A novel C-terminal degron identified in bacterial aldehyde decarbonylases using directed evolution

Yilan Liu, Jinjin Chen, Anna N. Khusnutdinova, Kevin Correia, Patrick Diep, Khorcheska A. Batyrova, Kayla Nemr, Robert Flick, Peter Stogios, Alexander F. Yakunin, Radhakrishnan Mahadevan

a. Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, ON, Canada M5S 3E5
b. Institute of Biomaterials and Biomedical Engineering, University of Toronto, University of Toronto, 200 College Street, Toronto, ON, Canada M5S 3E5
c. Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK

*Correspondence: krishna.mahadevan@utoronto.ca

Abstract:

Background: Aldehyde decarbonylase (AD), which converts acyl aldehydes into alkanes, supplies promising solution for producing alkanes from renewable feedstock. However the instability of AD impeded its further application. Therefore, the current study aimed to investigate the degradation mechanism of AD and engineer it towards high stability.

Results: Here, we describe the discovery of a degradation tag (degron) in the AD from marine cyanobacterium Prochlorococcus marinus via error-prone PCR based directed evolution system. Bioinformatic analysis revealed this C-terminal degron is common in the family of bacterial ADs and identified a conserved C-terminal motif, RMSAYGLAAA, representing the AD degron (ADcon). Furthermore, we demonstrated that the ATP-dependent proteases ClpAP and Lon are involved in the degradation of AD-tagged proteins in E. coli, thereby limiting alkane production. Deletion or modification of the degron motif increased alkane production in vivo.

Conclusions: This work revealed the presence of a novel degron in bacterial ADs responsible for its instability. The in vivo experiments proved eliminating or modifying the degron could stabilize AD, thereby producing higher titers of alkanes.

Keywords: Alkane, Aldehyde decarbonylase, Directed evolution, Degradation tag, Protease
Background

Rising energy consumption and the finite supply of fossil fuels are global challenges that demand sustainable alternative strategies [1-3]. Alkanes, the major components of conventional fuels [4], are mainly derived from non-renewable resources and can be naturally synthesized in cyanobacteria, green algae, insects and plants. These natural alkane/ene biosynthetic pathways are promising starting points for metabolic engineers to manipulate for biofuel production, serving as an ideal replacement for non-renewable fossil fuels [4-6]. Schirmer et al. (2010) identified two key enzymes responsible for alkane/ene biosynthesis in cyanobacteria: acyl-ACP reductase and aldehyde decarbonylase (AD), which catalyze the reduction of fatty acyl-ACP into the corresponding aldehyde and conversion of fatty aldehydes into alkane/enes and formate, respectively [7]. Since then, great progress has been made in producing alkanes in microbial cell factories via fatty acid synthases (FAS) and reverse beta-oxidation (RBO) pathways and in vitro enzymatic electrosynthesis assisted by AD (Fig. 1) [8-15]. To our knowledge, the highest reported value for long chain alkane production is 2.54 g/L (C13-C17) in engineered *Escherichia coli* [9], but this is insufficient for industrial application and for short length alkane (C<6) the titer drop significantly. It has been reported that the low efficiency of AD is the bottleneck of alkane production [4, 16, 17], which warrants further investigation of AD’s functionality in order to overcome potential biochemical limitations for alkane production.

Three distinct AD groups have been discovered (Fig. 1): AD in higher plants and green algae, that converts the aldehyde carbon to carbon monoxide [18-20]; AD in insects, a cytochrome P450 enzyme, which oxidizes the aldehyde carbon to carbon dioxide using the NADPH and molecular O₂ [21, 22]; AD in bacteria, that converts the aldehyde carbon to formate [7, 23, 24]. Among these, both plant and insect ADs are membrane proteins, and plant-type AD requires complexing with specific acyl CoA reductase and reducing equivalents from cytochrome b [25], overall complicating their characterization in vitro. In contrast, bacterial ADs are small soluble proteins, which allows them to be studied in vitro more readily, leading to their wider use in metabolic engineering. Hence structures, catalytic mechanisms, kinetic property and thermostability of bacterial Ads have been extensively studied [13, 17, 26-29]. Bacterial AD structures have a α-helical architecture with a ferritin-like four helix bundle as the catalytic domain. In the absence of oxygen, AD from cyanobacteria has recently been proven to catalyze the unusual hydrolysis of aldehydes to alkanes and formate in a reaction that requires electron transfer systems composed of
ferredoxin (Fd) and ferredoxin-NADP+ reductase (FNR) [29]. Other reports demonstrated that fusing AD with electron transfer systems (Fd/FNR) shows higher catalytic efficiency, which is two- to three-fold greater than the native AD in the presence of non-fused Fd/FNR [30]. Although new variants of AD have been discovered and engineered for alkane production [13, 24, 26, 31, 32], our experiments showed that AD has low abundance \textit{in vivo}. It was also reported that there is no significant accumulation of most bacterial AD in \textit{E. coli} during incubation[17], and even a decrease in AD abundance after 10 h of incubation [32]. The low levels of AD might be due to fast \textit{in vivo} degradation, which can limit the application of AD for alkane biosynthesis.

**Fig. 1.** Metabolic strategies for alkane production. There are three distinct AD groups catalyzing the formation of alkane via different mechanisms. Both plant and insect ADs are membrane proteins, whereas bacterial ADs are localized in cytosol. Abbreviations: AD, aldehyde decarbonylase; FAS, fatty acid synthesis; RBO, reverse beta-oxidation.

