Biotechnology for Biofuels

RESEARCH

Open Access

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

Yilan Liu[1] , Jinjin Chen[1], Anna N. Khusnutdinova[1], Kevin Correia[1], Patrick Diep[1], Khorcheska A. Batyrova[1], Kayla Nemr[1], Robert Flick[1], Peter Stogios[1], Alexander F. Yakunin[1,2] and Radhakrishnan Mahadevan[1,3*]

Abstract

Background: Aldehyde decarbonylases (ADs), which convert acyl aldehydes into alkanes, supply promising solution for producing alkanes from renewable feedstock. However the instability of ADs impedes their further application. Therefore, the current study aimed to investigate the degradation mechanism of ADs 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* using error-prone PCR-based directed evolution system. Bioinformatic analysis revealed that this C-terminal degron is common in 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.

Conclusion: 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. Since Schirmer et al. (2010) identified two key enzymes responsible for alkane/ene biosynthesis in cyanobacteria [5], significant 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) [6–13]. To our knowledge, the highest reported value for long chain alkane production is 2.54 g/L (C13-C17) in engineered *Escherichia coli* [7], but this is still insufficient for industrial scale-up. It has been reported that the low efficiency of AD is the bottleneck of alkane production [4, 14, 15], which warrants further investigation into how ADs behave in vivo.
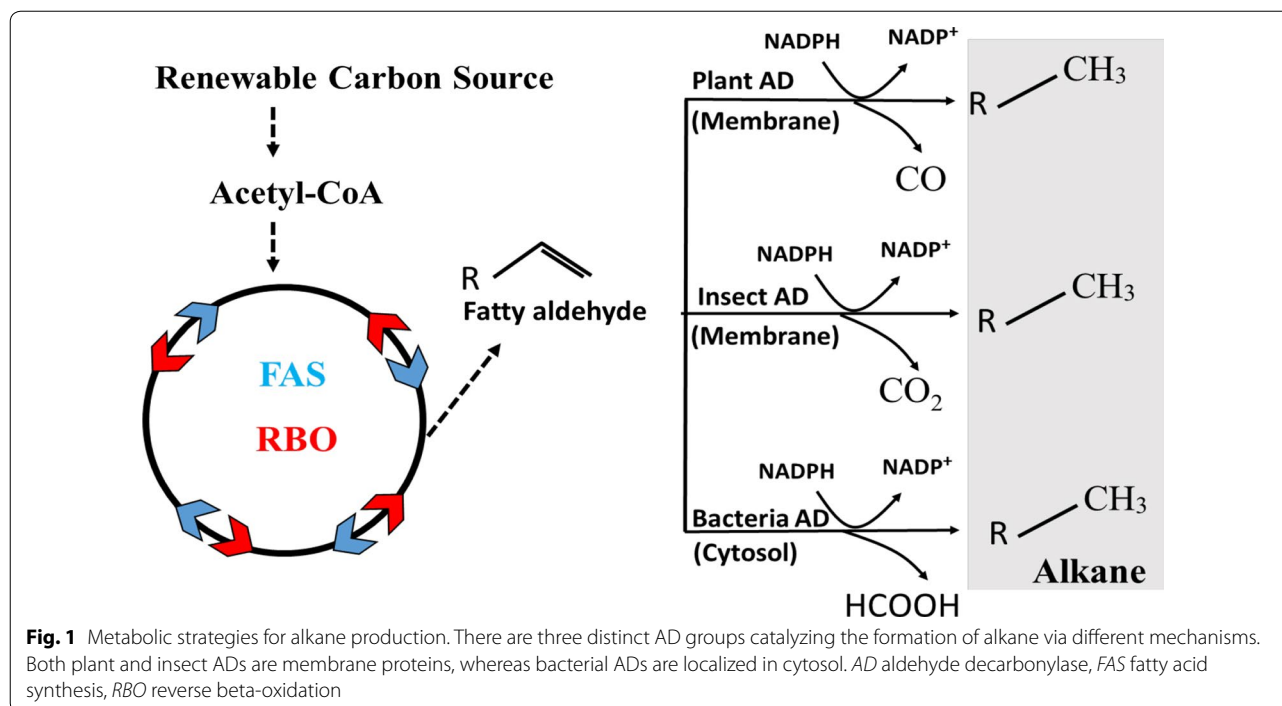
Three distinct groups of ADs have been discovered (Fig. 1): AD in higher plants and green algae, that converts the aldehyde carbon to carbon monoxide [16–18]; AD in insects, a cytochrome P450 enzyme, which oxidizes the aldehyde carbon to carbon dioxide using the

*Correspondence: krishna.mahadevan@utoronto.ca
[1] Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, ON M5S 3E5, Canada
Full list of author information is available at the end of the article

Liu *et al. Biotechnol Biofuels*   (2020) 13:114

Page 2 of 11

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. *AD* aldehyde decarbonylase, *FAS* fatty acid synthesis, *RBO* reverse beta-oxidation

NADPH and molecular O_2 [19, 20]; AD in bacteria, that converts the aldehyde carbon to formate [5, 21, 22]. Among these, both plant and insect ADs are membrane proteins, 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 [11, 15, 23–26]. Although new bacterial AD is constantly being characterized and engineered for alkane production [11, 22, 23, 27, 28], our experiments showed that protein accumulation of bacterial ADs is limited in vivo. It was also reported that there is no significant accumulation of most bacterial AD in *E. coli* during incubation [15], and even a decrease in AD abundance after 10 h of incubation [28]. The low levels of AD might be due to fast in vivo degradation, which can limit the application of AD for alkane biosynthesis. This prompted us to attempt to improve the performance of bacterial ADs.

