International Journal of Advanced Research in Engineering and Applied Sciences

ISSN: 2278-6252

INSILICO EXPLORATION OF PYC GENE FROM CORYNEBACTERIUM GLUTAMICUM ATCC 13032

Krishnarjuna Reddy. C V*

Rajeswara Reddy. E*

Muralidhara Rao. D**

Swamy. AVN***

Abstract: *Pyruvate carboxylase (EC: 6.4.1.1) is a crucial enzyme which involves in Glutamic acid production. The exploration of pyc gene with genomic and proteomics studies reveals that, the pyc gens has pyramidine rich nucleotide sequence with1140 aminoacids and 1043KD (ssDNA) of molecular weight. The primary structure reveals that it is an acidic, stable protein. Secondary structure defines that it mainly contains the random coils, alpha helix and beta turn. The gene contains both left & right primers, a hybridization probe and several restriction sites for restriction sites. The protein has 54 phosphorylation sites (Serine, Threonine & Tyrosine) along with one O-Glycosylation (Threonine at position 117) site.*

Keywords*: pyc gene, Corynebacterium glutamicum, Glutamic acid.*

*Department of Biotechnology, Jawaharlal Nehru Technological University Anantapur College of Engineering Pulivendula, Pulvendula 516390, Andhra Pradesh, INDIA.

**Departments of Biotechnology, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, INDIA.

***Department of Chemical Engineering, JNTUA College of Engineering, Anantapur, Andhra Pradesh. INDIA.

INTRODUCTION

Pyruvate caboxylase (EC 6.4.1.1) is an important member of the biotin-containing enzyme family, which catalyses the ATP-dependent carboxylation of pyruvate to oxaloacetate **[1]**. It catalyses a two-step reaction, involving the ATP- driven carboxylation of a covalently attached biotin moiety and the transfer of the bound carboxyl group to pyruvate, forming oxaloacetate. Pyruvate caboxylase has been conserved throughout evolution as it is found in a wide variety of prokaryotes and eukaryotes including fungi, bacteria, plants, and animals **[2]**. It has an anaplerotic role in the provision of oxaloacetate for both biosynthetic purposes and as a carrier for acetyl units into the citric acid cycle. Pyruvate caboxylase has a vital role in different metabolic pathways like gluconeogenesis, lipogenesis and biosynthesis of neurotransmitters which are often essential for the survival of a cell. Oxaloacetate is one of several important intermediates in Krebs cycle that are withdrawn for use in several biosynthetic pathways **[3 & 4]**. Deficiency of Pyruvate Carboxylase can cause lactic acidosis and malfunction of the citric acid cycle and gluconeogenesis, thereby depriving the body of energy; excess pyruvate is converted into lactic acid instead of oxaloacetate and also leads to hypoglycaemia **[5]**.

METHODOLOGY

Various genomic & proteomic offline and online tools were used to explore the *pyc* gene. The nucleic acid sequence of *pyc* gene of *Corynebacterium glutamicum* ATCC 13032 was retrieved from NCBI (gene ID is gi|62388892:706684-710106). Genomic studies were carried out by Bioedit, ORF finder, Primer 3.0 and Genscan followed by proteomic studies to predict primary structure with Bioedit, Protparam, secondary structure was predicted by GOR4 and SOPMA **[6 & 7]**. Post translational modifications are studied with NetOGlyc, NetNGlyc and NetPhos 2.0 tools.

RESULTS AND DISCUSSION

GENOMICS

Bioedit

DNA molecule: gi|62388892|ref|NC_006958.1|:706478-710311,
Corynebacterium glutamicum ATCC 13032, complete genome
Length = 3834 base pairs
Molecular Weight = 1167311.00 Daltons, single stranded
Molecular Weight = 2334482.00 Daltons, double stranded

