A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability

Tabinda Shakeel$^{1,3}$, Mayank Gupta$^{1,3}$, Zia Fatma$^{1,3}$, Rakesh Kumar$^2$, Raubins Kumar$^1$, Rahul Singh$^{1,3}$, Medha Sharma$^{1,3}$, Dhananjay Jade$^2$, Dinesh Gupta$^2$, Tasneem Fatma$^4$ and Syed Shams Yazdani$^{1,3}$*

From the $^1$Microbial Engineering Group, International Centre for Genetic Engineering and Biotechnology, New Delhi 110067; $^2$Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology, New Delhi 110067; $^3$DBT-ICGEB Centre for Advanced Bioenergy Research, International Centre for Genetic Engineering and Biotechnology, New Delhi 110067; $^4$Department of Biosciences, Jamia Millia Islamia, New Delhi 110025

**Running title:** Protein engineering for improved thermostability of ADO

*To whom correspondence should be addressed: Syed Shams Yazdani, Microbial Engineering Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110067, India. Email – shams@icgeb.res.in; Phone - +91 11 26742357; Fax - +91 11 26742316.

**Keywords:** protein engineering, structure-function, enzyme mechanism, enzyme mutation, molecular modelling, molecular dynamics, aldehyde deformylating oxygenase, biofuel, long-chain alkane, thermostable enzyme.

**ABSTRACT**

Aldehyde-deformylating oxygenase (ADO) is an essential enzyme for production of long-chain alkanes as drop-in biofuels, which are compatible with existing fuel systems. The most active ADOs are present in mesophilic cyanobacteria, especially *Nostoc punctiforme*. Given the potential applications of thermostable enzymes in biorefineries, here we generated a thermostable (Cts)-ADO based on a consensus of ADO sequences from several thermophilic cyanobacterial strains. Using an *in silico* design pipeline and a metagenome library containing 41 hot-spring microbial communities, we created Cts-ADO. Cts-ADO displayed a 3.8-fold increase in pentadecane production on raising the temperature from 30°C to 42°C, whereas ADO from *N. punctiforme* (Np-ADO) exhibited a 1.7-fold decline. 3D structure modeling and molecular dynamics simulations of Cts- and Np-ADO at different temperatures revealed differences between the two enzymes in residues clustered on exposed loops of these variants, which affected the conformation of helices involved in forming the ADO catalytic core. In Cts-ADO, this conformational change promoted ligand binding to its preferred iron, Fe2, in the di-iron cluster at higher temperature, but the reverse was observed in Np-ADO. Detailed mapping of residues conferring Cts-ADO thermostability identified four amino acids, which we substituted individually and together in Np-ADO. Among these substitution variants, A161E was remarkably similar to Cts-ADO in terms of activity optima, kinetic parameters, and structure at higher temperature. A161E was located in loop L6, which connects helices H5 and H6, and supported ligand binding to Fe2 at higher temperatures, thereby promoting optimal activity at these temperatures and explaining the increased thermostability of Cts-ADO.

Employing engineered microbes for the production of fuels provides an attractive alternative for mitigating the dependence on currently available fossil fuels (1,2). Research on alk(a/e)ne production gained tremendous attention recently due to its high energy density and compatibility with the existing fuel infrastructure, which makes it an ideal candidate for drop-in biofuels (3-7). Its production has been reported in
various microorganisms including algae, yeasts, fungi and bacteria (8,9,11), with the most consistent reports being from cyanobacteria and from natural habitats dominated primarily by cyanobacteria (12-15). Recently, two genes from cyanobacteria encoding an acyl-ACP reductase (AAR) and an aldehyde dehydrogenase (ADO) were identified which were proven to be responsible for the production of fatty alk(a/e)nes in cyanobacteria (3). This pathway involves the reduction of the fatty acyl-ACP into their corresponding aldehydes by the first enzyme, i.e., AAR, and the subsequent conversion of these fatty aldehydes into the respective one carbon less (C_{n-1}) alk(a/e)nes by ADO (3). There have been reported of enzymes functionally similar to the AAR, like the carboxylic acid reductase (CAR) from *Mycobacterium marinum* which was found to be capable of converting a wide range of aliphatic fatty acids (C_{6-C_{18}}) into their corresponding aldehydes (16). Another alternative route for the production of fatty aldehydes from free fatty acids was also explored by artificially assembling a metabolic pathway consisting of FA reductase (FAR) complex encoded by the genes *luxC, luxE*, and *luxD* (17,18) from the bioluminescent bacterium *Photobacterium luminescens* (19). The fatty aldehydes produced from both these routes (16,19) can be coupled to the ADOs from cyanobacterial sources for their subsequent conversion to alkanes, a very unique and chemically difficult process (20).

Due to the irreplaceable mechanism of action catalysed by ADO, this enzyme has gained tremendous attention. Several reports on the functionality of the ADO enzyme have elicited various conclusions (21). Crystal structures of both wild type and single point mutant ADOs from *Prochlorococcus marinus* MIT9313 have been reported by Khara *et al.* in 2013 (22). It has been proposed that ADOs belong to the superfamily of ferritin-like di-iron proteins and encompass a di-iron centre (23,24). The structures indicate that ADO adopts an α-helical folding with two iron atoms that are coordinated by two histidine and four glutamate residues acting as metal ligands (3,22,25,26). The ADO catalysed reactions invariably involve molecular oxygen and require an external reducing system for providing electrons which can be either chemical or biological (24,27-29). The chemical reducing system comprises of phenazine methosulfate and NADH (24) while the biological reducing system involves ferredoxin (Fd), ferredoxin reductase (FR), and NADPH (3,27) to reduce the diferric form of the enzyme into its active diferroso form at the beginning of each turn over. It has been proposed that the iron-peroxo species, generated by the entry of the oxygen, attacks the aldehyde substrate resulting in the formation of hemiacetal and causing the scission of the C1-C2 bond (28,30). Earlier it was hypothesized that the C1 derived product is carbon monoxide which was later proven to be formate (3,27,28,31). It was further observed that ADO is inhibited due to the formation of hydrogen peroxide (H_{2}O_{2}) and that supplementing the depleted assays with catalase could restore the ADO activity (32). Zhang *et al.* later showed that in *vitro* reconstituted ADO with an endogenous reducing system from cyanobacteria exhibited greater activity and resulted in a decreased formation of H_{2}O_{2} (33). Moreover, fusing ADO with cognate FR and Fd improved its activity by almost 3-fold (34). Mechanistic studies have demonstrated that a radical intermediate is involved in the ADO-catalyzed reaction, and a possible catalytic process has been proposed based on the crystal structures of ADO from *Synechococcus elongatus* PCC7942 strain (35-38). Recently ADO was engineered to improve specificity towards short to medium chain length aldehydes for the production of short chain alkanes(22). The role of cysteine residues has been investigated in ADO from *Nostoc punctiforme* PCC 73102 and it was shown that Cys71 is involved in maintaining the activity, structure, and stability of ADO (39). Certain amino acid residues present in the vicinity of the substrate channel were identified for successfully changing the substrate chain length selectivity of ADO (21). The ADO belongs to the non-heme dinuclear iron oxygenase family of enzymes that includes methane monoxygenase, type I ribonucleotide reductase and ferritin (3,35,40-42). The bound substrate analogs such as fatty acids or fatty alcohols or fatty aldehydes have been observed to occur in the vicinity of the di-iron centre in these structures (3,22,35,40). The active site of the enzyme is housed within an antiparallel 4-helix bundle in which the two iron atoms are each coordinated by a histidine and two glutamic acid residues from the protein. Few important amino
acids close to the di-iron centre have been identified which exerted beneficial effects on the ADO activity (43). However, so far there are no reports regarding studies on thermostability of ADO.

Thermostable enzymes having longer operational stability at higher temperature offer robust catalyst alternatives capable of withstanding the comparatively stringent environments of industrial processing (44). Given the biosynthetic potential of ADOs in industries for alkane production, we designed a robust metagenome pipeline to generate a synthetic thermostable consensus sequence and tested its activity in vivo and in vitro in comparison to an efficient mesophilic ADO enzyme adapted from earlier reports (3) (Figure 1). This was followed by a comparative computational evaluation of the mesophilic ADO and its thermophilic counterpart to identify amino acid residues responsible for imparting thermostability. Based on this, four single-mutant variants and their combination were generated and assessed for their impact on catalytic activity and thermostability of mesophilic ADO from *N. punctiforme*. The thermostable ADO generated in this study would ensure enzyme functionality and alkane formation for longer duration under in vivo conditions at higher temperatures.

**RESULTS**

Design of in silico pipeline for generation of consensus thermostable ADO. In the quest to identify thermostable ADO enzymes functional at higher temperature, we designed a pipeline to construct a consensus thermostable ADO using hot spring metagenome libraries of cyanobacterial strains (Figure 1). Metagenome libraries of 41 hot spring microbial communities were screened for the presence of *ado* gene homologs using the *ado* gene from *Synechococcus elongatus* PCC7942 (se-ado) as the query sequence. A total of 80 *ado* gene homologs were retrieved using this analysis. Out of these, only 45 *ado* gene homologs showed significant BLASTN alignment (45) with the *se-ado* gene, and thus were selected for further analysis. Translation of the sequences and consideration of only one sequence from every genus led to selection of 10 unique *ado* gene homologs. The BLAST results of these homologs indicated the closest hits being cyanobacterial species belonging to diverse genus (Supplementary Table 1). These homologs were then used for the generation of a consensus thermostable ADO protein (Cts-ADO) sequence using the Jalview software (Supplementary Figure 1A). The gene encoding this sequence (*cts-ado*) was synthesized following codon optimization (Supplementary Figure 1B) for optimal expression in *E. coli* and further expressed in *E. coli* to analyse its ability to produce alkanes in vivo.