Directed evolution has proven to be a powerful strategy for improving enzyme properties [33-36]. Error-prone PCR, developed and improved by Goeddel and co-workers [37], can introduce one to 16 mutations per kb into the target gene. Since high concentrations of aldehydes are toxic to \textit{E. coli} cells, they can be used to establish a selection pressure to select cells containing AD variants with increased performance. In this study, error-prone PCR and hexanal were used in a directed evolution system to discover a degron in the C-terminal region of AD from \textit{Prochlorococcus marinus}. Since degrons are normally used to regulate expression and activity in cells [38, 39], we use a bioinformatic and experimental approach to better understand bacterial AD
regulation and found that this C-terminal degron is conserved in the family of bacterial ADs. Knock-out strains were constructed and establish that AD degradation occurs through protein recognition by ATP-dependent proteases ClpAP and Lon. Besides, our work proves that deletion or modification of the AD degron represents a novel strategy for developing efficient microbial cell factories for alkanes.

**Results**

**Discovery of C-terminal Degron via Directed Evolution**

For directed evolution of ADs, the mutant variants of *Prochlorococcus marinus* AD (ADpm, Uniprot ID Q7V6D4) were generated using error-prone PCR, and *E. coli* cells were transferred to a chemostat cultivation system containing 2g/L hexanal which was found to inhibit the growth of *E. coli* cells without recombinant AD. We hypothesized that this system would select for AD variants with increased rates of hexanal decarbonylation to pentane, which is less toxic (Fig. 2A).

**Fig. 2.** Discovery of a degron in ADpm via directed evolution. (A) Schematic of directed evolution system designed for AD evolution. (B) Distribution analysis of the mutations across the α-helices of ADpm. (C) Crystal structure of ADpm (PDBID: 2OCS). There are 8 α-helixes in AD structure. α-helix 1, 2, 4 and 5 form the catalytic domain. (D) GFP degradation test with and without the ADpmC1-9 sequence. Fluorescence values were normalized to the number of cells by dividing by the OD$_{600}$. Data shows the mean value of five replicates.
After seven days of chemostat cultivation, 33 mutants were sequenced and the distribution of mutations in the AD coding sequence was analyzed. The highest mutation frequency was observed in the α4 and α8 helices (Fig. 2B), and the α8-truncated mutants, due to the nonsense mutations, were found to have the highest activity based on the amount of NADPH consumed (Supplementary Fig. 1). According to the crystal structure of ADpm, eight α-helices form the intact structure, among which α1, 2, 4 and 5 form the catalytic domain (Fig. 2C), therefore α8-truncated mutants might be filtered out due to a novel mechanism other than enhanced activity.

This was further probed in vitro, by expressing ADpm without the C-terminal helix, RMAAAALVS (ADpm\textsuperscript{C1-9}), herein termed ADpm-9. Enzymatic assays with purified ADpm-9 proteins revealed a decrease in specific activity and thermostability, whereas size-exclusion chromatography showed no change in the oligomeric state compared to ADpm (Supplementary Fig. 2). These results suggest that ADpm-9 might confer increased survival to \textit{E. coli} cells through increasing stability of enzyme. ADpm\textsuperscript{C1-9} shows similarity to \textit{E. coli} degron motif 1 (Supplementary Fig. 3) [40], so ADpm\textsuperscript{C1-9} was speculated to be a putative degron that shortens the half-life of ADpm. To test this hypothesis, the ADpm\textsuperscript{C1-9} was fused to green fluorescent protein (GFP) to monitor degradation over time \textit{in vivo}. The fluorescence increased slightly in the beginning four hours for \textit{E. coli} expressing ADpm\textsuperscript{C1-9}-tagged GFP, but decreased during the remainder of the incubation. In contrast, the fluorescence of untagged GFP increased till 19 h after inoculation and remained stable thereafter (Fig. 2D). Overall, ADpm\textsuperscript{C1-9} reduced the GFP fluorescence to 8.6% of the untagged GFP after 30 h of cultivation, which supports the hypothesis that the short C-terminal region of ADpm functions as a degron acting on degrading and reducing half-life of protein \textit{in vivo}.

Identification of the minimal AD degron sequence

To more precisely determine which part of the C-terminal region of ADpm is necessary for its degradation, GFP was tagged with different C-terminal segments of ADpm (3, 5, 10, 20, and 30 amino acids). Each tagged GFP was controlled by a constitutive synthetic promoter pJ23119. Compared to the untagged GFP, degradation was observed in all the tagged GFP strains (Fig. 3A). The degradation of GFP increased as the degron length increased from 3 (GFP+ADpm\textsuperscript{C1-3}) to 10 (GFP+ADpm\textsuperscript{C1-10}), but no further enhancement of degradation was observed beyond 10 residues.
These results indicate that the short C-terminal region of ADpm (10 amino acids) is the minimal AD degron sequence. 

**Fig. 3.** Degradation effect of C-terminal region of AD. (A) Fluorescence of GFP fused with different C-terminal sections of AD at 24 h after inoculation. (B) Western blot for GFP with and without 10 amino acids from the C terminus of AD. (C) Western blot for AD with and without elimination of 10 C-terminal sequence.

Western blot was used to analyze the protein degradation of GFP and ADpm triggered by the C-terminal degron. The results show that ADpmC1-10 present in either case resulted in a loss of band intensity, representing protein concentration (Fig. 3B-C). The protein concentration of untagged GFP increased with more incubation time, while the fluorescence of GFP+ADC1-10 decreased. Similarly, the protein concentration of wild-type ADpm decreased with more incubation time while protein accumulation was observed in ADpm-10 (Fig. 3C). These results confirm that a short C-terminal region of ADpm, specifically ADpmC1-10, leads to protein degradation *in vivo*.

