RESEARCH ARTICLE

AN AMINO ACID SEQUENCES BASED COMPUTATIONAL ANALYSIS OF ENZYME CYTIDYLATE KINASE

Nitin Kumar Verma[1, 2]*, Balwinder Singh[3], Vibha[4]

[1] *Department of Biotechnology and Bioinformatics, Uttaranchal College of Science and Technology, Dehardun, Uttarakhand, INDIA*
[2] *Faculty of Life Science, Uttarakhand Technical University, Dehradun, Uttarakhand, INDIA*
[3] *Department of Science, Ek Onkar Scholar Degree College, Shahbjnagar, Shahjahanpur, Uttar Pardesh, INDIA*
[4] *Genetics and Tree Propagation Division, Forest Research Institute, Dehradun, Uttarkhand, INDIA*
Email: nittinkumarverma@gmail.com

Received-01.08.2016, Revised-17.08.2016

Abstract: Computational analysis has been established for hypothetical study of amino acid sequences of the enzyme cytidylate kinase that derived from various programs and databases. Cytidylate kinase enzyme is widely distributed enzyme among bacteria and fungi. In the present study, thirteen full length amino acid sequences cytidylate kinase were retrieved, collected and subject to multiple sequence alignment (MSA), regular expression identification, domain identification, discovering individual amino acid composition, and construction of phylogenetic trees. Multiple sequence alignment revealed that three glycine, one lycine, one arginine and one valine were identically found in all the bacterial and fungal sources of cytidylate kinase. The two major sequence clusters were constructed by phylogenetic analysis. One cluster contains two species of fungi and six species of bacteria, where as other contain five species of only fungi. The amino acid composition results revealed that the average frequency of amino acid leucine is 9.29 % in fungi, where as alanine 13.61 % in bacteria. In addition, six unique motifs were also identified in the group analysis.

Keywords: Motif, Phylogentic analysis, Multiple sequence alignment, Cytidylate Kinase, Domain

INTRODUCTION

Kinases are a universal group of enzymes, which participate in a variety of cellular pathways. The name kinase is applied for enzymes, which catalyze the transfer of the terminal phosphate group from ATP to an acceptor that can be a small molecule, lipid, and protein substrate. The cellular and physiological roles of kinases are different (Cheek et al., 2002). Many kinases participate in signal transduction pathways, in which these enzymes are necessary components (Blenis, 1993). Other kinases are involved centrally in the metabolism of carbohydrates, lipids, nucleotides, amino acid residues, vitamins, and cofactors. Some kinases play roles in various other processes, such as gene regulation, muscle contraction, and antibiotic resistance. Their universal roles in cellular processes, kinases are the best-studied enzymes at the structural, biochemical, and cellular level. Although all kinases catalyze essentially the same phosphoryl transfer reaction, and display significant diversity in their structures, substrate specificity, and number of pathways in which they participate (Cheek et al., 2002).

Nucleoside monophosphate kinases (NMP kinases) are the key enzymes, which are involved in the metabolism of nucleotides. They act specifically on the various NMPs formed in *de novo* or salvage pathways of purine or pyrimidine nucleotides, by catalyzing the reversible transfer of a phosphoryl group from a nucleoside triphosphate to an NMP (Briozzo et al., 1998). CMP kinase is the key enzyme in the nucleotide metabolism that is connected to the family of nucleoside monophosphate kinase (NMK) (Leipe et al., 2003). Substrate specificity was studied on recombinant human UMP/CMP kinase (pyrimidine nucleoside polyphosphate kinase), which show that UMP and CMP (Verma et al., 2013) are far better substrates than dCMP (Liou et al., 2002).

Computational methods are used to analyze protein function that can be divided into three vast categories: sequence, expression and interaction based methods (Pellegrini, 2001). The success of computational approaches is used for solving important problems such as sequence alignment and comparisons (Altschul et al., 1990). The importance of this approach in research is used to annotate the proteome through functional and structural genomic efforts (Michalovich, 2002). Considering the above facts, a study of amino acid sequences of cytidylate kinase from different sources of organisms is really challenging. In the present study, the individual computational studies of amino acid sequences were performed, which were obtained from bacteria, fungi, and correlated them on the basis of some common feature.

