Asian Jr. of Microbiol. Biotech. Env. Sc. Vol. 22, No. (1) : 2020 : 140-148

ISSN-0972-3005

EVOLUTION-GUIDED OPTIMIZATION FOR SELECTING SITE-DIRECTED MUTAGENESIS SITES OF *AGROBACTERIUM TUMEFACIENS* URONATE DEHYDROGENASE

MURUGAN R., PRATHIVIRAJ R., DIPTI MOTHAY AND CHELLAPANDI P.*

Molecular Systems Engineering Lab, Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli 620 024, Tamil Nadu, India

(Received 19 October, 2019; accepted 26 November, 2019)

Key words : *Uronate dehydrogenase, Agrobacterium tumefaciens, Site-directed mutagenesis, Coevolution, Mutant stability, Substrate specificity, Molecular evolution*

Abstract – Uronate dehydrogenase (EC:1.1.1.203) belongs to the NAD-dependent epimerase/dehydratase (NDE/D) subfamily, which converts D-galacturonic acid and D-glucuronic acid into D-galactaric acid and D-glucaric acid, respectively. Uronate dehydrogenase-catalyzed reaction is reversible and no substrate-specific activity in nature. Therefore, evolution-guided optimization approach was employed for screening, selection, and evaluation of its mutants to increase the substrate specificity with NAD^+ and D-glucuronic acid. The phylogenetic analysis described that uronate dehydrogenase from *A. tumefaciens* evolved from the UDP-glucose-4-epimerase subfamily members and not related to closely related soil bacteria. Molecular conservation of its sequence-structure-function integrity was retained in this organism by imposing purifying selection and amino acid substitution patterns. A single amino acid substitution in its proton relay system or substrate-binding site found to bring several changes in the local structural environments. It hasenforced to optimize the substrate-binding site that recognizes the D-galacturonic acid or D-glucuronic acid. Hence, site-directed mutagenesis targets detectedinthis study would be useful for engineering uronate dehydrogenase subjected to be used in the biotransformation process of D-glucaric acid production.

INTRODUCTION

D-Glucarate and its derivatives have been used as detoxifying and natural anti-carcinogenic compounds as well as building blocks for polymer synthesis (Walaszek *et al.,* 1996; Werpy and Petersen, 2004; Bespalov and Aleksandrov, 2012). D-Glucarate has also been concerned as "top value-added chemicals from biomass" since it is a potential substitute for petroleum-derived chemicals (U.S. Department of Energy). It is synthesized from glucose by chemical oxidation using a strong oxidant such as nitric acid or nitric oxide (Smith *et al.,* 2012). However, it's extensive medical and food applications have been hindered by contamination of chemical moieties derived from the manufacturing process. The microbial transformation process is an alternative way to overwhelm such crises. Several bacteria and yeast have been reported to produce it in the production media via D-galacturonate catabolism (Yoon *et al.,* 2009; Groninger-Poe *et al.,* 2014; Pick *et al.,* 2015; Matsubara *et al.,* 2016). A recombinant *Escherichia coli* was employed to produce D-glucarate from glucose for which a synthetic pathway constructed with uronate dehydrogenase from *Pseudomonas syringae* (Moon *et al.,* 2009; Shiue and Prather, 2014; Reizman *et al.,* 2015).

Uronate dehydrogenase (EC: 1.1.1.203) belongs to the NAD-dependent epimerase/dehydratase (NDE/D) subfamily of short-chain dehydrogenase (SDH)superfamily. This enzyme catalyzes the oxidation of D-galacturonic acid and D-glucuronic acid with NAD^+ as a cofactor into D-galactaric acid and D-glucaric acid, respectively (Yoon *et al.,* 2009). It has been identified and characterized from *Pseudomonas syringae* pv. tomato strain DC3000, *Pseudomonas putida* KT2440 and *Agrobacterium tumefaciens* strain C58 (Pick *et al.,* 2015). Crystallographic structure of apo-form of uronate dehydrogenase (PDB id: 3RFX) from*A. tumefaciens* C58 strain and its ternary complex structures with co-factors and substrate were determined previously (Parkkinen *et al.,* 2011). Crystallographic

and mass spectrometric studies elucidated its catalytic mechanism on the respective substrates and cofactors. It has a monomeric structure consisting of a Rossmann fold that is essential for nucleotide binding. Enzyme kinetic studies investigated the bi-substrate binding mechanism of uronate dehydrogenases from *P. syringae*, *Polaromonas naphthalenivorans*, and *Chromohalobacter salixigens* (Wagschal *et al.,* 2015). The catalytic efficiency of this enzyme from *Fulvimarinapelagi* HTCC2506, *Streptomyces viridochromogenes* DSM 40736 and *Oceanicola granulosus* DSM 15982 was previously characterized by Pick *et al.,* (2015).

Short-chain dehydrogenases are very old and emerged early during the evolution. These superfamily members have shown great variability in the origin, but have conserved 3D structures (Sola-Carvajal *et al.,* 2012). Inferring evolutionary constraints that acting on the structure-function integrity of *A. tumefaciens* uronate dehydrogenase (AtuUdh) across the NDE/D subfamily is an optimistic approach for increased its substrate binding specificity and catalytic efficiency towards D-glucarate production. In the present work, we demonstrated how evolutionary restraints acting on the functional and structural diversity of NDE/D subfamily. Site-directed mutagenesis siteswere screened, selected and evaluated from the evolutionary imprints for engineering AtuUdh with increased substrate specificity. Perhaps, engineered AtuUdh could be applied as an inclusion in the biosynthetic pathway and also to design a microbial biofactory for efficient transformation of the pectin-based wastes.