**Recognition of C-terminal degron in bacterial ADs via bioinformatic analyses**

To expand our understanding in biostability of the bacterial AD, 607 protein sequences of bacterial AD homologues from the UniProtKB database were analyzed with multiple sequence alignments. This dataset was reduced to 371 sequences by removing redundant sequences. A phylogenetic tree was generated using MEGA 5.10’s NJ method with 100 bootstrap replications (Fig. 4A). For further analysis, eight representative sequences from different branches were
selected from the phylogenetic tree. A high degree of conservation was observed in C-terminal residues of the selected candidates (Fig. 4B), including basic amino acids (Arg or Lys) at the 10th position and non-polar Ala and Leu at the 7th and 4th positions, respectively. Ala-Ala dipeptides were also observed to be a common feature in the C terminus (Fig. 4B). These results suggest that the C-terminal degron detected in ADpm might be a conserved feature in the family of bacterial ADs. A statistical analysis was conducted for the last 10 amino acids of the C-terminal sequence from ADs in the reduced dataset. It was revealed that half of the positions have an amino acid frequency larger than 50% (Supplementary Table 1). The most conserved position is Ala at the 7th position with a frequency of 67.92% among the bacterial AD homologs (Fig. 4C). From this, RMSAYGLAAA appears to be a consensus sequence for the degron conserved in bacterial ADs, herein termed ADcon.

**Fig. 4.** Analysis of C-terminal sequences of AD homologs in bacteria. (A) Phylogenetic reconstruction of AD sequences. The phylogenetic tree was generated using MEGA 5.10’s NJ method with 100 bootstrap replications. (B) Multiple sequence alignment of the C-terminal regions of eight AD sequences, three representatives labelled A, B, C were used for subsequent investigation. (C) Statistical analysis of C-terminal sequence of 371 AD homologs. (D) GFP degradation test of three representative C-terminal peptides and the conserved degron sequence. GFP-A, GFP-B, GFP-C represent GFP fused with three C-terminal regions from (B). GFP-ADcon represent GFP fused with conserved sequence from (C). Fluorescence values were normalized to the number of cells by dividing by the OD$_{600}$. The data shown are from the mean of three biological replicates.
To test this hypothesis, the C-terminal sequence of three ADs from major branches of the phylogenetic tree, as well as the ADcon, were tagged onto GFP to test whether they would trigger protein degradation. All four C-terminal sequences caused marked GFP degradation (Fig. 4D), which proved that the C-terminus of bacterial AD homologs serve as degrons. Furthermore, it was observed that lack of some of the conserved features affected the degron efficacy. For example, the degron from *Euhalothece* (RMSAYGLREV), which lacked the Ala-Ala dipeptides, only caused 62% of GFP degradation after 25 h incubation compared to the ADcon which caused 94% GFP degradation. Besides, it was reported that AD from *Gloeobacter violaceus* PCC 7421 (7421ADO) shows higher protein level than AD from cyanobacterial strains [17]. Based on the sequence alignment, we observed significant amino acid change in the degron region of 7421ADO, which might be the reason for high protein level. These findings strongly suggest that the C-terminus of bacterial ADs is a conserved degron.

**Investigation of molecular mechanisms of AD degron-dependent protein degradation in *E. coli***

To pinpoint the mechanism of protein degradation triggered by the AD degron, ADcon was compared to reported C-terminal degrons in *E. coli* [40]. We found that ADcon shares sequence similarity to the ssrA degron (CAANDENYALAA) from *E. coli*, which is degraded by the ClpAP and ClpXP protease complexes [41]. To test the hypothesis that the degradation mechanism of AD degron is like ssrA’s, degron tagged GFP was transformed into constructed ∆ClpA, ∆ClpX, and ∆ClpP strains. In all four degron tagged GFP strains, fluorescence recovery was observed in both ∆ClpA and ∆ClpP strains, but not significant in the ∆ClpX strain, compared to WT *E. coli*. This suggests that AD degron triggers protein degradation using the ClpAP protease complex where ClpA is responsible for recognizing the signal peptide, unfolding the protein, and translocating it to ClpP for proteolysis (Fig. 5B).

However, GFP degradation was still observed in the ∆ClpP strain compared with the untagged GFP. Previous studies found that a single degron can be recognized by multiple protease complexes. This includes the Umu degron that is recognized by the Lon protease complex and ClpXP, as well as the ssrA degron that can be recognized by both ClpXP and ClpAP [42-44]. Lon protease was suspected to degrade AD degron-tagged GFP since it is known to be an
Fig. 5. Investigation of degradation mechanism of AD degron. (A) Effects of protease deletion on GFP degradation with data normalized by OD_{600}. (B) Diagram illustrating the recognition and degradation process of AD degron tagged protein. GFP-A, GFP-B, GFP-C and GFP-ADcon refer to Fig. 4D. AD degron can be recognized by the ClpA and Lon protease complexes. ClpA bound GFP will be unfolded and translocated into ClpP for subsequent degradation. Lon bound GFP will be unfolded and degraded simultaneously in itself. Fluorescence values were normalized to the number of cells by dividing by the OD_{600}. The data shown are from the mean of three biological replicates.

Efficient protease for non-native protein degradation [45]. To test whether Lon protease can recognize AD degron, we transformed the AD degron tagged-GFPs into a constructed ΔLon E. coli strain. The fluorescence was partially recovered in all tagged candidates when expressed in the E. coli ΔLon strain. Nonetheless, GFP degradation still occurred in all degron-tagged GFP samples even in the ΔClpP ΔLon strain. These results suggest that there are other proteases that can recognize AD degron in E. coli. The protease system in bacteria is complicated and vital for all biological pathways [46], which was supported by our observation that the physical appearance of the cell cultures bearing Lon and ClpP knock-outs had a stickier consistency compared to wild-type cells which are normally pasty. Hence, the need to maintain cell viability might preclude the possibility of deleting all proteases responsible for AD degron recognition in order to improve the biostability of AD, leaving only the option of managing the impact of the degradation tag.