Directed evolution has proven to be a powerful strategy for improving enzyme properties [29–32].Therefore, we built up a directed evolution system to evolve bacterial ADs. Since high concentrations of aldehydes are toxic to *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 and resulted in the discovery of C-terminal degron of AD from *Prochlorococcus marinus*. Since degrons are normally used to regulate protein expression and activity in cells [33, 34],

we use bioinformatic and experimental approaches 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 established 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 2 g/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).

After 7 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 (Additional file 1: Figure S1). According to the crystal structure of

Liu *et al. Biotechnol Biofuels*   (2020) 13:114

Page 3 of 11
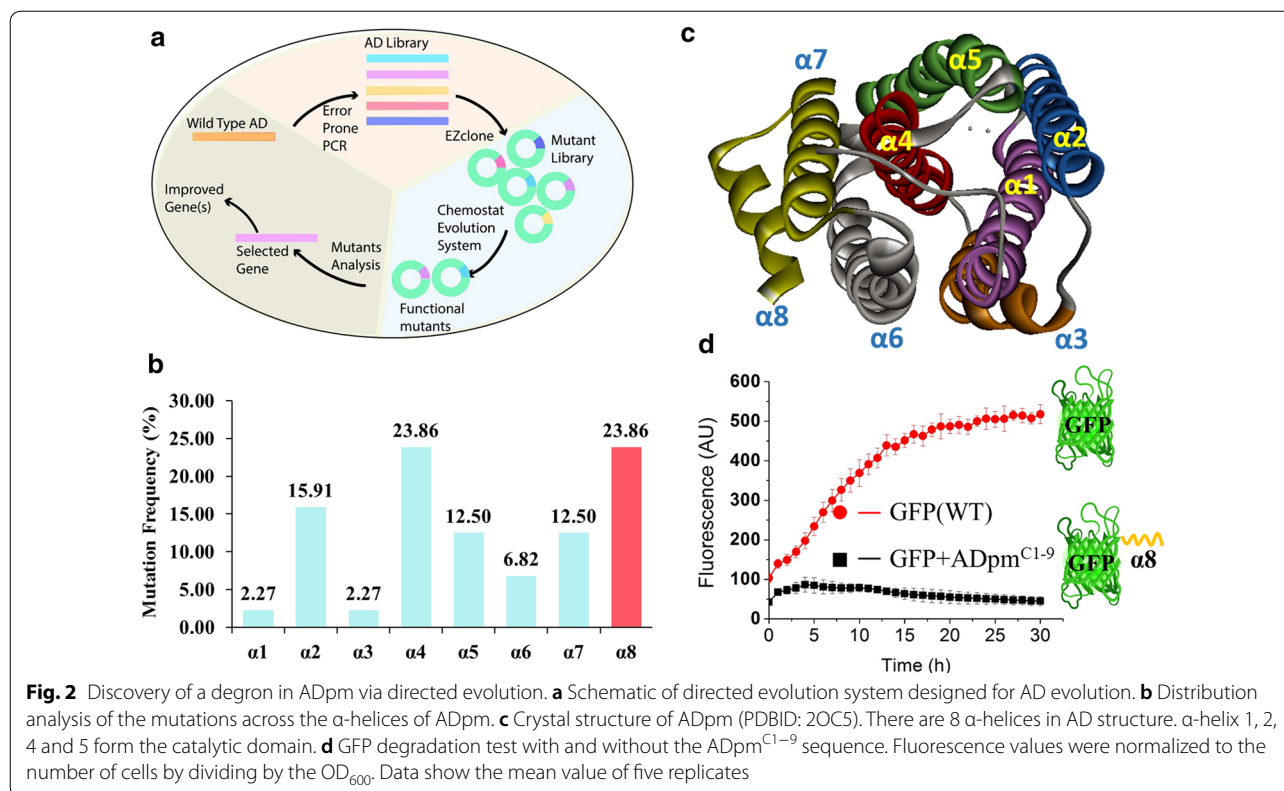
Fig. 2 Discovery of a degron in ADpm via directed evolution. **a** Schematic of directed evolution system designed for AD evolution. **b** Distribution analysis of the mutations across the α-helices of ADpm. **c** Crystal structure of ADpm (PDBID: 2OC5). There are 8 α-helices in AD structure. α-helix 1, 2, 4 and 5 form the catalytic domain. **d** GFP degradation test with and without the $ADpm^{C1-9}$ sequence. Fluorescence values were normalized to the number of cells by dividing by the OD_{600}. Data show the mean value of five replicates

