



## Analysis of Peptide Mass for Characterization of Common Enzymes in Shared Pathways of *Escherichia coli* and *Arabidopsis thaliana*

Prashant Ankur Jain<sup>1</sup>, A. K. Gupta<sup>2</sup>, Krishna Misra<sup>3</sup>, Satendra Singh<sup>1</sup>

<sup>1</sup>Department of Computational Biology & Bioinformatics, JSBB, SHIATS, Allahabad, India.

<sup>2</sup>Department of Forensic Science, SHIATS, Allahabad, India.

<sup>3</sup>Division of Bioinformatics, IITA, Allahabad, India.

\*Corresponding author's E-mail:

Accepted on: 26-06-2013; Finalized on: 31-08-2013.

### ABSTRACT

Peptide mass was calculated for *E. coli* and *A. thaliana* for peptide mass analysis and characterization of common enzymes according to same EC number in common metabolic pathways of *E. coli* and *A. thaliana*. Citrate pathway shows 5 enzymes of *E. coli* and 3 enzymes of *A. thaliana* with same E.C.no. Glycolysis pathway shows 5 enzymes of *E. coli* and 4 enzymes of *A. thaliana*. Pyruvate pathway gives the information about the 5 enzymes of *E. coli* and 3 enzymes of *Arabidopsis thaliana* having similarities. Galactose pathway has 5 enzymes of *E. coli* and 4 enzymes of *Arabidopsis thaliana* with common E.C.no. Methane pathway shows 5 enzymes of *E. coli* and 4 enzymes of *Arabidopsis thaliana* having similarities according to E.C.no. Nucleotide sugar pathway gives the information of 5 enzymes of *E. coli* and 4 enzymes of *Arabidopsis thaliana* which shares common E.C.no. Urea pathway has 5 enzymes of *E. coli* and 4 enzymes of *Arabidopsis thaliana* which are common. Pentose phosphate pathway shares 5 enzymes of *E. coli* and 3 enzymes of *Arabidopsis thaliana*. Riboflavin pathway shows 5 enzymes of *E. coli* and 3 enzymes of *Arabidopsis thaliana* having same E.C.no. Carboxylate pathway shows the similarities of common enzymes of 4 of *E. coli* and 3 of *Arabidopsis thaliana* between them.

**Keywords:** *A. thaliana*, *E. coli*, Metabolic pathway, Motif.

### INTRODUCTION

This present work studies how the enzyme sequence can be used to study metabolic pathways, the interrelationship over ranges of different organism and the comparative characteristics of specific enzyme within the pathways. The metabolic pathway was studied in the two organism *E. coli* and *Arabidopsis thaliana*. Bioinformatics tools were used for the identification of the similarity of the common enzymes between the two mentioned organisms in metabolic pathways.<sup>1-4</sup> *E. coli* is the most studied of all organisms in biology because of its occurrence, and the ease and speed of growing the bacterium in the laboratory. It has been used in hundreds of thousands of thousand of experiments in cell biology, physiology, and genetics and was among the first cell for which the entire chromosomal DNA base sequence (genome) was determined. A computer searching algorithm has been used to identify protein sequences in the Protein Information Resource (PIR) database with peptide mass information (mass map) obtained from proteolytic digests of proteins analyzed by micro capillary high-performance liquid chromatography electro spray ionization mass spectrometry.<sup>5-8</sup> A theoretical analysis of the cytochrome c family demonstrates the ability to identify protein sequences in the PIR database with a high degree of accuracy using a set of six predicted tryptic peptide masses.<sup>9-11</sup>

### MATERIALS AND METHODS

The bioinformatics tools were used to carry out of the comparative analysis of metabolic pathway in two organisms. Firstly two organism *E. coli* and *Arabidopsis*