Effects of AD degron engineering on alkane production

The three most commonly used ADs from Prochlorococcus marinus (ADpm, Uniprot ID – Q7V6D4), Nostoc punctiforme (ADnp, Uniprot ID B2J1M1) and Synechococcus elongatus (ADse,
Uniprot ID Q54764) were selected to investigate the effects of degrons modification on alkane production. First, degron parts were removed from all three ADs to create ADpm-9, ADnp-10, and ADse-10. Initial enzymatic screening revealed the specific activities of degron-free versions of ADpm, ADnp and ADse decreased from 14.9 to 9.6 (1.55 fold decrease), 90.9 to 16.6 (5.47 fold decrease), and 49.1 to 42.7 (1.14 fold decrease) mU/mg, respectively (Fig. 6A). In contrast, the degron-free versions increased in relative enzyme abundance 2.2, 2.65 and 3.3-fold in cells harboring ADpm, ADnp, ADse, respectively (Fig.6B). Consequently, pentane production increased from 3.1 to 3.9 mg/L with ADpm and from 5.9 to 9.7 mg/L with ADse in E. coli (Fig. 5C). We attribute this improvement in pentane production to the increase in AD protein abundance, which not only compensates for the enzyme activity dropping, but also benefits overall pentane accumulation. The pentane production decreased from 7.9 to 5.2 mg/L in E. coli harboring ADnp-9 compared with ADnp (Fig. 6C), because its 2.65-fold increase in enzyme abundance could not compensate the 5.47-fold decrease in activity. Our modelling results suggest the elimination of the C-terminal sequence has a negative effect on the substrate binding in all three candidates (Supplementary Fig. 4), which is consistent with the notion that residues far from the active site and the substrate binding site still contribute to the enzyme activity of AD [17].

Since deleting protease complexes poorly impacts cell viability and deleting the entire AD degron decreases enzyme activity, an alternative method to increase the half-life of AD is needed. We hypothesized that addition of amino acids after the degron would protect the AD from proteases. This was tested by adding a 6xHisTag after the native degron in wild-type ADse, which previously produced the most pentane (Fig. 6C). As shown in Fig. 6D, ADse with a 6xHisTag shows 1.9-fold and 2.1-fold increases in protein abundance and pentane production, respectively, compared to wild-type ADse. Surprisingly, the 6xHisTag also improved the specific activity, possibly because the addition of His tag affects the conformation of the C-terminal helix and somehow benefit the activity of ADse. Our docking results partially prove our hypothesis, one of the two highest scored interacting models shows ferredoxin can interact with the C-terminal helix (Fig. 6E), which indicate the C-terminal helix modification might benefit the recruitment of ferredoxin (Fig. 6D). Turning to pentane production, the highest titer come with overexpression of fdx and fpr (Fig. 6D), which further prove the availability of ferredoxin (fdx) and ferredoxin reductase (fpr) is necessary for electron transfer to AD for full activity. The protection principle was extended to GFP+ADcon in vivo to test the limit of the protective effect conferred by a
6xHisTag. Fig. 6E shows that the 6xHisTag protects GFP+ADcon from protein degradation during the cultivation (Fig. 6F). However, in comparison to wild-type GFP, the 6xHisTag did not totally block the degron from protease recognition (Supplementary Fig 5). In all, these results demonstrate that manipulating the AD degron can improve alkane production.

Fig. 6. Investigation of degron modification. Representative ADs from P. marinus (ADpm), N. punctiforme (ADnp) and S. elongatus (ADse). (A-C) Comparison of specific activities, enzyme abundance and pentane production in three representative ADs with and without degron. (D) Comparison of enzyme abundance and pentane production in ADse C-terminal his tag and N-
terminal his tag. (E) Modeling of ADse and ferredoxin form *E. coli* interaction. *E. coli* 2Fe2S ferredoxin (Uniprot ID P0A9R4) (colored pink) and ADSE decarbonylase (colored wheat yellow) interaction model predicted with GRAMMX protein-protein docking server. And C and N-terminal end of ADSE are designated. (F) C-terminal his tag effect on degron sealing. The data show the mean of three replicates, and all error bars show the S.D.

**Discussion**

Discussion

Given the importance of alkanes as the major liquid transportation fuel and the diminishing fossil fuel supplies, engineers have sought to find sustainable ways to produce alkane using the bacterial AD-based pathway [7, 9, 10]. AD was previously diagnosed as the bottleneck in the pathway [4]. It was reported that AD expressed in *E. coli* had low abundance and degradation problems [32]. However, the mechanism remained elusive, yet crucial for microbial fermentation application. In this study, we discovered a C-terminal degron in *P. marinus* AD via a directed evolution system, which has been successfully employed to improve activity of enzymes through the enhancement of solubility, thermostability, substrate affinity, and catalytic turnover [36]. For the first time, the directed evolution system was employed to discover a degron. Since degrons are ubiquitous in N-terminal, C-terminal, and internal regions of various proteins [40, 47, 48], our study marks new avenues to use directed evolution to explore protein degrons, and to correct them for improved enzyme abundance and potentially enzyme activity.

More importantly, a C-terminal degron (AD-con) was found to be conserved in the family of bacterial AD, which was surprising because degrons normally exist in stress proteins expressed in response to environmental changes [40]. This might be the result of its promiscuous substrate specificity that would convert metabolically vital aldehyde intermediates into volatile alkane/enes, ultimately resulting in a loss of carbon, which the cell would otherwise use for survival. Our product analysis showed various volatile compounds in the headspace (Supplementary Fig. 6). Therefore, we suggest that the existence of degron can protect microorganisms from losing carbon via volatile products. Growth curves of *E. coli* harboring ADse and ADse-10 in M9 medium containing different glucose concentrations were tested and showed a decrease in the growth rate in ADse-10 strain under all conditions (Supplementary Fig. 7), which supports the notion that the AD degron provide a fitness advantage during evolution. In addition, it was reported that the expression of the bacterial genes involved in the assimilation of alkanes is usually tightly regulated [49], which also suggests that there are regulatory processes present in the alkane metabolic pathway.
Our bioinformatics and experimental investigation strongly argue that virtually all bacterial ADs possess a C-terminal-degron and its elimination can increase alkane production \textit{in vivo}, while negatively affecting the catalytic efficiency. We also showed that one approach for mitigating the effect of the degron is the addition of extra amino acids after the native degron to protect it from protease recognition, which could confer moderate improvements in AD half-life and activity, ultimately improving alkane production. However, additional methods for managing the trade-off between activity and stability are needed in order to further enhance alkane production. Additionally, any increase in activity and stability also must address the specificity of the AD toward the desired aldehydes relative to the various aldehyde intermediates in central metabolism. In this study, we have taken the first step towards addressing this trade-off through the comprehensive characterization of the role of the degron in AD and its impact on their stability and activity. We anticipate that this work will contribute to transitioning AD-based biosynthesis of alkanes to industrial scales.

