Niger. J. Physiol. Sci. 29(June 2014) 047 –053
www.njps.com.ng

Niger. J. Physiol. Sci.

Bioinformatic analysis of dihydrofolate reductase predicted in the genome sequence of *Lactobacillus pentosus* KCA1

***Kingsley C. Anukam[1] and Uche Oge[2]**

[1]TWAS Genomic Research Unit, Department of Medical Laboratory Science, [2]Department of Physiology, School of Basic Medical Sciences, University of Benin, Benin City, Nigeria.

Summary: Physiologic studies of Lactobacillus species show that some species cannot synthesize folate *de novo*, which is required for growth. Folate plays a critical role in regulating the amount of tetrahydrofolate in the cell that is utilized for DNA replication, and proliferation of the erythropoietic system. We recently sequenced the genome of *Lactobacillus pentosus* KCA1, isolated from a Nigerian subject. The genome has open reading frames coding for the complete genes required for folate biosynthesis. Our previous study shows that rats fed with *L. pentosus* KCA1 led to enhancement of haematological parameters. Bioinformatic tool such as ClustalW algorithm was used to analyze dihydrofolate reductase (folA/dfrA) encoded in the genome sequence of *L. pentosus* KCA1 for comparative multiple sequence alignments. I-TASSER was used to predict the 3-D model structure of the protein and potential active binding site residues. Result show that two unique amino acid substitutions were found in KCA1_1610 sequence at position 85 with alanine (A-Ala85), while other strains have aspartic acid (D-Asp) for other *L. pentosus* and threonine (T-Thr) for *L. plantarum* strains at the same position. The result suggests that dihydrofolate reductase can be used as a distinguishing marker between *L. pentosus* KCA1 and other pentosus including *L. plantarum* strains. The secondary structure prediction with I-TASSER revealed 5 alpha helices and 8 beta-strands. Twelve binding site residues were predicted in KCA1_1610 relative to the template protein 2zzaA in protein database (PDB). The predicted structure of KCA1_1610 dihydrofolate reductase can serve as a new template as an addition to structural genomics and generation of models for use in drug screening and physiological function inference.

Keywords*: Lactobacillus pentosus*, folate biosynthesis, dihydrofolate reductase, probiotics

**Address for correspondence:* kanukam@gmail.com

Manuscript Accepted: May, 2014

INTRODUCTION

Lactobacilli are known to occur in a variety of microecology and several species colonize the human gut and vagina. Some species have been proven to confer health benefits on the host when given in adequate amounts, a concept known as probiotics (Lebeer, 2008). The mechanisms of action following these health-promoting characteristics are still investigated by several scientists. Probiotic bacteria have been shown to improve the immune system and biosynthesis of vitamins is one the suggested causal relationships of the health benefits. In many metabolic pathways of the body, folates represents an important nutritional constituent that are utilized for DNA replication, and proliferation of leucocytes, erythrocytes and enterocytes (Fuchs et al., 2002).

The microbiota of the human colon is known to produce vitamin K (menaquinones) and most of the water-soluble vitamins of group B, including biotin, nicotinic acid, folates, riboflavin, thiamine, pyridoxine, panthotenic acid, and cobalamin (Hill, 1997). In fact, the whole genetic information of the microbial community (microbiome) of the human distal gut revealed a variety of COGs (Clustered Orthologous Groups), which are involved in the synthesis of several essential vitamins (Gill et al., 2006). Unlike dietary vitamins, which are mainly absorbed in the proximal part of the small intestine, the uptake of microbial vitamins predominantly occurs in the colon (Said & Mohamed, 2006). Colonocytes appear to be able to absorb biotin, thiamin, folates, riboflavin, panthotenic acid, and menaquinones, indicating that the microbiota-produced vitamins may contribute to the systemic vitamin levels and especially to the homeostasis of the vitamins in the localized epithelial cells (Ichihashi et al., 1992). Absorption of folate occurs primarily in the duodenum and upper jejunum while the colon apparently represents a major repository of folate and the vitamin produced by the colonic microbiota

exceeds dietary intake and affects the folate status of the host. It is produced in large quantities by intestinal bacteria, mainly as monoglutamylated folate, the form that is absorbed at the highest rate (Kim et al., 2004). It has been demonstrated that the folate synthesized by intestinal bacteria can be absorbed across the colon and used by the host (Aufreiter et al., 2009).

The genome sequence of an increasing number of strains of Lactobacillus species has provided a major contribution to the knowledge of folate biosynthesis by these bacteria (De Crécy-Lagard et al., 2007). The analysis of genome sequences for predictable metabolic pathways using KEGG (Kyoto Encyclopedia of Genes and Genomes) database (http://www.genome.jp/kegg) suggests that the ability to synthesize pABA de novo is absent among all the sequenced members of the genus Lactobacillus. In fact, the enzymes which are necessary for chorismate conversion into pABA are lacking. Moreover, the shikimate pathway for chorismate production is complete only in the strains of *L. plantarum* WCFS1 (Kleerebezem et al., 2003) and *L. pentosus* KCA1 (Anukam et al., 2013), while it is absent or partial in all the other lactobacilli (Green et al., 1996).