*thaliana* have been selected. Different metabolic pathways had been searched like glycolysis pathway, citric acid pathway, urea pathway, riboflavin pathway, methane pathway, pentose phosphate pathway, galactose pathway, nucleotide sugar pathway, carboxylate pathway, pyruvate pathway of *E. coli* and *Arabidopsis thaliana* in KEGG database. The metabolic reference pathway indicates the metabolites involve in the pathway, enzymes involved in the pathway, with their EC classification and associated metabolic pathway. The EC number or the enzyme classification system describe the functions of an enzyme in terms of four numbers organized in a hierarchical manner, and their can be several enzymes that have the same EC number but are different enzyme present in different pathways. Different enzymes had been searched by clicking on individual EC number and collect their flat files to compare their enzymes in different metabolic pathways according to their same EC number. PEPTIDE MASS program was used to calculate the theoretical masses of peptides generated by the chemical or enzymatic cleavage of proteins, to assist the interpretation of peptide mass, fingerprinting and peptide mapping experiments ([http://web.expasy.org/peptide\\_mass/](http://web.expasy.org/peptide_mass/)). Protein sequences are retrieved from UniProt Knowledgebase (Swiss-Prot or TrEMBL). When proteins of interest are specified from UniProt KB/Swiss-Prot, the program considers all annotations for that protein in the UniProt KB/Swiss-Prot database; it is used in order to generate the correct peptide masses.<sup>12</sup>



## Mass Calculation

Masses have been calculated to 4 or 5 decimal places for all a.a. and post translational modifications. Average isotopic and monoisotopic mass values are available for all these modification.

## Enzymes

Consideration that has to be taken into account partial cleavages, it is possible to specify a maximum number (0, 1, 2, 3) of missed cleavages.

## Single sequences, propeptides and transit peptides

These all are removed from proteins before cleavage rule are applied. All will be removed before generating the masses of peptides from the mature protein.

## RESULTS AND DISCUSSION

The characterization of the common enzymes of metabolic pathway in *E.coli* and *Arabidopsis thaliana* was carried out. KEGG was employed to investigate the metabolic sequence of different metabolic pathways in two organisms *E.coli* and *Arabidopsis thaliana*. The KEGG reference page includes the amino acid and gene sequence for enzyme in the metabolic pathway. The reference cycle also indicates the EC number of the enzyme and the detail of the enzymes is searched for by using the EC number.

Different enzymes that are shared by both the organism in the metabolic pathway were selected for comparative analysis of the metabolic pathway according to EC number using computational tools. Various metabolic pathways such as glycolysis cycle, citric acid cycle, urea cycle, riboflavin cycle, methane cycle, pentose phosphate pathway, pyruvate, carboxylate cycle, nucleotide sugar cycle, galactose has been taken for the analysis of the peptide sequences of the various common enzymes.<sup>13-14</sup>

### Peptide mass analysis of enzymes in *E.coli*

The citrate pathway has 5 common enzymes having peptide analysis range of 88.9% to 96% of the sequence covered. The theoretical pl is of the range 5.15 -6.04, Average mass 38897.14- 105061.72and Mono isotopic mass of 38872.32- 104995.54 (Table 1).

The glycolysis pathway has 5 common enzymes having peptide analysis range of 84.8% to 95.7% of the sequence covered. The theoretical pl is of the range 6.21-5.46, Average mass 38897.14- 105061.72and Mono isotopic mass of 26954.82- 99605.95 (Table 1).

The pyruvate pathway has 5 common enzymes having peptide analysis range of 84.8% to 95.7% of the sequence covered. The theoretical pl is of the range 6.21-5.46, Average mass 38897.14- 105061.72and Mono isotopic mass of 26954.82- 99605.95 (Table 1).

The galactose pathway has 5 common enzymes having peptide analysis range of 88.6% to 93.8% of the sequence covered. The theoretical pl is of the range 5.51- 6.12,

Average mass 34218.27- 99062.61and Mono isotopic mass of 34196.71 - 99000.25 (Table 1).

The methane pathway has 5 common enzymes having peptide analysis range of 90.9% to 97.4% of the sequence covered. The theoretical pl is of the range 4.99- 6.23, Average mass 19543.28 - 51481.31 and Mono isotopic mass of 19531.02 - 61491.01 (Table 1).

The nucleotide sugar pathway has 5 common enzymes having peptide analysis range of 89.9% to 97.6% of the sequence covered. The theoretical pl is of the range 5.03- 6.03, Average mass 33102.73 - 84162.61 and Mono isotopic mass of 33081.78- 84110.43 (Table 1).

It was found that 92.1% of sequence for the enzyme alcohol dehydrogenase class III/glutathione-dependent formaldehyde dehydrogenase was covered for the analysis and the sequence has the theoretical pl-5.85, Average mass (Mw)-39358.99 and Mono isotopic mass(Mw)-39333.87 (Table 1).