\textbf{Conclusion}

For the first time, degradation tag (degron) was discovered in the AD from marine cyanobacterium \textit{Prochlorococcus marinus} \textit{via} error-prone PCR based directed evolution system. Bioinformatic analyses was carried out and revealed this C-terminal degron is common in the family of bacterial ADs and identified a conserved C-terminal motif, RMSAYGLAAA, representing the AD degron (ADcon). Furthermore, we demonstrated that the ATP-dependent proteases ClpAP and Lon are involved in the degradation of AD-tagged proteins in \textit{E. coli}, thereby limiting alkane production. Deletion or modification of the degron motif increased alkane production \textit{in vivo}. Thus, our work revealed the presence of a novel degron in bacterial ADs and paves way for its engineering for microbial production of sustainable fuels.

\textbf{Methods}

\textbf{General molecular Methods}

\textit{Escherichia coli} K-12 strains MG1655 was used for directed evolution and degradation mechanism investigation. RARE[4] strain was used for \textit{in vivo} fermentation. Random mutagenesis of AD by error-prone PCR was performed with the GeneMorphII Random Mutagenesis Kit (Agilent Technologies). ADs were cloned into pZa23MCS without codon optimization. Plasmids and DNA was purified using Monarch kits (NEB) according the manufacturer’s instructions. Restriction digests were carried out using standard protocols of NEB Restriction Endonucleases. For ligation of DNA fragments, ligaseT4 DNA ligase (Life Technologies) was used according to manufacturer’s instructions. Gene deletions were performed by-red mediated recombination [50]. The I-TASSER Suit was employed to generate structural model of ADs minus degrons[51].
Autodock Vina was used for molecular docking [52]. Protein stability was predicted by SCooP [53]. GRAMMX server was used for protein-protein interaction prediction [54].

**Culture conditions**

For fluorescence or fermentation, cells were culture in LB (Luria-Bertani) medium. For in vivo pentane synthesis, seed cultures were grown in LB medium at 30 °C overnight on a rotary shaker at 250 rpm, and were used to inoculate, at an inoculation volume of 1%, 1 ml LB in 15 mL tube for aerobic growth. Cultures were induced with 1 mM IPTG at OD of 0.6. Cells were transfer to anaerobic fermentation 4 h after IPTG induction and re-inoculated into 1 mL LB culture with 10 g/l glycerol, 1 mM IPTG and 1 g/L hexanal in sealed bottle for anaerobic fermentation 12 h. The head space was used for pentane quantification and the culture were used for protein level analysis. In general experiments were performed in triplicate, and the data are given as the averages and SDs.

For protein purification, starter cultures were grown in LB medium at 37 °C overnight on a rotary shaker at 250 rpm, and were used to inoculate, at an inoculation volume of 2%, 1 L TB (Terrific Broth) medium. Cultures were grown aerobically at 37 °C until OD$_{600}$ between 0.8-1, induced with 0.4 mM IPTG and left at 37 °C for 5 h. Biomass was collected by spinning at 3000 g, and frozen at -20 °C for purification.

**Western Blots**

Whole cell lysis was separated on a 18% SDS-PAGE gel. Sample amount was equaled base on culture OD. Protein on gel were transferred to a nitrocellulose membrane and detected by a primary 6x-His Tag Antibody (Thermofisher, USA) from mouse and goat anti-mouse conjugated to alkaline phosphatase or HRP as a secondary antibody (Bio-Rad). Protein with his tag on the nitrocellulose membrane was detected by using Opti-4CN kit (Bio-Rad).

**GC-MS and GC-FID Methods**

For the analysis of short chain alkanes produced, the sample headspace was analyzed by EI GC–MS. GC–MS analyses were performed on a Varian 2100T, with the detector in EI and operated at 10 eV. An Agilent VF-5ms column (30 m length, 0.25 mm inner diameter, 0.25 µm film thickness) was used at 50:1 split ratio. The ion source temperature was set to 200 °C. GC analyses were performed as follows. The oven temperature was maintained at 40 °C for 4 min. Temperature was increased at a rate of 50 °C/min up to 200 °C and maintained for 1 min, followed by an increase of 20 °C/min up to 270 °C and maintained for 3.3 min. 0.5 ml of headspace was injected. The retention times and fragmentation patterns were compared with those obtained from pure standards. Pentane were quantitatively analyzed by GC-FID. GC–FID analysis was performed with a Hewlett-Packard 5890 series II gas chromatograph equipped with a flash heat split inlet; and 30 m Long, 0.53 mm id capillary column (J&W Scientific). The GC program was as follows: initial temperature of 40 °C, hold for 4 min; then ramp to 190 °C at a rate of 40 °C per min and hold for 6 min.