METHODS

Dataset

The amino acid sequence of AtuUdh (UniProt ID:Q7CRQ0) and its crystallographic structural information (PDB id: 3RFT) were retrieved from the UniProt database (http://www.uniprot.org/) and Protein Data Bank (http://www.rcsb.org/pdb/home/home.do), respectively. The similarity sequences for this sequence were searched out against a National Collection for Biotechnology Information (NCBI) server (http://www.ncbi.nlm.nih.gov/) by BLASTp program (Altschul *et al.,* 1997). The e-value cut-off was set as 10^{-5}. The similarity sequences harboring two common functional motifs YxxxK and GxxGxxG (Yoon *et al.,* 2009) were included in the dataset of this study.

Evolution-based screening

The sequences in the dataset were aligned with multiple sequence alignment program with the CLUSTALX 2.01 software (Thompson *et al.,* 1997). Aligned sequences were inspected manually and non-aligned sequences removed from the dataset. Estimates of sequence-based phylogeny for the sequence alignment were obtained by MEGA 5.05 software (Tamura *et al.,* 2011). Evolutionary genetic analyses (R, I, d, S, π, D) were performed with programs in the MEGA. A standardized measure of segregating sites and the average number of mutations between pairs in the dataset was calculated by Tajima's neutrality statistic program (Tajima, 1989). Synonymous and non-synonymous substitution ratio (ω) was calculatedwith the HyPhy 1.0 program (Pond *et al.,* 2005). Global alignment of two sequences was performed by EMBOSS ALIGN tool using a Needleman-Wunsch algorithm (www.ebi.ac.uk/Tools/psa/) to determine the evolutionary rate across closely related sequences (Udaya Prakash *et al.,* 2010). Phylogenetic functional divergence and its parameters were computed with Splits Tree 4.0 software using Bio NJ algorithm (Huson and Bryant, 2006). Recombination events and scaled-recombination/mutation rate were detected by RDP 3.0 software using Recomb 2007 method (Martin *et al.,* 2010). The type I (θ_I) and type II (θ_{II}) functional divergence coefficients along with the rate for gamma distributions (α) were examined by DIVERGE 1.04 (Gu and Vander Velden, 2002).

Analysis of coevolved pairs

Conserved domain and functional motifs in the AtuUdh sequence were identified from the NCBI conserved domain database (Geer *et al.,* 2015) and literature (Parkkinen *et al.,* 2011). Secondary structural features including solvent accessibility and H-bonding were predicted with JOY server (Mizuguchi *et al.,* 1997). A coevolved pair in response to the stability of the local structural environment was predicted by the CMAT program (Jeong *et al.,* 2012). Coevolved pairs were validated by a confidence score, as described by Lee and Kim (Lee and Kim, 2009).

RESULTS

Phylogenetic analysis

A dataset of this study consists of 69 protein sequences related to the SDH superfamily. As

shown in Fig. 1, we found five major clusters in the phylogenetic tree. The members of NDE/D, nucleoside-diphosphate-sugar epimerase (NDSE), ADP-L-glycero-D-manno-heptose-6-epimerase (AGME), UDP-glucose-4-epimerase (UG4E) and dTDP-glucose 4,6-dehydratase (DG46D) subfamilies are grouped together along with AtuUdh and *Pseudomonas putida* uronate dehydrogenase (PpuUdh). Atu Udh has shared its phylogenetic relationships with NDE/D subfamily members of *Rhizobium* sp. and *R. lupine* and then with UG4E of *Salipigermucosus*. The members of NDE/D, AGME and DG46D subfamilies are distinctly clustered together across the diverse bacteria in cluster 2. Unlike AtuUdh, PpuUdh is a strain-specific enzyme related to NDE/D subfamily members of *Poloromonas* and *Achromobacter* genera. Interestingly, uronate dehydrogenases from *P. putida* and *Pseudomonas* sp. are more diverged from NDE/D subfamily than AtuUdh.

Analysis of functional coefficient

Estimates of the functional coefficient of each cluster show that cluster 2 with relevant subfamily highly diverges from cluster 5 (q_1: 0.148±0.074), but the rate of evolution was very low within NDE/D subfamily (Table 1). The rate of type I functional divergence is almost the same between cluster 2 and cluster 3, which varies slightly across cluster 3 (Ppu Udh) and cluster 4 (Atu Udh). Neutral selection acts on type I functional divergence of NDE/D family in order to purify function as AtuUdh. Negative selection imposeson the functional divergence of NDE/D, AGME and DG46D subfamilies. Population-scaled recombination rate (0.237) and its frequent recombination events are also detected as evolutionary forces to select the parental sites for purifying the function of AtuUdh.