MATERIAL AND METHOD

The full-length amino acid sequences of cytidylate kinase from bacteria and fungi were retrieved from protein databases available at NCBI (National Center for Biotechnology Information). The sequences were

*Corresponding Author

arranged as in bacterial and fungal profile, respectively. The multiple sequence alignment of the individual profiles was performed using CLUSTRALX (Bateman, 2007). Motifs were discovered in profiles using the expectation maximization approach implemented in multiple EM for motif elicitation server (Bailey et al., 2006). Further, the discovered motifs were used to search their protein family using Pfam at the sanger institute (Finn et al., 2010). The neighbor joining approach implemented in the MEGA (Molecular Evolutionary Genetics Analysis) program was employed for phylogenetic analysis (Tamura et al., 2011). The statistical reliability of the phylogenetic tree was tested by bootstrap analyses with 500 replications. MEGA program is also used for discovering individual amino acid composition.

RESULT AND DISCUSSION

Sequence retrieval and analysis

All the sequences belong to different families of bacteria and fungi were retrieved from genbank (National Center for Biotechnology Information) protein database and listed in Table 1 along with their accession number, organism name, family and source.

Table 1. list of retrieved sequences with their different sources

S.No.	Source	Name of Organisms	Family	Accession no.
1	Bacteria	*Streptococcus pneumoniae*	Streptococcaceae	KGI36288.1
2	Bacteria	*Staphylococcus aureus*	Staphylococcaceae	KII20889.1
3	Bacteria	*Salmonella enteric*	Enterobacteriaceae	CBY95005.1
4	Bacteria	*Escherichia coli*	Enterobacteriaceae	ACA78315.1
5	Bacteria	*Pseudomonas aeruginosa*	Pseudomonadaceae	KFL11511.1
6	Bacteria	*Mycobacterium tuberculosis*	Mycobacteriaceae	NP_216228.1
7	Fungi	*Trichoderma gamsii*	Hypocreaceae	KUE96443.1
8	Fungi	*Moesziomyces antarcticus*	Ustilaginaceae	XP_014653432.1
9	Fungi	*Fusarium langsethiae*	Nectriaceae	KPA47046.1
10	Fungi	*Rhizoctonia solani*	Ceratobasidiaceae	CUA68089.1
11	Fungi	*Puccinia sorghi*	Pucciniaceae	KNZ52000.1
12	Fungi	*Rhizopus microspores*	Mucoraceae	CEG68485.1
13	Fungi	*Mucor ambiguous*	Mucoraceae	GAN07890.1

Multiple sequence alignment

Multiple Sequence Alignment (MSA) showed the presence of some conserved residues in all the sequences from different sources while others were restricted only to their groups. Three glycine, one lycine, one arginine and one valine were found to be identically conserved residues in all analysed species in bacterial and fungal profile (Figure 1 and 2).

Conserved motif identification

Six conserved motifs were identified after the analysis of bacterial and fungal profile individually. Three conserved motifs were observed in bacterial and fungal profile (Table 2).

Conserveed motif family identification

The six identified conserved motifs were applied to their family identification in Pfam database using sequence search option. First three conserve motifs identified in bacterial profile belong to **Cytidylate_kin** domain family while last three conserved motifs, a single conserved motif identified in fungal profile belong to **AAA_17** domain family and in the Pfam entry of rest two conserved motifs no significant family was found (Table 2).

Table 2. Motif identified using MEME program and their pfam analysis using pfam database

S.No.	Motif	Width	Present in number of sequences	Family	Sources
1	[PG]G[IL][VI][AM]DGRD[IM]GTVV[FL]P DAP[LV]KIFL[DT]AS[ASV]EERA[EHR]R R[YM][LK]Q[LN]Q[AE]KG[FI][ES]V[DN]	49	6	Cytidylate_kin	Bacteria