**Enzyme kinetics**

Whole cell lysis activity analysis was carried in 100 mM PBS buffer pH 7.5, 0.1 M KCl, supplied with 0.2 mM FeCl$_2$, 8 mM hexanal and 2 µL of each cell lysis in 100 µL volume. Purified Enzyme assay was carried in 100 mM HEPES pH 7.5, 0.1 M NaCl, supplied with 2 mM NADH, 0.3 mM FeCl$_2$, 0.1 mM PMS, 10 mM hexanal and 50 µg of each decarbonylase in 0.2 ml liquid volume, 1.8 mL N$_2$ gas phase in gas tight vials. Reaction premix and enzymes were degassed and sparged.
with Argon. Water for reaction mixture dilution was degassed, sparged with argon and oxygen leftovers were removed by titanium citrate. Reaction was started in N₂ atmosphere glove box, vials were tightly capped and incubated on shaker at 200 rpm, 37 ºC. Reaction was stopped by heating at 100 ºC for 5 min. 200 µL of gas phase were analyzed using GC-MS protocol.

**Protein purification**

*E. coli* cell thawed biomass was diluted in binding buffer (50 mM HEPES, pH 7.5, 0.4 M NaCl, 5% glycerol, 5 mM imidazole) and sonicated on ice bath during 15 minutes following the 3 sec ON (120-140W) and 4 sec OFF regime using Qsonica Q700 equipped with dual-horn probe. Cell lysate was spun down at 37 000 g for 30 min, 4 ºC, and supernatant was applied to the Ni-agarose equilibrated with binding buffer. Resin was washed with 50 volumes of washing buffer (50mM HEPES, pH 7.5, 0.4 M NaCl, 5% glycerol, 30 mM imidazole) and eluted with 3-6 ml of elution buffer (50 mM HEPES, pH 7.5, 0.4 M NaCl, 5% glycerol, 250 mM imidazole). Proteins were frozen in droplets using liquid N₂ and stored at -80 ºC.

**Analytical size exclusion chromatography**

Analytical size exclusion chromatography was performed on Superdex 200 16/60 GL column. The equilibration buffer contained 50 mM HEPES pH 7.5, 0.1 M NaCl. For oligomeric state testing 1-2 mg of protein was loaded to the column and eluted for 2 column volumes. Relative protein size was estimated based on BioRad chromatography protein standards run.

**Fluorescent assay with whole cells**

The degradation of GFP was quantified using a fluorescent assay. For this purpose, cell were cultured in 24 well plates. The fluorescence was monitored by fluorescence in a microtiter plate reader (excitation wavelength, 488nm; emission wavelength, 535 nm; Infinite M1000 PRO, Tecan Group AG) at 37 °C and agitation (6-mm amplitude, orbital). Fluorescence signal in the linear range of the reaction normalized by the OD₆₀₀ of the respective culture.

**Thermal shift assay**

Protein thermostability (Tm, melting temperature) was determined using SYPRO Orange dye kit (ThermoFisher) on the CFX96 Touch Real-time PCR Detection System reading FRET channel on λ_ex 492 nm and λ_em 610 nm. Reaction volume was 20 µL, protein concentration 0.1 mg/mL. In order to optimize the fluorescence signal to noise ratio an assay optimization was performed.
Ethics declarations

Ethics approval and consent to participate
Not applicable—no human subjects involved in study.

Consent for publication
Not applicable—no human subjects involved in study.

Competing interests
The authors declare that they have no competing interests.

Acknowledgments
We thank Prof. Kristala L.J. Prather at Massachusetts Institute of Technology for supplying RARE strain. We thank Shen Guo at the University of Toronto for help with GC-FID.

Founding
This project was funded by Suncor Energy Inc, NSERC CRD grant (CRDPJ 517554-17) and Ontario Centres of Excellence VIP Grant (OCE-VIP29120).

Authors’ contributions
RM supervised the work and revised the manuscript. YL, JC and AK designed the experiments. YL and JC generated all the genetic constructs, analyzed the data and wrote the manuscript. AK, KB and PD carried out protein purification and characterization. AY and KC contributed to scientific discussions and comments on the manuscript. KN assisted fermentation. RF carried out GC-MS analysis, KB, PS and YL carried out protein modelling. PD, KC, AY and PS revised the manuscript. All authors read and approved the final manuscript.

Availability of data and material
All data are included in the manuscript.

Abbreviations
AD: Aldehyde decarbonylase; fdx: Ferredoxin; fpr: Ferredoxin reductase; PMS: phenazine methosulfate; clpA: ATP-dependent protease; clpP: serine protease; clpX: ATP-dependent protease; lon: Lon protease; Umu: umuC and umuD DNA polymerase; ssrA: transfer-messenger RNA; PMS: Phenazine methosulfate.

References
1. Jiang Y, Wu R, Zhou J, He A, Xu J, Xin F, Zhang W, Ma J, Jiang M, Dong W: Recent advances of biofuels and biochemicals production from sustainable resources using co-cultivation systems. Biotechnology for biofuels 2019, 12(1):155.
2. Bai W, Geng W, Wang S, Zhang F: Biosynthesis, regulation, and engineering of microbially produced branched biofuels. Biotechnology for biofuels 2019, 12(1):84.
3. Marella ER, Holkenbrink C, Siewers V, Borodina I: Engineering microbial fatty acid metabolism for biofuels and biochemicals. Current opinion in biotechnology 2018, 50:39-46.
21. Reed JR, Vanderwel D, Choi S, Pomonis JG, Reitz RC, Blomquist GJ: Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex
pheromone component (Z)-9-tricosene and CO2. Proceedings of the National Academy of Sciences 1994, 91(21):10000-10004.

22. Ellington BR, Paul B, Das D, Vitak AK, Zimmerman PM, Marsh ENG: An Unusual Iron-Dependent Oxidative Deformylation Reaction Providing Insight into Hydrocarbon Biosynthesis in Nature. ACS Catalysis 2016, 6(5):3293-3300.

23. Zhang J, Lu X, Li J-J: Conversion of fatty aldehydes into alk (a/e) nes by in vitro reconstituted cyanobacterial aldehyde-deformylating oxygenase with the cognate electron transfer system. Biotechnology for biofuels 2013, 6(1):86.