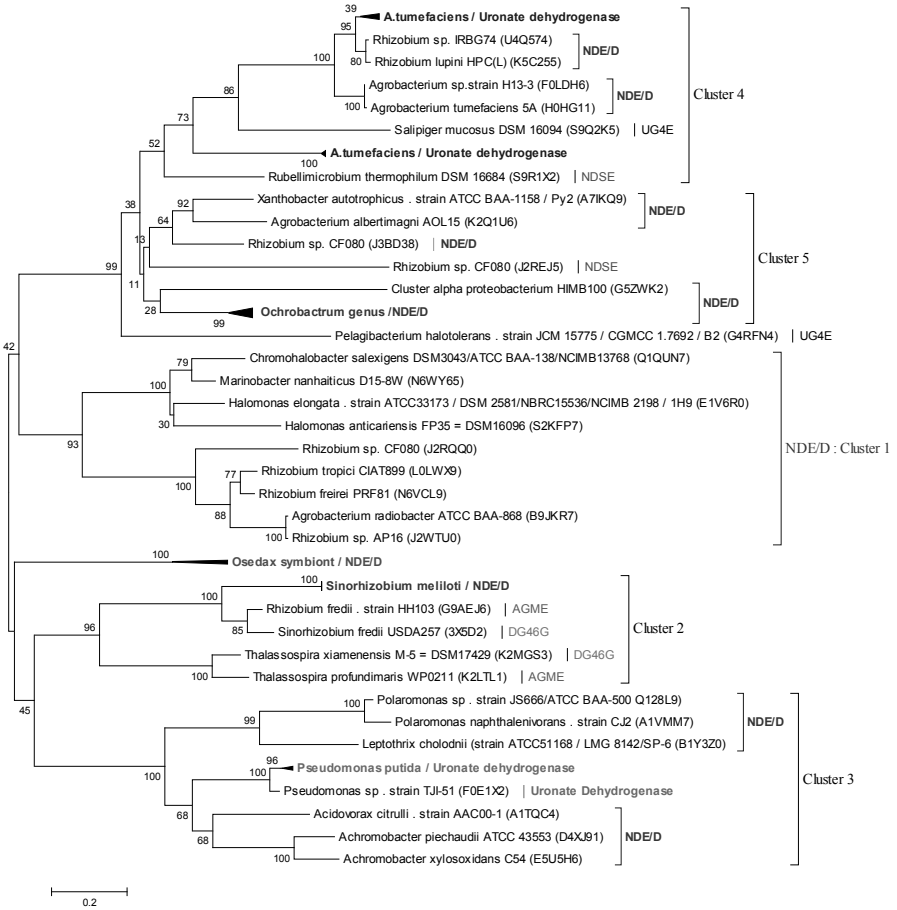

Fig. 1. Phylogeny for bacterial NDE/D subfamily, reconstructed by a neighbor-joining algorithm using the closely related protein sequences. Branch lengths are proportional to evolutionary distances. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed.

Table 1. Coefficients of functional divergence between homologous clusters of bacterial NDE/D subfamily, estimated by DIVERGE

Cluster	θ_1	α_1	θ_2	α_2
C1/C2	0.022±0.039	0.806	-0.328±0.120	0.939
C1/C3	0.047±0.042	0.771	-0.055±0.103	0.767
C1/C4	0.016±0.034	0.747	-0.144±0.103	0.745
C1/C5	0.082±0.052	0.654	-0.148±0.104	0.737
C2/C3	0.001±0.022	0.836	-0.182±0.113	0.911
C2/C4	0.036±0.044	0.879	-0.236±0.114	0.926
C2/C5	0.148±0.074	0.766	-0.204±0.113	0.919
C3/C4	-0.007±-0.64	1.006	-0.064±0.106	0.835
C3/C5	0.092±0.059	0.704	-0.088±0.108	0.847
C4/C5	0.010±0.033	0.682	-0.241±-.107	0.845

θ_I and θ_{II} are the coefficients of type-I and type-II functional divergence, respectively. The parameter á is the gamma shape parameter for rate variation among sites between clusters.

Genetic diversity and Darwinian selection

Gene/protein diversity is an important constraint to conclude the function of a protein family under selective pressure. Gene diversity (11.498) in NDE/D subfamily highly diverges than its protein diversity (0.433), as the results of significant segregating sites (508), mutability rate (0.237) and non-synonymous substitution rate (9.001) (Table 2). The results of the Tajima neutrality statistic (4.989/3.023) and nucleotide/amino acids diversity (0.426/0.374) pinpoint the positive selection acting on the recent ancestral gene sequences of NDE/D subfamily for the functional divergence of AtuUdh. We assume that the function of AtuUdh can be resolved consecutively from the NDE/D subfamily by the frequent occurrence of nucleotide substitution during protein evolution.

Table 2. Estimates of genetic diversity and the Darwinian selection of bacterial NDE/D subfamily

Genetic parameters	
Protein diversity	
Phylogenetic distance (d)	0.433
Invariant sites (+I)	0.685
Phylogenetic diversity	5.172
Number of segregating sites (S)	210
Nucleotide/amino acid diversity (ð)	0.426
Tajima test statistic (D)	4.894
Selective strength (Ka/Ks)	1.002
Gene diversity	
Transition/Transversion ratio (R)	1.387
Phylogenetic distance (d)	11.498
Recombination/mutability rate (D)	0.237
Number of segregating sites (S)	508
Nucleotide diversity (π)	0.374
Tajima test statistic (D)	3.023
Non-synonymous substitution rate(dN)	9.001
Synonymous substitution rate(dS)	8.749
Selective strength (dN – dS; &Ω)	0.252
Selective strength (dN – dS; normalized &Ω)	0.019

The general time reversible with gamma distribution found as an evolution model for all families.