	FE				
2	[PI]VI[AT]IDGP[AS]GAGK[GS]T[VL][AC]K[AR][LM]A[ER][AE]L[GQ]WH[LY]LD[ST]GA[MI]YR[AV]L[AT][LY]AAL[HK]H[GH]VD	48	6	Cytidylate_kin	Bacteria
3	L[LK]A[ED]I[KR][EA]RDDRD[SR]NR[AE]V[AS]PL[KV]PA[AD]DA[VL]VLD[ST]TG[LM]SIE[EQ]V[VI]EK[IA]L[AQ]Y[AV][ER][KQR][KR]	50	6	Cytidylate_kin	Bacteria
4	[SMP][KS][KD][ILV][FT][VR][IV][FA][VI][LD]G[GP]P[GA][AS]GK[GS]T[QT][CA][AK][RL]L[VA]E[DE][YL]GF[TV][HY][LI][SD][AS]G[DA][LM][LF]RA[EI][QT]Q[RK][ECP][GQ][SQ]QY	50	7	AAA_17	Fungi
5	[CT][PST]E[ED][VK][ML][LE][SKP]RL[LI]ERGKTSGR[ET]DDN[EAI]ESI[KR]KRF[RQ]TF[VAI][EQ]TSMPV	41	7	Pfam hit not found	Fungi
6	[FI]L[IVL]DGFPR[KER][ML][DE]QA[IVQ][KA]F[DE][EAR][ETS][VFI][CQV][PEIM][SAP][AKQSV][FL]VLF[FL]	29	7	Pfam hit not found	Fungi

Clustral analysis

Clustral analysis of bacterial profile

Clustral analysis of bacteria showed two major clusters as shown in Figure 4. Clustral A consist of four species namely *Salmonella enterica, Escherichia coli, Pseudomonas aeruginosa and Mycobacterium tuberculosis.* Clustral B consist of two species namely *Streptococcus pneumoniae* and *Staphylococcus aureus*, respectively.

Clustral analysis of fungal profile

Clustral analysis of fungi showed two major clusters as shown in Figure 3. Clustral A consist of five species namely *Moesziomyces antarcticus, Puccinia sorghi, Rhizoctonia solani, Fusarium langsethiae* and *Trichoderma gamsii*. Clustral B consist of two species namely *Rhizopus microsporus* and *Mucor ambiguous*, respectively.

Clustral analysis of joint bacterial and fungal profiles

Two major clusters were obtained by clustral analysis of joint bacterial and fungal profiles (Figure 5). Clustral A consists of eight species, which were further divided into two subclustral. Subclustral A contains three species of bacteria and two species of fungi. Subclustral B contains three species of bacteria and five species of fungi.

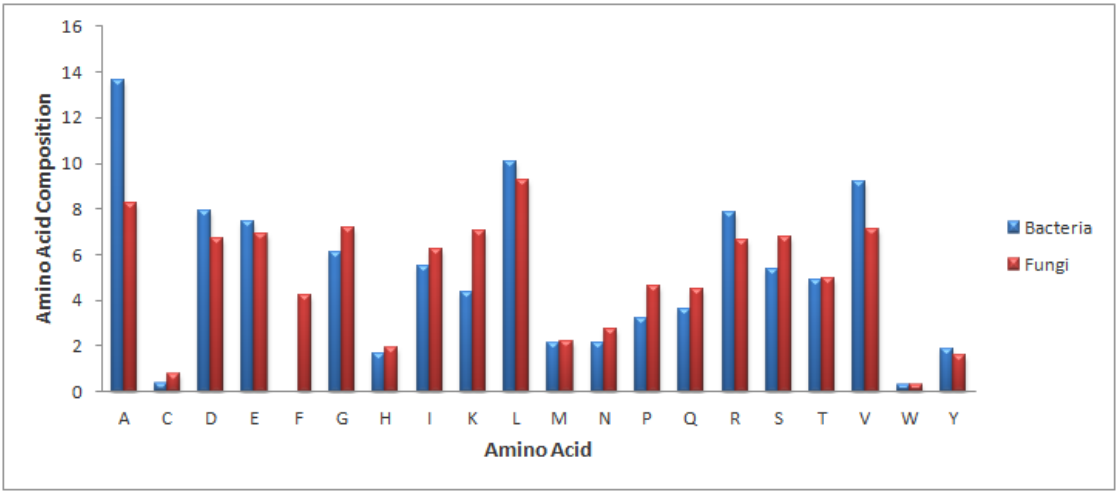

Fig. 1. Amino acid composition in different domain (Bacteria and Fungi)

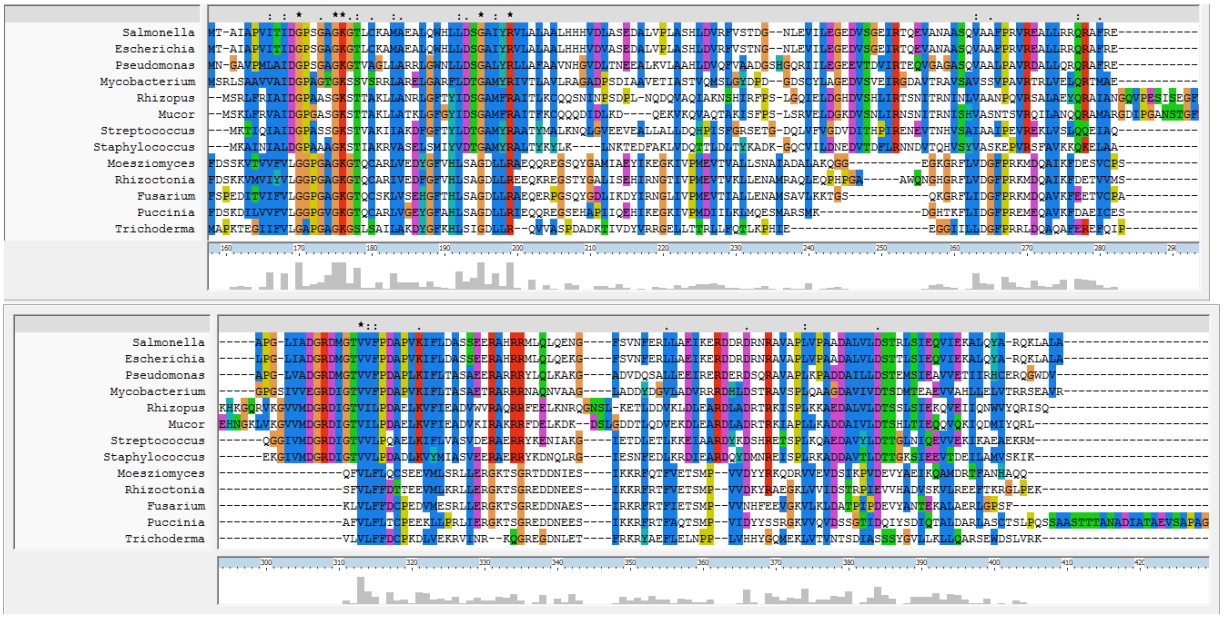

Fig. 2. The conservation study of enzyme cytidylate kinase between bacteria and fungi

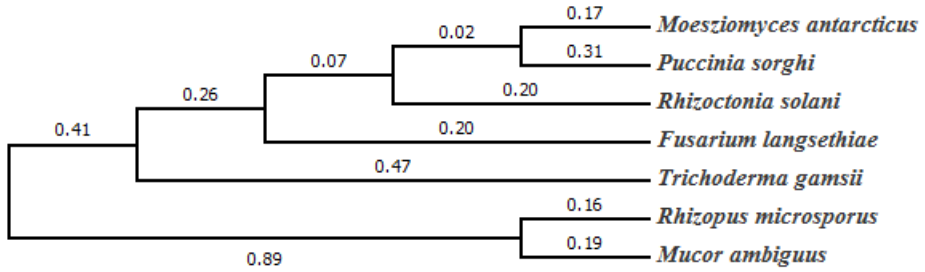

Fig. 3. Phylogenetic tree of fungal profile using neighbor joining method

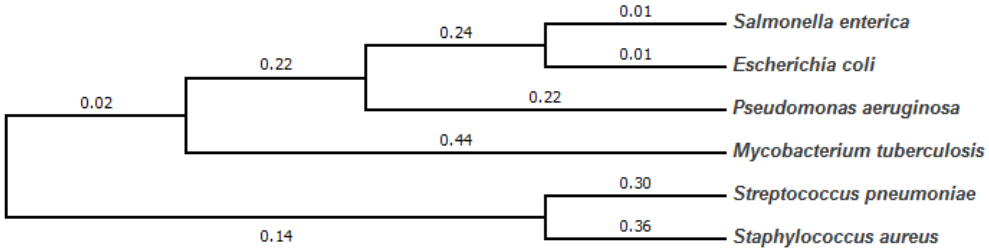

Fig. 4. Phylogenetic tree of bacterial profile using neighbor joining method

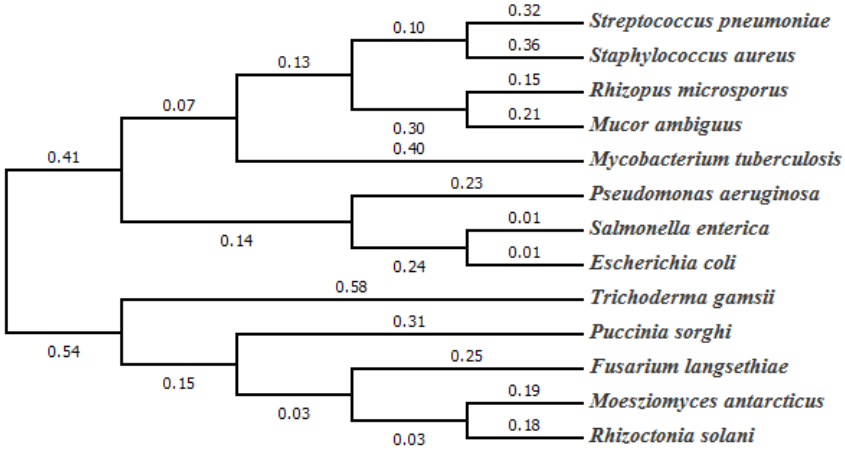

Fig. 5. Phylogenetic tree of joint profile of bacteria and fungi using neighbor joining method

CONCLUSION

Computational analysis of the cytidylate kinase sequences showed sequence-based similarities depending on their source organism. Three glycine, one valine, one lysine and one arginine residues were identically conserved in all analyzed species. These results suggested that the conserved amino acid residues have an important function in cytidylate kinase sequences. Six domains were identified in this research work, but three domains belong to Cytidylate_kin family that were found in all analyzed sequences of bacteria and fungi. These domains were found to be responsible for the functional activity of cytidylate kinase enzyme in different source organisms. These domains were conserved during the evolution of lower organism and their existence is important for the functional activity of this enzyme. In all species of bacteria and fungi an average frequency of amino acid leucine is 9.29 % in fungi, where as alanine 13.61 % in bacteria, which was higher in comparison to other amino acid average frequency. The amino acid alanine and leucine play an important role in the composition of cytidylate kinase. Two major sequences cluster were obtained by phylogenetic analysis. These phylogenetic analysis results suggested classification significance, which contributes the understanding of evolutionary relationship between the species at molecular level.

Competing interests

The authors declare no funding for this project, and no competing interests exits.

ACKNOWLEDGEMENT

Authors are grateful to Director, UCST, Dehradun, Uttarakhand, India for his generous laboratory support and throughout inspiration for accomplishment of this study.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Bailey, T.L., Williams, N., Misleh, C., Li, W.W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34, 369–373.

Bateman, A. (2007). ClustalW and ClustalX version 2.0. Bioinformatics 21, 2947–8.

Blenis, J. (1993). Signal transduction via the MAP kinases: proceed at your own RSK. Proc. Natl. Acad. Sci. U. S. A. 90, 5889–5892.

Briozzo, P., Golinelli-Pimpaneau, B., Gilles, A.-M., Gaucher, J.-F., Burlacu-Miron, S., Sakamoto, H., Janin, J., Bârzu, O. (1998). Structures of Escherichia coli CMP kinase alone and in complex with CDP: a new fold of the nucleoside monophosphate binding domain and insights into cytosine nucleotide specificity. Structure 6, 1517–1527.

Cheek, S., Zhang, H., Grishin, N. V. (2002). Sequence and structure classification of kinases. J. Mol. Biol. 320, 855–881.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A. (2010). The Pfam protein families database. Nucleic Acids Res. 38, D211–D222.

Leipe, D.D., Koonin, E. V., Aravind, L. 2003. Evolution and classification of P-loop kinases and related proteins. J. Mol. Biol. 333, 781–815.

Liou, J., Dutschman, G.E., Lam, W., Jiang, Z., Cheng, Y. (2002). Characterization of Human UMP / CMP Kinase and Its Phosphorylation of d- and l-Form Deoxycytidine Analogue Monophosphates Characterization of Human UMP / CMP Kinase and Its Phosphorylation of D - and L -Form Deoxycytidine Analogue Monophosphates. Cancer Res. 62, 1624–1631.

Michalovich, D. (2002). Protein sequence analysis in silico: application of structure-based bioinformatics to genomic initiatives. Curr. Opin. Pharmacol. 2, 574–580.

Pellegrini, M. (2001). Computational methods for protein function analysis. Curr. Opin. Chem. Biol. (5)1,46-50

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 2731-9

Verma, K. N., Singh, B. (2013). Insight from the structural molecular model of cytidylate kinase from *mycobacterium tuberculosis*. Bioinformation. (9)13, 680-684