The genome of *Lactobacillus pentosus* KCA1 dedicates 121 genes for the biosynthesis of cofactors including twenty-four open reading frames (ORF) coding genes involved in the biosynthesis of folate (Anukam et al., 2013). Our previous study provided useful information on the immuno-regulatory potentials of *Lactobacillus pentosus* KCA1 and suggest that ingestion of the strain will not cause any deleterious effect on the haematological parameters in healthy subjects as the red cell indices of Sprague-Dawley rats were enhanced (Anukam et al., 2014).

The objectives of the present study are to use bioinformatic tools to analyze the dihydrofolate reductase (folA or dfrA) encoded in the genome sequence of Lactobacillus pentosus KCA1 and second to determine the 3-D model structure of the protein and potential active binding site residues.

MATERIALS AND METHODS

Location of folate biosynthetic genes in *L. pentosus* KCA1

Folate biosynthesis repertoire was analyzed from RAST SEED database (Aziz et al., 2008). The Ensembl genome annotation system developed jointly by the European Buoinformatic Institute (EBI) and the Wellcome Trust Sanger Institute was used for the location, extraction of the nucleotide base sequence or open reading frame (ORF) and the amino acid translation of the dihydrofolate reductase from *L. pentosus* KCA1 (KCA1_1610) (http://ensemblgenomes.org/id/EIW13833).

Sequences similar to dihydrofolate reductase of *L. pentosus* KCA1 were searched for in UniProt® database using BLASTp algorithm.

Multiple sequence alignments

The amino acid translations from the nucleotide bases of 16 bacterial organisms were selected from the BLASTp of UniProt® database (http://www.uniprot.org/) based on product annotation hit (dihydrofolate reductase), gene name, % identity, matrix score and E-value. These 16 amino acids sequences including the sequence of *L. pentosus* KCA1 (http://www.uniprot.org/uniprot/I8R625) were imported into the ClustalW algorithm for multiple sequence alignments.

Prediction of secondary structure, 3-D model, similarity structure in PDB, functional and binding sites predictions using I-TASSER.

The iterative threading assembly refinement (I-TASSER) server is a four stage integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm (Roy *et al.*, 2010). The amino acid sequence of *L. pentosus* KCA1_1610 was submitted online (Yang, 2008) for the prediction of the 3D structure, similar structures in PDB, function and the binding site by this integrated algorithm.

RESULTS

The RAST subsystem identified all the genes involved in the pterin branch and *de novo* biosynthesis of folate in *L. pentosus* KCA1 (**Figure 1**). Three dihydrofolate reductase dfrA or FolA (KCA1_1610, KCA1_2309, KCA1_2335) genes were located in the *L. pentosus* KCA1 chromosome. Dihydrofolate reductase KCA1_1610 was found in contig AKAO01000040.1 from the DNA assembly and located between 1,801,463 and 1,801,954 along the *L. pentosus* KCA1 chromosome.

Basic Local Alignment Search Tool for proteins (BLASTp) from the UNIPROT® database yielded 250 hits with KCA1_1610 amino acid sequence. Sixteen Lactobacillus species were selected based on the same number of amino acid sequence, annotation name- dihydrofolate reductase, percentage identity (83-100 %), matrix score and e-value cutoff of $4.0\times10e^{-98}$ (**Table 1**). Two Lactobacillus species, notably, *L. pentosus* MP-10 and *L. pentosus* IG1 had 98.0 % amino acid sequence identity to KCA1_1610 with e-value 4.0×10e-115 and 2.0×10e-114 respectively. The remaining Lactobacillus species had 83.0% amino acid sequence identity.

ClustalW multiple sequence alignments showed that *L. pentosus* KCA1_1610 dihydrofolate reductase had significant amino acid sequence identity with the dihydrofolate reductase from the selected