24. Bao L, Li J-J, Jia C, Li M, Lu X: Structure-oriented substrate specificity engineering of aldehyde-deformylating oxygenase towards aldehydes carbon chain length. Biotechnology for biofuels 2016, 9(1):185.

25. Marsh ENG, Waugh MW: Aldehyde decarbonylases: enigmatic enzymes of hydrocarbon biosynthesis. In.: ACS Publications; 2013.

26. Khara B, Menon N, Levy C, Mansell D, Das D, Marsh ENG, Leys D, Scrutton NS: Production of Propane and Other Short-Chain Alkanes by Structure-Based Engineering of Ligand Specificity in Aldehyde-Deformylating Oxygenase. ChemBioChem 2013, 14(10):1204-1208.

27. Rajakovich LJ, Nørgaard H, Warui DM, Chang W-c, Li N, Booker SJ, Krebs C, Bollinger Jr JM, Pandelia M-E: Rapid reduction of the diferric-peroxyhemiacetal intermediate in aldehyde-deformylating oxygenase by a cyanobacterial ferredoxin: Evidence for a free-radical mechanism. Journal of the American Chemical Society 2015, 137(36):11695-11709.

28. Shakeel T, Gupta M, Fatma Z, Kumar R, Kumar R, Singh R, Sharma M, Jade D, Gupta D, Fatma T: A consensus-guided approach yields a heat-stable alkane-producing enzyme and identifies residues promoting thermostability. Journal of Biological Chemistry 2018; jbc. RA117. 000639.

29. Eser BE, Das D, Han J, Jones PR, Marsh ENG: Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. Biochemistry 2011, 50(49):10743-10750.

30. Wang Q, Huang X, Zhang J, Lu X, Li S, Li J-J: Engineering self-sufficient aldehyde deformylating oxygenases fused to alternative electron transfer systems for efficient conversion of aldehydes into alkanes. Chemical Communications 2014, 50(33):4299-4301.

31. Kang M-K, Zhou YJ, Buijs NA, Nielsen J: Functional screening of aldehyde decarbonylases for long-chain alkane production by Saccharomyces cerevisiae. Microbial cell factories 2017, 16(1):74.

32. Patrikainen P, Carbonell V, Thiel K, Aro E-M, Kallio P: Comparison of orthologous cyanobacterial aldehyde deformylating oxygenases in the production of volatile C3-C7 alkanes in engineered E. coli. Metabolic engineering communications 2017, 5:9-18.

33. Jeschek M, Reuter R, Heinisch T, Trindler C, Klehr J, Panke S, Ward TR: Directed evolution of artificial metalloenzymes for in vivo metathesis. Nature 2016, 537(7622):661.

34. Bali AP, Genee HJ, Sommer MO: Directed Evolution of Membrane Transport Using Synthetic Selections. ACS synthetic biology 2018, 7(3):789-793.

35. Zeymer C, Hilvert D: Directed evolution of protein catalysts. Annual review of biochemistry 2018(0).

36. Packer MS, Liu DR: Methods for the directed evolution of proteins. Nature Reviews Genetics 2015, 16(7):379.

37. Drummond DA, Iversen BL, Georgiou G, Arnold FH: Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. Journal of molecular biology 2005, 350(4):806-816.

38. Sauer RT, Baker TA: AAA+ proteases: ATP-fueled machines of protein destruction. Annual review of biochemistry 2011, 80:587-612.
39. Varshavsky A: N-degron and C-degron pathways of protein degradation. Proceedings of the National Academy of Sciences 2019, 116(2):358-366.
40. Flynn JM, Neher SB, Kim Y-I, Sauer RT, Baker TA: Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. Molecular cell 2003, 11(3):671-683.
41. Flynn JM: Substrate selection by the ClpXP protease: a tail of destruction. Massachusetts Institute of Technology; 2004.
42. Frank EG, Ennis DG, Gonzalez M, Levine AS, Woodgate R: Regulation of SOS mutagenesis by proteolysis. Proceedings of the National Academy of Sciences 1996, 93(19):10291-10296.
43. Olivares AO, Kotamarthi HC, Stein BJ, Sauer RT, Baker TA: Effect of directional pulling on mechanical protein degradation by ATP-dependent proteolytic machines. Proceedings of the National Academy of Sciences 2017, 114(31):E6306-E6313.
44. Gottesman S, Roche E, Zhou Y, Sauer RT, Baker TA: The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes & development 1998, 12(9):1338-1347.
45. Hansen J, Gregersen N, Bross P: Differential degradation of variant medium-chain acyl-CoA dehydrogenase by the protein quality control proteases Lon and ClpXP. Biochemical and biophysical research communications 2005, 333(4):1160-1170.
46. Mahmoud SA, Chien P: Regulated proteolysis in bacteria. Annual review of biochemistry 2018(0).
47. Striebel F, Kress W, Weber-Ban E: Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. Current opinion in structural biology 2009, 19(2):209-217.
48. Olivares AO, Baker TA, Sauer RT: Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. Nature Reviews Microbiology 2016, 14(1):33.
49. Moreno R, Rojo F: Genetic Features and Regulation of n-Alkane Metabolism in Bacteria. Aerobic Utilization of Hydrocarbons, Oils, and Lipids 2019:521-542.
50. Liu Y, Yang M, Chen J, Yan D, Cheng W, Wang Y, Thygesen A, Chen R, Xing J, Wang Q: PCR-based seamless genome editing with high efficiency and fidelity in Escherichia coli. PLOS one 2016, 11(3):e0149762.
51. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y: The I-TASSER Suite: protein structure and function prediction. Nature methods 2015, 12(1):7.
52. Trott O, Olson AJ: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry 2010, 31(2):455-461.
53. Pucci F, Kwasigroch JM, Rooman M: SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. Bioinformatics 2017, 33(21):3415-3422.
54. Tovchigrechko A, Vakser IA: GRAMM-X public web server for protein–protein docking. Nucleic acids research 2006, 34(suppl_2):W310-W314.
**Supplementary Fig. 1.** Whole cell lysis activity analysis of different mutants. M2 give the highest activity, which has a premature stop codons in the C-terminal region. Note, WT, wild type ADpm, M1-M9 different mutants.
Supplementary Fig. 2 Comparison of ADpm and ADpm-9. (A) Specific activity of ADpm and ADpm-9. (B) Relative protein thermal stability estimated using SYPRO Orange dye. (C) Investigation of oligomeric state of ADpm and ADpm-9 using size exclusion chromatography.
**C-motif 1**