In order to test the activity of the recombinant Cts-ADO under in vivo conditions at different temperatures, the cells were grown at 30 °C, 37 °C and 42 °C in the culture medium supplemented with 100 mg/L exogenous aldehyde hexadecanal (to serve as substrate for the ADO enzyme), and the subsequent alkane formed was analysed in the extracellular medium using GC-MS/MS. A synthetic gene encoding an efficient ADO from a mesophilic cyanobacterial species, *Nostoc punctiforme* (*np-ado*) (3), was taken as control. It was observed that pentadecane level of the *E. coli* strain expressing *np-ado* was significantly higher than the one expressing *cts-ado* at 30 °C (Figure 2A). This could be attributed to a higher in vivo soluble expression level of *np-ado* as compared to *cts-ado*, as monitored by Western blot analysis (Figure 2B). However, while *np-ado* exhibited 1.7-fold decline in pentadecane production by raising growth temperature from 30 °C to 42 °C, the *cts-ado* showed a remarkable 3.8-fold increase in pentadecane production under similar condition. We also noticed that while the soluble expression of *np-ado* declined at 37 °C, the soluble expression of *cts-ado* improved at this temperature as compared to that at 30 °C (Figure 2). Interestingly, although soluble expression of ADO at 42 °C did not change as compared to 37 °C in both the *E. coli* strains (Figure 2B), the alkane titer increased only in the case of *cts-ado*. This finding indicated higher specific activity of the ADO encoded by *cts-ado* at higher temperature.

Nevertheless, in addition to the impact of temperature on the alkane titer, the soluble expression of ADOs under in vivo condition also had a major impact on the alkane titer (Figure 2A and 2B). We thus decided to monitor the specific activities of these ADOs under in vitro condition. This would also circumvent the limitation of testing the activity of the mutants beyond 42 °C...
under *in vivo* conditions in *E. coli*. Each of these proteins was recombinantly expressed in *E. coli*, purified by metal affinity chromatography (Supplementary Figure 2) and *in vitro* assay was conducted as described in the Experimental Procedures section. From the *in vitro* assay it was inferred that Cts-ADO had the maximum activity at 320 K (~47 °C), a 2.5-fold higher activity than at 300 K (~27 °C) (Figure 3, Supplementary Table 2). It also demonstrated significant activities at much higher temperatures of 330 K (~57 °C) and 333 K (~60 °C), which were close to that of 300 K (~27 °C). On contrary, Np-ADO showed minor increase in activity from 300 K to 310 K and then continuous fall in activity on further rise in temperature, eventually leading to no detectable activity at 330 K (Figure 3, Supplementary Table 2).

**Computational assessment of operational stability of ADOS at different temperatures.** Based on the observed *in vivo* and *in vitro* functional variations between the mesophilic ADO and its thermophilic counterpart, we were motivated to understand the molecular events which could possibly contribute to the protein stability and activity at higher temperature using computational modelling and simulation methods (46-51). Experimentally determined crystal structures of ADO are available at the PDB database (35,52,53) with different bound substrate analogs such as fatty acids, fatty alcohols or fatty aldehydes (3,22,35,40). From the available protein structures in the PDB database, 20C5 and 4TW3 from *Prochlorococcus marinus* MIT9313 and 4RCS from *Synechococcus elongatus* PCC7942 were selected to serve as templates and the amino acid sequences corresponding to these PDB entries were retrieved and subsequently aligned to the target protein sequences with ClustalW to prepare for homology modelling. From the 20 models generated through the MODELLER 9.15 (54), the best model (Figure 4) was selected based on the least DOPE score and its corresponding Φ and Ψ distributions of the non-glycine, non-proline residues are summarized on the Ramachandran plot (Supplementary Figure 3). As expected from the previous reports (3,22,35,40), the protein adopts an α helical folding with two iron atoms (namely Fe1 and Fe2) being coordinated by histidine residues and carboxylate ligands. The modelled structures of both Np-ADO and Cts-ADO appeared quite similar and consisted of 8 α helices connected by loops. Helices H1, H2, H4 and H5 formed a compact structure and housed two iron atoms forming the catalytic core of the protein. The iron atoms are coordinated by six conserved amino acid residues (Glu32 from helix H1, Glu115 from helix H4, Glu60 & His63 from helix H2, and Glu144 & His147 from helix H5) which act as metal ligands. Some notable differences in the loop L2 connecting helices H1 and H2 and loop L6 connecting helices H5 and H6 were visible between the two structures (Figure 4), which needed to be explored further.

The modelled structures were used as starting points for an explicit 100 ns molecular dynamics simulation to observe significant changes in the conformational trends of the different proteins with temperature. MD simulations performed at different temperatures could potentially provide valuable information relating to protein unfolding. For each simulation, the root mean square deviations (RMSD) of the protein backbones were calculated and the results are plotted in Figure 5. Our results indicated that both the mesophilic and thermophilic proteins were mostly stable between 300 K - 320 K (~27 °C – 47 °C) with their average RMSDs in general up to 2-2.5 Å units under 100 ns timescale simulations. However, increasing the temperature to 330 K (~57 °C) resulted in a significant increase in the RMSD fluctuations of the mesophilic ADO while the thermophilic ADO RMSD continued to show an average deviation of less than 2.5 Å units from the starting point thus indicating the relative stability of the thermophilic ADO. Further increasing the temperature to 340 K (~67 °C) resulted in complete loss of stability in the mesophilic ADO while the thermophilic ADO still remained stable with its RMSD values averaged between 1-3 Å units under 100 ns simulations.

It was postulated earlier that rigidity is a prerequisite for improved thermostability based on comparisons between mesophilic and thermophilic proteins (55,56). The large thermal fluctuations within flexible regions potentially expose the hydrophobic core of a protein to water penetration, triggering protein unfolding (57). In order to identify thermally sensitive regions of ADO, the secondary structures of Np-ADO and Cts-ADO retrieved from the cluster representatives of the
Protein engineering for improved thermostability of ADO

largest clusters at different temperatures were plotted and analysed (Figure 6). Structural analysis revealed ten distinct changes spread across the entire protein structure of the Np-ADO when the temperature was increased from 300 K to 320 K and to 340 K (Figure 6A, Panel 300K, 320 K and 340 K with changes marked in red). Majority of these changes involved elongation or shortening of loops or deletion and formation of additional loops. Changes were also observed in the structure of Cts-ADO when temperature was increased from 300 K to 320 K, and then further to 340 K, however, those changes were far less intense than that of Np-ADO under similar condition (Figure 6A and 6B). A unique pattern was observed in the loops L2 and L6 in both Np-ADO and Cts-ADO on increasing the temperature. While these loops appeared to get distorted constantly in Np-ADO when the temperature was being raised from 300 K onward, they got structured in Cts-ADO when the temperature was raised from 300 K to 320 K and to a certain extent to 340 K as well. Changes in loop L6 which connects the helices H5 and H6 may impact flexibility of the helix H5 which has been shown to be an important parameter for the optimal activity of ADO (35). The loop L2 that connects the helices H1 and H2 was also found to be important for the enzyme activity as these helices are directly involved in aiding the substrate binding (35). These differences in the behaviour of structures of Np-ADO and Cts-ADO upon increase in temperature may well explain the reasons for increase in catalytic activity of Cts-ADO at higher temperature under both in vivo and in vitro conditions while reverse in case of Np-ADO.

We further wanted to assess the impact of these dynamic variations on the positioning of active site residues and their subsequent interactions with iron atoms (Fe1 and Fe2) and substrate ligand upon increasing the temperature. These changing interactions were visualised using LigPlot+ v.2.1 (Figure 7) (58). The differences in the interactions of the substrate ligand and the residues lining the substrate channel arising due to changes in the confirmation of the helices forming the binding pocket have been further summated in Figure 8. It has been shown in previous studies that the ligand approaches the di-iron cluster from opposite side of the H148 residue with its head group binding directly to the Fe2, which is the preferred iron for substrate binding (21,35). We also observed similar interactions in case of the Np-ADO at 300 K where ligand appeared to be bound to the Fe2 (Figure 7 – Np-ADO, 300 K; Figure 8 – Np-ADO, 300 K). The distance between the head group of the ligand and the Fe2 was 2.1 Å as against 4.0 Å between the ligand and the Fe1 (Supplementary Figure 4A, 300 K). On increasing the temperature to 320 K and 340 K, stearic hindrance from a tyrosine residue Y123 present on helix H4 appeared to alter the substrate binding, leading to a shift in substrate binding towards Fe1 (Figure 8 – Np-ADO, 320 K; Supplementary Figure 4A, 320 K and 340 K). This could lead to loss in activity since Fe2 is the preferred iron for substrate binding. Interestingly, a reverse trend was observed in case of Cts-ADO. Here the head group of the ligand was interacting with the Fe1 at 300 K, with distance between the Fe1 and the substrate ligand was only 2.1 Å against 3.9 Å with the Fe2 (Figure 7 - Cts-ADO, 300 K; Figure 8 – Cts-ADO, 300 K; and Supplementary Figure 4B). On analysing the active site residues, it was found that at 300 K the residue Tyr40 (located on helix H1 and not part of active site residues) was also interacting with the Fe2 thus possibly hindering the binding of the substrate (Figure 7 – Cts-ADO, 300 K). However, on increasing the temperature to 320 K and 340 K, there was a change in positioning of the substrate which appeared to shift towards Fe2 from 3.9 Å to 2.2 Å (Figure 7 - Cts-ADO; Figure 8 – Cts-ADO; and Supplementary Figure 4B, 320 K and 340 K). These phenomena explain why activity of Cts-ADO increased with increase in temperature while activity of Np-ADO decreased under similar condition.

