

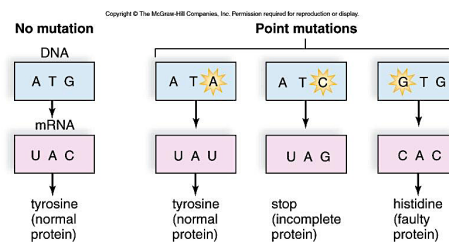
Bioinformatics II: PAM matrices

Dr Manaf A Guma
University of Anbar- college of applied science-Heet.
Department of chemistry

1

Before we start, what is the difference between point mutation and frameshift mutation?

- Point mutation is an alteration of **a single nucleotide** in a gene whereas frameshift mutation involves **one or more nucleotide changes** of a particular gene.
- Point mutations are mainly nucleotide substitutions, which lead to silent, missense or nonsense mutations. Frameshift mutations occur by insertion or deletion of nucleotides.



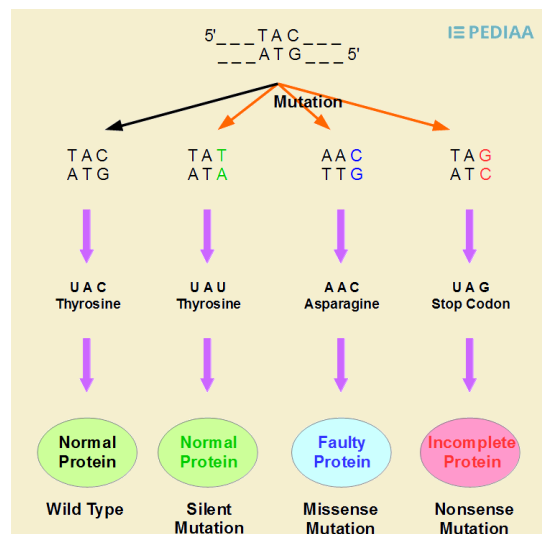
2

Define?

- Nonsense Mutations: the alteration of a nucleotide in a particular codon may introduce a stop codon to the gene. This stops the translation of the protein at halfway of the complete protein.
- Silent mutations, a single base pair has changed in a particular **codon**, the same **amino acid** is coded by the altered codon as well.
- Missense mutations, once the alteration occurs in a particular codon by a nucleotide substitution, the codon is altered in such a way to code a different amino acid.

3

Point accepted mutation



4

PAM matrices: Background and concepts

- How the PAM work?
 1. Only mutations are allowed.
 2. Sites evolve independently.
 3. Evolution at each site occurs according to a Markov equation.
- It follows Markov process.? How?

5

5

What is Markov concept?

- Markov process:
- (The substitution is independent from their past history!).
- Meaning:
 - Next mutation depends only on current state and is independent of previous mutations.
 - It is derived from global alignment. do you remember?

6

What are PAM matrices ?

- **Point accepted mutation matrix** known as a PAM.
- It is also called **P**ercent **A**ccepted **M**utation.
- Dayhoff and colleagues defined the PAM1 matrix as that which produces 1 accepted point mutation per 100 amino acid residues.
- **PAM matrix** is designed to compare two sequences which are a specific number of PAM units apart.
- <https://www.youtube.com/watch?v=F8WdDfpQqCM>
- <https://www.youtube.com/watch?v=UCtP5-KtB94>

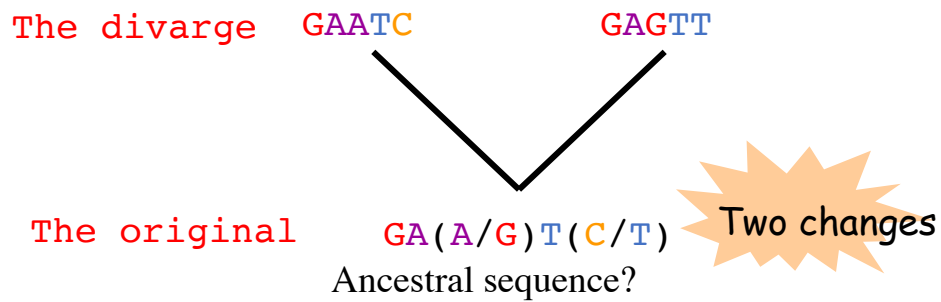
7

What are the different types of PAM?