ADpm, eight α-helices form the intact structure, among which α1, 2, 4 and 5 form the catalytic domain (Fig. 2c). Mutations in these domains might benefit the activity of AD. Since α8 helix is far from the catalytic domain, we hypothesized that α8-truncated mutants might be filtered out due to a novel mechanism other than enhanced activity.
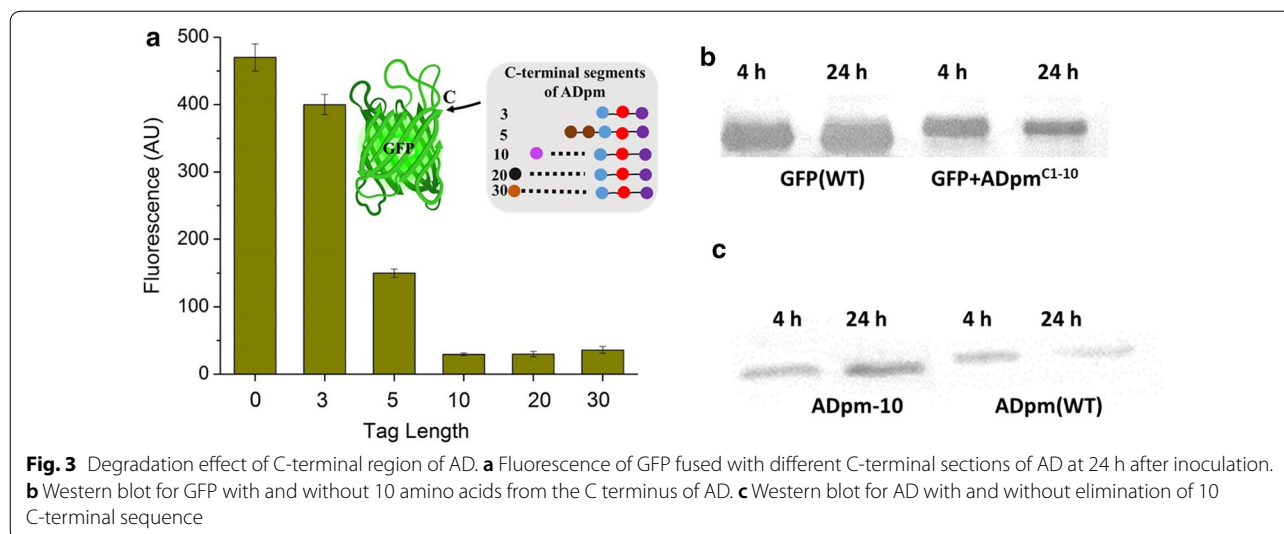
This was further probed in vitro, by expressing ADpm without the C-terminal helix, RMAAAALVS ($ADpm^{C1-9}$), herein termed ADpm-9. Enzymatic assays with purified ADpm-9 proteins revealed a decrease in specific activity and thermostability, whereas size-exclusion chromatography showed no change in the oligomeric state compared to ADpm (Additional file 1: Figure S2). The results lead us to hypothesize that *E. coli* expressing ADpm-9 exhibited improved alkane production because the biostability of the enzyme was improved, thus increasing its abundance. As $ADpm^{C1-9}$ shows rich of L, A, V and S as *E. coli* degron motif 1 (Additional file 1: Figure S3) [35], so $ADpm^{C1-9}$ was speculated to be a degron that shortens the half-life of ADpm. To test this hypothesis, the $ADpm^{C1-9}$ was fused to green fluorescent protein (GFP) to monitor degradation over time in vivo. The fluorescence increased slightly in the beginning four hours for *E. coli* expressing $ADpm^{C1-9}$-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, $ADpm^{C1-9}$ 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 reducing protein half-life 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 ($GFP + ADpm^{C1-3}$) to 10 ($GFP + ADpm^{C1-10}$), but no further enhancement of degradation was observed beyond 10 residues (Fig. 3a). These results indicate that the short C-terminal region of ADpm (10 amino acids) is the minimal AD degron sequence.

Western blot was used to analyze the protein degradation of GFP and ADpm triggered by the C-terminal degron. Fluorescence assays revealed a 90% decrease in fluorescence of $GFP + ADpm^{C1-10}$ compared to