We further analysed the change in the surface hydrophobicity of both the Np-ADO and Cts ADO on different temperatures using the preset option in the Chimera software which determines the hydrophobicity employing the Kyte and Doolittle scale (Supplementary Figure 5) (59). It was noted that hydrophobic residues increased on the surface of Np-ADO when the temperature was raised from 300 K to 320K and further to 340 K, indicating unfolding of the protein. On the other hand, the overall hydrophilic nature of the thermostable Cts-ADO remained consistent throughout the increase in the temperature.
(Supplementary Figure 5). This phenomenon also provides rational for our observation of in vivo study (Figure 2B) where the solubility of Np-ADO decreased at higher temperature while it increased in the case of Cts-ADO.

Identification and functional validation of amino acid residues conferring thermostability to ADO. In order to identify amino acid residues conferring thermostability to ADO, the regions undergoing fluctuations with increase in temperature in the mesophilic ADO but not in the thermophilic ADO were noted (Figure 6). The protein sequences of Cts-ADO and Np-ADO were further aligned with ADO protein sequence of a thermophile Thermosynechococcus elongatus and a mesophile Synechococcus elongatus PCC7942 ADO (Supplementary Figure 6). The amino acid residues present in the regions corresponding to high fluctuations were analysed and the residues which were found to be not conserved between the mesophilic and thermophilic ADOs in these regions were noted. This predicted 4 residues, i.e., histidine (H), aspartate (D), alanine (A) and threonine (T), present at 51st, 132nd, 161st and 191st position of Np-ADO, which have been mutated in Cts-ADO to lysine (H51K), proline (D132P), glutamate (A161E) and isoleucine (T191I), respectively. Structural analysis revealed that all these residues were present on surface exposed regions of Np-ADO and would possibly not interfere with the active site of the protein directly (Supplementary Figure 7).

The four naturally occurring mutations, i.e., H51K, D132P, A161E and T191I, were analysed for their role in imparting thermostability by introducing these single amino acid substitutions in the mesophilic protein, Np-ADO, using site directed mutagenesis. All mutants were successfully overexpressed in E.coli DH5α to estimate the in vivo activity of the enzymes at different temperatures (Figure 2), as described in the earlier section. It was observed that while the wildtype showed a decline in pentadecane formation from 3.25 mg/g DCW at 30 °C to 1.96 mg/g DCW at 42 °C, all the four mutants recorded improvement in activity at higher temperatures. While improvement was prominent from 30 °C to 37 °C in all mutants, A161E and H51K mutants seemed to have further improved activity at 42 °C.

The activities of all the mutants were further analysed under in vitro condition after purifying them by metal affinity chromatography (Supplementary Figure 2). Consistent with the in vivo results it was observed that the A161E and H51K mutant exhibited a 2- and 1.6-fold increase in the pentadecane production at 320 K (~47 °C), respectively, as compared to their activities at 300 K (~27 °C) (Figure 3, Supplementary Table 2). These two mutants shared temperature optima for their specific activity with Cts-ADO at 320 K. The other two mutants and wild type mesophilic Np-ADO behaved similarly and showed small improvement in activity at 310 K (~37 °C ) followed by decline at 320 K and beyond (Figure 3, Supplementary Table 2). Thus, these results indicated that mutations at 161st and 51st position were mainly responsible for imparting thermostability to Cts-ADO. Replacement of alanine (A) at 161st position with glutamate (E) indicated that substitution of a neutral amino acid with an acidic one helped in enhancing the thermostability. Replacement of histidine with lysine at 51st position led to substitution of one basic amino acid with another. However, different pKa values for the side chains of these two amino acids (lysine pKa ~ 10.53 and histidine pKa ~6.0) and the absence of an imidazole ring in lysine possibly confers a different property to the surface and thus may be responsible for the thermostability conferred to Cts-ADO.

Evaluation of kinetic parameters of ADO variants at different temperatures. The steady state kinetic studies of Np-ADO, Cts-ADO, and A161E mutant that resulted in highest improvement in specific activity of Np-ADO at higher temperature were performed at different temperature. We also constructed for evaluating kinetic parameters the double substitution (H51K-A161E) mutant of Np-ADO combining the two amino acid change leading to the most favourable thermostability experimentally, as well as the quadruple substitution (H51K-D132P-A161E-T191I) mutant of Np-ADO combining all the four amino acids predicted to confer thermostability in silico. As expected, the kinetic efficiency (kcat/Km) of Np-ADO declined by 1.7-fold upon increasing the temperature from 27 °C to 47 °C, while it increased by 1.6-fold in case of Cts-ADO under similar condition (Table 1, Supplementary Figure 8). Also, while significant activity was observed for Cts-ADO at 60 °C, no activity was observed in case of Np-ADO. Interestingly, A161E mutant of
Np-ADO showed similar kinetic efficiency as Cts-ADO at 27 °C and 47 °C and higher at 60 °C, which could be attributed to lower affinity of Cts-ADO at 60 °C. The double mutant of Np-ADO showed further higher kinetic efficiency at 60 °C as compared to mutant A161E, while the quadruple mutant showed higher kinetic efficiency at 27 °C and 47 °C, but lower at 60 °C (Table 1, Supplementary Figure 8). These results indicated that combining all four mutations had no further advantage over the double mutant having A161E and H51K mutation at higher temperature. This also shortlisted two major residues responsible for imparting thermostability to the enzyme, i.e., H51K and A161E. In order to understand how these residues contributed individually to the thermal tolerance, their structural implication needed to be explored further.

Probing the structural significance of the beneficial mutations A161E and H51K in silico

The performance of the ADO mutants under both in vivo and in vitro conditions pointed out two mutations endowing the ADO enzyme with improved thermostability. As a single mutation is unlikely to cause a significant change to the overall stability of a protein, it was important to analyse the changes in the secondary structure caused at the mutation site. For this purpose, wild type Np-ADO was remodelled after introducing the single amino acid substitutions. The positions of the four residues which were mutated have been indicated on the wild type Np-ADO in the Supplementary Figure 9. The two mutations, i.e., A161E and H51K, which were positioned on loop L6 and L2, respectively, and showed improved activity at or above 320 K were taken up for further analysis. 100 ns molecular dynamic simulations were applied to examine the flexibility changes of the variants A161E and H51K. The protein backbone RMSDs were calculated in order to validate the stability of these mutants (Supplementary Figure 10). At 300 K both the mutants displayed similar RMSD values as the wild type. It was observed that though the A161E mutant exhibited a relatively stable conformation up to 330 K with its RMSD values ranging between 1-3 Å units, the RMSD fluctuations at 340 K were stable initially for the first 40 ns beyond which the RMSD fluctuations appeared to increase. It was noted that the A161E mutant exhibited highest stability at 320 K while in case of the mutant H51K the RMSD values appeared stable up to 310 K after which the fluctuations increased. Surprisingly, the structure for H51K again appeared to stabilise at 340 K. The RMSD result for A161E correlated well with the in vitro enzyme assay results (Figure 3) discussed in the previous section where it showed almost 2-fold higher product formation at 320 K (~47 °C). However, the RMSD result for H51K mutant, where it appeared to have the lowest fluctuations from the starting point at 340 K near the 100 ns time scale simulations, was not apparent in the in vitro results where the activity continuously declined beyond 320 K.

In order to understand the variation in the secondary structures with changes in the temperatures, cluster representatives of the largest clusters at different temperatures were analysed for the mutants (Supplementary Figure 11) and compared with those of the Np-ADO and Cts-ADO (Figure 6). The secondary structure analysis revealed that this naturally occurring mutation, A161E, in the thermophile was present at the transitions between helices, on the loop 6 (L6) connecting helices H5 and H6 (Supplementary Figure 9). Also, the site of the second most effective substitution, i.e., H51K, was present at the beginning of the helix H2 where the loop L2 connecting the helices H1 and H2 was present (Supplementary Figure 9). It was observed that incorporation of these single amino acid substitutions in the wild type Np-ADO resulted in significant variations in the overall secondary structures of the mutants (Supplementary Figure 11). As observed in the Figure 6, the shortening of the loop L6 at 340 K in the thermophile was important for the stability of the protein and it appears that in the mutant A161E (Supplementary Figure 11 Panel A, 320 K), the shortening of the loop with the presence of an additional turn promoted the stability while maintaining the active confirmation. In order to understand how these substitutions impacted the substrate binding, the active site interactions were probed using LigPlot+ v.2.1 and the distances between the substrate and the iron atoms were recorded (Supplementary Figure 12). It was observed that the ligand appeared to be bound to the Fe1 in the mutant A161E at 300 K which shifted to Fe2 on increasing the temperature to 320 K. Further increment in the temperature resulted in the
shifting of the ligand back to Fe1 thus probably resulting in low activity. The mutant H51K appeared to show a pattern similar to the wild type Np-ADO at 300 K where the substrate showed a preferred binding to the Fe2 (Supplementary Figure 12B and Figure 8A). While the ligand appeared to shift to Fe1 in the wild type Np-ADO at 320K (Figure 8A), it remained bound to Fe2 in the mutant H51K thereby managing to retain the activity as shown during in vitro study (Figure 3).