Analysis of evolutionary patterns

Finding a suitable substitution matrix can provide a description of the understanding of evolutionary patterns or events of NDE/D subfamily. As shown in Fig. 2, a high sequence similarity score is observed

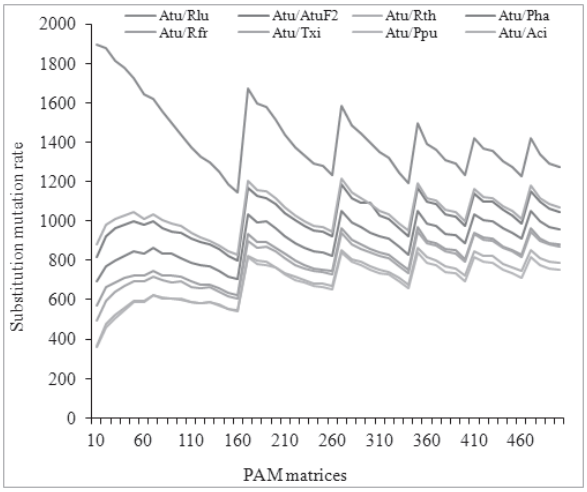

Fig. 2. Graph showing the substitution score (reflecting conservation patterns) obtained while aligning two sequences with respect to each PAM matrix for the evolution of Atu Udh and its subfamily. The Atuudh sequence was aligned against the members in the NDE/D subfamily of Aci: *Acidovoraxcitruli*, Ppu: *Pseudomonas putida*, Txi: *Thalassospira xiamenensis*, Rfr: *Rhizobium fredii*, Pha: *Pelagibacterium haltolerans*, Rth: *Rubellimicrobium thermophilum*, Atu F2: *Agrobacterium tumefaciens*F2 and Rlu; *Rhizobium lupine*

between AtuUdh and NDE/D subfamily member (*R. lupine*). It indicates that a low amino acid substitution rate and acceptable mutation are to be anticipated between them. Uronate dehydrogenase acquires the sites/residues from the parental sequences through five major evolutionary events. It states that significant sequence homologies and amino acid substitution patterns are found across the NDE/D subfamily that purifies the function of Atu Udh during the evolutionary process.

Coevolved pairs on the local structural environments

Coevolved sites are one of the evolutionary constraints to know how a particular residue-coupled mutation (coevolved pair) can convince the local structural environments. Coevolution process has significantly acted on the residues (23%) in the coils or loops of AtuUdh (Fig. 3). About 11% of changes are observed between residues in helices and no radical changes are in other secondary structural elements. We detected 56% of first residues showing to mutate the second residues conferring solvent accessibility/ solvent in accessibility ratio. Coevolved sites bring coupled changes in its local structural environment in the coiled and buried regions. About 20% of coevolved pairs do not contribute to the H-bonding patterns and 38% of first mutated residues destabilize the H-bonding of the coevolved pairs. H-bonding frequency (2-4 H-bonds per sites) is also compensated between coevolved pairs. It indicates that coevolution events fundamentally rearrange the H-bonding patterns with non-hydrogen bonding