Liu *et al. Biotechnol Biofuels* (2020) 13:114

Page 4 of 11

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
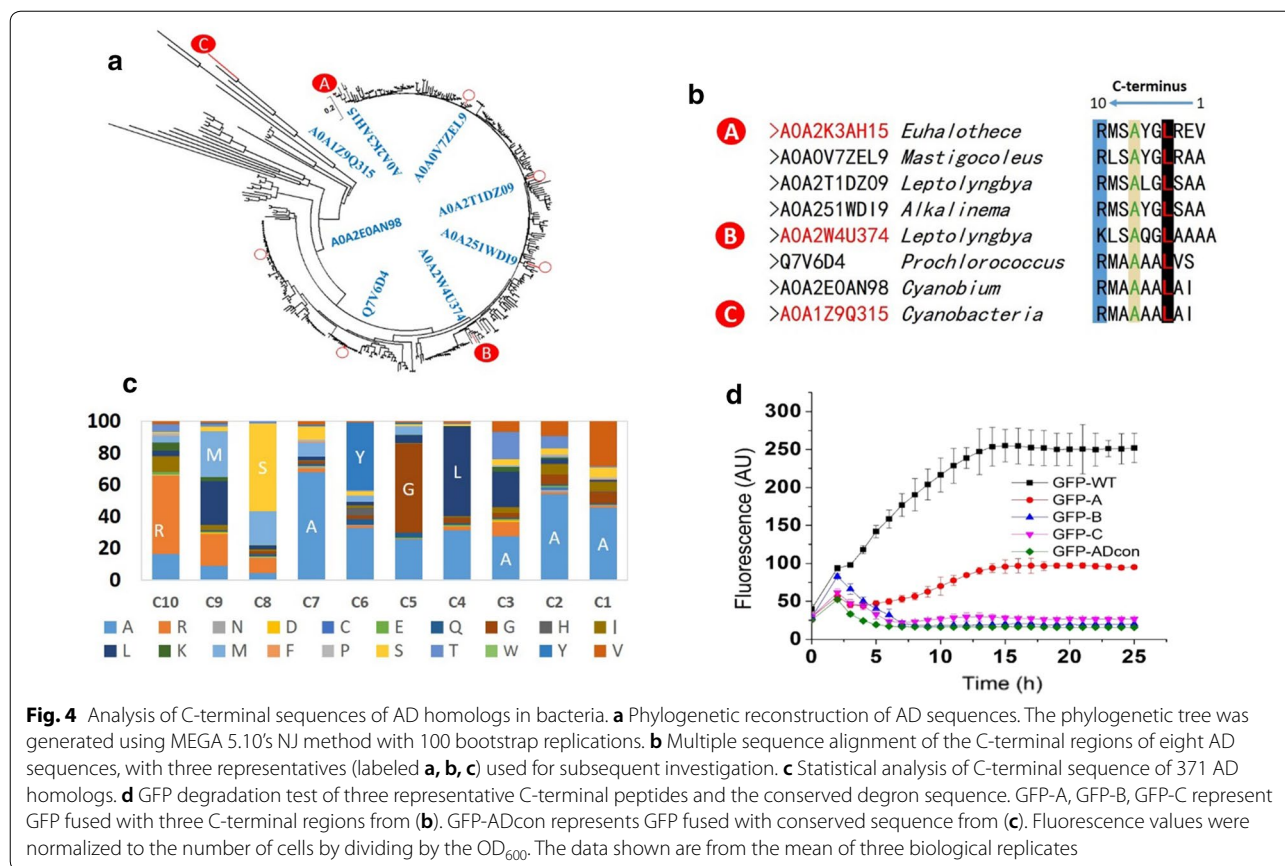
the untagged GFP, which is in line with approximately 90% reduction in the GFP protein level for the GFP + ADpm^{C1-10} construct (Fig. 3a,b). These results demonstrate fluorescence from the GFP protein is directly proportional to the abundance of the GFP protein. The protein concentration of untagged GFP increased with more incubation time, while the fluorescence of GFP + ADpm^{C1-10} decreased. Similarly, the protein concentration of wild-type ADpm decreased with more incubation time while protein accumulation was observed in ADpm-10 (with 10 C-terminal amino acids missing) (Fig. 3c). These results indicate that a short C-terminal region of ADpm, ADpm^{C1-10}, represents a degron whose presence leads to proteolytic degradation in vivo.

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

To expand our understanding of biostability of the bacterial ADs, over 600 protein sequences of bacterial AD homologues from the UniProtKB database were analyzed using 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 nonpolar 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 of ADs (Fig. 4b). These results suggest that the C-terminal

degron detected in ADpm appears to be a conserved motif 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% (Additional file 1: Table S1). 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 a degron conserved in bacterial ADs, herein termed ADcon.

To test this hypothesis and evaluate the functionality of ADcon, the C-terminal sequences of three ADs from major branches of the phylogenetic tree, as well as the ADcon, were fused to 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-termini of bacterial AD homologs serve as degrons. Furthermore, it was observed that lack of some of the conserved residues 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 in *E. coli* than AD from cyanobacterial strains [15]. Based on sequence alignment, we observed significant amino acid difference in the degron region of 7421ADO, suggesting that 7421ADO is less susceptible to proteolytic degradation explaining a higher protein level. These findings strongly support our hypothesis that the C-termini of bacterial ADs are a conserved degron.
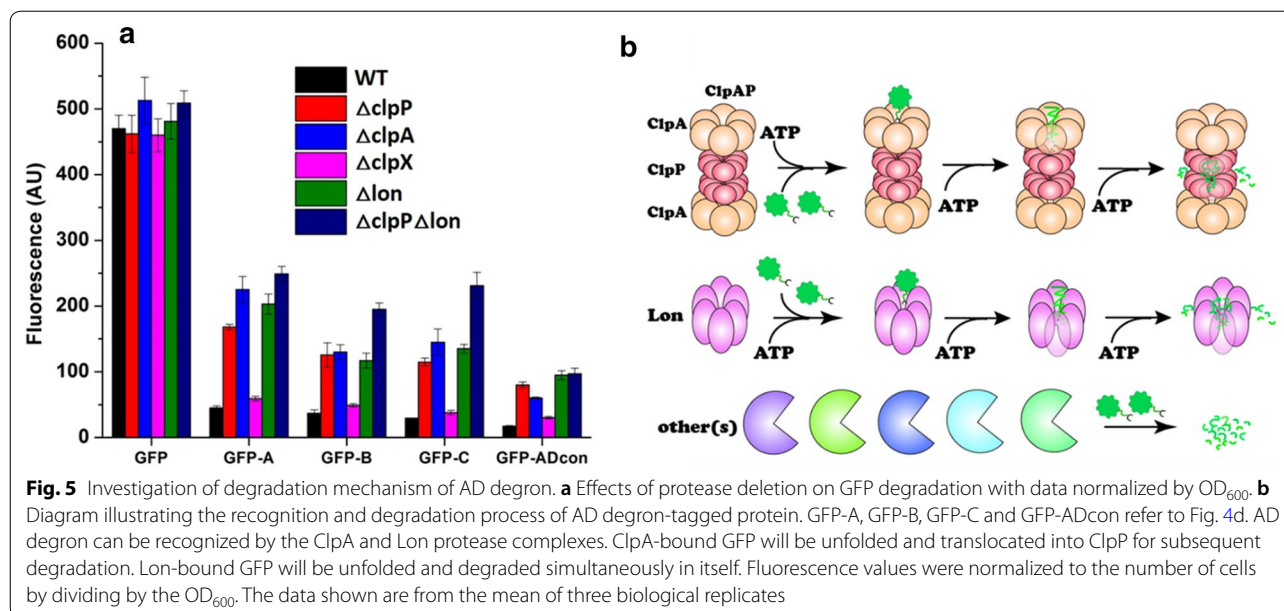
Liu *et al. Biotechnol Biofuels*   (2020) 13:114