Characterization of the thermostable mutants using differential scanning fluorimetry and circular dichroism spectroscopy. Differential scanning fluorimetry employing SYPRO Orange was used to determine the Tm of the wild type Np-ADO, Cts-ADO and the most effective single mutant variant A161E. SYPRO Orange is an environmentally sensitive dye and is known to bind to the hydrophobic patches of the proteins exposed due to the unfolding process and results in a large increase in fluorescence, which is used to monitor the protein unfolding transition (60). In our studies (Table 2 and Figure 9), we observed that the Tm of the thermostable mutant A161E showed a remarkable improvement by 5 °C over wildtype. The observed Tm values for the thermostable mutant A161E and the thermostable consensus Cts-ADO were found to be 58 °C and 60 °C respectively, which was in concordance with the in silico and the in vitro studies. To further assess the impact of protein engineering on overall secondary structure of the proteins, circular dichroism (CD) spectra of the wild type Np-ADO, Cts-ADO and the mutant A161E were compared. The CD spectra results indicated that all three proteins possessed similar structures at 25 °C, which was consistent with the fact that all of these proteins were active at physiological conditions (Figure 10, Supplementary Table 3). The secondary structure analysis revealed the predominance of alpha-helices (87%) in both wild type and its mutant as well as in the thermostable consensus protein, which was in concordance with the homology modelling results.

DISCUSSION

Given the importance of consolidated bioprocessing where in a single microbial platform an enzyme blend that work synergistically under similar environmental condition for production of industrially relevant fuels and chemicals, it is important to rationally engineer proteins for improved thermal stability which can withstand the harsh conditions of biorefineries while conserving their catalytic efficiency (61). For this purpose it is crucial to identify key structural regulators that govern the maintenance of the protein structure at high temperatures while preserving its overall activity. Considering the importance of long chain alkanes as ideal drop-in biofuel candidates, the engineering of the alkane producing enzyme, ADO is essential for the feasibility of this process.

There have been various reports which involve the improvement of ADO activity by employing protein engineering tools (43). However, no work has been done on improving the thermostability of this enzyme. High temperature aquatic ecosystems, such as hot spring, harbour various thermophilic microorganisms (62), and thus could be an excellent reservoir to look for thermostable enzymes. Since less than 1% microbial community from the environment is cultivable under the laboratory condition, the preparation of metagenome library by culture-independent method provides comprehensive representation of majority of microorganisms present in the environmental sample. We therefore used metagenome libraries of 41 hot spring microbial communities and designed a thermostable ADO using an in silico pipeline constructed specifically for generating a consensus thermostable sequence (Cts-ADO).

An in-depth analysis of Cts-ADO showed its higher thermal tolerance as compared to mesophilic Np-ADO both under in vivo and in vitro condition. Interestingly, Cts-ADO showed much lower in vivo soluble expression level at 30 °C, which subsequently increased at 37 °C and 42 °C, while a reverse trend was observed for Np-ADO (Figure 2B). This suggested that the consensus thermostable ADO has preference for soluble expression at higher temperature. There have been reports suggesting higher soluble expression of recombinant thermostable proteins from thermophiles, however, the exact correlation between thermostability of proteins and their soluble expression remains to be understood (63,64). While we did observe increase in product formation with increase in temperature under in
Protein engineering for improved thermostability of ADO

vivo condition in case of Cts-ADO (Figure 2A), this could be due to concomitant improvement in the solubility. Nevertheless, the specific activity and kinetic efficiency measurement of purified Cts-ADO under in vitro condition indicated a temperature maxima at 320 K (~47 °C), in addition to showing significant activity until 333 K (60 °C). On the contrary, the specific activity and kinetic efficiency of Np-ADO declined at 320 K and no activity was detected beyond 330 K. These observations indeed demonstrated the thermostable nature of Cts-ADO and prompted us to probe the structural significance contributing to its thermal stability.

The experimentally determined structures of ADO from the PDB database served as template to prepare homology models for Cts-ADO and Np-ADO. As expected, the modelled structures for both proteins were similar and consisted of 8 α helices connected by loops, and two iron atoms present at the catalytic core of the protein. Further computational assessment of the modelled structures using molecular dynamics simulation showed distinctly stable regions in the structure of Cts-ADO at higher temperature possibly imparting thermostability to the protein. Examining these regions highlighted the role of residues clustered on the exposed loops in conferring thermal stability to the protein. Previous reports have also indicated the role of loops in enhancing protein thermostability (65). The role of these changes on substrate binding and positioning of the active site residues also needed to be analysed. It has been shown earlier that among the two iron atoms Fe1 and Fe2, it is the preferential binding of the substrate to the Fe2 which triggers the decarboxylation reaction of fatty aldehyde (35). Based on the interactions between the active site residues, substrate ligand and the Fe atoms, it was observed that the substrate ligand showed varied binding with the iron atoms on changing the temperature. In case of mesophilic NP-ADO, the substrate ligand appeared to be bound to Fe2 at 300 K and shifted to Fe1 as the temperature was increased beyond 320 K. Interestingly, however reverse dynamics were observed in Cts-ADO where the ligand was initially bound to Fe1 at 300 K but on raising the temperature to 320 K it resulted in shifting of the ligand interactions to Fe2. Based on these observations and keeping in view the changes occurring in the secondary structures of Np-ADO and Cts-ADO with increase in temperature, the regions appearing to show highest fluctuations in the Np-ADO while maintaining a stable confirmation in the Cts-ADO were probed further for prediction of key residues responsible for imparting thermal stability. This analysis led to identification of four single amino acid substitutions, namely A161E, D132P, H51K, and T191I, all of them lying on the surface of the protein (Supplementary Figure 7). When all these substitutions were incorporated in the Np-ADO individually, A161E showed the highest improvement in thermostability, followed by H51K. Combining these two mutations had favourable impact on the thermostability, but when all the four mutations were incorporated in the Np-ADO, it increased the enzyme efficiency up to 47 °C (320 K) but declined on further increasing the temperature to 60 °C (340 K).

Structural insights indicated that A161 was present at the transitions between helices, on the loop 6 and H51 was present at the beginning of the helix H2 where the loop L2 connecting the helices H1 and H2 was present (Supplementary Figure 9). Flexible sites such as loops connecting different secondary structure elements are potential targets for engineering stability of enzymes. Various studies have highlighted the importance of loops in modulating enzyme catalysis, specificity, stability, and protein-protein interactions (66-68). There are numerous reports which have exhibited the successful alteration of enzyme stability by engineering loop regions (65). Substitution of A161E was particularly found to be effective in improving the thermostability of mesophilic Np-ADO and the interaction of ligand with its Fe2 atom at higher temperature behaved similar to that in case of thermostable Cts-ADO. This was further validated by the melting curve where incorporating A161E substitution in Np-ADO shifted its T_m value close to that of Cts-ADO (Figure 9).

This study reports for the first time construction of thermostable ADO of cyanobacterial origin through structure-guided engineering with net retention of its catalytic activity. Moreover, the role of different residues in contributing to the thermostability was identified employing MD simulation dynamics and validation of the predicted residues was performed through in vivo and in vitro assays. A similar
approach using this rationale can help in improving the thermostability of the first enzyme, AAR of the alkane biosynthesis pathway which can aid in the assembly of this pathway in thermostable hosts such as Geobacillus for optimal alkane production at higher temperatures.

EXPERIMENTAL PROCEDURES

Microbial strains, media and reagents - E. coli strains and plasmids used in this study have been listed in Supplementary Table 4. E. coli DH5α was used as host for molecular cloning. Fatty aldehydes and fatty alcohols used in enzyme assays and as standards for GCMS-MS were purchased from TCI Chemicals (India) Pvt. Ltd. Ethyl acetate and methanol used in the extraction of hydrocarbons were procured from Merck. Codon optimized genes for the aldehyde decarboxylating oxygenase of *Nostoc punctiforme* PCC 73102 (np-ado), the consensus thermostable aldehyde deformylating oxygenase (cts-ado) and the quadruple mutant of Np-ADO (H51K-D132P-A161E-T191I) were commercially synthesized at Genscript, USA. The kits and enzymes used for molecular biology purpose and their sources were as follows: Genomic DNA isolation kit (ZR Fungal/Bacterial DNA Miniprep from Zymo Research), Plasmid Miniprep kit, PCR and gel extraction kit (Qiagen), Phusion DNA polymerase (Finnzymes), Taq DNA polymerase (Himedia Pvt. Ltd.), dNTPs (Fermentas, USA), DNA fast digest restriction endonuclease enzymes and T4 DNA ligase (Thermo Scientific). The site directed mutagenesis was carried out using the QuikChange II Site Directed Mutagenesis Kit by Agilent in accordance with the protocols described in the kit manual. pQE30-Np-ADO plasmid DNA was isolated using the Plasmid Miniprep Kit by Qiagen to serve as the template for the mutagenesis experiment. Site-directed mutants were constructed using the primers listed in Supplementary Table 5. For constructing the double mutant H51K-A161E, the single mutant A161E was used as the template to which the second mutation H51K was introduced. The desired mutations were confirmed by DNA sequencing carried out at Macrogen, Korea.