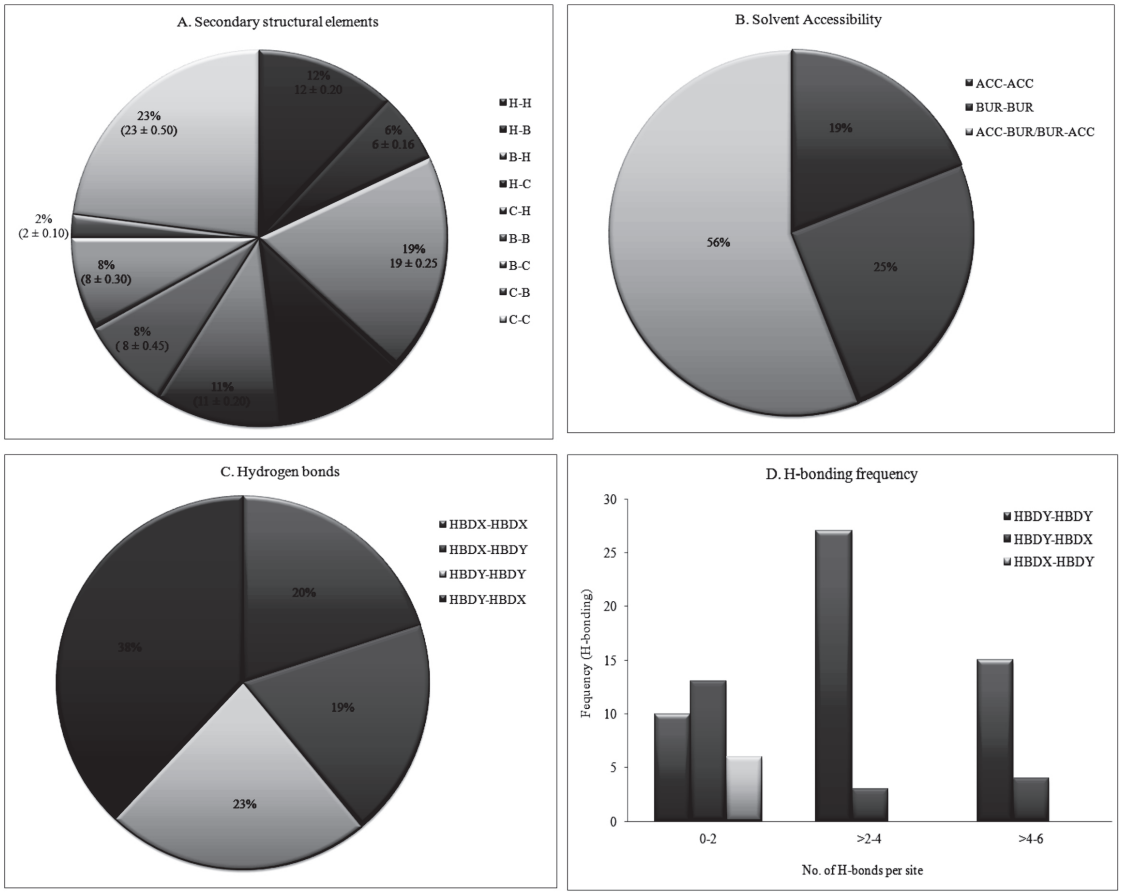

Fig. 3. Calculation of secondary structural types (A), solvent accessibility (B), hydrogen bonds (C) and hydrogen bonding frequency (D) for coevolved sites detected in the structure of AtuUdh. Within the coevolved pairs, if both residues are buried or accessible, they are shown as 'BUR-BUR' or 'ACC-ACC', respectively. 'HBDY-HBDY' and 'HBDX-HBDX' indicate cases where both coevolved residues are involved or not involved in hydrogen bonding correspondingly. 'HBDX-HBDY/HBDY-HBDX' indicates cases where at least one residue is involved in hydrogen bonding. Values in the parenthesis showthe mean and standard error of estimated from the distribution of structural property values for randomly selected non-coevolved residue pairs.
H: α-Helix; B: β-Strand: C: Coil

Table 3. Identification of functional sites presents in AtuUdh using conserved domain search and literature survey.

Binding site for	Amino acid residues
Proton relay system	N87-S111-Y136-K140
Catalytic site	S111-Y136
NAD^+	G10-[A12-G13-Q14-L15]-D34-S36-[C50-D51-L52-A53]-[L71-G73]-S75-G90-A109-S110-I163-[S165-C166]
NADH	[D34-L35-S36]-[D51-L52]-L71-G73-Y136-K140
$NADH_2$	Q14- L15-D34-S36-[D51-L52[]-L71-G73-S75-Y136-K140-C166
Substrate	S75-[S111-N112]-Y136-G164-R174-D184-A208-D210-D245
G15L	S75-S111-Y136-H113-R174-S165
Nicotinamide ribose	K140-Y136-I163-C166
Pyrophosphate	[Q14-L15]-S75
Sulfate	[K4-R5]-[N112-H113]-S165-R174
Homodimer	F15-F80-[L84-Q85]-[I88-I89]-[Y92-N93]-R99-L135-V138-F142-[N145-L146]-K153

pairs. Thus, the coevolved site takes a key role in its structural stability and functional fidelity.

Analysis of functional sites in AtuUdh

The proton relay system (N87-S111-Y136-K140) is an important catalytic mechanism of AtuUdh mainly contributed to cofactors (NADH, $NADH_2$) recognition and substrate binding (Table 3). Two conserved motifs such as [A12-G13-Q14-L15] and [G50-D51-L52-A53] are identified for the NAD^+ binding, whereas a motif [D34-L35-S36] detected for the NADH binding. The residues S111 and Y136 recognize nicotinamide ribose moiety in $NAD^+$and substrate. It implies the mutational changes in any one of these sites destabilize the structure-function integrity of AtuUdh.

DISCUSSION

All of the members in the SDR superfamily are not related to closely related organisms, providing evidence for convergent evolution (Rat *et al.,* 1991; Penning *et al.,* 1997; Hubbard *et al.,* 1998; Kavanagh *et al.,* 2008; Groninger-Poe *et al.,* 2014). Phylogenetic analysis of 66 bacterial species suggested two different origins for SDH in β-Proteobacteria and four origins for γ-Proteobacteria (Sola-Carvajal *et al.,* 2012). In our study, we found strong phylogenetic proximity between AtuUdh and NDE/D subfamily members of *Rhizobium* sp. and *R. lupine* and not related to closely related bacteria. UG4E subfamily suggested as a phylogenetic origin for the functional evolution of AtuUdh, which was more diverged from PupUdh and other members in NDE/D subfamily.

SDR superfamily members contain a common dinucleotide-binding Rossmann-fold domain and have a highly conserved 3D structure. Early origin has allowed them to diverge into several subfamilies and enzymatic activities in accordance with earlier studies (Tarrío *et al.,* 2011; Sola-Carvajal *et al.,* 2012; Groninger-Poe *et al.,* 2014; Martinez Cuesta *et al.,* 2014). The functional core of AtuUdh was duplicated slowly from the NDE/D subfamily members and subsequently purified its molecular function from closely related UG4E subfamily members by imposing type I functional divergence and neutral selection. It indicated that a purifying selection acted on its functional diversity from the UG4E subfamily remains unchanged for a given characteristic, despite the continuous process of mutagenesis.

Gene diversity of NDE/D subfamily is higher than its protein diversity upon non-synonymous substitutions. Genetic diversity analysis revealed the establishment of new consensus sequences and specific fingerprints for the lineages and sub-lineages of this subfamily (Sola-Carvajal *et al.,* 2012). However, a certain level of divergence in the NDE/D subfamily could make as AtuUdh in accordance with Choi and Hannehall, (2013). The sequence conservation analysis described the existence of significant sequence homologies and amino acid substitution patterns across the NDE/D subfamily that act on the functional convergence of AtuUdh.

