ISSN No:-2456-2165

Phylogenetic Analysis of Phenylalanine Hydroxylase Enzyme and Its Future Aspect in Treatment of Phenylalanine Hydroxylase Enzyme Deficiency (Phenylketonuria)

Samran Shahid
Department of Science
India International School In Japan , Japan

Abstract:- PAH enzyme is one of the most vital enzymes in protein metabolism of the body. The enzyme has been found in various organisms and thus proves it has evolved along with speciation.PAH catalyses hydroxylation of the aromatic side of the phenylalanine to generate Tyrosine (4-hydroxyphenylalanine), one of the 20 standard amino acids that exist. The buildup of excess phenylalanine in the body due to deficiency of PAH causes a condition called Phenylketonuria which causes significant nerve damage. The condition Phenylketonuria is caused due to genetic mutation in PAH gene (Cr.12)in an individual which can cause PAH enzyme deficiency. The purpose of this analysis was to use the existing Bioinformatics databases to draw relevant similarities of PAH of *Homo sapiens* and other organism using BLAST , MSA(Multiple Sequence Alignment) and phylogenetic relation while proposing the use of gene therapy using the data derived to cure Phenylketonuria.

Keywords:- *PAH enzyme , PKU disease , Gene therapy for PKU.*

I. INTRODUCTION

Mammalian phenylalanine hydroxylase (PAH) catalyzes the rate-limiting step in the phenylalanine catabolism, consuming about 75% of the phenylalanine input from the diet and protein catabolism under physiological conditions [1] .It converts phenylalanine to 4-Hydroxyphenylalanine (Tyrosine). PAH is one of three members of the biopterin-dependent aromatic amino acid hydroxylases, a class of monooxygenase that uses tetrahydrobiopterin (BH_4, a pteridine cofactor) and a non-heme iron for catalysis. During the reaction, molecular oxygen is heterolytically cleaved with sequential incorporation of one oxygen atom into BH_4 and phenylalanine substrate [2]. PAH is extensively expressed in the hepatocytes of the liver and the tubules of the kidney mainly due to its catabolic .The occurrence in rest of body remains quite low. The low occurrence of PAH enzyme can lead to high Phenylalanine level in the blood which is toxic to the Central Nervous System (Phenylketonuria).

II. STRUCTURE

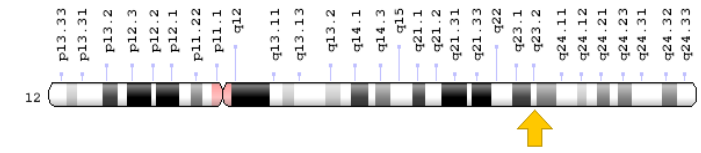
Most of the enzymes that exist in human body are in form of oligomers or in form form of polymers and a significant subset of these can disassociate or associate in response to an effector lignand bringing a change in subunit which accompanies change in function/activity of the enzyme. Therefore this creates a regulatory mechanism in various enzymes in the body[3].

III. MUTATION

Diminished activity of the PAH enzyme which catalyzes iron , dioxygen ,BH4 dependent oxidation of phenylalanine to tyrosine is the most common anomaly in the metabolic system of an individual which in severest form is known as Phenylketonuria (PKU) disease.

PAH Gene : The *PAH* gene provides instructions for making an enzyme called phenylalanine hydroxylase. This enzyme is responsible for the first step in processing phenylalanine, which is a building block of proteins (an amino acid) obtained through the diet.

Chromosomal location of PAH gene : Cytogenetic Location: 12q23.2, which is the long (q) arm of chromosome 12 at position 23.2 Molecular Location: base pairs 102,836,889 to 102,958,441 on chromosome 12 (Homo sapiens Updated Annotation Release 109.20200522, GRCh38.p13) (NCBI)



Credit: Genome Decoration Page/NCBI

Fig 1:- showing location of PAH gene on the 12th chromosome

ISSN No:-2456-2165

IV. METHODS AND MATERIALS

The sequence of the Phenylalanine hydroxylase was acquired from NCBI , alongside sequences of Phenylalanine hydroxylase from several other species.

➢ *Sample used from Homo sapiens :*

Protein sequence name : Phenylalanine-4-hydroxylase

Database:nr (All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects)

Ascession No. : AAA60082 ; **No of Amino Acids**: 452

Amino Acid sequence : 1 MSTAVLENPG LGRKLSDFGQ ETSYIEDNCN QNGAISLIFS LKEEVGALAK VLRLFEENDV NLTHIESRPS RLKKDEYEFF THLDKRSLPA LTNIIKILRH DIGATVHELS RDKKKDTVPWFPRTIQELDRFANQILSYGA ELDADHPGFK DPVYRARRKQ FADIAYNYRH GQPIPRVEYM EEEKKTWGTV FKTLKSLYKT HACYEYNHIF PLLEKYCGFH EDNIPQLEDV SQFLQTCTGFRLRPVAGLLS SRDFLGGLAF RVFHCTQYIR HGSKPMYTPE PDICHELLGH VPLFSDRSFA QFSQEIGLAS LGAPDEYIEK LATIYWFTVE FGLCKQGDSI KAYGAGLLSS FGELQYCLSEKPKLLPLELE KTAIQNYTVT EFQPLYYVAE SFNDAKEKVR NFAATIPRPF SVRYDPYTQR IEVLDNTQQL KILADSINSE IGILCSALQK IK

Program of NCBI (nr Database) and found the following results.