Design of synthetic thermostable ADO using metagenome pipeline - Metagenome analysis was conducted where a pipeline for generating consensus sequence for designing a thermostable *ado* gene was articulated. Here numerous metagenome libraries were screened using IMG/MER database and only sequences of organisms belonging to hot springs were retrieved. Metagenome libraries were analysed for presence of the *ado* gene homologs using the *ado* gene from *Synechococcus elongatus* PCC7942 as the query sequence. Only those homologs which showed coverage score ≥ 300 in BLASTN alignment were considered further. The sequences were translated using ExPASy translate tool (69) and aligned using MEGA6 software (70). Multiple polypeptide sequences which were assigned the same genus based on the BLAST results and which showed complete alignment on MEGA6 software were identified and only one sequence each for every genus was selected for further analysis. The Jalview software (71) was then employed for generating a consensus sequence which was sent for synthesis to GenScript.

Bacterial strains and plasmid construction - *E.coli* DH5α was used for routine DNA cloning and protein expression. Codon optimized aldehyde deformylating oxygenase gene from *Nostoc punctiforme* PCC73102 was used to construct the wild type pQE30-Np-ADO. The codon optimized consensus thermostable cts-ado gene synthetically designed using the metagenome pipeline as discussed in the later section was cloned at BamHI/SacI restriction site in pQE30 to obtain pQE30-Cts-ADO. Engineered *E.coli* strains were constructed by transformation of *E. coli* DH5α with these plasmids. All the engineered *E. coli* strains were grown in LB broth containing antibiotics at standard concentrations. 100 µg/ mL Ampicillin was added in the culture when required.

Site-directed mutagenesis - The construction of single or double amino acid mutant was carried out using Quick Change site-directed mutagenesis kit by Agilent in accordance with the protocols described in the kit manual, pQE30-Np-ADO plasmid DNA was isolated using the Plasmid Miniprep Kit by Qiagen to serve as the template for the mutagenesis experiment. Site-directed mutants were constructed using the primers listed in Supplementary Table 5. For constructing the double mutant H51K-A161E, the single mutant A161E was used as the template to which the second mutation H51K was introduced. The desired mutations were confirmed by DNA sequencing carried out at Macrogen, Korea.
Protein engineering for improved thermostability of ADO

substrate bound forms of PDB templates were chosen as criteria to select templates to model these enzymes. The PDB entries 4RC5, 20C5 and 4TW3 (PDB IDs- 4RC5 from S. elongatus PCC7942 and 20C5 and 4TW3 from P. marinus MIT9313) were chosen as consensus templates for homology modelling of Np-ADO and Cts-ADO enzymes. The hexadecanone bounded in crystal structure of 4RC5 was used as substrate form in all three enzymes. Total 20 models were generated for each enzyme in substrate bound (hexadecanone) form using MODELLER v9.15 software separately (54,73). The models were ranked according to the DOPE statistical potential score. The PROCHECK webserver (74) was used for performing quality assessments including Ramachandran plots for the best models of three enzymes. All models were aligned using sequence similarities and secondary structure information through ESPript3 (75). The modelled structures were visualized using Delano PyMOL molecular graphics system (76). The two separate (H51K and A161E) single residue mutations were introduced into the Np-ADO enzyme model using python script available in MODELLER to bring the desired mutation effects. Total five models were used to study different temperatures dynamics on respective structures. All models were subjected to five different temperatures starting from 300 K - 340 K at intervals of 10 K. All molecular dynamics on different temperatures were performed using AMBER14 software and modelling and data analysis were performed by AmberTools14 suite of programs (77,78). The topology and parameter files were constructed using the force field leaprc.ff99SB for the enzymes Np-ADO and its mutants and Cts-ADO. The parameters and atom-types of hexadecanone substrate were generated through ANTECHAMBER module and charges calculated using AM1-BCC method. The systems were first solvated with triangulated 3-point water molecules using the using the 12 Å pad of TIP3P water mode in the five systems. Neutralizing counter ions Na⁺ species were added in respective systems. The energy minimization procedure included initial 500 cycles of steep descent (SD) algorithm followed by remaining 19500 cycles of conjugate gradient algorithm (CG) totalling 20000 steps. After the energy minimization, the systems were heated from 0 K to 300 K, 310 K, 320 K, 330 K, and 340 K over 50 ps with a collision frequency of 2.0 ps⁻¹, and weak harmonic restraints of 2 Kcal mol⁻¹ Å⁻² on all atoms using Langevin thermostat for temperature regulation. The next step was accompanied by density equilibration (50 ps) and constant pressure equilibration at 5 ns. The final production molecular dynamics simulations were performed for 100 ns time interval at different temperatures. All molecular simulations were performed using PMEMD module in AMBER14. The time step 2 fs was used for all molecular dynamics stages and all atoms free bond stretching freedoms involving hydrogen atoms were constrained using the SHAKE algorithm. All simulation co-ordinates were saved in single trajectory in all the six systems and the trajectory was saved every 2ps. Total 2.7 μs duration time was completed in all production molecular dynamics. The RMSD and RMSF and cluster analysis of trajectories were performed using CPPTRAJ module of AmberTools (79). Cluster representative of the largest cluster was taken for each enzyme at different temperatures from their respective MD trajectories using CPPTRAJ module of AmberTools for generating structures plotted in Figures 6, 7 and 8 and supplementary figures 5, 6, 11 and 12.

Based on the RMSD and RMSF analysis, the amino acid residues lying in the thermally sensitive regions were identified. Multiple sequence alignment of different mesophilic and thermophilic ADOs was performed using Jalview. The mutations identified from these approaches were visually assured using Chimera to be on surface of the ADO enzyme, away from the substrate binding site, and outside the catalytic centre of the enzyme. LigPlot⁺ v.2.1 was used to view the interaction between the active site residues, the Fe atoms and the substrate ligands. All the protein structures depicted in the study were finally plotted using Chimera (80).

Protein overexpression and purification - Wild-type Np-ADO and its variants were overexpressed in E.coli DH5α while the Cts-ADO was overexpressed in Shuffle due to its low expression in DH5α. The DH5α cells were grown at 37 °C in LB broth medium containing antibiotics ampicillin (100 μg/ml) while Shuffle cells were cultured at 30 °C in LB broth medium containing antibiotics ampicillin (100 μg/ml) and spectinomycin (100 μg/ml). The cultures were
induced with 1 mM IPTG when OD_{600} reached around 0.5-0.6. Following induction, the DH5α cells were incubated at 37 °C, 180 rpm, while the shuffle cells were incubated at 20 °C, 180 rpm, for 3.5 hours before being harvested. Cells were collected through centrifugation at 8000 rpm at 4 °C, re-suspended in lysis buffer (50 mM KH_{2}PO_{4}, 50 mM K_{2}HPO_{4}, 400 mM NaCl, 100 mM KCl, 10% Glycerol, 0.5% Triton X-100 and 10 mM Imidazole) and lysed by adding lysozyme (1 mg/ml, USB Corporation, USA) followed by sonication. The proteins were purified on Ni-NTA agarose matrix (Qiagen) as previously described method with some modifications (40). Elutes having the pure protein were pooled and dialysed against final assay buffer (100 mM potassium phosphate buffer pH 7.2, 100 mM KCl and 10% glycerol). For Circular Dichroism spectroscopy and differential scanning fluorimetry, proteins were dialysed against 10 mM potassium phosphate buffer pH 7.2, concentrated on Vivaspin 20 sample concentrator (5 kDa MWCO, GE healthcare) and used. Protein concentration was determined by BCA kit (G-Biosciences, USA) using bovine serum albumin as a standard.

In vivo and in vitro enzyme assays - For screening of thermostable mutants in vivo, E. coli strains harbouring the wild type and mutant plasmids were grown in 5 ml LB broth medium with appropriate antibiotics at a final concentration of 100 μg/ml ampicillin in 55 ml Borosil culture tubes. The culture tubes were incubated overnight in an incubator-shaker at 180 rpm at 37 °C. 50 μl of this grown culture was used to inoculate 3 ml secondary culture having M9 modified medium supplemented with 2% glucose, 1 mg/L thiamine and 100 μg/ml ampicillin (3). The culture was induced with 0.01 mM IPTG at the time of secondary inoculation in M9 medium. 100 mg/L hexadecanal was added exogenously to the culture medium at the time of secondary inoculation and the culture tubes were sealed with Parafilm to avoid evaporation at high temperatures. The culture tubes were then kept on incubator shaker at different temperatures (30 °C, 37 °C, and 42 °C) for 48 hours at 120 rpm. The hydrocarbon analysis was done after extraction of hydrocarbons from the culture medium by adding equal volume of ethyl acetate and vortexing the mixture for 20 minutes following which the samples were centrifuged and the upper phase was collected in vials for GC analysis (3). 1-Octadecene was used as an internal standard. For each temperature, the experiments were performed in triplicates and the data has been presented as average and standard deviation of three independent biological replicates.