Page 5 of 11

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, with three representatives (labeled **a, b, c**) 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 represents 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

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 previously reported C-terminal degrons in *E. coli* [35]. 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 [36]. To test the hypothesis that the degradation mechanism of AD degron is like ssrA's, degron-tagged GFP was transformed into constructed ΔClpA, ΔClpX, and ΔClpP strains. In all four degron-tagged GFP strains, fluorescence recovery was observed in both ΔClpA and ΔClpP strains, but not significant in the ΔClpX strain, compared to WT *E. coli* (Fig. 5a). 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 ΔClpP strain compared with the untagged GFP (Fig. 5a). 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 [37–39]. Lon protease was suspected to degrade AD degron-tagged GFP since it is known to be an efficient protease for non-native protein degradation [40]. 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 ΔClpPΔLon strain (Fig. 5a). 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 [41], 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 prevents the deletion of 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.

Liu *et al. Biotechnol Biofuels*   (2020) 13:114

Page 6 of 11

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

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 relative enzyme abundance of degron-free versions increased 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 reduction in enzyme activity, but also increases overall pentane accumulation. The pentane production decreased from 7.9 to 5.2 mg/L in *E. coli* harboring ADnp-10 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 modeling results suggest that the elimination of the C-terminal sequence has a negative effect on the substrate binding in all three candidates (Additional file 1: Figure S4), 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 [15].

Since deleting protease complexes negatively 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 proteolytic degradation. This was tested by adding a 6xHis-tag after the native degron in wild-type ADse, which previously produced the most pentane (Fig. 6c). As shown in Fig. 6d, ADse with a 6xHis-tag 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 also provide some support for this hypothesis. One of the two highest scored interacting models shows that ferredoxin can interact with the ADse C-terminal helix (Fig. 6e), indicating that the C-terminal helix modification might improve the recruitment of ferredoxin (Fig. 6d). Turning to pentane production, the highest titer comes with over-expression of fdx and fpr (Fig. 6d), suggesting that ferredoxin (fdx) and ferredoxin reductase (fpr) are necessary for electron transfer to AD and for high enzyme activity. The protection principle was extended to GFP+ADcon in vivo to test the limit of the protective effect conferred by a 6xHis-tag. Figure 6e shows that the 6xHis-tag