Many functionally important residues that do not have apparent conservation patterns are evolutionarily connected with each other (Chakraborty and Chakrabarti, 2015; Zhang and Yang, 2015). Hence, coevolution process is an important constraint for inferring a variety of biological knowledge as a cooperative mechanism between interconnected residues plays a critical role in a protein function (Marks *et al.,* 2012; Sandler *et al.,* 2013). Compensating mutations might have an influence on the solvent accessibility/solvent

inaccessibility ratio of our mutant proteins than secondary structures. It suggested that substrate binding specificity of our mutants can be determined by coevolved pairs, which was agreed with the previous studies (Sarabojia *et al*., 2005; Wu and Cheng, 2014). H-bonding patterns in the first residues of AtuUdh are rearranged with non-hydrogen bonding pairs. It reflected that H-bonding frequency of it compensated between coevolved pairs in order to optimally select a specific substrate or cofactor.

Extensive knowledge of the structure of an enzyme can often provide crucial importance for its molecular function and regulation. The structure and function of a protein are highly correlated together. A single amino acid substitution in a protein may even change the function that the protein carries out (Alberts *et al.,* 2002; Engelhardt *et al*., 2011). If the change in function becomes advantageous, it is being subjected to the processes of natural selection and the point mutation has been accepted into the genetic pool (Arenas *et al*., 2013; Williams *et al*., 2013). Uronate dehydrogenase consists of two primary sequence motifs, YxxxK (Tyr145 and Lys149) and GxxGxxG (Gly18-to-Gly24), related to conserved domains (Zajic, 1959; Thomas *et al*., 2002; Hoffmann *et al*., 2007). GxxxG or Gx1-2GxxG motif is found in the NAD-binding domain of SDH superfamily (Kleiger and Eisenberg, 2002; Yoon *et al*., 2009). Our motif analysis indicated that AtuUdh has a proton relay system (N87-S111-Y136-K140) exhibiting to contribute to the binding of NAD, NADH, and substrates. If a single amino acid substitution exists on these sites, the structure-function integrity and catalytic function of AtuUdh become undesirable. We observed that overall secondary structural stability of AtuUdh is not affected upon point mutations, which was agreed with earlier works (Kajander *et al*., 2000; Kleiger and Eisenberg, 2002; Hoffmann *et al*., 2007). Sequence variability in the substrate-binding regions could have an effect on the potential Van der Waal's surfaces of the binding pocket and H-bonding frequency with aminoacyl side chains (Jordan and Goldstein, 1995). Accordingly, a point mutation in the binding sites destabilizes the secondary structural environments, particularly H-bonding and solvent accessibility (Sarabojia *et al*., 2005; Shaw *et al*., 2006).

CONCLUSION

Enzyme engineering is a powerful technique to modify the biological function of this enzyme for industrial interest. Evolution-guided optimization of the host cellular metabolic machinery is required for engineering biosynthetic pathway included this enzyme for glucaric acid production (Raman *et al*., 2014; Murugan *et al*., 2019). However, it is time-consuming, cost-effective experiments and limited to reveal structure-function-evolution integrity. Evolutionary forces are taking an influential role in the molecular diversity of NDE/Dsubfamily and also imposed on the sequence-structure-function link of NDE/Dsubfamily. The present approach has gained importance for screening, selection,and evaluation of Atu Udh mutants based on their evolutionary blueprints. Our approach is being a great interest to discover site-directed mutagenesis targets for rational designing and engineering of AtuUdh, in accordance with the previous works on *Clostridium botulinum* C2 and C3 toxins (Chellapandi *et al*., 2013; Prathiviraj *et al*., 2016; Prisilla *et al*., 2017; Chellapandi *et al*., 2018; Chellapandi *et al*., 2019; Murugan *et al*., 2019). Nevertheless, substrate-imprinted docking is a recent attempt to computationally evaluate the substrate specificity and molecular recognition at the binding cleft of AtuUdh. It provides a better understanding of the structural and functional aspects of selected mutant proteins for gearing site-directed mutagenesis experiments. Moreover, a high-level expression and crystallographic structural studies will provide an advance of using this enzyme to meet the requirements of the biotechnology industry for the production of D-glucaric acid.

REFERENCES

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. 2002. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science ISBN-10: 0-8153-3218-1, 0-8153-4072-9.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z. and Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 25 : 3389-3402.

Arenas, M., Dos Santos, H.G., Posada, D. and Bastolla, U. 2013. Protein evolution along phylogenetic histories under structurally constrained substitution models. *Bioinformatics.* 29 : 3020-3028.

Bespalov, V.G. and Aleksandrov, V.A. 2012. Anticarcinogenic effect of potassium salts of glucaric and glucuronic acid in induced models of cervical and esophageal tumors. *Voprosyonkologii.* 58: 537-540.

Chakraborty, A. and Chakrabarti, S. 2015. A survey on

prediction of specificity-determining sites in proteins. *Briefings in Bioinformatics.* 16 : 71-88.

Chellapandi, P., Prathiviraj, R. and Prisilla, A. 2018. Molecular evolution and functional divergence of IspD homologs in malarial parasites. *Infection, Genetics and Evolution*. 65 : 340-349.

Chellapandi, P., Prathiviraj, R. and Prisilla, A. 2019. Deciphering structure, function and mechanism of Plasmodium IspD homologs from their evolutionary imprints. *Journal of Computer-Aided Molecular Design*. 33 : 419-436.

Chellapandi, P., Shree, S.S. and Bharathi, M. 2013. Phylogenetic approach for inferring the origin and functional evolution of bacterial ADP-ribosylation superfamily. *Protein and Peptide Letters*. 20 : 1054-1065.