G+C content = 55.76%
A+T content = 44.24%

Nucleotide	Number	Mol%
A	838	21.86
C	1077	28.09
G	1061	27.67
T	58	22.38

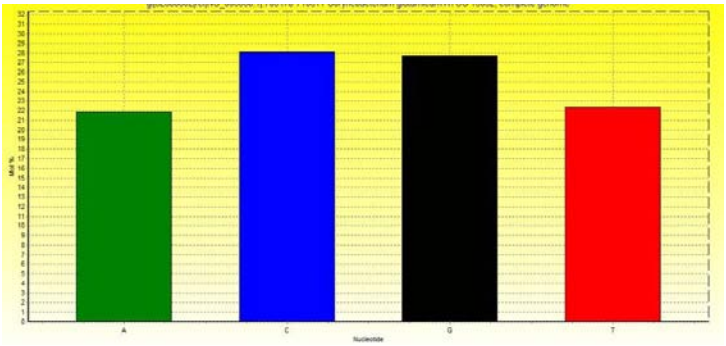

Figure 1 : Nucleotide composition of *pyc* gene

Nucleotide composition of pyc gene from Corynebacterium glutamicum ATCC 13032 has been predicted by Bioedit & the results reveals that the sequence is pyramidine rich with themolecular weight of more than 1167KDa (ssDNA) containing 3834 base pairs. The CG & AT content was found to be **55.76%** and **44.24 %** respectively.

ORF Finder

gi|62388892:706684-710106 Corynebacterium glutamicum ATCC 13032, complete genome

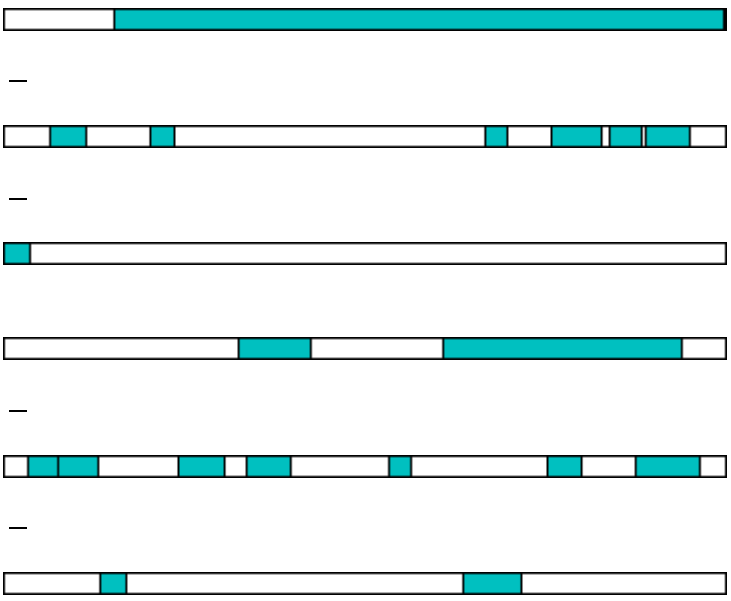

Figure 2: Presence of the possible protein coding region sequences in *pyc* gene

Frame	from	to	Length
+1	532..	3s422	2892
-1	2086..	3222	1137
-1	1117..	1458	342
-2	3003..	3305	303
-3	2183..	2455	273
+2	2600..	2842	243
-2	831..	1055	225
+2	3047..	3256	210
-2	1155..	1361	207
-2	267..	452	186
+2	224..	394	171
-2	2586..	2741	156
+2	2882..	3034	153
-2	123..	266	144
+3	3..	134	132
-3	464..	583	120
-2	1830..	1940	111
+2	704..	811	108
+2	2288..	2392	105

Open Reading Frame Finder predicts the presence of the possible protein coding region sequence.

1. It was been identified that the *pyc* codes for 19 encoded proteins present in both the + and – strands.
2. The largest protein-coding region (exon) was identified in the 1st frame of the direct strand from the position 532 to 3422 of length 2892 bases.

Primer 3.0

Primer picking results for gi|62388892:706684-710106 *Corynebacterium glutamicum* ATCC 13032, complete genome

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	2759	20	59.99	55.00	5.00	2.00	AGCTTGGTAACCCTCCAGGT
RIGHT PRIMER	2952	20	59.98	55.00	6.00	1.00	CTCGAGGAACTCTTCGGTTG
HYB OLIGO	2894	20	60.28	50.00	4.00	0.00	AACGTCGCAATAGCCTCAAC

Sequence size: 3423

Included region size: 3423

Product size: 194, pair any compl: 5.00, pair 3' compl: 1.00

PRIMER3 predicts the presence of the left and the right primers of length 20 residues in the oligonucleotide query.