The in vitro enzyme assay was carried out according to the published protocol (33). Enzyme assays were performed in 100 mM HEPES buffer, containing 100 mM KCl and 10% glycerol at pH 7.2. The reaction mixtures contained the following components; NADH (750 μM), ferrous ammonium sulfate (80 μM), PMS (75 μM), hexadecanal (150 μM), cADO (20 μM) in a total 500 μL reaction volume. For enzyme assays at different temperatures, both the reaction components and the enzymes were pre-incubated at the specific temperature before starting the reaction. Equal volumes of ethyl acetate was used for termination of the reaction. The sample was then vortexed at high speed for 20 minutes and centrifuged. A 400 μL extractant was then analysed by GC–MS/MS. For determining the kinetic parameters the enzymatic reaction was set up as mentioned above at different temperatures (300 K, 320 K and 333 K) for 15 minutes on an incubator shaker. All kinetic assays were carried out in triplicates. The Michaelis-Menten plots were constructed using reaction rates obtained at different substrate concentrations at each temperature. The values for the kinetic parameters K_{m}, V_{max}, k_{cat} and k_{cat}/K_{m} were obtained by curve fitting using GraphPad Prism 5.

Quantification of alkanes using gas chromatography. Gas chromatography analysis was performed on an Agilent 7890 A system equipped with HP-5 column of 30 m length, 0.32 mm internal diameter and 0.25 mm film thickness and FID detector was used for analysis (81-83). The oven temperature programme was set as follows: initial 100 °C held for 3 min, to 250 °C at 10 °C min⁻¹ and held at 250 °C for additional 10 minutes. The total run-time of the programme was 28 min. The inlet and detector temperature were maintained at 150 °C and 280 °C, respectively. Further confirmation of products was done on a 7890 A gas chromatography system equipped with a 7000 GC/MS triple quadrupole system (Agilent) with HP-5 MS column using the same programme as mentioned above. However, the final oven temperature was ramped up to 300 °C at a rate of
20 °C min⁻¹ and was then held at 300 °C for an additional 5 minutes. The MS quadrupole scanned from 50 to 550 m/z. Retention times and fragmentation patterns of product peaks were compared with authentic references to confirm peak identity. For confirmation and quantification, pentadecane, hexadecane, 1-octadecene, and hexadecanol standards from TCI America were used.

**Differential Scanning Fluorimetry** - The thermal stability of the purified wild type and thermostable mutant proteins was evaluated by differential scanning fluorimetry. The assay was performed using a CFX96 Real-Time PCR System (Bio-Rad, USA). Briefly, a 5µL of 25× SYPRO orange dye (Invitrogen, USA) was added to 10 µL of protein at 1 mg/mL in 10 mM potassium phosphate buffer at pH 7.2. SYPRO orange dye with the buffer served as the control to correct background fluorescence. The reaction volume was made up to 50 µL in a Bio-Rad 96-well plate and the samples were heated at 0.5 °C per five seconds from 25 to 95 °C. The fluorescence intensity (excitation/emission: 450 to 490 nm/560 to 580 nm) was measured every 0.5 °C. Thermal midpoint (Tₘ) values of proteins were determined with the CFX Manager Program (Bio-Rad) based on calculation of the negative first derivative (84). The melt curve RFU (relative fluorescence unit) values were entered against the respective temperature to generate the denaturation curve by non-linear regression, fitting the Boltzmann equation to the denaturation curves using GraphPad Prism 5. Experiments were carried out in triplicates.

**Characterisation using CD spectroscopy** - Far-UV CD spectra were measured for the wild type ADO and its thermostable mutant using circular dichroism spectroscopy. The proteins were concentrated to a final concentration of 0.12 mg/ml in 10 mM potassium phosphate buffer at pH 7.2. CD spectrum was measured by JASCO J-810 (JASCO, Japan) using 1 mm path-length cuvette. Full wavelength data was collected at 25 °C, with the wavelengths ranging from 190 nm to 250 nm at 0.5 nm intervals. Data was averaged over three runs and the background was corrected using a buffer blank. Secondary structure analysis was performed with K2D2 method (85).

**Acknowledgments**: The authors would like to thank PCC culture collections for providing the cyanobacterial strains for study and Mr. Nitin Yadav for his assistance in performing CD spectroscopy. The research was funded by the grants from the Department of Biotechnology, Government of India to SSY (Grant No. BT/PB/Center/03/2011 and BT/IN/INDO-UK/SuBB/21/SSY/2013) and to DG (Grant No. BT/BI/25/066/2012). Senior Research Fellowship to RK is duly acknowledged.

**Conflict of Interest**: The authors declare that they have no conflicts of interest with the contents of this article.
References

1. Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., and Keasling, J. D. (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Current opinion in biotechnology 19, 556-563
2. Timilsina, G. R. (2014) Biofuels in the long-run global energy supply mix for transportation. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 372, 20120323
3. Schirmer, A., Rude, M. A., Li, X., Popova, E., and Del Cardayre, S. B. (2010) Microbial biosynthesis of alkanes. Science 329, 559-562
4. Lu, X. (2010) A perspective: photosynthetic production of fatty acid-based biofuels in genetically engineered cyanobacteria. Biotechnology advances 28, 742-746
5. Tan, X., Yao, L., Gao, Q., Wang, W., Qi, F., and Lu, X. (2011) Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. Metabolic engineering 13, 169-176
6. Knothe, G. (2010) Biodiesel and renewable diesel: a comparison. Progress in Energy and Combustion Science 36, 364-373
7. Speight, J. G. (2016) Handbook of Petroleum Refining, Crc Press
8. Gelpi, E., Oro, J., Schneider, H., and Bennett, E. (1968) Olefins of high molecular weight in two microscopical algae. Science 161, 700-701
9. Merdinger, E., and Frye, R. H. (1966) Distribution of C14 from Glucose-1-C14 in the Lipid Fractions of Debaryomyces hansenii. Journal of bacteriology 91, 1831-1833
10. Oro, J., Laseter, J., and Weber, D. (1966) Alkanes in fungal spores. Science 154, 399-400
11. Jones, J. (1969) Studies on lipids of soil micro-organisms with particular reference to hydrocarbons. Microbiology 59, 145-152
12. Winters, K., Parker, P., and Van Baalen, C. (1969) Hydrocarbons of blue-green algae: geochemical significance. Science 163, 467-468
13. McInnes, A. G., Walter, J. A., and Wright, J. L. (1980) Biosynthesis of hydrocarbons by algae: Decarboxylation of stearic acid to N-heptadecane inAnacystis nidulans determined by 13 C-and 2 H-labeling and 13 C nuclear magnetic resonance. Lipids 15, 609-615
14. Dembitsky, V., and Srebnik, M. (2002) Variability of hydrocarbon and fatty acid components in cultures of the filamentous cyanobacterium Scytonema sp. isolated from microbial community “black cover” of limestone walls in Jerusalem. Biochemistry (Moscow) 67, 1276-1282
15. Shiea, J., Brassell, S. C., and Ward, D. M. (1990) Mid-chain branched mono-and dimethyl alkanes in hot spring cyanobacterial mats: a direct biogenic source for branched alkanes in ancient sediments? Organic Geochemistry 15, 223-231
16. Akhtar, M. K., Turner, N. J., and Jones, P. R. (2013) Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. Proceedings of the National Academy of Sciences 110, 87-92
17. Westerlund-Karlsson, A., Saviranta, P., and Karp, M. (2002) Generation of thermostable monomeric luciferases from Photorhabdus luminescens. Biochemical and biophysical research communications 296, 1072-1076
18. Winson, M. K., Swift, S., Hill, P. J., Sims, C. M., Griesmayr, G., Bycroft, B. W., Williams, P., and Stewart, G. S. (1998) Engineering the luxCDABE genes from Photorhabdus luminescens to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn5 constructs. FEMS microbiology letters 163, 193-202
19. Howard, T. P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D. M., Taylor, G. N., Parker, D. A., Lee, R., Smirnoff, N., and Aves, S. J. (2013) Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli. Proceedings of the National Academy of Sciences 110, 7636-7641
20. Buist, P. H. (2007) Exotic biomodification of fatty acids. Natural product reports 24, 1110-1127
21. Bao, L., Li, J.-J., Jia, C., Li, M., and Lu, X. (2016) Structure-oriented substrate specificity engineering of aldehyde-deformylating oxygenase towards aldehydes carbon chain length. *Biofuels Technology* 9, 185
22. Khara, B., Menon, N., Levy, C., Mansell, D., Das, D., Marsh, E. N. G., Leys, D., and Scrutton, N. S. (2013) Production of Propane and Other Short-Chain Alkanes by Structure-Based Engineering of Ligand Specificity in Aldehyde-Deformylating Oxygenase. *ChemBioChem* 14, 1204-1208
23. Krebs, C., Bollinger, J. M., and Booker, S. J. (2011) Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like ‘di-iron-carboxylate’ proteins. *Current opinion in chemical biology* 15, 291-303
24. Das, D., Eser, B. E., Han, J., Sciore, A., and Marsh, E. N. G. (2011) Oxygen-Independent Decarbonylation of Aldehydes by Cyanobacterial Aldehyde Decarbonylase: A New Reaction of Diiron Enzymes. *Angewandte Chemie International Edition* 50, 7148-7152
25. Kurtz Jr, D. M. (1997) Structural similarity and functional diversity in diiron-oxo proteins. *Journal of Biological Inorganic Chemistry* 2, 159-167
26. Merkx, M., Kopp, D. A., Sazinsky, M. H., Blazyk, J. L., Müller, J., and Lippard, S. J. (2001) Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: a tale of two irons and three proteins. *Angewandte Chemie International Edition* 40, 2782-2807
27. Warui, D. M., Li, N., Nørgaard, H., Krebs, C., Bollinger Jr, J. M., and Booker, S. J. (2011) Detection of formate, rather than carbon monoxide, as the stoichiometric co-product in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *Journal of the American Chemical Society* 133, 3316
28. Li, N., Nørgaard, H., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger Jr, J. M. (2011) Conversion of Fatty Aldehydes to Alk (a/e) nes and Formate by a Cyanobacterial Aldehyde Decarbonylase: Cryptic Redox by an Unusual Di-metal Oxygenase. *Journal of the American Chemical Society* 133, 6158
29. Eser, B. E., Das, D., Han, J., Jones, P. R., and Marsh, E. N. G. (2011) Oxygen Independent Alkane Formation by Non-Heme Iron-Dependent Cyanobacterial Aldehyde Decarbonylase: Investigation of Kinetics and Requirement for an External Electron Donor. *Biochemistry* 50, 10743
30. Li, N., Chang, W.-c., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger Jr, J. M. (2012) Evidence for only oxygenative cleavage of aldehydes to alk (a/e) nes and formate by cyanobacterial aldehyde decarbonylases. *Biochemistry* 51, 7908-7916
31. Dennis, M., and Kolattukudy, P. (1992) A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Proceedings of the National Academy of Sciences* 89, 5306-5310
32. Andre, C., Kim, S. W., Yu, X.-H., and Shanklin, J. (2013) Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H2O2 to the cosubstrate O2. *Proceedings of the National Academy of Sciences* 110, 3191-3196
33. Zhang, J., Lu, X., and Li, J.-J. (2013) Conversion of fatty aldehydes into alk (a/e) nes by in vitro reconstituted cyanobacterial aldehyde-deformylating oxygenase with the cognate electron transfer system. *Biofuels Technology* 6, 86
34. Wang, Q., Huang, X., Zhang, J., Lu, X., Li, S., and Li, J.-J. (2014) Engineering self-sufficient aldehyde deformylating oxygenases fused to alternative electron transfer systems for efficient conversion of aldehydes into alkanes. *Chemical Communications* 50, 4299-4301
35. Jia, C., Li, M., Li, J., Zhang, J., Zhang, H., Cao, P., Pan, X., Lu, X., and Chang, W. (2015) Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases. *Protein & cell* 6, 55-67
36. Das, D., Ellington, B., Paul, B., and Marsh, E. N. G. (2014) Mechanistic insights from reaction of α-oxiranyl-aldehydes with cyanobacterial aldehyde deformylating oxygenase. *ACS chemical biology* 9, 570
37. Paul, B., Das, D., Ellington, B., and Marsh, E. N. G. (2013) Probing the mechanism of cyanobacterial aldehyde decarbonylase using a cyclopropyl aldehyde. *Journal of the American Chemical Society* **135**

