2	-3	-1	-2	-5	6	1	0	-6	-5	-1	-1	-2	0	-2
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2	1	-2	-3	-1	0	-2	0	-2
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-1	0	1	3	-5	-3	0	0	-1	-1	-2
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	-6	-5	-3	-6	-2
Y	-3	-4	-2	-4	0	-4	-5	0	-1	-1	-4	-2	7	-5	-1	-1	0	10	-2	3	-1	-4	-1	-1
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	-2	-2	-1	-1	-1	0	-6	-2	4	-2	-2	-2	-1	-1
B	0	-1	2	3	-4	1	3	0	1	-2	3	1	-2	-4	-1	0	0	-5	-1	-2	3	-1	2	-1
J	-1	-3	-3	-3	-5	-2	-3	-4	-2	3	5	-3	3	-2	-2	-1	-1	-1	2	3	5	-2	-5	-5
Z	0	0	1	3	-5	3	3	0	-2	-3	0	-2	-5	0	0	-1	-4	-4	-2	-2	3	-1	-1	-1
X	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

PAM250 matrix with both halves filled in.

- Molecular Basis of Inheritance". *Biology: Australian Version* (8th ed.). Pearson Education Australia. pp. 307–325. ISBN 9781442502215.
2. Campbell NA, Reece JB, Meyers N; Urry LA; Cain ML; Wasserman SA; Minorsky PV; Jackson RB (2009). "From Gene to Protein". *Biology: Australian Version* (8th ed.). Pearson Education Australia. pp. 327–350. ISBN 9781442502215.
  3. Pal JK, Ghaskadbi SS (2009). "DNA Damage, Repair and Recombination". *Fundamentals of Molecular Biology* (1st ed.). Oxford University Press. pp. 187–203. ISBN 9780195697810.
  4. Campbell NA, Reece JB, Meyers N; Urry LA; Cain ML; Wasserman SA; Minorsky PV; Jackson RB (2009). "The Structure and Function of Large Biological Molecules". *Biology: Australian Version* (8th ed.). Pearson Education Australia. pp. 68–89. ISBN 9781442502215.
  5. Lobo, Ingrid (2008). "Mendelian Ratios and Lethal Genes" (<http://www.nature.com/scitable/topicpage/mendelian-ratios-and-lethal-genes-557>). *Nature*. Retrieved 19 October 2013.
  6. Pevsner J (2009). "Pairwise Sequence Alignment". *Bioinformatics and Functional Genomics* ([https://archive.org/details/bioinformaticsfu00pevs\\_0](https://archive.org/details/bioinformaticsfu00pevs_0)) (2nd ed.). Wiley-Blackwell. pp. 58–68. ISBN 978-0-470-08585-1.
  7. Dayhoff, M.O., Schwartz, R. and Orcutt, B.C. (1978). "A model of Evolutionary Change in Proteins". *Atlas of protein sequence and structure* (volume 5, supplement 3 ed.). Nat. Biomed. Res. Found. pp. 345–358. ISBN 978-0-912466-07-1.
  8. Wing-Kin Sung (2010). *Algorithms in Bioinformatics: A Practical Introduction*. CRC Press. pp. 51–52. ISBN 978-1-4200-7033-0.
  9. Kosiol C, Goldman N (2005). "Different versions of the Dayhoff rate matrix" (<http://mbe.oxfordjournals.org/cgi/content/full/22/2/193>). *Molecular Biology and Evolution*. **22** (2): 193–9. doi:10.1093/molbev/msi005 (<https://doi.org/10.1093%2Fmolbev%2Fmsi005>). PMID 15483331 (<https://www.ncbi.nlm.nih.gov/pubmed/15483331>).
  10. Liò P, Goldman N (1998). "Models of molecular evolution and phylogeny" (<http://pixfunlobd.59.to/content/8/12/1233.full>). *Genome Research*. **8** (12): 1233–44. doi:10.1101/gr.8.12.1233 (<https://doi.org/10.1101%2Fgr.8.12.1233>). PMID 9872979 (<https://www.ncbi.nlm.nih.gov/pubmed/9872979>).
  11. Gusfield, Dan (1997). *Algorithms on String, Trees, and Sequences -Computer Science and Computational Biology*. Cambridge University Press. pp. 383–384. ISBN 978-0521585194.
  12. Boeckenhauer, Hans-Joachim; Dirk Bongartz (2010). *Algorithmic Aspects of Bioinformatics*. Springer. pp. 94–96. ISBN 978-3642091001.
  13. Pevsner J (2009). "Molecular Phylogeny and Evolution". *Bioinformatics and Functional Genomics* ([https://archive.org/details/bioinformaticsfu00pevs\\_0](https://archive.org/details/bioinformaticsfu00pevs_0)) (2nd ed.). Wiley-Blackwell. pp. 221–227. ISBN 978-0-470-08585-1.
  14. Motwani, Rajeev; Prabhakar Raghavan (1995). *Randomized Algorithms* (<https://books.google.com/books?id=QKVY4mDivBEC&pg=PA94#v=onepage&q&f=false>). Cambridge University Press. p. 94. ISBN 978-0521474658.
  15. "The Statistics of Sequence Similarity Scores" (<https://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html#head9>). *National Centre for Biotechnology Information*. Retrieved 20 October 2013.
  16. Henikoff S, Henikoff JG (1992). "Amino acid substitution matrices from protein blocks" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC50453>). *Proceedings of the National Academy of Sciences of the United States of America*. **89** (22): 10915–10919. Bibcode:1992PNAS...8910915H (<https://ui.adsabs.harvard.edu/abs/1992PNAS...8910915H>). doi:10.1073/pnas.89.22.10915 (<https://doi.org/10.1073%2Fpnas.89.22.10915>). PMC 50453 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC50453>). PMID 1438297 (<https://www.ncbi.nlm.nih.gov/pubmed/1438297>).
  17. Saud, Omama (2009). "PAM and BLOSUM Substitution Matrices" (<https://web.archive.org/web/20130309015755/http://birec.org/sandbox/omamasaudtutorial>). *Birec*. Archived from the original (<http://www.birec.org/sandbox/omamasaudtutorial>) on 9 March 2013. Retrieved 20 October 2013.

## External links

---

- <http://www.inf.ethz.ch/personal/gonnet/DarwinManual/node148.html>
- <http://www.bioinformatics.nl/tools/pam.html> For quickly calculating a PAM matrix.
- <http://web.expasy.org/docs/relnotes/relstat.html> The most recent statistics from the Swiss-Prot protein knowledgebase. Section 6.1 contains the most up-to-date amino acid frequencies

---

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Point\\_accepted\\_mutation&oldid=918315932](https://en.wikipedia.org/w/index.php?title=Point_accepted_mutation&oldid=918315932)"

---

**This page was last edited on 28 September 2019, at 02:22 (UTC).**

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.