1. It gives the presence of the left primer starting from 2759th position with GC content of 55% and the right primer at starting from 2952th position with GC content of 55%.
2. It also predicts the presence of the hybridization probe starting from 2894th position with GC content of 50%

Genscan

Sequence /tmp/10_10_12-05:35:51.fasta: 3468 bp : 56.76% C+G : Isochore 3 (51 - 57 C+G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P...	. Tscr..
1.01	Term	+	77	3468	3392	0	2	53	41	4471	0.951	425.37

GenScan predicts the presence of the exon in the direct strand from the position 77 to 3468.

1. The probability of the predicted output for the exon is **0.951.**
2. The predicted exon is categorized as a strong exon as the exon score is greater than 100 i.e., **425.37.**

PROTEOMICS

Primary Structure Prediction

Bioedit

Protein: peptide sequence

Length = 1145 amino acids

Molecular Weight = 123614.87 Daltons

Amino Acid	Number	Mol%	Amino Acid	Number	Mol%
Ala A	148	12.93	Met M	17	1.48
Cys C	7	0.61	Asn N	29	2.53
Asp D	79	6.90	Pro P	63	5.50
Glu E	83	7.25	Gln Q	29	2.53
Phe F	40	3.49	Arg R	82	7.16
Gly G	88	7.69	Ser S	55	4.80
His H	25	2.18	Thr T	73	6.38
Ile I	53	4.63	Val V	96	8.38
Lys K	48	4.19	Trp W	4	0.35
Leu L	101	8.82	Tyr Y	24	2.10

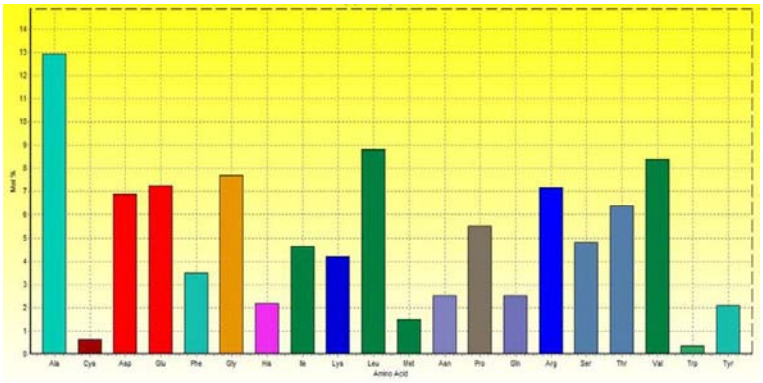

Figure 3 : Aminoacid composition of *pyc* gene

Amino acid composition of *pyc* protein has been predicted by Bioedit & the results reveal that the protein is rich in Ala, Leu and Val. The protein has a molecular weight protein of 123.614KDa containing 1145 amino acids.

PROTPARAM

Number of amino acids: 1145

Theoretical pI: 5.39

Amino acid composition

Ala (A)	148	13.0%	Lys (K)	48	4.2%
Arg (R)	82	7.2%	Met (M)	18	1.6%
Asn (N)	29	2.5%	Phe (F)	40	3.5%
Asp (D)	79	6.9%	Pro (P)	63	5.5%
Cys (C)	7	0.6%	Ser (S)	55	4.8%
Gln (Q)	29	2.5%	Thr (T)	72	6.3%
Glu (E)	83	7.3%	Trp (W)	4	0.4%
Gly (G)	88	7.7%	Tyr (Y)	24	2.1%
His (H)	25	2.2%	Val (V)	95	8.3%
Ile (I)	51	4.5%	Pyl (O)	0	0.0%
Leu (L)	100	8.8%	Sec (U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 162

Total number of positively charged residues (Arg + Lys): 130

Atomic composition:

Carbon	C	5422
Hydrogen	H	8672
Nitrogen	N	1546
Oxygen	O	1674
Sulfur	S	25

Formula: $C_{5422}H_{8672}N_{1546}O_{1674}S_{25}$

Total number of atoms: 17339

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 58135

Abs 0.1% (=1 g/l) 0.472, assuming all pairs of Cys residues form cystines

Ext. coefficient 57760

Abs 0.1% (=1 g/l) 0.469, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 34.94

This classifies the protein as stable.