Choi, S.S. and Hannehall, S. 2013. Three independent determinants of protein evolutionary rate. *Journal of Molecular Evolution*. 76 : 98-111.

Engelhardt, B.E., Jordan, M.I., Srouji, J.R. and Brenner, S.E. 2011. Genome-scale phylogenetic function annotation of large and diverse protein families. *Genome Research.* 21 : 1969-1980.

Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C. and Bryant, S.H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research.* 43 : D222-D226.

Groninger-Poe, F.P., Bouvier, J.T., Vetting, M.W., Kalyanaraman, C., Kumar, R.,Almo, S.C.,Jacobson, M.P. and Gerlt, J.A. 2014. Evolution of enzymatic activities in the enolase superfamily: galactaratedehydratase III from *Agrobacterium tumefaciens* C58. *Biochemistry*. 53 : 4192-4203.

Gu, X. and Vander Velden, K. 2002. DIVERGE phylogeny-based analysis for the functional-structural divergence of a protein family. *Bioinformatics.* 18 : 500-501.

Hoffmann, F., Sotriffer, C., Evers, A., Xiong, G. and Maser, E. 2007. Understanding oligomerization in 3-alpha-hydroxysteroid dehydrogenase/carbonyl reductase from *Comamonastestosteroni*: an *in silico* approach and evidence for an active protein. *Journal of Biotechnology.* 129 : 131-139.

Hubbard, B.K., Koch, M., Palmer, D.R., Babbitt, P.C. and Gerlt, J.A. 1998. Evolution of enzymatic activities in the enolase superfamily: characterization of the (D)-glucarate/galactarate catabolic pathway in *Escherichia coli. Biochemistry.* 37 : 14369-14375.

Huson, D.H. and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution.* 23 : 254-267.

Jeong, J.S. and Kim, D. 2012. Reliable and robust detection of coevolving protein residues.*Protein Engineering, Design and Selection*. 25 : 705-713.

Jordan, E.T. and Goldstein, I.J. 1995. Site-directed mutagenesis studies on the lima bean lectin. Altered carbohydrate-binding specificities result from single amino acid substitutions. *European Journal of Biochemistry.* 230 : 958-964.

Kajander, T., Kahn, P.C., Passila, S.H., Cohen, D.C., Lehtiö, L. and Adolfsen, 2000. Buried charged surface in proteins, *Structures.* 8: 1203-1214.

Kavanagh, K.L., Jörnvall, H., Persson, B. and Oppermann, U. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cellular and Molecular Life Sciences.* 65: 3895-3906.

Kleiger, G. and Eisenberg, D. 2002. GXXXG and GXXXA motifs stabilize FAD and NAD (P)-binding Rossmann folds through C(alpha)-H...O hydrogen bonds and van der Waals interactions.*Journal of Molecular Biology.* 323 : 69-76.

Lee, B.C. and Kim, D. 2009. A new method for revealing correlated mutations under the structural and functional constraints in proteins. *Bioinformatics.* 25: 2506-2513.

Marks, D.S., Hopf, T.A. and Sander, C. 2012. Protein structure prediction from sequence variation. *Nature Biotechnology.* 30 : 1072-1080.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D. and Lefeuvre, P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 26: 2462-2463.

Martinez Cuesta, S., Furnham, N., Rahman, S.A., Sillitoe, I. and Thornton, J.M. 2014. The evolution of enzyme function in the isomerases. *Current Opinion in Structural Biology.* 26 : 121-130.

Matsubara, T., Hamada, S., Wakabayashi, A.and Kishida, M. 2016. Fermentative production of l-galactonate by using recombinant *Saccharomyces cerevisiae* containing the endogenous galacturonatereducta;e gene from *Cryptococcus diffluens. Journal of Bioscience and Bioengineering.* 122 : 639-644.

Mizuguchi, K., Deane, C.M., Blundell, T.L., Johnson, M.S. and Overington, J.P. 1998. JOY: protein sequence-structure representation and analysis. *Bioinformatics.* 14: 617-623.

Moon, T.S., Yoon, S.H., Lanza, A.M., Roy-Mayhew, J.D. and Prather, K.L. 2009. Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli. Applied and Environmental Microbiology.* 75: 589-595.

Murugan, A., Prathiviraj, R., Mothay, D. and Chellapandi, P. 2019. Substrate-imprinted docking of *Agrobacterium tumefaciens*uronate dehydrogenase for increased substrate selectivity. *International Journal of Biological Macromolecules*. 140 : 1214-1225.

Parkkinen, T., Boer, H., Jänis, J., and berg, M., Penttilä, M. and Koivula, A. and Rouvinen, J. 2011. Crystal structure of uronate dehydrogenase from *Agrobacterium tumefaciens. The Journal of Biological Chemistry.* 286 : 27294-27300.

Penning, T.M., Bennett, M.J., Smith-Hoog, S., Schlegel, B.P., Jez, J.M. and Lewis, M. 1997. Structure and function of 3 alpha-hydroxysteroiddehydrogenase. *Steroids.* 62 : 101-111.

Pick, A., Schmid, J. and Sieber, V. 2015. Characterization

of uronate dehydrogenases catalysing the initial step in an oxidative pathway. *Microbial Biotechnology.* 8: 633-643.

Pond, S.L., Frost, S.D. and Muse,S.V. 2005.HyPhy: hypothesis testing using phylogenies.*Bioinformatics.* 21: 676-679.