| ssrA  | LAA-COOH |
|-------|-----------|
| YdaM* | LAA       |
| GlpD  | LAS       |
| Crt†  | LTA       |
| PepB  | LTA       |
| Tpx   | LKA       |
| RpiU  | ISA       |
| DksA  | MAG       |
| IscR  | LRA       |
| RpiJ  | EAA       |
| LdD†  | NAA       |
| Gcp†  | PAA       |
| KatE  | IPA       |
| AldA  | LCS       |
| NrdH  | ASA       |
| CysA  | QSA       |
| AcmB  | TAV       |
| GlyA  | VYA       |
| CheW  | EVA       |
| Fnr   | NVA       |
| YcbW  | QAV       |
| ExbB  | RAG       |

\[ \text{ADpm}^{C1-9}:\text{RMAAAALVS} \]

**Supplementary Fig. 3.** Comparison of C-terminal of ADpm and C-motif 1\[40\] of protease trapped proteins
Supplementary Fig. 4. Modelling results of degron subtraction in three representative ADs from *P. marinus* (ADpm), *N. punctiforme* (ADnp) and *S. eLongates* (ADse). (A) Thermostability and hexanal docking analysis of three ADs and their C-terminal truncations. (B-D) Examples of docking analysis in ADpm and ADpm-9. (B) ADpm, (C) ADpm-9, (D) Superimposed structures.
Supplementary Fig. 5. Effect of C-terminal his tag on GFP degradation. C-terminal 6His tag was added to GFP with conserved degron.
Supplementary Fig. 6. Effects of AD degron elimination on fermentation. Headspace analysis of ADse (Red) and ADse-10 (Blue). Note, U1-U7, unidentified peaks.
Supplementary Fig. 7. Effects of AD degron elimination on cell growth. (A-D) Growth curves of ADse (black) and ADse-10 (red) in M9 medium containing different concentrations (0.5, 1, 2 and 10 g/L) of glucose.
Supplementary Table 1. Statistical analysis of amino acid usage frequency (%) at C-terminus in 371 bacterial ADs.

|   | C10 | C9  | C8  | C7  | C6  | C5  | C4  | C3  | C2  | C1  |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A | 16.17 | 8.89 | 4.31 | 67.92 | 32.61 | 25.61 | 31.27 | 27.49 | 54.18 | 45.55 |
| R | 49.60 | 19.95 | 9.43 | 2.16 | 1.62 | 0.27 | 2.16 | 8.89 | 1.08 | 1.08 |
| N | 0.27 | 0.00 | 0.27 | 1.35 | 0.27 | 0.00 | 0.27 | 0.27 | 1.35 | 0.27 |
| D | 0.54 | 1.08 | 0.54 | 0.27 | 0.00 | 0.27 | 0.27 | 0.81 | 0.27 | 0.00 |
| C | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.27 | 1.35 | 0.27 | 0.00 |
| E | 1.62 | 0.54 | 0.27 | 0.00 | 0.27 | 0.54 | 0.81 | 0.54 | 1.08 | 0.54 |
| Q | 0.27 | 0.27 | 1.35 | 1.35 | 3.50 | 2.96 | 1.08 | 0.81 | 0.81 | 0.27 |
| G | 0.27 | 0.27 | 2.16 | 1.62 | 2.43 | 56.06 | 3.23 | 2.96 | 6.20 | 7.55 |
| H | 0.00 | 0.81 | 0.00 | 0.54 | 4.85 | 0.27 | 0.27 | 0.00 | 0.54 | 0.27 |
| I | 9.16 | 2.70 | 0.81 | 0.27 | 1.35 | 0.27 | 1.08 | 3.77 | 6.47 | 5.93 |
| L | 3.50 | 27.76 | 2.43 | 2.16 | 2.16 | 4.85 | 56.33 | 22.64 | 3.23 | 1.08 |
| K | 5.39 | 2.70 | 0.54 | 0.00 | 0.00 | 0.54 | 0.27 | 2.70 | 0.81 | 0.00 |
| M | 4.04 | 28.84 | 21.29 | 9.16 | 4.04 | 5.12 | 0.00 | 0.54 | 1.08 | 0.00 |
| F | 0.27 | 0.00 | 0.00 | 0.54 | 0.00 | 0.00 | 0.00 | 0.27 | 0.27 | 0.00 |
| P | 1.35 | 0.00 | 0.00 | 1.08 | 0.27 | 0.27 | 0.00 | 0.81 | 0.81 | 1.35 |
| S | 0.81 | 2.96 | 55.26 | 8.36 | 2.43 | 1.08 | 1.08 | 3.77 | 3.50 | 6.20 |
| T | 4.85 | 1.62 | 1.35 | 0.27 | 0.81 | 0.27 | 0.81 | 16.98 | 7.82 | 0.81 |
| W | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Y | 0.00 | 0.54 | 0.00 | 0.81 | 42.86 | 1.08 | 0.00 | 0.27 | 0.27 | 0.00 |
| V | 1.89 | 1.08 | 0.00 | 2.16 | 0.54 | 0.54 | 1.08 | 6.47 | 9.16 | 28.30 |