- By Dayhoff meaning, a PAM0 matrix is the identity matrix, so that no amino acid can change.
- Since the PAM1 matrix was based on closely related protein sequences that share more than 85 % sequence identity, its use is limited for the protein sequences that are less than 85 % identical.
- For this, other types of PAM matrices were derived from PAM1 matrix by multiplying PAM1 by itself.
- PAM100 matrix was derived by multiplying PAM1 by itself 100 times.
- Similarly, the PAM250 matrix is used for proteins that share about 20 % sequence identity

8

What are based on? PAM matrices are based on a simple evolutionary model



- Only mutations are allowed
- Sites evolve independently

9

9

Is the replacement of amino acids with other accepted always?

- The replacement of an amino acid by another with similar biochemical properties is sometimes accepted in a protein.
- These replacements are known as conservative substitutions.
- For example, replacement of serine with threonine, glutamic acid with aspartic acid, and isoleucine with valine are some of the most common amino acid substitutions that are readily accepted.

10

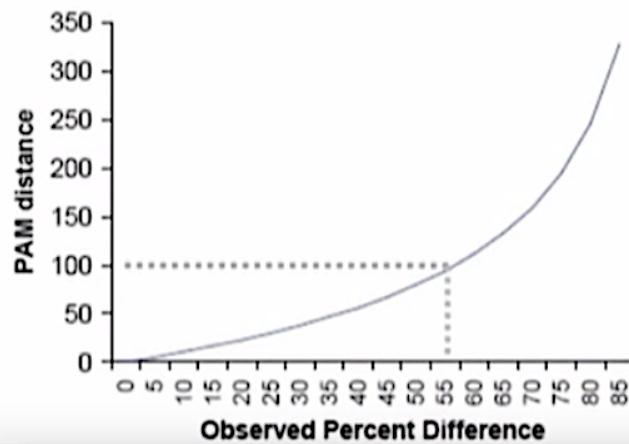
Explain the PAM unit?

- 1PAM – Unit for measuring the similarity of two amino acid sequences.
- We say two sequences are n PAMs apart if every 100 residues contain, on average, n actual changes (including multiple substitutions) between them.
- A 100% PAM will have 100% variation in the sequence because the same site can be changed more than one time.
- A PAM unit is the amount of evolution that will on average change 1% of the amino acids within a protein sequence.

11

What do you get from this graph?

Sequence Difference vs. PAM Distance



12

How to explain the graph?

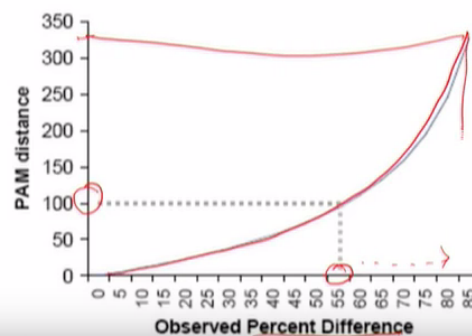
- To understand the graph:
- We have a figure in which the sequence variance against PAM distance to compare.
- If we do not know the idea, then we can think that a PAM distance of 100 will mean a 100 % change on the seq. But that is not the case.
- Because on a site, mutation may further mutated and thereby get and accumulate multiple mutations.

13

If you go further !!!

- Then therefore the experimental data show that if you have a 100 PAM distance, only 55-60 % sites of the protein are actually mutated.
- So, if you go further, the PAM distance may increase to over 300 for an 85% variations

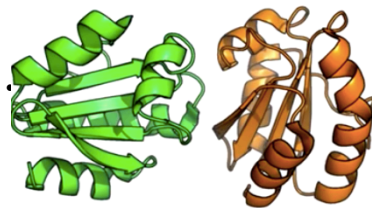
Sequence Difference vs. PAM Distance



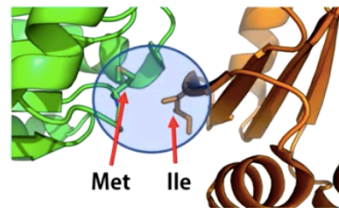
14

How can different amino acids change the scoring?

- different amino acids are partially match in chemicals properties.
- So, if you assume 1 for match and 0 for mis-match, then that is not enough.
- Because it depends on the side chain of the amino acid which may not



- **Met** → **Leu** substitution does not alter the hydrophobic interaction but
- **Met** → **Arg** substitution alters the hydrophobic interaction



Hydrophobic interactions between this Met and Ile stabilizes this binding