Aliphatic index: 88.81

Grand average of hydropathicity (GRAVY): -0.204

The protein was acidic in nature, as it has more negatively charged residues than the positively charged ones. The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo) and >10 hours (Escherichia coli, in vivo). The protein is classified as stable, as the instability index is computed to be 34.94. The aliphatic index predicts the volume occupied by the aliphatic residue side chains and the index is 88.81. The protein is highly hydrophilic as the Grand Average of Hydropathicity value is –0.204, which is very much lesser than 0.05.

SECONDARY STRUCTURE PREDICTION

GOR 4

Sequence length : 1130

Alpha helix (Hh) : 502 is 44.42%

3_{10} helix (Gg) : 0 is 0.00%

Pi helix (Ii) : 0 is 0.00%

Beta bridge (Bb) : 0 is 0.00%

Extended strand (Ee) : 143 is 12.65%

Beta turn (Tt) : 0 is 0.00%

Bend region (Ss) : 0 is 0.00%

Random coil (Cc) : 485 is 42.92%

Ambigous states (?) : 0 is 0.00%

Other states : 0 is 0.00%

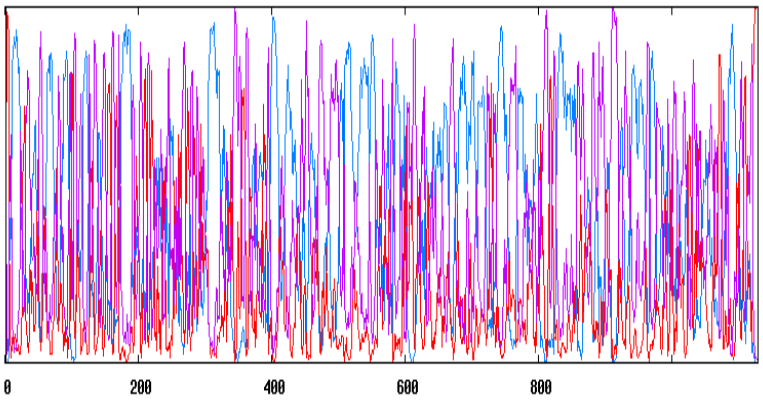

Figure 4: Secondary structure of peptide encode by *pyc* gene

GOR4 (Garnier Method) methods was used to predict the secondary structure of PC. It was found that 44.42% of amino acids fall on Alpha helix region, 12.65% of amino acid was found to be lays in beta sheet remaining 42.92% tends to form random coil.

SOPMA

Sequence length : 1140

Alpha helix (Hh) : 478 is 41.93%

3_{10} helix Gg) : 0 is 0.00%

Pi helix (Ii) : 0 is 0.00%

Beta bridge (Bb) : 0 is 0.00%

Extended strand(Ee) : 187 is 16.40%

Beta turn (Tt) : 113 is 9.91%

Bend region (Ss) : 0 is 0.00%

Random coil (Cc) : 362 is 31.75%

Ambigous states(?) : 0 is 0.00%

Other states : 0 is 0.00%

SOPMA (Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments) method was used to predict the secondary structure of PC. It was found that 41.93% of amino acids fall on Alpha helix region, 16.40% of amino acid was found to be lays in beta sheet remaining 31.75% tends to form random coil.

POST TRANSLATIONAL MODIFICATIONS

NetOGlyc

Name: Sequence Length: 1140

MSTHTSSTLPAFKKILVANRGEIAVRAFRAALETGAATVAIYPREDRGSFHRSFASEAVRIGTEGSPVKAYLDIDEIIGAAKKVKAD
AIYPGYGFLSENAQLARECAENGITFIGPTPEVLDLTGDKSRAVTAAKKAGLPVLAESTPSKNIDEIVKSAEGQTYPIFVKAVAGG
GGRGMRFVASPDELRKLATEASREAEAAFGDGAVYVERAVINPQHIEVQILGDHTGEVVHLYERDCSLQRRHQKVVEIAPAQH
LDPELRDRICADAVKFCRSIGYQGAGTVEFLVDEKGNHVFIEMNPRIQVEHTVTEEVTEVDLVKAQMRLAAGATLKELGLTQDK
IKTHGAALQCRITTEDPNNGFRPDTGTITAYRSPGGAGVRLDGAAQLGGEITAHFDSMLVKMTCRGSDFETAVARAQRALAEF
TVSGVATNIGFLRALLREEDFTSKRIATGFIADHPHLLQAPPADDEQGRILDYLADVTVNKPHGVRPKDVAAPIDKLPNIKDLPLP
RGSRDRLKQLGPAAFARDLREQDALAVTDTTFRDAHQSLLATRVRSFALKPAAEAVAKLTPELLSVEAWGGATYDVAMRFLFE
DPWDRLDELREAMPNVNIQMLLRGRNTVGYTPYPDSVCRAFVKEAASSGVDIFRIFDALNDVSQMRPAIDAVLETNTAVAEV
AMAYSGDLSDPNEKLYTLDYYLKMAEEIVKSGAHILAIKDMAGLLRPAAVTKLVTALRREFDLPVHVHTHDTAGGQLATYFAAA
QAGADAVDGASAPLSGTTSQPSLSAIVAAFAHTRRDTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQLSN
LRAQATALGLADRFELIEDNYAAVNEMLGRPTKVTPSSKVVGDLALHLVGAGVDPADFAADPQKYDIPDSVIAFLRGELGNPPG
GWPEPLRTRALEGRSEGKAPLTEVPEEEQAHLDADDSKERRNSLNRLLFPKPTEEFLEHRRRFGNTSALDDREFFYGLVEGRETLI
RLPDVRTPLLVRLDAISEPDDKGMRNVVANVNGQIRPMRVRDRSVESVTATAEKADSSNKGHVAAPFAGVVTVTVAEGDEVK
AGDAVAIIEAMKMEATITASVDGKIDRVVVPAATKVEGGDLIVVVS.....
.....T.....
.....
.....
.....
.....
.....

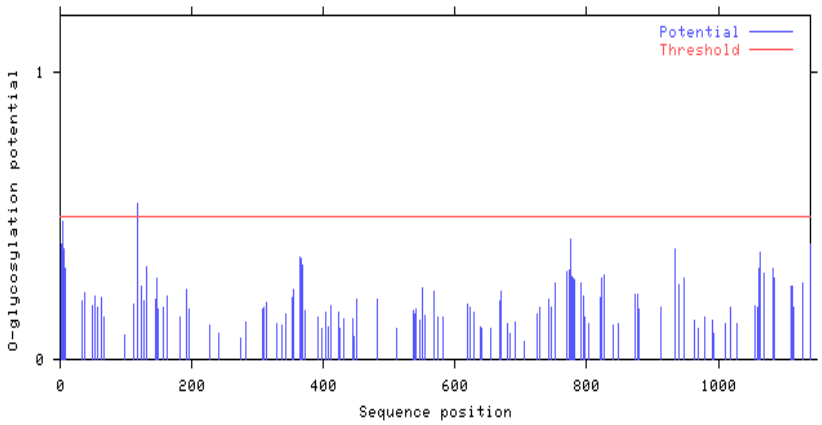

Figure 5 : O Glycosylation sites of protein

NetOGlyc results states that in the given protein contains only one O-Glycosylation site for Threonine.