Name of Protein Sequence	No.of Amino Acids	Organism	Ascession Number	Database
phenylalanine-4-hydroxylase	452	Pan troglodytes	XP_001156919.1	nr
phenylalanine-4-hydroxylase	452	Gorilla gorilla	XP_004053827.1	nr
phenylalanine-4-hydroxylase	452	Nomascus leucogenys	XP_003269981.1	nr
phenylalanine-4-hydroxylase	452	Trachypithecus francoisi	XP_033076760.1	nr
phenylalanine-4-hydroxylase	452	Theropithecus gelada	XP_025258879.1	nr
phenylalanine-4-hydroxylase isoform X1	452	Pongo abelii	XP_024111837.1	nr
phenylalanine-4-hydroxylase	452	Rhinopithecus roxellana	XP_010355577.1	nr
phenylalanine-4-hydroxylase	452	Macaca mulatta	XP_001094859.1	nr

Table 1

V. TOOLS

➢ *Pairwise Alignment:*

A Sequence Alignment is a way of arranging the sequences of DNA to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences.

Sequences of PAH protein of the organisms under study were aligned with Human PAH enzyme in order to find the percent identity which describes the similarity of query sequence with the target sequence.

❖ *Results of Pairwise Allignments:*

Homo Sapiens vs Pan troglodytes(99.78%)
Homo sapiens vs Gorilla gorilla(99.56%)
Homo Sapiens vs Nomascus leucogenys(98.45%)
Homo Sapiens vs Trachypithecus francoisi(98.23%)
Homo Sapiens vs Theropithecus gelada(98.01%)
Homo Sapiens vs Pongo abelii(98.01%)
Homo Sapiens vs Rhinopithecus roxellana(98.23%)
Homo Sapiens vs Macaca mulatta(98.01%)

Thus from the pairwise alignments we can conclude that Pan trogolodytes are a major candidate for correction of PKU disease in human.

Thus The DOTTPLOT analysis of Homo sapiens vs Pan trogolodytes PAH enzyme further strengthens our argument.

➢ *EMBOSS dottup:*

Dottup is a commonly used tool for drawing dotplots between two sequences. It looks for places where words of a specified length have an exact match in both sequences and draws a diagonal line over the position of these words. Shorter sequences are more sensitive to shorter regions of similarity but also display random points of similarity and run slower as compared to longer sequences which run faster, display minimum random points of similarity but are less sensitive.

ISSN No:-2456-2165

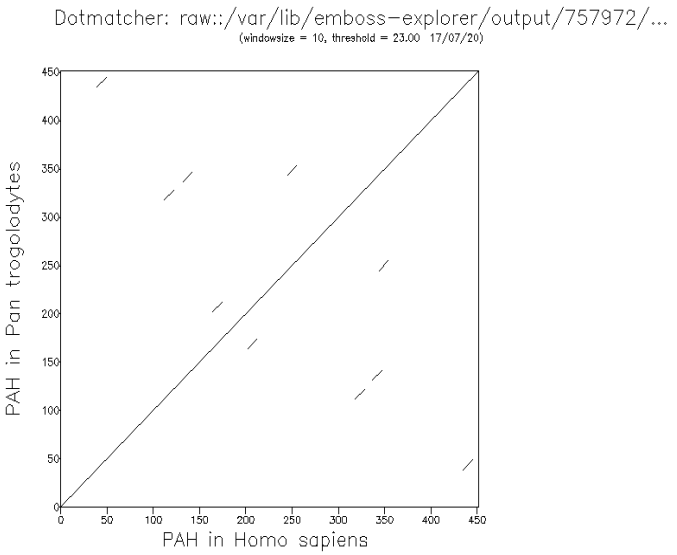


Fig 2:- DOTTPLOT of PAH in Homo sapiens vs PAH in Pan trogolodytes

➢ *Multiple sequence Alignment:*

A multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. In many cases, the input set of query sequences are assumed to have an evolutionary relationship by which they share a linkage and are descended from a common ancestor. From the resulting MSA, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences; shared evolutionary origins.