Liu *et al. Biotechnol Biofuels* (2020) 13:114

Page 7 of 11
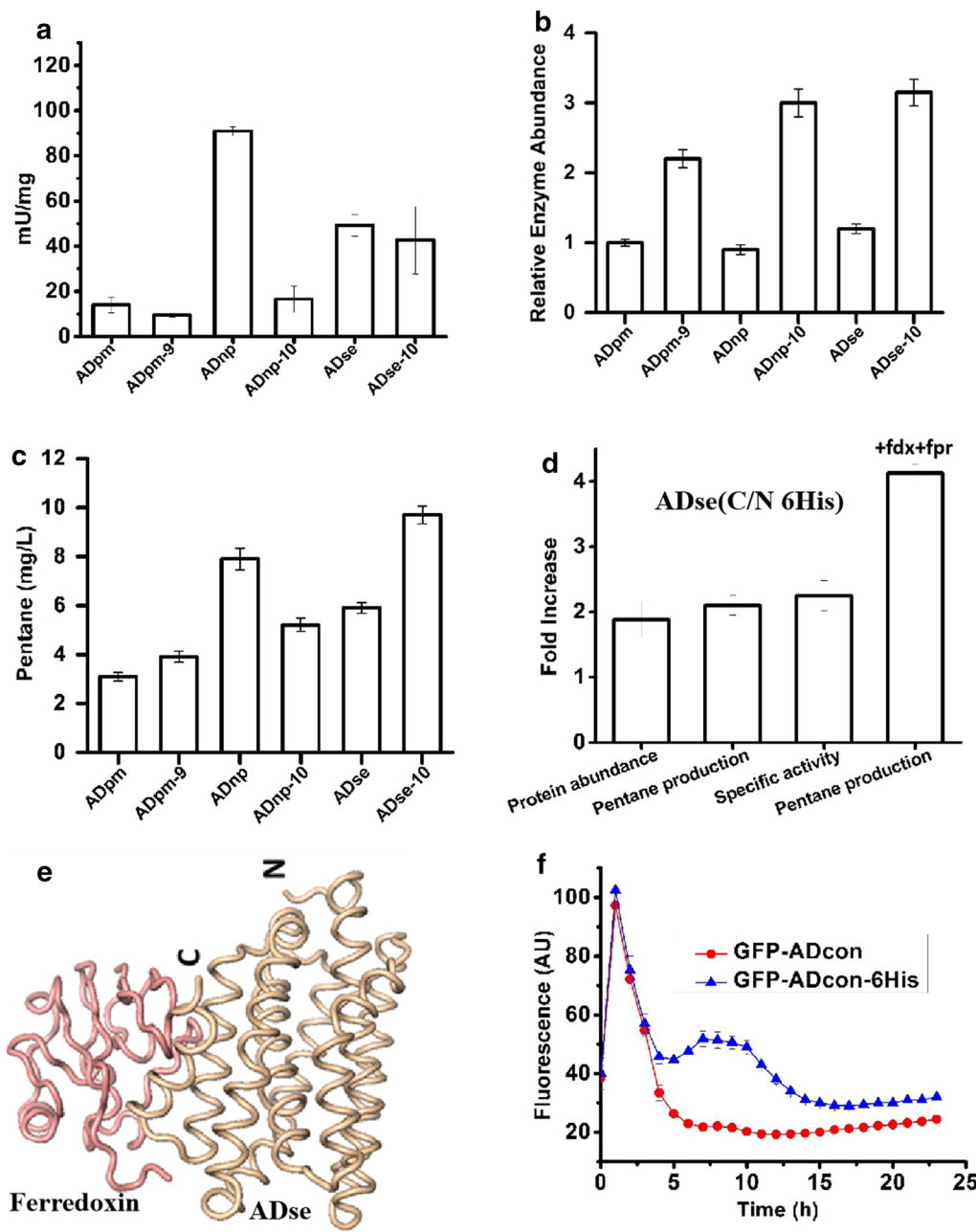
Fig. 6 Investigation of degron modification. Representative ADs from *P. marinus* (ADpm), *N. punctiforme* (ADnp) and *S. elongatus* (ADse). **a-c** Comparison of specific activities, enzyme abundance and pentane production in three representative ADs with and without degron. Enzyme abundance was normalized to ADpm based on peak area. **d** Comparison of enzyme abundance and pentane production in ADse C-terminal His-tag and N-terminal His-tag. **e** Modeling of protein–protein interaction between ADse and ferredoxin from *E. coli*. *E. coli* 2Fe2S ferredoxin (Uniprot ID P0A9R4) (colored pink) and ADse decarbonylase (colored wheat) interaction model predicted with GRAMMX protein–protein docking server. The C- and N-terminal ends of ADse are designated. **f** C-terminal His-tag effect on degron sealing. The data shows the mean of three replicates, and all error bars show the S.D

protects GFP + ADcon from protein degradation during the cultivation (Fig. 6f). However, in comparison to wild-type GFP, the presence of 6xHis-tag did not totally prevent the degron recognition by protease (Additional file 1: Figure S5). Overall, these results demonstrate that manipulating the AD degron can improve enzyme activity and alkane production.

Liu *et al. Biotechnol Biofuels*   (2020) 13:114

Page 8 of 11

Discussion

Given the importance of alkanes for liquid transportation fuels, engineers have sought to find sustainable ways to produce alkane using the bacterial AD-based pathway [5, 7, 8]. 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 [28]. 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 [32]. 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 [35, 42, 43], 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 [35]. This might be the result of its promiscuous substrate specificity that would convert metabolically vital aldehyde intermediates into non-recyclable alkane/enes in bacteria, ultimately resulting in a loss of carbon. Our product analysis showed the presence of various volatile compounds in the culture headspace (Additional file 1:Figure S6). 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 (Additional file 1: Figure S7), which supports the notion that the AD degron provides a fitness advantage during evolution. In addition, it was reported that the expression of bacterial genes involved in the assimilation of alkanes is usually tightly regulated [44], which also suggests that there are regulatory processes present in the alkane metabolic pathway.

Our bioinformatics and experimental investigation strongly suggest that virtually all bacterial ADs possess a C-terminal degron, and its elimination can increase alkane production in vivo, albeit negatively affecting the catalytic efficiency. We also showed that one of the approaches for mitigating the proteolytic degradation of the degron is adding extra amino acids after the native degron to protect it from protease recognition. This modification 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 AD toward the desired aldehydes relative to other 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 C-terminal degron in AD and its impact on enzyme stability and activity. We anticipate that this work will contribute to transitioning AD-based biosynthesis of alkanes to industrial scales.

Conclusion

For the first time, a degradation tag (degron) was discovered in the AD from marine cyanobacterium *Prochlorococcus marinus* using error-prone PCR-based directed evolution system. Bioinformatic analyses were carried out and revealed that 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. 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.

Methods

General molecular methods

Escherichia coli K-12 strains MG1655 was used for directed evolution and degradation mechanism investigation. RARE [4] strain was used for in vivo fermentation. Random mutagenesis of AD by error-prone PCR was performed with the GeneMorphII Random Mutagenesis Kit (Agilent Technologies) with initial template amount 100 ng, PCR cycle number 25 to achieve low mutation frequency (0–4.5%). ADs were cloned into pZa23MCS without codon optimization. Plasmids and DNA were 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, T4 DNA ligase (Life Technologies) was used according to manufacturer's instructions. Gene deletions were performed by red-mediated recombination [45]. The I-TASSER Suite was employed to generate structural model of ADs minus degrons [46]. Autodock Vina was used for molecular docking [47]. Protein stability was predicted by SCooP

Liu *et al. Biotechnol Biofuels* (2020) 13:114

Page 9 of 11

[48]. GRAMMX server was used for protein–protein interaction prediction [49].