Prathiviraj, R., Prisilla, A. and Chellapandi, P. 2015. Structure-function discrepancy in *Clostridium botulinum* C3 toxin for its rational prioritization as a subunit vaccine. *Journal of Biomolecular Structure and Dynamics.* 34 : 1317-1329.

Prisilla, A., Prathiviraj, R. and Chellapandi, P. 2017. Molecular evolutionary constraints that determine the avirulence state of *Clostridium botulinum* C2 toxin. *Journal of Molecular Evolution*. 84 : 174-186.

Raman, S., Rogers, J.K., Taylor, N.D. and Church, G.M. 2014. Evolution-guided optimization of biosynthetic pathways. *Proc Natl Acad Sci USA.* 111 : 17803-17808.

Rat, L., Veuille, M. and Lepesant, J.A. 1991. Drosophila fat body protein P6 and alcohol dehydrogenase are derived from a common ancestral protein. *Journal of Molecular Evolution*. 33 : 194-203.

Reizman, I.M., Stenger, A.R., Reisch, C.R., Gupta, A., Connors, N.C. and Prather, K.L. 2015. Improvement of glucaric acid production in *E. coli* via dynamic control of metabolic fluxes.*Metabolic Engineering Communications.* 2 : 109-116.

Sandler, I., Abu-Qarn, M. and Aharoni, A. 2013. Protein co-evolution: how do we combine bioinformatics and experimental approaches? *Molecular Bio Systems.* 9: 175-181.

Sarabojia, K., Gromihab, M.M. and Ponnuswamy, M.N. 2005. Relative importance of secondary structure and solvent accessibility to the stability of protein mutants.A case study with amino acid properties and energetics on T4 and human lysozymes. *Computational Biology and Chemistry.* 29 : 25-35.

Shaw, B., Durazo, A., Nersissian, A.M., Whitelegge, J.P., Faull, K.F. and Valentine, J.S. 2006. Local unfolding in a destabilized, pathogenic variant of superoxide dismutase 1 observed with H/D exchange and mass spectrometry. *The Journal of Biological Chemistry.* 281: 18167-18176.

Shiue, E. and Prather, K.L. 2014. Improving D-glucaric acid production from myo-inositol in *E. coli* by increasing MIOX stability and myo-inositol transport. *Metabolic Engineering*. 22: 22-31.

Smith, T.N., Hash, K., Davey, C.L., Mills, H., Williams, H. and Kiely, D.E. 2012. Modifications in the nitric acid oxidation of D-glucose. *Carbohydrate Research.* 350 : 6-13.

Sola-Carvajal, A., García-García, M.I.,García-Carmona, F. and Sánchez-Ferrer, A. 2012. Insights into the evolution of sorbitol metabolism: phylogenetic analysis of SDR196C family. *BMC Evolutionary Biology.* 12 : 147.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 123 : 585-595.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution.* 28: 2731-2739.

Tarrío, R., Ayala, F.J. and Rodríguez-Trelles, F. 2011. The Vein Patterning 1 (VEP1) gene family laterally spread through an ecological network. *PLoS One*. 6: e22279.

Thomas, J.L., Mason, J.I., Brandt, S., Spencer, B.R. and Norris, W. 2002. Structure/function relationships responsible for the kinetic differences between human type 1 and type 2 3beta-hydroxysteroid dehydrogenase and for the catalysis of the type 1 activity. *The Journal of Biological Chemistry.* 277 : 42795-42801.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research.* 25 : 4876-4882.

Udaya Prakash, N.A., Jayanthi, M., Sabarinathan, R. ,Kangueane, P., Mathew, L. and Sekar, K. 2010. Evolution, homology conservation, and identification of unique sequence signatures in GH19 family chitinases. *Journal of Molecular Evolution.* 70: 466-478.

Wagschal, K., Jordan, D.B., Lee, C.C., Younger, A., Braker, J.D. and Chan, V.J. 2015. Biochemical characterization of uronate dehydrogenases from three Pseudomonads, *Chromohalobacter salixigens*, and *Polaromonas naphthalenivorans. Enzyme and Microbial Technology.* 69: 62-68.

Walaszek, Z., Szemraj, J., Hanausek, M., Adams, A.K. and Sherman, U. 1996. D-Glucaric acid content of various fruits and vegetables and cholesterol-lowering effects of dietary D-glucarate in the rat. *Nutrition Research*. 16 : 673-681.

Werpy, T. and Petersen, G. 2004. Top value added chemicals from biomass, Volume I: Results of screening for potential candidates from sugars and synthesis gas. N.R.E.L. (NREL) and P.N.N.L. (PNNL) (ed.) US Department of Energy (US).

Williams, B.P., Johnston, I.G., Covshoff, S. and Hibberd, J.M. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *Elife*. 2: e00961.

Wu, H. and Cheng, Y.S. 2014. Combining secondary-structure and protein solvent-accessibility predictions in methionine substitution for anomalous dispersion. *Acta Crystallographica Section F Structural Biology.* 70 : 378-383.

Yoon, S.H., Moon, T.S., Iranpour, P.Lanza, A.M. and Prather, K.J. 2009. Cloning and characterization of uronate dehydrogenases from two pseudomonads and *Agrobacterium tumefaciens* strain C58. *Journal of Bacteriology.* 191: 1565-1573.

Zajic, J.E. 1959. Hexuronic dehydrogenase of *Agrobacterium tumefaciens. Journal of Bacteriology.* 78: 734-735.

Zhang, J. and Yang, J.R. 2015. Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics.* 16: 409-420.