The urea sugar pathway has 5 common enzymes having peptide analysis range of 87.8% to 96.1% of the sequence covered. The theoretical pl is of the range 5.47- 5.96, Average mass 35241.54 - 72093.59 and Mono isotopic mass of 35219.30 - 72048.25 (Table 1).

The pentose phosphate pathway has 5 common enzymes having peptide analysis range of 90.9% to 95.2% of the sequence covered. The theoretical pl is of the range 5.34- 6.61, Average mass 50729.42 - 99668.49 and Mono isotopic mass of 50697.26- 99605.95 (Table 1).

The riboflavin pathway has 5 common enzymes having peptide analysis range of 88.2% to 95.8% of the sequence covered. The theoretical pl is of the range 5.52 - 9.34, Average mass 21835.98 -99062.61 and Mono isotopic mass of 21822.18- 99000.25 (Table 1).

The carboxylate pathway has 5 common enzymes having peptide analysis range of 89.6% to 94.9% of the sequence covered. The theoretical pl is of the range 5.46 – 6.11, Average mass 35570.97 -49195.19 and Mono isotopic mass of 35548.31- 49164.49 (Table 1).

### Peptide mass analysis of enzymes in *A. Thaliana*

The citrate pathway has 3 common enzymes having peptide analysis range of 89.8% to 91.7% of the sequence covered. The theoretical pl is of the range 6.47 – 7.58, Average mass 47234.10 -115221.51 and Mono isotopic mass of 47203.96 - 115221.51 (Table 2).

The glycolysis pathway has 4 common enzymes having peptide analysis range of 90.3% to 94.9% of the sequence covered. The theoretical pl is of the range 6.11 – 9.19, Average mass 35570.97 -108495.05 and Mono isotopic mass of 35548.31 - 108427.07 (Table 2).

The pyruvate pathway has 3 common enzymes having peptide analysis range of 90.9% to 93.8% of the sequence covered. The theoretical pl is of the range 5.41– 7.58, Average mass 47234.10 -110216.75and Mono isotopic mass of 47203.96- 110216.75 (Table 2).



The galactose pathway has 3 common enzymes having peptide analysis range of 90.9% to 97.2% of the sequence covered. The theoretical pl is of the range 5.32– 7.02,

Average mass 32306.53 -62307.94 and Mono isotopic mass of 32286.68- 62268.86 (Table 2).

**Table 1:** Peptide mass analysis of enzymes in *E.coli*

Sequence Similarity	Enzyme EC no.	Theoretical pl-	Average Mass (Mw)-	Mono isotopic Mass (Mw)-
<b>Citrate pathway</b>				
93.6%	EC:4.2.1.3	5.24	93498.11	93438.96.
93.2%	EC:1.2.4.2	6.04	105061.72	104995.54
88.9%	EC:2.3.1.61	5.58	44011.39	43984.01
96.1%	EC:1.1.1.37	5.96	38897.14	38872.32
91.1%	EC:1.1.1.42	5.15	45756.71	45727.4
<b>Glycolysis pathway</b>				
92.1%	EC:1.2.4.1	5.46	99668.49	99605.95
84.8%	EC:5.4.2.1	5.85	28556.40	28538.75
95.7%	EC:5.3.1.1	5.64	26971.81	26954.82
<b>Pyruvate pathway</b>				
95.4%	EC:5.1.3.3	4.84	38190.46	38166.77
94.5%	EC:6.2.1.1	5.50	72093.59	72048.25
93.6%	EC:4.2.1.3	5.24	93498.11	93438.96
96.1%	EC:1.1.1.37	5.96	38897.14	38872.32
94.5%	EC:6.2.1.1	5.50	72093.59	72048.25
<b>Galactose pathway</b>				
91.1%	EC:1.1.1.42	5.15	45756.71	45727.48
91.0%	EC:1.1.1.49	5.56	55704.44	55669.34
<b>Methane pathway</b>				
92.5%	EC:1.1.1.44	5.05	51481.31	51449.31
92.4%	EC:5.1.3.1	5.13	24554.25	24538.65
90.9%	EC:2.7.1.12	6.23	19543.28	19531.02
97.4%	EC:2.7.1.15	4.99	32290.52	32270.86
93.1%	EC:5.3.1.9]	5.85	61529.77	61491.01
<b>Nucleotide sugar pathway</b>				
93.0%	EC:2.1.2.1	6.03	45316.59	45287.97
89.9%	EC:1.5.1.20	6.00	33102.73	33081.78
<b>Urea pathway</b>				
94.5%	EC:6.2.1.1	5.50	72093.59	72048.25
91.4%	EC:2.3.3.13	5.47	57297.91	57262.10
96.1%	EC:1.1.1.37	5.96	38897.14	38872.32
<b>Pentose phosphate pathway</b>				
95.2%	EC:2.3.1.9	6.61	40352.48	40326.98
93.9%	EC:1.1.1.40	5.34	82417.38	82364.79
92.1%	EC:1.2.4.1	5.46	99668.49	99605.95
<b>Riboflavin pathway</b>				
88.2%	EC:2.7.1.26 2.7.7.2	9.34	34734.27	34712.53
93.5%	EC:3.5.4.26 1.1.1.193	7.22	40338.31	40312.82
95.8%	EC:2.5.1.9	5.64	23444.90	23430.03
<b>Carboxylate pathway</b>				
96.7%	EC:2.7.2.11	6.07	39056.52	39032.47
89.6%	EC:2.3.1.1	6.09	49195.19	49164.49
95.3%	EC:3.5.1.16	5.46	42266.20	42239.33
94.9%	EC:6.2.1.1	6.11	35570.97	35548.31