Here we used Clustal Omega for Multiple Sequence Allignment:

```
CLUSTAL O(1.2.4) multiple sequence alignment

NP_000268.1       MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDV   60
XP_001156919.1    MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDI   60
                  ************************************************************;

NP_000268.1       NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKKKDTVPW   120
XP_001156919.1    NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKKKDTVPW   120
                  ************************************************************

NP_000268.1       FPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIAYNYRHGQPIPRVEYM   180
XP_001156919.1    FPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIAYNYRHGQPIPRVEYM   180
                  ************************************************************

NP_000268.1       EEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCTGF   240
XP_001156919.1    EEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCTGF   240
                  ************************************************************

NP_000268.1       RLRPVAGLLSSRDFLGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFA   300
XP_001156919.1    RLRPVAGLLSSRDFLGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFA   300
                  ************************************************************

NP_000268.1       QFSQEIGLASLGAPDEYIEKLATIYWFTVEFGLCKQGDSIKAYGAGLLSSFGELQYCLSE   360
XP_001156919.1    QFSQEIGLASLGAPDEYIEKLATIYWFTVEFGLCKQGDSIKAYGAGLLSSFGELQYCLSE   360
                  ************************************************************

NP_000268.1       KPKLLPLELEKTAIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQR   420
XP_001156919.1    KPKLLPLELEKTAIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQR   420
                  ************************************************************

NP_000268.1       IEVLDNTQQLKILADSINSEIGILCSALQKIK   452
XP_001156919.1    IEVLDNTQQLKILADSINSEIGILCSALQKIK   452
                  ********************************
```

Fig 3:- Clustal Omega MSA program please note **":"** is used where there is no match in amino acid sequences and **"*"** is used where the amino acids match

(NP_00268.1 is PAH protein of Homo sapiens whereas XP_00115919.1 is PAH protein of Pan troglodytes)

➢ *Cladogram:*

It is a diagram that depicts evolutionary relationships among groups. It is based on PHYLOGENY, which is the study of evolutionary relationships. We used Clustal Omega for cladogram analysis[4].

Cladogram of PAH of Homo sapiens and PAH of Pan troglydytes:

Fig 4:- Cladogram

VI. CORRECTION OF PHENYLALANINE HYDROXYLASE DEFICIENCY

Gene therapy is designed to introduce genetic material into cells to compensate for abnormal genes or to make a beneficial protein. If a mutated gene causes a necessary protein to be faulty or missing, gene therapy may be able to introduce a normal copy of the gene to restore the function of the protein.

Now we compared the DNA sequences of PAH gene found in Homo sapiens (NG_008690.2) and Pan troglodytes (NC_036891.1) using Clustal Omega Multiple Sequence Alignment tool for nucleotides (DNA).

Results: There were 80356 Nucleotides in Homo sapiens PAH gene whereas PAH gene of Pan troglodytes contained around 78784 Nucleotides , of the 78784 nucleotides 77020 nucleotides perfectly aligned with the sequence of Homo sapiens PAH DNA. Thus PAH gene of Homo sapiens and Pan troglodytes is 95.81% similar in nature, therefore PAH gene extracted from Pan troglodytes can be successful for use in gene therapy of hepatocyte of Homo sapiens to correct the PKU disorder.

VII. CONCLUSION

To understand the structural and evolutionary similarities of human PAH enzyme with other species, we compared the same protein in different organism's Pairwise alignment using BLAST showed that human PAH protein was most closely related with the same protein present in Chimpanzee (Pan troglodytes). The reason for comparing human PAH protein with the same protein of other species was to find similarity between them. As the results indicate, PAH of Chimpanzee (Pan troglodytes) is most similar to human's. For this reason, Phenylketonuria disease may be cured by replacing or inserting the gene encoding PAH from a closely related specie, i.e. Chimpanzee (Pan troglodytes) into humans in order to overcome the PAH deficiency.

The hypothesis that this way of curing PKU would be effective is based on results on Multiple Sequence Alignment of human PAH gene and protein with the PAH gene and protein of Chimpanzees (Pan troglodytes) showing 99.78% and 95.81% alignment. Thus by means of viral transfer of DNA encoding into defective cell nucleus

(in terms of PAH production in humans) using Adenoviruses or etc can help eradicate PKU disease.

REFERENCES

[1]. Flydal, M.I. and Martinez, A. (2013), Phenylalanine hydroxylase: Function, structure, and regulation. IUBMB Life, 65: 341-349. doi:10.1002/iub.1150
[2]. Fitzpatrick PF (1999). "Tetrahydropterin-dependent amino acid hydroxylases". *Annual Review of Biochemistry*.
[3]. Traut TW. Dissociation of enzyme oligomers: a mechanism for allosteric regulation. *Crit Rev Biochem Mol Biol*. 1994;29(2):125-163. doi:10.3109/10409239409086799
[4]. Georgios A Pavlopoulos, T. G.-S. (2010). A reference guide for tree analysis and visualization