NetNGlyc

```
Name:  Sequence    Length:  1140
MSTHTSSTLPAFKKILVANRGEIAVRAFRAALETGAATVAIYPREDRGSFHRSFASEAVRIGTEGSPVKAYLDIDEIIGA      80
AKKVKADAIYPGYGFLSENAQLARECAENGITFIGPTPEVLDLTGDKSRAVTAAKKAGLPVLAESTPSKNIDEIVKSAEG     160
QTYPIFVKAVAGGGGRGMRFVASPDELRKLATEASREAEAAFGDGAVYVERAVINPQHIEVQILGDHTGEVVHLYERDCS     240
LQRRHQKVVEIAPAQHLDPELRDRICADAVKFCRSIGYQGAGTVEFLVDEKGNHVFIEMNPRIQVEHTVTEEVTEVDLVK     320
AQMRLAAGATLKELGLTQDKIKTHGAALQCRITTEDPNNGFRPDTGTITAYRSPGGAGVRLDGAAQLGGEITAHFDSMLV     400
KMTCRGSDFETAVARAQRALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFIADHPHLLQAPPADDEQGRILDYLADV     480
TVNKPHGVRPKDVAAPIDKLPNIKDLPLPRGSRDRLKQLGPAAFARDLREQDALAVTDTTFRDAHQSLLATRVRSFALKP     560
AAEAVAKLTPELLSVEAWGGATYDVAMRFLFEDPWDRLDELREAMPNVNIQMLLRGRNTVGYTPYPDSVCRAFVKEAASS     640
GVDIFRIFDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYYLKMAEEIVKSGAHILAIKDMAGLLR     720
PAAVTKLVTALRREFDLPVHVHTHDTAGGQLATYFAAAQAGADAVDGASAPLSGTTSQPSLSAIVAAFAHTRRDTGLSLE     800
AVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQLSNLRAQATALGLADRFELIEDNYAAVNEMLGRPTKVTPSS     880
KVVGDLALHLVGAGVDPADFAADPQKYDIPDSVIAFLRGELGNPPGGWPEPLRTRALEGRSEGKAPLTEVPEEEQAHLDA     960
DDSKERRNSLNRLLFPKPTEEFLEHRRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTPLLVRLDAISEPDDKGMRNVV    1040
ANVNGQIRPMRVRDRSVESVTATAEKADSSNKGHVAAPFAGVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGKID    1120
RVVVPAATKVEGGDLIVVVS                                                                 1200
..............................................................................          80
..............................................................................         160
..............................................................................         240
..............................................................................         320
..............................................................................         400
..............................................................................         480
..............................................................................         560
..............................................................................         640
..............................................................................         720
..............................................................................         800
..............................................................................         880
..............................................................................         960
..............................................................................        1040
..............................................................................        1120
..................                                                                     1200
```

pyc protein has no N-Glycosylation Units.

NetPhos 2.0

```
MSTHTSSTLPAFKKILVANRGEIAVRAFRAALETGAATVAIYPREDRGSFHRSFASEAVRIGTEGSPVKAYLDIDEIIGA      80
AKKVKADAIYPGYGFLSENAQLARECAENGITFIGPTPEVLDLTGDKSRAVTAAKKAGLPVLAESTPSKNIDEIVKSAEG     160
QTYPIFVKAVAGGGGRGMRFVASPDELRKLATEASREAEAAFGDGAVYVERAVINPQHIEVQILGDHTGEVVHLYERDCS     240
LQRRHQKVVEIAPAQHLDPELRDRICADAVKFCRSIGYQGAGTVEFLVDEKGNHVFIEMNPRIQVEHTVTEEVTEVDLVK     320
AQMRLAAGATLKELGLTQDKIKTHGAALQCRITTEDPNNGFRPDTGTITAYRSPGGAGVRLDGAAQLGGEITAHFDSMLV     400
KMTCRGSDFETAVARAQRALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFIADHPHLLQAPPADDEQGRILDYLADV     480
TVNKPHGVRPKDVAAPIDKLPNIKDLPLPRGSRDRLKQLGPAAFARDLREQDALAVTDTTFRDAHQSLLATRVRSFALKP     560
AAEAVAKLTPELLSVEAWGGATYDVAMRFLFEDPWDRLDELREAMPNVNIQMLLRGRNTVGYTPYPDSVCRAFVKEAASS     640
GVDIFRIFDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYYLKMAEEIVKSGAHILAIKDMAGLLR     720
PAAVTKLVTALRREFDLPVHVHTHDTAGGQLATYFAAAQAGADAVDGASAPLSGTTSQPSLSAIVAAFAHTRRDTGLSLE     800
AVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQLSNLRAQATALGLADRFELIEDNYAAVNEMLGRPTKVTPSS     880
KVVGDLALHLVGAGVDPADFAADPQKYDIPDSVIAFLRGELGNPPGGWPEPLRTRALEGRSEGKAPLTEVPEEEQAHLDA     960
DDSKERRNSLNRLLFPKPTEEFLEHRRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTPLLVRLDAISEPDDKGMRNVV    1040
ANVNGQIRPMRVRDRSVESVTATAEKADSSNKGHVAAPFAGVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGKID    1120
RVVVPAATKVEGGDLIVVVS                                                                 1200
.................................................S.............T.................          80
.....................................................................T..........S...     160
.......................S.........................Y.............................     240
.....................................................................................T......     320
.........T.........................TT..........T.......S.............................     400
......S.....................................T.....T.....................................     480
.............................S.............................TT.....................     560
........T....S.......T.....................................T...T....S...........S     640
.........................................S............Y.....................          720
....T.....................................................S.S.........T...T..S..     800
..S...................T...T.....................................Y..........T..T.SS     880
.........................Y.....................................S......T.............     960
..S.....S.........................S..............T..................S............    1040
...............S..S.........SS.....................................................    1120
.......T............                                                                 1200
```

Phosphorylation sites predicted:

Ser: 25 Thr: 25 Tyr: 4

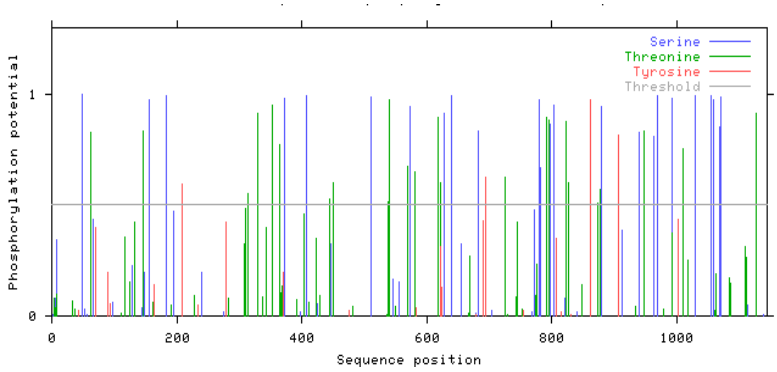

Figure 6 : N Glycosylation sites of protein

Netphos results say that it contains the phosphorylation sites of 25, 25 and 4 for Serine, Threonine & Tyrosine respectively.

CONCLUSION

Pyruvate Carboxylase is an important enzyme catalyses the ATP-dependent carboxylation of pyruvate to oxaloacetate. The pyc gene of *Corynebacterium glutamicum* strain ATCC 13032 (gi|62388892:706684-710106) codes for the Pyruvate carboxylase enzyme. Genomics states that the pyramidine rich gene codes for 3834 aminoacid peptide with the molecular weight of approximately 1167KD (ss DNA). The gene contains both left & right primers, a hybridization probe and several restriction sites for restriction sites. In proteomics primary structure says that it is an acidic, stable protein. Secondary structure reveals that it mainly contains the random coils & alpha helix. The protein has a total of 54 phosphorylation sites (Serine-25, Threonine-25 & Tyrosine-4) along with one O-Glycosylation site and no N-Glycosylation Units. This insilico exploration studies are very much useful for designing the *Corynebacterium glutamicum* for the enhancement of glutamic acid production.

REFERENCES

1. Scrutton, M.C., Young, M.R. and Utter, M.F. Pyruvate carboxylase from baker's yeast. The presence of bound zinc. J. Biol. Chem. **245**: 6220-6227 (1970).
2. Keech, D. B. & Wallace, 1. C. (1985). Pyruvate carboxylase. CRC Series in Enzyme Biology. Boca Raton, Florida: CRC Press.

3. Owen OE, Kalhan SC, Hanson RW. The key role of anaplerosis and cataplerosis for citric acid cycle function. J. Biol. Chem. 2002;277:30409–30412.
4. Jitrapakdee S, Nezic MG, Cassady AI, Khew-Goodall Y, Wallace JC (July 2002). "Molecular cloning and domain structure of chicken pyruvate carboxylase". *Biochem. Biophys. Res. Commun.* **295** (2): 387–93.
5. García-Cazorla A, Rabier D, Touati G, Chadefaux-Vekemans B, Marsac C, de Lonlay P, Saudubray JM (January 2006). "Pyruvate carboxylase deficiency: metabolic characteristics and new neurological aspects". *Ann. Neurol.* **59** (1): 121–7.
6. Geourjon C, Deléage G, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, PMID: 8808585, Medline.
7. C. Geourjon & G. Deleage, 1995, "SOPMA : Significant improvement in protein secondary structure prediction by c prediction from alignments and joint prediction", CABIOS, 11, 681-684.