**Table 2:** Peptide mass analysis of enzymes in *A. Thaliana*

Sequence Similarity	Enzyme EC no.	Theoretical pl-	Average Mass (Mw)-	Monoisotopic Mass (Mw)-
<b>Citrate acid pathway</b>				
90.9%	EC:1.1.1.42	7.58	47234.10	47203.96
89.8%	EC:1.2.4.2	6.47	115221.51	115149.02
91.7%	EC:4.2.1.3	6.71	108495.05	108427.07
<b>Glycolysis pathway</b>				
90.3%	EC:2.3.1.61	9.19	50133.51	50102.14
94.9%	EC:1.1.1.37	6.11	35570.97	35548.31
91.7%	EC:4.2.1.3	6.71	108495.05	108427.07
<b>Pyruvate pathway</b>				
93.4%	EC:6.2.1.1	5.41	76733.11	76684.09
90.9%	EC:1.1.1.42	7.58	47234.10	47203.96
<b>Galactose pathway</b>				
91.4%	EC:5.4.2.1	5.32	60579.57	60541.70
90.9%	EC:5.1.3.3	5.88	37196.78	37173.50
91.8%	EC:5.3.1.1	7.02	32306.53	32286.68
<b>Methane pathway</b>				
89.5%	EC:1.2.4.1	5.67	39175.88	39151.10
93.4%	EC:6.2.1.1	5.74	81889.05	81836.89
93.9%	EC:3.2.1.23	8.64	81998.87	81945.96
<b>Nucleotide sugar pathway</b>				
97.9%	EC:3.2.1.22	4.73	48362.63	48331.56
90.5%	EC:1.11.1.6	6.40	66646.89	66604.23
91.6%	EC:1.5.1.20	6.23	45612.26	45583.05
<b>Urea pathway</b>				
88.9%	EC:5.1.3.2	6.14	39157.71	39132.80
90.9%	EC:4.2.1.46	7.09	38389.13	38364.69
<b>Pentose phosphate pathway</b>				
90.9%	EC:2.7.7.9	5.72	51919.64	51887.17
92.6%	EC:1.1.1.49	6.02	70202.45	70158.76
<b>Riboflavin pathway</b>				
92.4%	EC:2.7.1.15]	5.45	37676.91	37653.26
87.8%	EC:2.7.6.1	8.48	43332.92	43305.53
94.7%	EC:5.1.3.1	5.72	24110.96	24095.16
<b>Carboxylate pathway</b>				
86.8%	EC:2.7.1.12	5.52	20969.20	20955.78
89.9%	EC:1.1.1.44	5.34	53377.51	53344.08
97.2%	EC:5.3.1.9	5.62	62307.94	62268.86

The methane pathway has 4 common enzymes having peptide analysis range of 89.5% to 93.9% of the sequence covered. The theoretical pl is of the range 5.74– 8.64, Average mass 39175.88 -81998.87 and Mono isotopic mass of 39151.10 - 81945.96 (Table 2).

The nucleotide sugar pathway have 4 common enzymes having peptide analysis range of 90.5% to 97.9% of the

sequence covered. The theoretical pl is of the range 4.73– 6.40, Average mass 45612.26 -67988.91 and Mono isotopic mass of 45583.05- 81945.96 (Table 2).