38. Waugh, M. W., and Marsh, E. N. G. (2014) Solvent isotope effects on alkane formation by cyanobacterial aldehyde deformylating oxygenase and their mechanistic implications. *Biochemistry* **53**, 5537-5543

39. Hayashi, Y., Yasugi, F., and Arai, M. (2015) Role of cysteine residues in the structure, stability, and alkan producing activity of cyanobacterial aldehyde deformylating oxygenase. *PloS one* **10**, e0122217

40. Buer, B. C., Paul, B., Das, D., Stuckey, J. A., and Marsh, E. N. G. (2014) Insights into substrate and metal binding from the crystal structure of cyanobacterial aldehyde deformylating oxygenase with substrate bound. *ACS chemical biology* **9**, 2584

41. Shanklin, J., Guy, J. E., Mishra, G., and Lindqvist, Y. (2009) Desaturases: emerging models for understanding functional diversification of diiron-containing enzymes. *Journal of Biological Chemistry* **284**, 18559-18563

42. Sazinsky, M. H., and Lippard, S. J. (2006) Correlating structure with function in bacterial multicomponent monoxygenases and related diiron proteins. *Accounts of chemical research* **39**, 558-566

43. Wang, Q., Bao, L., Jia, C., Li, M., Li, J.-J., and Lu, X. (2017) Identification of residues important for the activity of aldehyde-deformylating oxygenase through investigation into the structure-activity relationship. *BMC biotechnology* **17**, 31

44. Turner, P., Mamo, G., and Karlsson, E. N. (2007) Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial cell factories* **6**, 9

45. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008) NCBI BLAST: a better web interface. *Nucleic acids research* **36**, W5-W9

46. Papaleo, E., Olufsen, M., De Gioia, L., and Brandsdal, B. O. (2007) Optimization of electrostatics as a strategy for cold-adaptation: a case study of cold-and warm-active elastases. *Journal of Molecular Graphics and Modelling* **26**, 93-103

47. Gribenko, A. V., Patel, M. M., Liu, J., McCallum, S. A., Wang, C., and Makhatadze, G. I. (2009) Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proceedings of the National Academy of Sciences* **106**, 2601-2606

48. Vinther, J. M., Kristensen, S. M., and Led, J. J. (2010) Enhanced stability of a protein with increasing temperature. *Journal of the American Chemical Society* **133**, 271-278

49. Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M. M., Iimura, S., Gajiwala, K., and Scholtz, J. M. (2011) Contribution of hydrophobic interactions to protein stability. *Journal of molecular biology* **408**, 514-528

50. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *Journal of molecular biology* **327**, 1135-1148

51. Thomas, A. S., and Elcock, A. H. (2004) Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *Journal of the American Chemical Society* **126**, 2208-2214

52. Crystal structure of hypothetical protein (NP_895059.1) from Prochlorococcus marinus MIT9313 at 1.68 A resolution, [http://www.rcsb.org/pdb/explore/explore.do?structureId=2OC5](http://www.rcsb.org/pdb/explore/explore.do?structureId=2OC5)

53. Buer, B. C., Paul, B., Das, D., Stuckey, J. A., and Marsh, E. N. (2014) Insights into substrate and metal binding from the crystal structure of cyanobacterial aldehyde deformylating oxygenase with substrate bound. *ACS Chem Biol* **9**, 2584-2593

54. Webb, B., and Sali, A. (2014) Protein structure modeling with MODELLER. *Protein Structure Prediction*, 1-15
Protein engineering for improved thermostability of ADO

55. Mamonova, T. B., Glyakina, A. V., Galzitskaya, O. V., and Kurnikova, M. G. (2013) Stability and rigidity/flexibility—Two sides of the same coin? *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1834**, 854-866

56. Paul, M., Hazra, M., Barman, A., and Hazra, S. (2014) Comparative molecular dynamics simulation studies for determining factors contributing to the thermostability of chemotaxis protein “CheY”. *Journal of Biomolecular Structure and Dynamics* **32**, 928-949

57. Wintrode, P. L., Zhang, D., Vaidehi, N., Arnold, F. H., and Goddard, W. A. (2003) Protein dynamics in a family of laboratory evolved thermophilic enzymes. *Journal of molecular biology* **327**, 745-757

58. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. ACS Publications

59. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology* **157**, 105-132

60. Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical biochemistry* **332**, 153-159

61. Olson, D. G., McBride, J. E., Shaw, A. J., and Lynd, L. R. (2012) Recent progress in consolidated bioprocessing. *Current opinion in biotechnology* **23**, 396-405

62. Chan, C. S., Chan, K.-G., Tay, Y.-L., Chua, Y.-H., and Goh, K. M. (2015) Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Frontiers in microbiology* **6**, 177

63. Idicula-Thomas, S., and Balaji, P. V. (2005) Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*. *Protein Science* **14**, 582-592

64. Liao, H. H. (1991) Effect of temperature on the expression of wild-type and thermostable mutants of kanamycin nucleotidyltransferase in *Escherichia coli*. *Protein expression and purification* **2**, 43-50

65. Yu, H., Yan, Y., Zhang, C., and Dalby, P. A. (2017) Two strategies to engineer flexible loops for improved enzyme thermostability. *Scientific Reports* **7**, 41212

66. Malabanan, M. M., Amyes, T. L., and Richard, J. P. (2010) A role for flexible loops in enzyme catalysis. *Current opinion in structural biology* **20**, 702-710

67. Gunasekaran, K., Ma, B., and Nussinov, R. (2003) Triggering loops and enzyme function: identification of loops that trigger and modulate movements. *Journal of molecular biology* **332**, 143-159

68. Damnjanović, J., Nakano, H., and Iwasaki, Y. (2014) Deletion of a dynamic surface loop improves stability and changes kinetic behavior of phosphatidylinositol-synthesizing *Streptomyces phospholipase D*. *Biotechnology and bioengineering* **111**, 674-682

69. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., and Gasteiger, E. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic acids research* **40**, W597-W603

70. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* **30**, 2725-2729

71. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191

72. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., and Söding, J. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539

73. Eswar, N., Webb, B., Marti-Renom, M., Madhusudhan, M., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2007) Current protocols in protein science, editorial board, John E. *Coligan et al*
Protein engineering for improved thermostability of ADO

74. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* 26, 283-291

75. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* 42, W320-W324

76. DeLano, W. (2002) The PyMOL molecular graphics system, version 1.8. Schrödinger LLC. *Doyle, DA, Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon*, 1067-1076

77. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *Journal of computational chemistry* 26, 1668-1688

78. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E., Duke, H. G., A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. L., T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. S., C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X., and Kollman, W. a. P. A. (2014) AMBER 14.