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_{600} of 0.6. Cells were transferred 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 headspace was used for pentane quantification and the culture was 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 and 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

The protein abundance analysis, Western blot was carried out. Whole cell lysis was separated on a 18% SDS-PAGE gel. Sample amount was equaled based on culture OD. Proteins on gel were transferred to a nitrocellulose membrane and detected by a primary 6x-His-tag Antibody (ThermoFisher, USA) from mouse and goat antimouse 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). The intensity of western blot bands was analyzed using ImageJ.

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-5 ms 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. And 0.5 mL of headspace was

injected. The retention times and fragmentation patterns were compared with those obtained from pure standards. Pentane was 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 uL of each cell lysis in 100 uL 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 N_2 atmosphere glove box, vials were tightly capped and incubated on shaker at 200 rpm, 37 °C. Reaction was stopped by heating at 100 °C for 5 min. 200 μL of gas phase were analyzed using GC–MS protocol.

Protein purification

Escherichia 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 min following the 3-s ON (120-140 W) and 4-s OFF regime using Qsonica Q700 equipped with dual-horn probe. Cell lysate was spun down at 37,000 g for 30 min, 4 °C, and supernatant was applied to the Ni-agarose equilibrated with binding buffer. Resin was washed with 50 volumes of washing buffer (50 mM 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 N_2 and stored at − 80 °C.

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.

Liu *et al. Biotechnol Biofuels*   (2020) 13:114

Page 10 of 11

Fluorescent assay with whole cells

The degradation of GFP was quantified using a fluorescent assay. For this purpose, cells were cultured in 24-well plates. The fluorescence was monitored by fluorescence in a microtiter plate reader (excitation wavelength, 488 nm; 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 was normalized by the OD_{600} 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 around 0.1 mg/mL. In order to optimize the fluorescence signal-to-noise ratio, an assay optimization was performed.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13068-020-01753-5.

Additional file 1: Figure S1. Whole cell lysis activity analysis of different mutants. M2 give the highest activity, which has a premature stop codon in the C-terminal region. *WT* wild type ADpm, M1–M9 different mutants. **Figure S2**. 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. **Figure S3.** Comparison of C-terminal of ADpm and C-motif 1 [35] of protease trapped proteins. **Figure S4.** 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. Note, residues shown in red indicate the interact amino acids keep consistent in the truncated and wild-type ones. **Figure S5.** Effect of C-terminal His-tag on GFP degradation. C-terminal 6His-tag was added to GFP with conserved degron. **Figure S6.** Effects of AD degron elimination on fermentation. Headspace analysis of ADse (Red) and ADse-10 (Blue). U1–U7, unidentified peaks. **Figure S7.** 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. **Table S1.** Statistical analysis of amino acid usage frequency (%) at C-terminus in 371 bacterial ADs.

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.

Availability of data and materials

All data are included in the manuscript.

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.

Author details

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

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. Biotechnol Biofuels. 2019;12(1):155.
2. Bai W, Geng W, Wang S, Zhang F. Biosynthesis, regulation, and engineering of microbially produced branched biofuels. Biotechnol Biofuels. 2019;12(1):84.
3. Marella ER, Holkenbrink C, Siewers V, Borodina I. Engineering microbial fatty acid metabolism for biofuels and biochemicals. Curr Opin Biotechnol. 2018;50:39–46.
4. Sheppard MJ, Kunjapur AM, Prather KL. Modular and selective biosynthesis of gasoline-range alkanes. Metab Eng. 2016;33:28–40.
5. Schirmer A, Rude MA, Li X, Popova E, Del Cardayre SB. Microbial biosynthesis of alkanes. Science. 2010;329(5991):559–62.
6. Choi YJ, Lee SY. Microbial production of short-chain alkanes. Nature. 2013;502(7472):571.
7. Fatma Z, Hartman H, Poolman MG, Fell DA, Srivastava S, Shakeel T, Yazdani SS. Model-assisted metabolic engineering of Escherichia coli for long chain alkane and alcohol production. Metab Eng. 2018;46:1–12.
8. Zhu Z, Zhou YJ, Kang M-K, Krivoruchko A, Buijs NA, Nielsen J. Enabling the synthesis of medium chain alkanes and 1-alkenes in yeast. Metab Eng. 2017;44:81–8.
9. Song X, Yu H, Zhu K. Improving alkane synthesis in Escherichia coli via metabolic engineering. Appl Microbiol Biotechnol. 2016;100(2):757–67.
10. Sinha M, Weyda I, Sørensen A, Bruno KS, Ahring BK. Alkane biosynthesis by Aspergillus carbonarius ITEM 5010 through heterologous expression of Synechococcus elongatus acyl-ACP/CoA reductase and aldehyde deformylating oxygenase genes. Amb Express. 2017;7(1):18.
11. Abdellaoui S, Macazo FC, Cai R, De Lacey AL, Pita M, Minteer SD. Enzymatic electrosynthesis of alkanes by bioelectrocatalytic decarbonylation of fatty aldehydes. Angew Chem Int Ed. 2018;57(9):2404–8.
12. Howard TP, Middelhaufe S, Moore K, Edner C, Kolak DM, Taylor GN, Parker DA, Lee R, Smirnoff N, Aves SJ. Synthesis of customized petroleum-replica

fuel molecules by targeted modification of free fatty acid pools in Escherichia coli. Proc Natl Acad Sci. 2013;110(19):7636–41.

13. Tseng H-C, Prather KL. Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. Proc Natl Acad Sci. 2012;109(44):17925–30.

14. Xu P, Qiao K, Ahn WS, Stephanopoulos G. Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. Proc Natl Acad Sci. 2016;113(39):10848–53.

15. Kudo H, Hayashi Y, Arai M. Identification of non-conserved residues essential for improving the hydrocarbon-producing activity of cyanobacterial aldehyde-deformylating oxygenase. Biotechnol Biofuels. 2019;12(1):89.

16. Dennis M, Kolattukudy P. A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. Proc Natl Acad Sci. 1992;89(12):5306–10.

17. Schneider-Belhaddad F, Kolattukudy P. Solubilization, partial purification, and characterization of a fatty aldehyde decarbonylase from a higher plant, Pisum sativum. Arch Biochem Biophys. 2000;377(2):341–9.

18. Bernard A, Domergue F, Pascal S, Jetter R, Renne C, Faure J-D, Haslam RP, Napier JA, Lessire R, Joubès J. Reconstitution of plant alkane biosynthesis in yeast demonstrates that Arabidopsis ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. Plant Cell. 2012;24(7):3106–18.

19. 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. Proc Natl Acad Sci. 1994;91(21):10000–4.

20. Ellington BR, Paul B, Das D, Vitek AK, Zimmerman PM, Marsh ENG. An Unusual Iron-Dependent Oxidative Deformylation Reaction Providing Insight into Hydrocarbon Biosynthesis in Nature. ACS Catalysis. 2016;6(5):3293–300.

21. 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. Biotechnol Biofuels. 2013;6(1):86.

22. 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. Biotechnol Biofuels. 2016;9(1):185.

23. 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):1204–8.

24. Rajakovich LJ, Nørgaard H, Warui DM, Chang W, Li N, Booker SJ, Krebs C, Bollinger JM, Pandelia ME. Rapid reduction of the diferric-peroxyhemiacetal intermediate in aldehyde-deformylating oxygenase by a cyanobacterial ferredoxin: evidence for a free-radical mechanism. J Ame Chem Soc. 2015;137(36):11695–709.

25. 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. J Biol Chem. 2018;RA117:000639.

26. 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):10743–50.

27. Kang M-K, Zhou YJ, Buijs NA, Nielsen J. Functional screening of aldehyde decarbonylases for long-chain alkane production by *Saccharomyces cerevisiae*. Microb Cell Fact. 2017;16(1):74.

28. Patrikainen P, Carbonell V, Thiel K, Aro EM, Kallio P. Comparison of orthologous cyanobacterial aldehyde deformylating oxygenases in the production of volatile C3-C7 alkanes in engineered *E coli*. Metab Eng Commun. 2017;5:9–18.

29. 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.

30. Bali AP, Genee HJ, Sommer MO. Directed evolution of membrane transport using synthetic selections. ACS synthetic Biol. 2018;7(3):789–93.

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

32. Packer MS, Liu DR. Methods for the directed evolution of proteins. Nat Rev Genet. 2015;16(7):379.

33. Sauer RT, Baker TA. AAA + proteases: ATP-fueled machines of protein destruction. Annu Rev Biochem. 2011;80:587–612.

34. Varshavsky A. N-degron and C-degron pathways of protein degradation. Proc Natl Acad Sci. 2019;116(2):358–66.

35. 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. Mol Cell. 2003;11(3):671–83.

36. Flynn JM: Substrate selection by the ClpXP protease: a tail of destruction. Massachusetts Institute of Technology; 2004.

37. Frank EG, Ennis DG, Gonzalez M, Levine AS, Woodgate R. Regulation of SOS mutagenesis by proteolysis. Proc Natl Acad Sci. 1996;93(19):10291–6.

38. Olivares AO, Kotamarthi HC, Stein BJ, Sauer RT, Baker TA. Effect of directional pulling on mechanical protein degradation by ATP-dependent proteolytic machines. Proc Natl Acad Sci. 2017;114(31):E6306–13.

39. Gottesman S, Roche E, Zhou Y, Sauer RT. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev. 1998;12(9):1338–47.

40. Hansen J, Gregersen N, Bross P. Differential degradation of variant medium-chain acyl-CoA dehydrogenase by the protein quality control proteases Lon and ClpXP. Biochem Biophys Res Commun. 2005;333(4):1160–70.

41. Mahmoud SA, Chien P: Regulated proteolysis in bacteria. *Annual review of biochemistry* 2018(0).

42. Striebel F, Kress W, Weber-Ban E. Controlled destruction: AAA + ATPases in protein degradation from bacteria to eukaryotes. Curr Opin Struct Biol. 2009;19(2):209–17.

43. Olivares AO, Baker TA, Sauer RT. Mechanistic insights into bacterial AAA + proteases and protein-remodelling machines. Nat Rev Microbiol. 2016;14(1):33.

44. Moreno R, Rojo F: Genetic Features and Regulation of n-Alkane Metabolism in Bacteria. Aerobic Utilization of Hydrocarbons, Oils, and Lipids. 2019:521-542.

45. 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.

46. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015;12(1):7.

47. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. J Comput Chem. 2010;31(2):455–61.

48. 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):3415–22.

49. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein–protein docking. Nucleic Acids Res. 2006;34:W310–4.

Publisher's Note