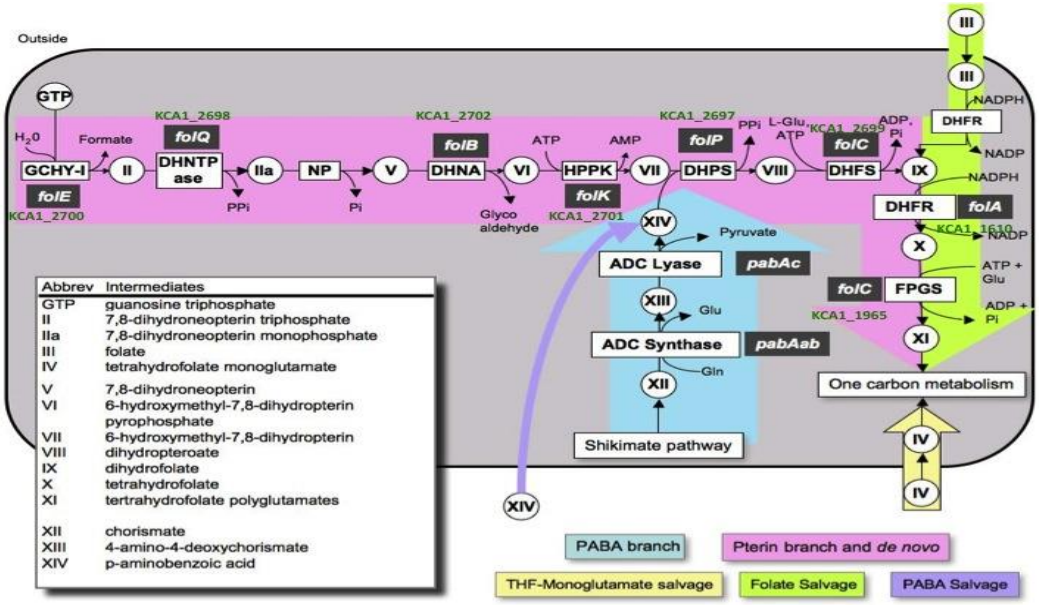

Figure 1: Folate biosynthesis pathways as identified with RAST Subsystems. Gene names are white on grey and the corresponding *L. pentosus* KCA1 genes are in green. ADC, aminodeoxychorismate; DHFR, dihydrofolate reductase; DHFS, dihydrofolate synthase; DHNA, dihydroneopterin aldolase; DHNTPase, dihydroneopterin triphosphate pyrophosphatase; DHPS, dihydropteroate synthase; FPGS, folylpolyglutamyl synthase; HPPK, hydroxymethyldihydropterin pyrophosphokinase.

```
                                                    1
F9UPJ9-L_plantarum_WCFS1                            MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
S2VAP5-L_plantarum_IPLA88                           MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
U6Z0T8-L_plantarum_WJL                              MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
R4PQX9-L_plantarum_P-8                              MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
D7V9P3-L_plantarum_ATCC_14917                       MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
M4KHV3-L_plantarum_ZJ316                            MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
U2INY3-L_plantarum_2165                             MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHIVMGRKTFAGF 50
C6VQJ3-L_plantarum_JDM1                             MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
R9X2L3-L_plantarum_16                               MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
H3P3R2-L_plantarum_NC8                              MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
M7CV21-L_plantarum_UCMA_3037                        MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
U2I5L6-L_plantarum_2025                             MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
E1TSF5-L_plantarum_ST-III                           MIALIWAEDQNGLIGNQGQLPWHLPADMQRFKALTTGHHVVMGRKTFAGF 50
F6IRR5-L_pentosus_MP-10                             MIALIWAEDQNGLIGKQGQLPWHLPADMHRFKTLTTGHHVVMGRKTFMGF 50
G0LZK9-L_pentosus_IG1                               MIALIWAEDQNGLIGKQGQLPWHLPADMHRFKTLTTGHHVVMGRKTFMGF 50
I8R625-L_pentosus_KCA1                              MIALIWAEDQNGLIGKQGQLPWHLPADMHRFKTLTTGHHVVMGRKTFMGF 50
                                                    **************:************:***:******:*******:**

F9UPJ9-L_plantarum_WCFS1                            KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
S2VAP5-L_plantarum_IPLA88                           KRPLPRRTNWVLSRQSDLKLPPEAHQLADVAAIQTLAAAHPDEPIFVIGG 100
U6Z0T8-L_plantarum_WJL                              KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
R4PQX9-L_plantarum_P-8                              KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
D7V9P3-L_plantarum_ATCC_14917                       KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
M4KHV3-L_plantarum_ZJ316                            KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
U2INY3-L_plantarum_2165                             KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
C6VQJ3-L_plantarum_JDM1                             KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
R9X2L3-L_plantarum_16                               KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
H3P3R2-L_plantarum_NC8                              KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
M7CV21-L_plantarum_UCMA_3037                        KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
U2I5L6-L_plantarum_2025                             KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
E1TSF5-L_plantarum_ST-III                           KRPLPRRTNWVLSRQSDLKLPPEVHQLADVAAIQTLAAAHPDEPIFVIGG 100
F6IRR5-L_pentosus_MP-10                             KRPLPHRTNWVLSQQSDLALPAGVQQLTDLAALKDLAAKNPDELIFVIGG 100
G0LZK9-L_pentosus_IG1                               KRPLPHRTNWVLSQQSDLALPAGVQQLTDLAALKDLAAKNPDELIFVIGG 100
I8R625-L_pentosus_KCA1                              KRPLPHRTNWVLSQQSDLALPAGVQQLTDVAALKALAAKNPDELIFVIGG 100
                                                    *****:*******:*****:**::.:**:*:**:: ***:*** ******

F9UPJ9-L_plantarum_WCFS1                            AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
S2VAP5-L_plantarum_IPLA88                           AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
U6Z0T8-L_plantarum_WJL                              AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
R4PQX9-L_plantarum_P-8                              AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
D7V9P3-L_plantarum_ATCC_14917                       AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
M4KHV3-L_plantarum_ZJ316                            AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
U2INY3-L_plantarum_2165                             AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
C6VQJ3-L_plantarum_JDM1                             AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
R9X2L3-L_plantarum_16                               AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
H3P3R2-L_plantarum_NC8                              AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
M7CV21-L_plantarum_UCMA_3037                        AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
U2I5L6-L_plantarum_2025                             AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
E1TSF5-L_plantarum_ST-III                           AVVFEAVLPVADYLYRTRINARFDGDTWMPAVDYTQWQLVSQQIGTVDEK 150
F6IRR5-L_pentosus_MP-10                             AVVFGAVLPIADYLYRTRINARFDGDTWMPAIDYSQWQLVTKQVGTVDDK 150
G0LZK9-L_pentosus_IG1                               AVVFGAVLPIADYLYRTRINARFDGDTWMPAIDYSQWQLVAKQVGTVDDK 150
I8R625-L_pentosus_KCA1                              AVVFGAVLPIADYLYRTRINARFDGDTWMPTIDYSQWQLVTKQVGTVDDK 150
                                                    ****:****:**********************.:**:*****::*:*****:*

F9UPJ9-L_plantarum_WCFS1                            NQYPYEFDDFRRR 163
S2VAP5-L_plantarum_IPLA88                           NQYPYEFDDFRRR 163
U6Z0T8-L_plantarum_WJL                              NQYPYEFDDFRRR 163
R4PQX9-L_plantarum_P-8                              NQYPYEFDDFRRR 163
D7V9P3-L_plantarum_ATCC_14917                       NQYPYEFDDFRRR 163
M4KHV3-L_plantarum_ZJ316                            NQYPYEFDDFRRR 163
U2INY3-L_plantarum_2165                             NQYPYEFDDFRRR 163
C6VQJ3-L_plantarum_JDM1                             NQYPYEFDDFRRR 163
R9X2L3-L_plantarum_16                               NQYPYEFDDFRRR 163
H3P3R2-L_plantarum_NC8                              NQYPYEFDDFRRR 163
M7CV21-L_plantarum_UCMA_3037                        NQYPYEFDDFRRR 163
U2I5L6-L_plantarum_2025                             NQYPYEFDDFRRR 163
E1TSF5-L_plantarum_ST-III                           NQYPYEFDDFRRR 163
F6IRR5-L_pentosus_MP-10                             NEYPYEFDDFRRR 163
G0LZK9-L_pentosus_IG1                               NEYPYEFDDFRRR 163
I8R625-L_pentosus_KCA1                              NEYPYEFDDFRRR 163
                                                    *:***********
```

Figure 2:ClustalW multiple alignment of the amino acid sequence of dihydrofolate reductase. Shown are the 16 dihydrofolate reductases indicating the gene entry name and the position of dihydrofolate reductase from *L. pentosus* KCA1 (I8R625). Residues in red show unique amino acid substitution relative to other strains. Green colour shows similar amino acid residues in L. pentosus strains. Overall homology is indicated under with *.

Lactobacillus strains especially *L. pentosus* MP-10 and IG1 (**Figure 2 and 3**). Two unique amino acid substitutions were noted in KCA1_1610 sequence at position 85 with alanine (A-Ala85) and valine at position Val131, while other strains have aspartic acid (D-Asp) for other *L. pentosus* and threonine (T-Thr) for *L. plantarum* strains at the 85 position.

The secondary structure prediction with I-TASSER revealed 5 alpha helices and 8 beta-strands as shown in **figure 4.** The first α-helix started at position 25 (P-Pro25) and ends at position 34 (L-Leu). It appears the shortest α-helix has only three amino acid residues and occurred between position 134 (Y-Tyr134) and 136 (Q-Gln136). The 3-D model of KCA1_1610 has a C-score of 1.68 and an estimated accuracy of 0.95±0.05 (TM-score-template modeling score) and 1.8±1.5Å (RMSD-root mean square deviation) based on the 10 templates used for

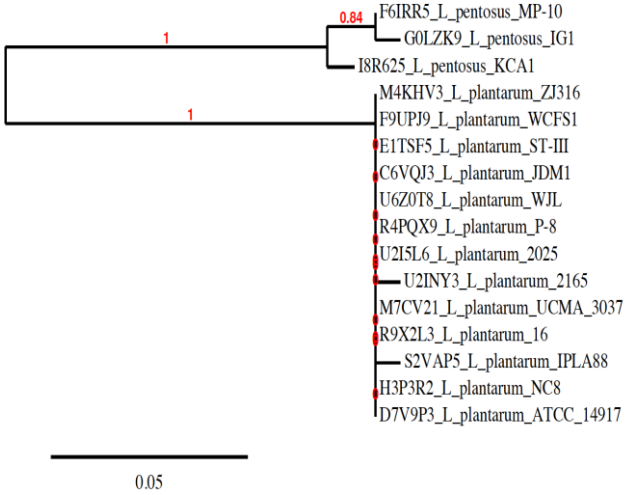

Figure 3. Dihydrofolate reductase gene tree. Confidence values for the branching order were generated by bootstrapping (based on 100 replications). The number at the nodes indicates the bootstrap values. The scale bar indicates 1 amino acid substitution per 100 amino acids

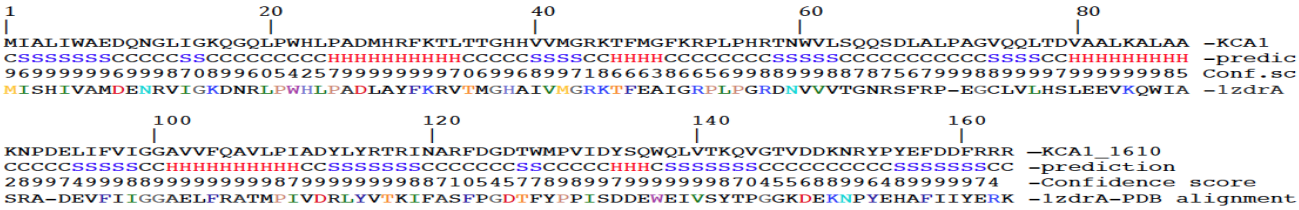

Figure 4: Predicted secondary structure with I-TASSER showing 5 alpha helices (red color) and 8 beta-strands (blue) and protein template (1zdrA) alignment with the top rank normalized Z-score

Table 1. Folate biosynthesis gene identity matrices predicted in *L. pentosus* KCA1

Gene entry ID	Organism	Length(AA)	Identity	Score	E-value	Gene name
I8R625	Lactobacillus pentosus KCA1	163	100.00%	871	4.0×10-117	dfrA KCA1_1610
F6IRR5	Lactobacillus pentosus MP-10	163	98.00%	858	4.0×10-115	LPE_00223
G0LZK9	Lactobacillus pentosus IG1	163	98.00%	853	2.0×10-114	LPENT_02812
M4KHV3	Lactobacillus plantarum ZJ316	163	83.00%	750	1.0×10-98	dfrA zj316_1848
F9UPJ9	Lactobacillus plantarum WCFS1	163	83.00%	750	1.0×10-98	dfrA lp_1869
E1TSF5	Lactobacillus plantarum (strain ST-III)	163	83.00%	750	1.0×10-98	dfrA LPST_C1499
C6VQJ3	Lactobacillus plantarum (strain JDM1)	163	83.00%	750	1.0×10-98	dfrA JDM1_1571
U6Z0T8	Lactobacillus plantarum WJL	163	83.00%	750	1.0×10-98	LPLWJ_30090
U2I5L6	Lactobacillus plantarum 2025	163	83.00%	750	1.0×10-98	N876_12100
R9X2L3	Lactobacillus plantarum 16	163	83.00%	750	1.0×10-98	Lp16_1445
R4PQX9	Lactobacillus plantarum subsp. plantarum P-8	163	83.00%	750	1.0×10-98	dfrA LBP_cg1422
M7CV21	Lactobacillus plantarum UCMA 3037	163	83.00%	750	1.0×10-98	H073_01658
H3P3R2	Lactobacillus plantarum subsp. plantarum NC8	163	83.00%	750	1.0×10-98	dfrA nc8_1602
D7V9P3	Lactobacillus plantarum subsp. plantarum ATCC 14917	163	83.00%	750	1.0×10-98	folA HMPREF0531_10866
U2INY3	Lactobacillus plantarum 2165	163	83.00%	749	1.0×10-98	N574_11250
S2VAP5	Lactobacillus plantarum IPLA88	163	83.00%	746	4.0×10-98	L103_11005

Table 2: Top 10 proteins with highly similar structural analogs in PDB (as identified by ™-align)

Rank	PDB-Hit	TM-score	RMSD[a]	IDEN[a]	Coverage
1	1zdrA	0.971	0.65	0.435	0.988
2	3ix9B	0.928	1.33	0.36	0.988
3	3e0bB	0.924	1.27	0.356	0.982
4	2w9sB	0.919	1.1	0.376	0.963
5	3tq8A	0.918	1.24	0.399	0.969
6	3f0bX	0.917	1.13	0.35	0.963
7	2zzaB	0.91	1.52	0.381	0.982
8	2d0kA	0.894	1.63	0.377	0.976
9	3jsuA	0.89	1.87	0.298	0.988
10	2blcA	0.889	1.85	0.298	0.988

(a) Ranking of proteins is based on TM-score of the structural alignment between the KCA1_1610 query structure and known structures in the PDB library. (b) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align. (c) IDEN[a] is the percentage sequence identity in the structurally aligned region. (d) Coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Table 3. Top 5 enzyme homologs in PDB

Rank	$Cscore^{EC}$	PDB-Hit	TM-score	$RMSD^a$	$IDEN^a$	Cov.	EC Number	Predicted Active Site Residues
1	0.812	1zdrA	0.971	0.65	0.435	0.988	1.5.1.3	20,27
2	0.811	3fyvX	0.916	1.14	0.35	0.963	1.5.1.3	20,27
3	0.794	1ao8A	0.88	1.82	0.381	0.982	1.5.1.3	20,27
4	0.776	1ddrA	0.88	1.73	0.371	0.976	1.5.1.3	5,27,31,54,98
5	0.725	3ia4A	0.909	1.49	0.381	0.982	1.5.1.3	5,31,54,98

(a) $Cscore^{EC}$ is the confidence score for the Enzyme Classification (EC) number prediction. CscoreEC values range in between (0-1); where a higher score indicates a more reliable EC number prediction. (b) TM-score is a measure of global structural similarity between query and template protein. (c) $RMSD^a$ is the RMSD between residues that are structurally aligned by TM-align. (d) $IDEN^a$ is the percentage sequence identity in the structurally aligned region. (e) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

Table 4: Template proteins with similar binding site:

Rank	$Cscore^{LB}$	PDB-Hit	TM-score	$RMSD^a$	$IDEN^a$	Cov.	BS-score	Lig. Name	Predicted binding site residues
1	0.65	2zzaA	0.911	1.51	0.381	0.982	1.55	FOL	5,6,7,20,27,28,31,32,50,54,57,98
2	0.49	2anoA	0.909	1.43	0.371	0.976	1.56	817	5,6,20,27,30,31,46,49,50,104,117
3	0.49	3kfyA	0.913	1.39	0.371	0.976	1.48	JZM	5,14,20,27,28,31,98,99,100,104,117
4	0.48	2w9sD	0.911	1.18	0.376	0.963	1.48	TOP	5,6,7,18,20,27,31,98,104
5	0.39	3tq8A	0.918	1.24	0.399	0.969	1.51	NDP	6,7,14,15,18,19,20,43,44,45,46,62,63,64,65,78,79,98,99,100,101,102,103,106,127
6	0.35	2kgkA	0.776	2.6	0.301	0.957	1.1	N22	5,50,52,55,98
7	0.26	3ia5B	0.892	1.59	0.38	0.969	1.74	PO4	43,45,46,100,103
8	0.24	1ddrB	0.887	1.65	0.371	0.976	1.35	URE	57,58,59,72,74
9	0.07	3fl9D	0.915	1.31	0.358	0.976	1.02	CA	109, 110, 111, 112, 162

(a) $Cscore^{LB}$ is the confidence score of predicted binding site. $Cscore^{LB}$ values range in between 0 and 1; where a higher score indicates a more reliable ligand-binding site prediction. (b) BS-score is a measure of local similarity (sequence & structure) between template binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, we have observed that a BS-score >1 reflects a significant local match between the predicted and template binding site. (c) TM-score is a measure of global structural similarity between query and template protein. (d) $RMSD^a$ the RMSD between residues that are structurally aligned by TM-align. (e) $IDEN^a$ is the percentage sequence identity in the structurally aligned region. (f) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

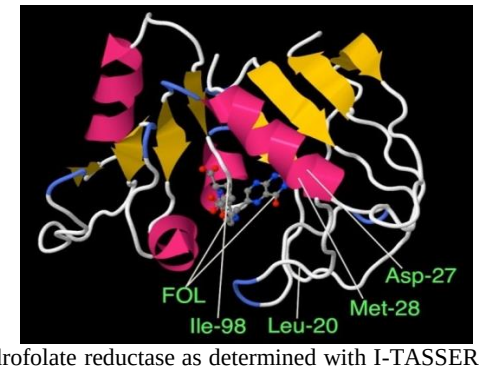

Figure 5: 3D model of KCA1_1610 dihydrofolate reductase as determined with I-TASSER based on alignments with 2zzaA-PDB. The co-ordinate file model in PDB format was visualized with Jmol molecular visualization program showing the position of the predicted binding site residues. Red color indicates the α-helices, while yellow indicates the β-pleated sheet. Magenta (+3 turns) and White (+2 turns).

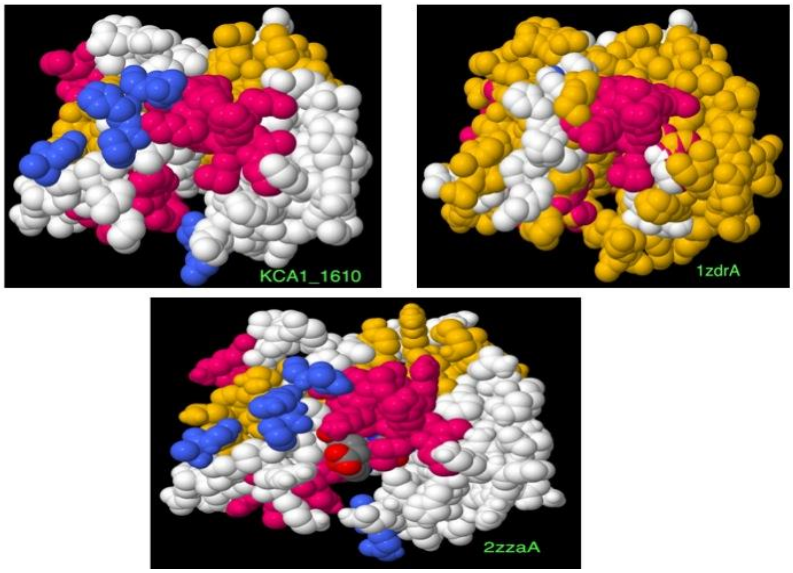

Figure 6: Comparative structure of 100% van der waal forces between KCA1_1610 dihydrofolate reductase and the top rank proteins in the PDB for similar binding site ligands (2zzaA), functional enzyme homologs and highly similar structure (1zdrA)

alignments with 1zdrA-PDB having the top rank normalized Z-score of 4.37. The co-ordinate file of the *L. pentosus* KCA1_1610 dihydrofolate reductase model was downloaded in PDB format and Jmol molecular visualization program (Hanson, 2010) was used to view the predicted structure as shown in **figure 5**. Proteins with highly similar structure in PDB as identified by TM-align are shown in **Table 2**. The protein 1zdrA PDB-hit as the top rank has a TM-score of 0.971 and a coverage of 0.988.

Five enzyme homologs were identified in PDB as having similar functions to the predicted KCA1_1610 sequence (**Table 3**). Notably, 1zdrA PDB-hit has the top rank with confidence score of 0.812 for the Enzyme Classification (EC) number (1.5.1.3). The predicted active site residues were identified as residues at position 20 and 27 in the KCA1_1610 amino acid sequence relative to three proteins in the PDB (1zdrA, 3fyvX, and 1ao8A). Predicted gene ontology (GO) terms associated with the KCA1_1610 query sequence identified 1zdrA from the PDB-hit with 6 GO terms. This protein has the top GO confidence score of 0.75 and TM-score of 0.9708.

Nine template proteins with similar binding site residues were predicted relative to KCA1_1610 sequence (**Table 4**). The template protein 2zzaA from

PDB-hit predicted binding site residues at position 5,6,7,20,27,28,31,32,50,54,57,98 relative to KCA1_1610 and has FOL (Folic Acid) as the ligand name and ligand-binding site prediction top rank with confidence score of 0.65. **Figure 6** shows the comparative structure of 100 % van der waal forces between KCA1_1610 dihydrofolate reductase and the top rank proteins in the PDB for similar binding site ligands (2zzaA), functional enzyme homologs and highly similar structure (1zdrA).

DISCUSSION

DHFR is a key enzyme in folate metabolism and play a critical role in regulating the amount of tetrahydrofolate in the cell. Tetrahydrofolate and its derivatives are essential for purine and thymidylate synthesis, which are important for cell proliferation and cell growth (Schnell, 2004). *L. pentosus* KCA1 was found to encode three genes coding for dihydrofolate reductase (EC 1.5.1.3) while *L. plantarum* WCSF1 has only one, suggesting that *L. pentosus* is equipped with more enzyme capabilities to produce folate. The amino acid composition of the gene coding for the dihydrofolate reductase with 163 amino acid residues from *L. pentosus* KCA1_1610 has a calculated molecular mass of 18, 665 daltons and it belongs to the protein family number PF00186 specific for DHFR (Myllykallio et al., 2003)

Clustal W multiple sequence alignment matrices (%) of the selected organisms clearly show that all the *L. plantarum* strains, with the exception of *L. plantarum*_2165 and *L. plantarum*_IPLA88 strains have the same sequence identity. This can be visualized from the phylogenetic gene tree (NJT) demonstrating the relatedness of the *L. plantarum* strains (**Figure 3**). The gene tree clearly shows that *L. pentosus* KCA1_1610 dihydrofolate reductase can be distinguished from other pentosus strains as the relationships at the node is apparently resolved. This suggests that KCA1_1610 can be used as a distinguishing marker between *L. pentosus* KCA1 and other pentosus including *L. plantarum* strains. However, dihydrofolate reductase of *L. pentosus* KCA1_1610 is closely related to L. *pentosus* IG1 and MP-10 than to *L. plantarum* strains. This is similar to our previous findings where the gene trees of the three conserved (housekeeping) genes (recA, dnaK, pheS) suggests that L. *pentosus* KCA1 is closer to *L. pentosus* IG1 and *L. pentosus* MP- 10 with higher percentage identity than to *L. plantarum* WCFS1 housekeeping genes (Anukam et al., 2013). The impact of the unique amino acid substitution at position 85 with alanine is yet to be determined and it would be interesting to know if there would be any difference on the biosynthesis of folate between *L. pentosus* KCA1 and *L. pentosus* MP-10 and IG1.

The secondary structure prediction was based on the alignment of KCA1_1610 to 1zdrA-PDB having top rank normalized Z-score. KCA1_1610 protein is in the same class with 1zdrA with EC number of 1.5.1.3. This class of enzyme from the physical characterization of the protein from *Bacillus stearothermophilus* indicates that it is a monomeric protein with a molecular mass of 18,694.6 Dalton, coincident with the mass of 18 694.67 Da calculated from the primary sequence (Kim et al., 2005). Determination of the X-ray structure of KCA1_1610 will provide an insight on whether the structure will turn out to be monomeric and if the calculated molecular mass of 18,665 will tally with the physical molecular mass.

Similarly, KCA1_1610 produced a hit to 1zdrA having the top rank confidence score in the enzyme homologs. C-score^{EC} is the confidence score for the Enzyme Classification (EC) number prediction. C-score^{EC} values range between 0 and 1; where a higher score indicates a more reliable EC number prediction. The predicted active binding site shows that residues Leu20 and Asp27 are found to be involved in KCA1_1610 protein, similar to three proteins (1zdrA, 3fyvX, 1ao8A) in the PDB. Recent study has shown that the nature of C-H→C transfer, and a phylogenetic analysis of DHFR sequences are consistent with evolutionary preservation of the protein dynamics to enable H-tunneling from well re-organized active sites (Francis et al., 2013).

The predicted gene ontology (GO) terms for KCA1_1610 identified 1zdrA PDB with 6 GO terms for biological functions, of which the structurally based sequence alignment of DHFRs indicates the following levels of sequence identity for KCA1_1610; 43% with 1zdrA, 37% and 35% with 3q10A and 3m08A PDB respectively. The implication of large number of GO terms associated with KCA1_1610 suggests that the protein may have biological attributes high in dihydrofolate reductase activities with a capacity to

optimize H-tunneling from donor (NADPH) to acceptor (DHF) substrates (Kim et al., 2005).

The template protein (2zzaA) with similar binding site residues was predicted to occur at 12 positions (Ile5, Trp6, Ala7, Leu20, Asp27, Met28, Phe31, Lys32, Phe50, Leu54, Arg57, Ile98) of the KCA1_1610 sequence. The template protein 2zzaA has the top rank confidence score for ligand binding site represented as FOL (Folic Acid). X-RAY DIFFRACTION with resolution of 2.00 Å shows that the template protein has three ligands including FOL (DOI:10.2210/pdb2zza/pdb).

In conclusion, bioinformatics tools have characterized the dihydrofolate reductase predicted in the genome sequence of *Lactobacillus pentosus* KCA1 as a protein with 5 alpha helices and 8 beta-strands. The protein putatively employs 12 amino acid residues as the ligand binding sites and has two unique amino acid substitutions at position ala85 and at position val131 relative to other lactobacillus species in the same clad. The structure of KCA1_1610 dihydrofolate reductase may serve as a new template that may be an addition to structural genomics and generation of models for use in drug screening and physiological function inference.

REFERENCES

Anukam, K. C., Macklaim, J. M., Gloor , G. B., Reid, G., Boekhorst , J., Renckens, B., van Hijum S. A., Siezen, R.J. (2013). Genome sequence of Lactobacillus pentosus KCA1: vaginal isolate from a healthy premenopausal woman. PLoS One. 8:e59239.

Anukam, K. C., Okungbowa, M., Mezue, C., Olise, N.A. (2014). *Lactobacillus pentosus* KCA1 does not alter haematological parameters of Sprague-Dawley albino rats, but significantly affects red cell indices pre-exposure to clinical isolate of *Escherichia coli*. Ann Biomed Sci, 13 (1): 43-52.

Aufreiter, S., Gregory, J.F., III, Pfeiffer, C.M., Fazili, Z., Kim, Y.I., Marcon, N., Kamalaporn, P., Pencharz, P.B., O'Connor, D.L. (2009). Folate is absorbed across the colon of adults: evidence from cecal infusion of (13)C-labeled (6S)-5-formyltetrahydrofolic acid. Am. J. Clin. Nutr. 90: 116–123

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9: 75.

De Crécy-Lagard, V., El Yacoubi, B., de la Garza, R.D., Noiriel, A., Hanson, A.D. (2007). Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genomics, 8: 245.

Francis, K., Stojkovic, V., Kohen, A. (2013). Preservation of protein dynamics in dihydrofolate reductase evolution. J Biol Chem;288(50):35961-35968.

Fuchs, C.S., Willett, W.C., Colditz, G.A., Hunter, D.J., Stampfer, M.J., Speizer, F.E., Giovannucci, E.L. (2002). The influence of folate and multivitamin use on the familial risk of colon cancer in women. Cancer Epidemiol. Biomarkers Prev, 11: 227–234.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., Nelson, K.E. (2006). Metagenomic analysis of the human distal gut microbiome. Science, 312, 1355–1359.

Green, J. C., Nichols, B.P., Matthews, R. G. (1996). Folate biosynthesis, reduction, and polyglutamylation. In Escherichia coli and Salmonella: Cellular and Molecular Biology Volume 1. Second edition. Edited by: Neidhardt FC, Curtiss R 3rd, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE. Washington DC, ASM Press :665-673.

Hill, M.J. (1997). Intestinal flora and endogenous vitamin synthesis. Eur. J. Cancer Prev, 6: S43–S45.

Ichihashi, T., Takagishi, Y., Uchida, K., Yamada, H. (1992). Colonic absorption of menaquinone-4 and menaquinone-9 in rats. J. Nutr., 122: 506–512.

Kim, T.H., Yang, J., Darling, P.B., O'Connor, D.L. (2004). A large pool of available folate exists in the large intestine of human infants and piglets. J. Nutr. 134:1389–1394.

Kim, H.S., Damo, S.M., Lee, S.Y., Wemmer ,D., Klinman, J.P. (2005). Structure and hydride transfer mechanism of a moderate thermophilic dihydrofolate reductase from Bacillus stearothermophilus and comparison to its mesophilic and hyperthermophilic homologues. Biochem. 44(34):11428-39.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O.P., Leer, R., Tarchini, R., Peters, S.A., Sandbrink, H.M., Fiers, M.W ., Stiekema, W., Lankhorst, R.M., Bron P.A., Hoffer, S.M., Groot, M.N., Kerkhoven, R., de Vries, M., Ursing, B., de Vos, W.M., Siezen R.J. (2003). Complete genome sequence of Lactobacillus plantarum WCFS1. Proc Natl Acad Sci USA, 100:1990-1995.

Lebeer, S., Vanderleyden, J., De, K., Sigrid, C. J. (2008). Genes and molecules of lactobacilli supporting probiotic action. Microbiol Mol Biol Rev, 72:728-64.

Myllykallio, H., Leduc, D., Filee, J., Liebl, U. (2003). Life without dihydrofolate reductase FolA. Trends Microbiol, 11:220-223.

Said, H.M., and Mohammed, Z.M (2006). Intestinal absorption of water-soluble vitamins: an update. Curr. Opin. Gastroenterol., 22, 140–146.

Schnell, J.R., Dyson, H.J, Wright, P.E. (2004). Structure, dynamics, and catalytic function of dihydrofolate reductase. Ann Rev Biophysics Biomolecular Structure 33 (1): 119–140.