The urea sugar pathway has 3 common enzymes having peptide analysis range of 88.9% to 90.9% of the sequence covered. The theoretical pl is of the range 6.10 – 7.09, Average mass 38389.13 -42529.13 and Mono isotopic mass of 38364.69- 42501.76 (Table 2).

The pentose phosphate pathway has 3 common enzymes having peptide analysis range of 90.9% to 92.7% of the sequence covered. The theoretical pI is of the range 5.72–6.02, Average mass 51919.64 -70202.45 and Mono isotopic mass of 51887.17- 70158.76 (Table 2).

The riboflavin pathway has 3 common enzymes having peptide analysis range of 87.8% to 92.4% of the sequence covered. The theoretical pI is of the range 5.45– 8.48, Average mass 24110.96 -43332.92 and Mono isotopic mass of 24095.16- 43305.53 (Table 2).

The carboxylate pathway has 3 common enzymes having peptide analysis range of 86.8% to 97.2% of the sequence covered. The theoretical pI is of the range 5.34– 5.62, Average mass 20969.20 -62307.94 and Mono isotopic mass of 20955.78- 62268.86 (Table 2).

### CONCLUSION

The peptide mass analysis provides important information about the enzymes of the different metabolic pathways. Peptide mass analysis gives the details of the peptide sequences of the enzymes which are covered, also shows the mass cleavage of the enzymes with the real mass value. Characterization of enzymes of metabolic pathways of *E.coli* and *Arabidopsis thaliana* shows that this central set of pathway is largely conserved in terms of pathways present. Thus, with the help of computational tools the metabolic pathway can be interpreted in terms of the protein involved.

### REFERENCES

1. Agrawal, Dunhum IV, Bioinformatics: An undergraduate research/teaching tool, department of biology, Coastal Carolina University, Conway sc29528, Journal of the south caroline academy of science, 2(1), 1-17.
2. Bjorkland M, The impendent contrast method in comparative biology, cladistics, 10, 1994, 425-433.
3. Bjorkland M, Are comparative methods always necessary, oikos, 80, 1997, 607-612.
4. Beeckmans, Kanarek L, Enzyme –enzyme interactions as modulators of the metabolic flux through the citric acid cycle, Biochem soc symp, 54, 1987, 163-172. PMID:3332992
5. Haladia S, Alberto Santos Z, Socorro Gama S, BMC Bioinformatics, 7, 2006, 5. doi:10.1186;(1471-2105).
6. Harold j, Morowitz, Jennifer D, Kostelnik Jeremy Yang, George D, Cody, The origin of intermediary metabolism, Proceedings of National Academy of Science, 97(14), 2000, 7704-7708.
7. Holms WH, Control of flux through the citric acid cycle and the glyoxylate bypass in *Escherichia coli*, Biochem Soc Symp, 54, 1987, 17-31. PMID :3332993.
8. Laporte DC, Walsh K, Koshland DE, The branch point effect Ultrasensitivity and subsensitivity to metabolic control, J Biol Chem, 259(22), 1984, 14068-75. PMID:6389540
9. Lazo, Pascal AS, Robert AL, doi:10.1038/nbt1091-963, Department of Biology, Sinsheimer Laboratories, University of California, Santa Cruz, CA 95064, Bio/Technology, 9, 1991, 963-967.
10. Mansi EM, Nimmo HG, Holms WH, Pyruvate metabolism and the phosphorylation state of isocitrate dehydrogenase in *Escherichia coli*, J Gen Microbiol, 132, 1986, 797 -806. PMID:3525743
11. Michael WK, Ph.d/IU school of medicine /university of Brescia, TCA cycle and its regulation, G.W 77-89.
12. Milburn D, Laskowski RA, Thornton JM, Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis, Prot. Eng., 11, 1998, 855-859.
13. Magasanik B, reznikoff WS, Riley M, Schaechter M, Umbarger HE, *Escherichia coli* and cellular and molecular biology , second edition, American society of microbiology, Washington, D.C. 1996.
14. Neidhardt FC, curtiss iii R, Ingraham JL, Lin ECC, Low Jr KB, Comprehensive approach to computational biology, 5, 2003, 546-611.

Source of Support: Nil, Conflict of Interest: None.