79. Roe, D. R., and Cheatham III, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* 9, 3084-3095

80. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605-1612

81. Shakeel, T., Fatma, Z., Fatma, T., and Yazdani, S. S. (2015) Heterogeneity of alkane chain length in freshwater and marine cyanobacteria. *Frontiers in bioengineering and biotechnology* 3

82. Fatma, Z., Jawed, K., Mattam, A. J., and Yazdani, S. S. (2016) Identification of long chain specific aldehyde reductase and its use in enhanced fatty alcohol production in E. coli. *Metabolic engineering* 37, 35-45

83. Fatma, Z., Hartman, H., Poolman, M. G., Fell, D. A., Srivastava, S., Shakeel, T., and Yazdani, S. S. (2018) Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production. *Metabolic engineering* 46, 1-12

84. Huynh, K., and Partch, C. L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Current protocols in protein science*, 28.29. 21-28.29. 14

85. Perez-Iratxeta, C., and Andrade-Navarro, M. A. (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC structural biology* 8, 25
**Table 1.** Kinetic parameters of different ADOs at varying temperature

|                  | Kₘ (μM) *10² | Vₘₐₓ (μM min⁻¹) | k-cat (min⁻¹) *10⁻² | k-cat / Kₘ (μM⁻¹ min⁻¹) *10⁻³ |
|------------------|--------------|------------------|---------------------|-------------------------------|
| **Np-ADO**       |              |                  |                     |                               |
| 27 °C            | 0.76 ± 0.15  | 0.37 ± 0.02      | 3.74 ± 0.25         | 0.48                          |
| 47 °C            | 0.61 ± 0.12  | 0.18 ± 0.01      | 1.86 ± 0.12         | 0.30                          |
| 60 °C            | ND           | ND               | ND                  | ND                            |
| **Cts-ADO**      |              |                  |                     |                               |
| 27 °C            | 0.81 ± 0.17  | 0.27 ± 0.02      | 2.73 ± 0.21         | 0.33                          |
| 47 °C            | 0.45 ± 0.19  | 0.24 ± 0.03      | 2.48 ± 0.30         | 0.54                          |
| 60 °C            | 1.79 ± 0.62  | 0.29 ± 0.05      | 2.90 ± 0.50         | 0.16                          |
| **Mutant A161E** |              |                  |                     |                               |
| 27 °C            | 0.62 ± 0.19  | 0.20 ± 0.02      | 2.07 ± 0.21         | 0.33                          |
| 47 °C            | 0.84 ± 0.19  | 0.46 ± 0.03      | 4.67 ± 0.38         | 0.55                          |
| 60 °C            | 0.50 ± 0.11  | 0.29 ± 0.01      | 2.91 ± 0.19         | 0.58                          |
| **Double Mutant**|              |                  |                     |                               |
| (H51K-A161E)     |              |                  |                     |                               |
| 27 °C            | 1.50 ± 0.22  | 0.42 ± 0.02      | 4.22 ± 0.29         | 0.28                          |
| 47 °C            | 0.66 ± 0.13  | 0.38 ± 0.02      | 3.86 ± 0.26         | 0.58                          |
| 60 °C            | 0.24 ± 0.07  | 0.17 ± 0.01      | 1.71 ± 0.11         | 0.70                          |
| **Quadruple**    |              |                  |                     |                               |
| (H51K-D132P-A161E-T191I) | |         |                     |                               |
| 27 °C            | 0.42 ± 0.21  | 0.21 ± 0.03      | 2.15 ± 0.30         | 0.51                          |
| 47 °C            | 0.26 ± 0.05  | 0.19 ± 0.00      | 1.96 ± 0.08         | 0.76                          |
| 60 °C            | 0.62 ± 0.16  | 0.17 ± 0.01      | 1.70 ± 0.14         | 0.27                          |

*ND- No activity detected

**Table 2.** Melting temperature of wild type Np-ADO, Cts-ADO and mutant A161E as determined by differential scanning fluorimetry.

|                  | Np-ADO     | Cts-ADO    | Mutant A161E |
|------------------|------------|------------|--------------|
| Tₘ              | 53 ± 0.5 °C| 60 ± 0.35 °C| 58 ± 0.53 °C|

*± The Michaelis-Menten graphs used for determination of kinetic parameters have been given in Supplementary Figure 4.
Protein engineering for improved thermostability of ADO

Figure 1. Schematic representation of the study design. A metagenome pipeline was constructed for generation of the consensus thermostable ADO (cts-ado), which was further compared with its mesophilic counterpart for the prediction of amino acid residues responsible for imparting thermostability. Computational assessment and in vivo and in vitro activity measurements were incorporated in the study to evaluate the mutants at higher temperatures.
Figure 2. Comparison of *in vivo* production of pentadecane for mesophilic Np-ADO, thermostable Cts-ADO and four single amino acid substituted variants of Np-ADOs. (A) The bar graph represents the amount of pentadecane produced by each *E. coli* construct harbouring the pQE30 plasmid carrying the desired ADO gene. Hexadecanal (100 mg/L) was exogenously added to the growing bacterial culture and its conversion to pentadecane was analyzed on GC-MS/MS after 48 hr of cultivation. Data represents average and standard deviation of three independent biological replicates. (B) Western blot of the soluble and insoluble fractions of the proteins collected at different temperatures after 48 hr of cultivation of *E. coli* strain expressing various ADO variants. Full length Western blot are included as Supplementary Figure 13.
Figure 3. Fold change in specific activities of wild type and thermostable mutants at different temperatures with respect to 300 K. The activity optima of the mutants D132P and T191I remain unchanged as of the wild type at 310 K while the mutants A161E and H51K showed improved specific activities at 320 K, exhibiting similar activity optima as the thermophile, Cts-ADO.
Figure 4. Modelled structure of (A) mesophilic *Nostoc* Np-ADO and (B) consensus thermostable Cts-ADO with the substrate ligand hexadecanal bound. The best model was selected from 20 models generated using Modeller 9.15 software employing PDB entries 2OC5, 4TW3 and 4RC5 as templates. The secondary structures are indicated as helices H1 - H8 connected by loops L1 - L9 and the iron atoms are represented as red spheres.
Figure 5. Comparison of the Root Mean Square Deviation (RMSD) over the 100 ns simulations of (A) mesophilic *Nostoc* Np-ADO and (B) consensus thermostable Cts-ADO in the presence of bound ligand hexadecanal.
Figure 6. Secondary structures of (A) Np-ADO and (B) Cts-ADO at 300 K, 320 K and 340 K. Helices are labelled as H1 – H8 and loops are labelled as L1 – L9. Loops and helices in addition to those present in the Np-ADO structure at 300K are indicated with a suffix α, β, γ, δ, φ or ω, and are highlighted in red.
Protein engineering for improved thermostability of ADO

Figure 7. Interactions of the Fe atom with the active site residues and substrate ligand (C16 aldehyde – hexadecanal shown as LC16) at 300 K, 320 K and 340 K for (A) Np-ADO and (B) Cts-ADO. Y40 residue (coloured green) appeared to interact with the Fe2 in Cts-ADO at 300 K, which at higher temperature was removed due to shift in the position of the helix H1. This rearrangement perhaps led to moving the ligand close to Fe2 in Cts-ADO at 320 K and 340 K. Cluster representatives of the largest cluster were taken for each enzyme at different temperatures from their respective MD trajectories using CPPTRAJ module of AmberTools for generating structures. The interactions between Fe atoms, active site residues and substrate ligand (LC16) were viewed using LigPlot+ v.2.1.
Figure 8. Interactions of the substrate ligand (hexadecanal) with active site residues, residues lining the substrate canal and Fe1 and Fe2 atoms (indicated as orange spheres), at 300 K, 320 K and 340 K for Np-ADO and Cts-ADO. Residues which were found to be hindering the binding of the ligand to the preferred Fe2 atom are shown as cyan coloured sticks and labelled in red.
Figure 9. Differential scanning fluorimetry assay performed with SYPRO orange to determine the melting temperatures of the purified proteins. Grey dotted grid lines represent T_m of the denaturation curves.
Figure 10. CD spectra of the purified Np-ADO, Cts-ADO and mutant A161E proteins to determine the structural effects of the point mutation.
A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability
Tabinda Shakeel, Mayank Gupta, Zia Fatma, Rakesh Kumar, Raubins Kumar, Rahul Singh, Medha Sharma, Dhananjay Jade, Dinesh Gupta, Tasneem Fatma and Syed Shams Yazdani

J. Biol. Chem. published online April 9, 2018

Access the most updated version of this article at doi: 10.1074/jbc.RA117.000639

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts