Aldehyde dehydrogenases are versatile enzymes that serve a range of biochemical functions. Although traditionally considered metabolic housekeeping enzymes because of their ability to detoxify reactive aldehydes, like those generated from lipid peroxidation damage, the contributions of these enzymes to other biological processes are widespread. For example, the plant pathogen *Pseudomonas syringae* strain *Pto*DC3000 uses an indole-3-acetaldehyde dehydrogenase to synthesize the phytohormone indole-3-acetic acid to evade host responses. Here we investigate the biochemical function of AldC from *Pto*DC3000. Analysis of the substrate profile of AldC suggests that this enzyme functions as a long-chain aliphatic aldehyde dehydrogenase. The 2.5 Å resolution X-ray crystal of the AldC C291A mutant in a dead-end complex with octanal and NAD⁺ reveals an apolar binding site primed for aliphatic aldehyde substrate recognition. Functional characterization of site-directed mutants targeting the substrate- and NAD(H)-binding sites identifies key residues in the active site for ligand interactions, including those in the “aromatic box” that define the aldehyde-binding site. Overall, this study provides molecular insight for understanding the evolution of the prokaryotic aldehyde dehydrogenase superfamily and their diversity of function.

In all organisms, the diversification of large superfamilies of enzymes provides a foundation for the evolution of biochemical capacity and the ability to metabolize varied small molecules (1). Typically, across a superfamily, the core chemical mechanism is retained with substrate profiles becoming either specialized or promiscuous, depending on evolution and the metabolic needs of the organism (2). Classic examples of enzyme superfamilies found across all kingdoms include the cytochromes P450, alcohol dehydrogenases, aldo-keto reductases, and aldehyde dehydrogenases (3–10). For example, aldehyde dehydrogenases are NAD(P)(H)-dependent enzymes that metabolize a wide range of aldehydes to their corresponding carboxylic acids in prokaryotes and eukaryotes and tend to be encoded by multiple genes in the genome of a given species (9, 10) (Fig. 1).

Aldehyde dehydrogenases are generally associated with the detoxification of aldehydes, which are highly reactive compounds generated through cellular metabolism (11). For example, these enzymes can scavenge aldehydes, such as malondialdehyde resulting from lipid peroxidation, and convert them to a less chemically reactive carboxylic acid (12). Aldehyde dehydrogenases and their biochemical functions are also linked to a wide variety of biochemical processes ranging from ethanol metabolism via oxidation of acetaldehyde into acetate (13–15), polyamine metabolism (16), and plant cell wall ester biosynthesis (17, 18) to protective cellular responses to stresses such as dehydration, osmotic shock, and temperature changes (19, 20). Recent work has also linked aldehyde dehydrogenase activity to the synthesis of indole-3-acetic acid, the primary plant hormone auxin, in the plant pathogenic microbe *Pseudomonas syringae* strain *Pto*DC3000 (21).

To suppress host defenses and promote diseases development, *Pseudomonas syringae* produces a variety of virulence factors, including phytohormones or chemical mimics of hormones, to manipulate hormone signaling in its host plants (22–24). *P. syringae* and many other plant-associated microbial pathogens can synthesize the major auxin indole-3-acetic acid (IAA), whose production is implicated in pathogen virulence (21, 25–29). *Pto*DC3000 was shown to synthesize IAA using an uncharacterized pathway requiring indole-3-acetaldehyde dehydrogenase activity (21). Previously, a mutation in *Azospirillum brasilense* (aldA) was identified that decreased IAA production and was linked to a gene and annotated as encoding an aldehyde dehydrogenase (30). Bioinformatic analysis of *Pto*DC3000 identified a set of aldehyde dehydrogenases (AldA–C) sharing 30–40% amino acid identity with each other. Subsequent metabolic, biochemical, and in planta analyses of AldA (UniProt: PSPTO_0092), AldB (UniProt: PSPTO_2673), and AldC (UniProt: PSPTO_3644) demonstrated that AldA functions as an indole-3-acetaldehyde dehydrogenase, is essential for IAA synthesis, and contributes to virulence of *Pto*DC3000 (21). AldB may also contribute to...
IAA synthesis but is not as metabolically or kinetically efficient as AldA for IAA production (21). Analysis of AldC indicates that it lacks a role in IAA synthesis (21); however, its potential biochemical role in PtoDC3000 was not fully examined.

Here we investigate the substrate profile of AldC from PtoDC3000, which suggests that this enzyme functions as a long-chain aliphatic dehydrogenase. The 2.5 Å resolution X-ray crystal of the AldC C291A mutant in complex with octanal and NAD\(^+\) and biochemical analysis of a set of site-directed mutants provide insight on substrate recognition in this enzyme. Comparison of the three-dimensional structures of AldA and AldC from PtoDC3000 reveals the sequence and structural changes that lead to distinct substrate profiles of these aldehyde dehydrogenases.

Results and discussion
Biochemical analysis of AldC as a long-chain aliphatic dehydrogenase

Previously, three aldehyde dehydrogenases (AldA–C) from the plant pathogen *P. syringae* strain PtoDC3000 were identified and examined for their contribution to the synthesis of IAA and virulence (21). Each protein was expressed and purified to homogeneity for enzyme assays using indole-3-acetaldehyde as a substrate (21). Unlike AldA and AldB, which functioned as physiological tetramers, AldC was shown to be a homodimer (21). Steady-state kinetic parameters for AldA using indole-3-acetaldehyde showed this enzyme to be 130- and 710-fold more efficient than AldB and AldC, respectively (21).

Bioinformatic analysis predicted that AldC contains the PF00171 domain, a signature domain of the aldehyde dehydrogenase superfamily, ranging from Ser\(^{159}\) to Ile\(^{382}\) (Fig. 2A). Specifically, the 4 catalytic residues (Asn\(^{159}\), Glu\(^{257}\), Gly\(^{288}\), and Cys\(^{291}\)) and 19 residues in the NAD\(^+\)-binding site (Ile\(^{155}\)-Asn\(^{159}\), Lys\(^{182}\), Gly\(^{219}\), Ile\(^{223}\)-Ser\(^{226}\), Ala\(^{239}\), Leu\(^{242}\), Glu\(^{257}\), Leu\(^{258}\), Gly\(^{259}\), Cys\(^{291}\), Glu\(^{391}\), and Phe\(^{399}\)) define the aldehyde dehydrogenase consensus sequence motifs of AldC. The NCBI Conserved Protein Domain Family server places AldC in the cd01738/ALDH_CddD_SSP0762 family, 1 of 42 aldehyde dehydrogenase families found in the *Pseudomonas* (31, 32).

The *Pseudomonas* orthologous groups classification system in the *Pseudomonas* Genome Database (RRID:SCR_006590) found orthologs of AldC (group ID POG018413) from 93 *Pseudomonas* species and strains, including *P. aeruginosa*, *P. putida*, *P. fluorescens*, and *P. savastanoi* (Fig. 2B and Table S1). AldC belongs to a clade consisting of aldehyde dehydrogenases, which share ~90% sequence identity, from various plant pathogenic *Pseudomonas* spp.: *P. syringae*, *P. viridiflava*, *P. savastanoi*, and *P. amygdali* in the *P. syringae* phylogenetic group. None of AldC-related enzymes from *Pseudomonas* had previously been experimentally characterized, and their substrates and functions had yet to be described.

To determine the substrate preference of AldC, a panel of 23 molecules (Table S2), including short- to long-chain aliphatic aldehydes (acetaldehyde, propionaldehyde, butyaldehyde, valeraldehyde (pentanal), isovaleraldehyde, 3-methylcrotanoldehyde, hexanal, trans-2-hexen-1-ol, heptanal, octanal, nonanal, trans-2-nonenal), betaine aldehyde, and aromatic aldehydes (benzaldehyde, m-anisaldehyde, p-anisaldehyde, phenylacetalddehyde, cinnamaldehyde, hydrocinnamaldehyde, coniferylaldehyde, sinapaldehyde, 4-pyridinecarboxaldehyde, and indole-3-acetaldehyde) was used to screen for enzymatic activity. Spectrophotometric assays of AldC identified aliphatic aldehydes of 5–9-carbon length, as well as hydrocinnamaldehyde and 4-pyridinecarboxaldehyde, as substrates (Fig. 3A) with octanal having the highest specific activity (Fig. 3B). These data suggest that short 2–4-carbon aldehydes, branched aliphatic aldehydes, and larger aromatic aldehydes are poor substrates for AldC. To evaluate the nicotinamide cofactor preference of AldC, the activity of the enzyme was tested with octanal and either NAD\(^+\) or NADP\(^+\), which showed a distinct preference for NAD\(^+\) (Fig. 3C).

Steady-state kinetic analysis of AldC with valeraldehyde, hexanal, heptanal, octanal, and nonanal indicates that the 8-carbon substrate (i.e. octanal) is the preferred aliphatic aldehyde substrate (Table 1). Although the *k_cat* values of AldC vary less than 3-fold between these substrates, the *K_m* value for octanal is the lowest (1.2 mM) with that of valeraldehyde as the highest (48.8 mM). The overall effect is a marked difference in catalytic efficiency (*k_cat/K_m*) between octanal and the other aliphatic substrates ranging from use of nonanal exhibiting a 10-fold reduction in *k_cat/K_m* to ~80-fold less efficient use of the 5-carbon substrate valeraldehyde. Generally, the aromatic aldehyde substrates hydrocinnamaldehyde and indole-3-acetaldehyde were poorly used by AldC with catalytic efficiencies comparable with valeraldehyde and 4-pyridinecarboxaldehyde used with a *k_cat/K_m* value 6-fold lower than octanal (Table 1). Overall, the biochemical analysis of AldC suggests that this enzyme functions primarily as a long-chain aliphatic
aldehyde dehydrogenase (Fig. 3D). Moreover, conservation of related homologs in 93 different Pseudomonas species and strains (Fig. 2B) indicates a likely conserved function across these organisms.

Three-dimensional structure of AldC C291A mutant in complex with octanal and NAD$^+$

To understand how AldC recognizes aliphatic aldehyde substrates, crystals of the AldC C291A mutant were grown in the presence of octanal and NAD$^+$. The analogous cysteine to Cys$^{291}$ of AldC in aldehyde dehydrogenases, including AldA from P. syringae, is the essential catalytic residue (9, 10, 14–16, 21). The AldC C291A point mutant was generated by PCR mutagenesis, and the resulting protein was expressed, purified, and crystallized with the goal of obtaining a dead-end complex of the enzyme with octanal and NAD$^+$ bound in the active site.

The 2.5 Å resolution X-ray crystal structure of the AldC (C291A)·octanal·NAD$^+$ complex was solved by molecular replacement using the three-dimensional structure of P. syringae AldA as a search model (21) (Table 2). Each of the two unique protein chains in the asymmetric unit form a corresponding physiological dimer through crystallographic symmetry (Fig. 4A). The secondary structure features and domains of the AldC monomer are similar to those of other aldehyde dehydrogenase family members (Fig. 4, B and C). The N-terminal
Rossmann-fold domain contains a central β-sheet ($\beta_9$-$\beta_8$-$\beta_7$-$\beta_{10}$-$\beta_{11}$) surrounded by α-helices to form the NAD(H)-binding site (Fig. 4, B and C, blue). The C-terminal region consists of a mixed α/β domain, which includes the catalytic cysteine residue and forms the aldehyde-binding site (Fig. 4, B and C, red). An interdomain linker region (Fig. 4, B and C, green) connects the N- and C-terminal domains of AldC. A small three-stranded β-sheet domain (Fig. 4, B and C, gold) facilitates oligomerization.

The overall fold of AldC shares structural similarity with multiple aldehyde dehydrogenases, including P. syringae AldA, from a variety of microbes and eukaryotes (Fig. S1 and Table S3). Overlays of the AldC structure with twelve aldehyde, retinal, betaine aldehyde, and indole-3-acetaldehyde (i.e. AldA) dehydrogenases (14, 15, 21, 33–37), which are related by 38–46% amino acid identity, show the conservation of the three-dimensional fold with 0.7–1.2 Å RMSD for 400–450 $C_\alpha$-atoms in these enzymes from widely varied organisms (Fig. S1). Amino acid sequence comparisons of these proteins highlight how the catalytic cysteine and residues of the NAD(H)-binding site are highly conserved with major variations in the substrate-binding site leading to functional differences in these proteins (Fig. S2).

Although the specific determinants of oligomerization in aldehyde dehydrogenases can be highly variable (38), the X-ray crystal structure of AldC provides possible insight on its dimer structure compared with the tetramer reported for AldA (21). The P. syringae AldA tetramer is formed from a dimer of dimers with an extensive interaction surface, shown in Fig. 5 (A and B) with the oligomerization region highlighted in the close-ups. Fig. 5 (C and D) shows the same region of the AldC dimer for comparison. Multiple amino acid differences between AldA and AldC in the oligomerization domain alter both the electrostatic surface charge, which is strongly basic in AldA, and the surface topology, especially in the positioning of $\beta_5$, $\beta_6$, and the...
### Analysis of AldC from P. syringae DC3000

#### Table 1

**Steady-state kinetic parameters of AldC with aldehyde substrates**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_{cat}$</th>
<th>$K_m$</th>
<th>$K_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valeraldehyde/pentanal</td>
<td>34.3 ± 4.2</td>
<td>48.8 ± 17.1</td>
<td>12</td>
</tr>
<tr>
<td>Hexanal</td>
<td>22.9 ± 1.0</td>
<td>16.4 ± 2.3</td>
<td>23</td>
</tr>
<tr>
<td>Heptanal</td>
<td>22.1 ± 1.5</td>
<td>7.4 ± 2.0</td>
<td>30</td>
</tr>
<tr>
<td>Octanal</td>
<td>67.4 ± 1.4</td>
<td>1.2 ± 0.1</td>
<td>49</td>
</tr>
<tr>
<td>Nonanal</td>
<td>27.7 ± 5.0</td>
<td>4.9 ± 2.3</td>
<td>23</td>
</tr>
<tr>
<td>Hydrocinnamaldehyde</td>
<td>16.8 ± 1.6</td>
<td>14.2 ± 3.1</td>
<td>10</td>
</tr>
<tr>
<td>4-Pyridinecarboxaldehyde</td>
<td>38.7 ± 1.9</td>
<td>4.4 ± 0.6</td>
<td>147</td>
</tr>
<tr>
<td>Indole-3-acetaldehyde</td>
<td>3.6 ± 0.2</td>
<td>1.3 ± 0.3</td>
<td>14</td>
</tr>
<tr>
<td>NAD$^+$ (octanal)</td>
<td>30.0 ± 0.3</td>
<td>0.5 ± 0.1</td>
<td>1,050</td>
</tr>
</tbody>
</table>

*Steady-state kinetic parameters of AldC with indole-3-acetaldehyde were previously reported (21).*

#### Table 2

**Summary of crystallographic statistics for AldC structures**

<table>
<thead>
<tr>
<th>Crystal</th>
<th>AldC (C291A) octanal NAD$^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C222,1</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>92.19</td>
</tr>
<tr>
<td>Resolution range (highest shell)</td>
<td>92.17</td>
</tr>
<tr>
<td>Reflections (total/unique)</td>
<td>156.137/32,956</td>
</tr>
<tr>
<td>Completeness (highest shell)</td>
<td>97.8% (99.1%)</td>
</tr>
<tr>
<td>$&lt;\langle I/\sigma(I)\rangle$</td>
<td>12.8 (1.8)</td>
</tr>
<tr>
<td>$R_{free}$ (highest shell)</td>
<td>6.0% (52.9%)</td>
</tr>
</tbody>
</table>

**Refinement**

- $R_{cryst} = \Sigma |F_o| - |F_c|/\Sigma |F_o|$, where $|F_c|$ is the average intensity over symmetry.
- $R_{free} = \Sigma |F_o| - |F_c|/\Sigma |F_o|$, where summation is over the data used for refinement.
- $R_{free}$ is defined as the same as $R_{cryst}$ but was calculated using 5% of data excluded from refinement.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of protein atoms</td>
<td>7,180</td>
</tr>
<tr>
<td>No. of waters</td>
<td>197</td>
</tr>
<tr>
<td>No. of ligand atoms</td>
<td>106</td>
</tr>
<tr>
<td>RMSD</td>
<td>0.008</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.93</td>
</tr>
<tr>
<td>Average B-factor (Å$^2$)</td>
<td>45.9</td>
</tr>
<tr>
<td>Protein</td>
<td>43.7</td>
</tr>
<tr>
<td>Water</td>
<td>55.9</td>
</tr>
<tr>
<td>Stereoechemistry (%)</td>
<td>95.7</td>
</tr>
<tr>
<td>Allowed</td>
<td>4.3</td>
</tr>
<tr>
<td>Outliers</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Active site structure of AldC

Unambiguous electron densities for NAD$^+$ and octanal define how these ligands bind to the AldC C291A mutant and indicate the location of the substrate and cofactor-binding sites (Fig. 6A). Forming a dead-end complex, NAD$^+$ and octanal occupy two separate pockets, which are at the opposite ends of a ~45-Å-long tunnel (Fig. 6B). The Rossmann fold of the NAD (H)-binding domain provides extensive polar and apolar interactions that position the nicotinamide ring of NAD$^+$ in proximity to the C291A point mutation, which would be the invariant catalytic cysteine in WT AldC (Fig. 6C). The nicotinamide ring is mainly held in place by van der Waals contacts with Leu$^{258}$, Leu$^{349}$, and Phe$^{356}$ and a critical hydrogen bond from the backbone carbonyl of Leu$^{258}$ to the NH2 group of the cofactor carboxamide. The nicotinamide-ribose of NAD$^+$ establishes multiple polar interactions with the backbone carbonyl of Gly$^{235}$ and the carboxylate side chain of Glu$^{185}$. In addition, the adenine ring of NAD$^+$ is mainly stabilized by multiple van der Waals interactions with Pro$^{216}$, Ile$^{233}$, Leu$^{342}$, and Val$^{243}$, along with a hydrogen bond with a water molecule in an apolar pocket. As with nicotinamide-ribose binding, polar interactions between the adenine-ribose ring and the side-chains of Lys$^{182}$ and Glu$^{185}$ contribute to NAD$^+$ binding. Interaction of Glu$^{185}$ with the 2'-hydroxyl-group of the adenine-ribose is the key structural determinant of the cofactor specificity, as observed in other aldehyde dehydrogenases (39). AldC is not able to accommodate the 2'-phosphate of NADP(H) sterically and electrostatically, which is consistent with its preference for NAD(H) as cofactor (39). Sequence comparison of AldC and structurally related aldehyde dehydrogenases (Fig. S2) underscores the conservation of the NAD(H)-binding site across these enzymes from distantly related organisms.

In contrast to the NAD(H)-binding site, apolar interactions dominate the octanal binding in the hydrophobic substrate-binding pocket (Fig. 6, B and C). A cluster of aromatic residues (Trp$^{160}$, Tyr$^{163}$, Trp$^{450}$, Phe$^{456}$, and Tyr$^{469}$) and two apolar residues (Met$^{114}$ and Leu$^{118}$) provide the hydrophobic environment that accommodates octanal and other aliphatic aldehydes. As described for other aldehyde dehydrogenases (40–42), the substrate-binding site forms an “aromatic box” for adaptable apolar ligand interaction.

### Active site structure of AldC

- β5-β6 loop of the oligomerization domain of the two proteins.
- In addition, the length of the β5-β6 loop differs between AldA (Ala$^{143}$–Val$^{152}$) and AldC (Val$^{308}$–Thr$^{317}$), which contributes to alterations in the interaction region (Fig. 5, B and D, and Fig. S2, orange box). In AldC, these changes appear to alter what would be the corresponding tetramer (i.e. dimer–dimer) interface of AldA but may only be one contributing factor to different oligomerization.

### Active site structure of AldC

Unambiguous electron densities for NAD$^+$ and octanal define how these ligands bind to the AldC C291A mutant and indicate the location of the substrate and cofactor-binding sites (Fig. 6A). Forming a dead-end complex, NAD$^+$ and octanal occupy two separate pockets, which are at the opposite ends of a ~45-Å-long tunnel (Fig. 6B). The Rossmann fold of the NAD (H)-binding domain provides extensive polar and apolar interactions that position the nicotinamide ring of NAD$^+$ in proximity to the C291A point mutation, which would be the invariant catalytic cysteine in WT AldC (Fig. 6C). The nicotinamide ring is mainly held in place by van der Waals contacts with Leu$^{258}$, Leu$^{349}$, and Phe$^{356}$ and a critical hydrogen bond from the backbone carbonyl of Leu$^{258}$ to the NH2 group of the cofactor carboxamide. The nicotinamide-ribose of NAD$^+$ establishes multiple polar interactions with the backbone carbonyl of Gly$^{235}$ and the carboxylate side chain of Glu$^{185}$. In addition, the adenine ring of NAD$^+$ is mainly stabilized by multiple van der Waals interactions with Pro$^{216}$, Ile$^{233}$, Leu$^{342}$, and Val$^{243}$, along with a hydrogen bond with a water molecule in an apolar pocket. As with nicotinamide-ribose binding, polar interactions between the adenine-ribose ring and the side-chains of Lys$^{182}$ and Glu$^{185}$ contribute to NAD$^+$ binding. Interaction of Glu$^{185}$ with the 2'-hydroxyl-group of the adenine-ribose is the key structural determinant of the cofactor specificity, as observed in other aldehyde dehydrogenases (39). AldC is not able to accommodate the 2'-phosphate of NADP(H) sterically and electrostatically, which is consistent with its preference for NAD(H) as cofactor (39). Sequence comparison of AldC and structurally related aldehyde dehydrogenases (Fig. S2) underscores the conservation of the NAD(H)-binding site across these enzymes from distantly related organisms.

In contrast to the NAD(H)-binding site, apolar interactions dominate the octanal binding in the hydrophobic substrate-binding pocket (Fig. 6, B and C). A cluster of aromatic residues (Trp$^{160}$, Tyr$^{163}$, Trp$^{450}$, Phe$^{456}$, and Tyr$^{469}$) and two apolar residues (Met$^{114}$ and Leu$^{118}$) provide the hydrophobic environment that accommodates octanal and other aliphatic aldehydes. As described for other aldehyde dehydrogenases (40–42), the substrate-binding site forms an “aromatic box” for adaptable apolar ligand interaction.

In the AldC(C291A)-octanal-NAD$^+$ complex, the 8-carbon chain of octanal extends toward the solvent exposed, but highly apolar, opening of the aldehyde-binding site. This places the reactive aldehyde group, which hydrogen bonds with Asn$^{159}$ and the backbone nitrogen of Ala$^{291}$, in proximity to the C4 position of the nicotinamide ring that undergoes hydride transfer in the chemical reaction and the location of the catalytic cysteine. Multiple sequence alignment of AldC and other aldehyde dehydrogenases (Fig. S2) shows that the residues forming the aldehyde-binding site are highly variable, which account for substrate preferences in these enzymes.

Molecular docking of the valeraldehyde (pentanal), hexanal, heptanal, nonanal, hydrocinnamaldehyde, and 4-pyridinecarboxaldehyde (Fig. S3) provides insight on likely binding modes for each substrate that place the reactive aldehyde near the catalytic center of AldC and the aliphatic or aromatic portions of each into the apolar tunnel leading toward the solvent opening. Although the size of the site would allow short-chain aliphatic substrates, such as acetaldehyde, propionaldehyde, and butyraldehyde, to bind, the region of the substrate-binding site where these molecules would fit is more polar than the largely hydrophobic region of the tunnel accessible to longer-chained aldehydes.
aliphatic substrates (i.e. heptanal, octanal, and nonenal), which is consistent with the biochemical assays of AldC. This suggests that substrate preference is governed by van der Waals contacts that maximize nonpolar surface–ligand interactions.

Site-directed mutagenesis of AldC active site residues

To examine the contribution of active site residues, a series of site-directed mutants targeting residues in the NAD(H)-binding site (i.e. Trp\textsuperscript{158}, Lys\textsuperscript{182}, Glu\textsuperscript{185}, Thr\textsuperscript{234}, Ser\textsuperscript{236}, Glu\textsuperscript{257}, and Glu\textsuperscript{391}) and the octanal-binding site (i.e. Met\textsuperscript{114}, Leu\textsuperscript{118}, Trp\textsuperscript{160}, Tyr\textsuperscript{163}, Gln\textsuperscript{164}, Arg\textsuperscript{285}, Ser\textsuperscript{292}, Leu\textsuperscript{419}, Trp\textsuperscript{450}, Phe\textsuperscript{456}, and Tyr\textsuperscript{468}) were generated. All of the 31 AldC point mutants (M114A, L118A, W158A, W158H, N159A, W160A, W160Y, Y163A, Q164A, Q164N, K182A, K182Q, E185A, E185Q, T234A, T234V, S236A, S236T, E257A, E257Q, R285A, S290A, S292A, E391A, E391D, L419A, L419V, W450A, F456A, F456W, and Y468A) were expressed in \textit{Escherichia coli} and purified using nickel-affinity and size-exclusion chromatographies. The solubility and purity of each mutant was comparable with WT protein. Under standard assay conditions, enzymatic activities of WT AldC, the AldC C291A mutant (as described above), and the additional 31 point mutants were tested with fixed concentrations of octanal (5 mM) and NAD\textsuperscript{+} (2 mM) using up to 100 \textmu g of each protein (Fig. 7). The initial enzyme activity screening assays showed that mutation of the catalytic cysteine (C291A) and certain residues in the NAD(H)-binding site (K182A, K182Q, T234V, E257A, and E391A), and octanal-binding site (N159A and L419A) resulted in enzymes with less than 1% of WT specific activity. For example, no activity was detected with the N159A, K182A, and C291A mutants.

Kinetic analysis of mutants with changes in the NAD(H)-binding site of AldC reveals critical residues for biochemical function. As noted above, low activity of the K182A, K182Q, T234V, E257A, and E391A mutants did not allow for accurate
determination of their kinetic parameters (Fig. 7); however, the other mutant proteins with changes in the NAD(H) site were analyzed (Table 3). Mutation of Lys182 (K182A and K182Q) removes interaction of the lysine side-chain with a hydroxyl group of the adenine ribose of NAD$^+$.

Similarly, alanine substitutions of Glu257 and Glu391 severely disrupted AldC activity (Fig. 7), whereas the E257Q and E391D mutations, which introduced subtler alterations, displayed comparable or less than a 2-fold change in catalytic efficiency (i.e. $k_{\text{cat}}/K_m$) with NAD$^+$ (Table 3). Loss of the Glu257 side-chain would remove an interaction that helps position the amide group of the nicotinamide and changes at this residue could alter the orientation of the nicotinamide group for hydride transfer (Fig. 6). Similar effects have been observed in other NAD(P)(H)-dependent enzymes (43). Mutation of Glu391 to alanine eliminates interaction of the carboxylate group to the nicotinamide ribose group, which can be partially complemented by an aspartate, as seen in the E391D mutant (Table 3). The E185A mutant displayed a 21-fold decrease in $k_{\text{cat}}/K_m$ with NAD$^+$, and the E185Q mutant is similar to that of WT AldC (Table 3).

Substitutions of Trp$^{158}$, Thr$^{234}$, and Ser$^{236}$ in the NAD(H)-binding site also reduced activity. The loss of steric bulk with the W158A mutant, which may change how the pyrophosphate backbone of NAD$^+$ binds in the site, resulted in a 10-fold decrease in $k_{\text{cat}}/K_m$ with NAD$^+$, but the W158H mutant modestly reduced efficiency 3-fold (Table 3). Mutation of Thr$^{234}$, which is positioned with its methyl group to provide surface contacts to the nicotinamide ring and its hydroxyl group oriented toward the amine of Lys168 (Fig. 6), led to a slight 5-fold reduction of $k_{\text{cat}}/K_m$ with NAD$^+$, but nearly inactive protein with a valine mutation (T234V). It is possible that local conformational changes impact the positioning of the nicotinamide ring to impact activity of this mutant protein. The kinetic parameters of S236A and S236T mutations with NAD$^+$ were consistent with the loss of a hydrogen bond or changes in the side-chain position for interaction with the pyrophosphate bridge of NAD$^+$ (Table 3).

Within the octanal-binding site (Fig. 6), mutations of Asn$^{159}$, Trp$^{160}$, Ser$^{292}$, Leu$^{419}$, and Phe$^{456}$ were the most disruptive with other changes having more modest effects (Table 3). These residues are located closer to the catalytic site of AldC than those residues with lesser impact on AldC activity. The N159A and L419A mutants lacked significant activity, as noted above. Although the W160A, S292A, L419V, and F456A mutants
displayed activity in the initial screen experiment (Fig. 7), the estimated $K_m$ values for these proteins with octanal exceeded the solubility limit of this substrate ($\sim 10$ mM) in assays, which prevented accurate assessment of kinetic parameters. The N159A mutant abrogates the hydrogen bond between Asn159 and the carbonyl of octanal; the loss of this interaction likely alters the position of the reactive group for catalysis and results in the undetected activity of this mutant. Ser$^{292}$ is also positioned in proximity to the reactive aldehyde group of octanal with mutation to an alanine, resulting in a 30-fold decrease in specific activity (Fig. 7).

Trp$^{160}$ provides surface contacts along the central portion of octanal carbon chain with the W160A mutation significantly altering the topology of the binding tunnel. Replacement of
Trp^{160} with a tyrosine resulted in less than 2-fold change in catalytic efficiency (Table 3). Leu^{118} is also positioned at the inter- nal hydrophobic surface as the major determinant of substrate interaction. Two other residues in the octanal-binding site, Arg^{285}, have only slight 3-fold decreases in catalytic efficiency compared with WT AldC (Table 3), which is consistent with the interior hydrophobic surface as the major determinant of substrate interaction. Two other residues in the octanal-binding site, Glu^{164} and Ser^{290}, also play lesser roles in substrate binding, as suggested by the modest effects of the Q164A and Q164N mutations and the S290A mutant (Table 3).

**Conclusion**

The genus *Pseudomonas* is one of the most ubiquitous and complex among the Gram-negative bacteria (22). Because many *Pseudomonas* species evolved to grow under unfavorable environmental conditions (i.e., severe nutrient limitation, extreme temperatures, high salinity, and low oxygen or water availability), they also evolved metabolic diversity and plasticity to use a variety of nutrient sources (i.e., carbon, nitrogen, and sulfur), to detoxify toxic organic chemicals, and to produce multiple specialized metabolites, including polymers and small molecule compounds (31, 44). In particular, *P. syringae*, a species that includes many plant pathogenic strains, developed diverse bacterial virulence mechanisms to survive in the adverse environmental conditions of the phyllosphere (45). Further, recent studies suggest that *Pto*DC3000 uses several strategies to manipulate the auxin biology in its host plants to promote pathogenicity, including using an indole-3-acetaldehyde dehydrogenase for IAA synthesis (21). Recently, a genome-wide study identified a total of 6,510 aldehyde dehydrogenase sequences in 258 *Pseudomonas* strains belonging to 46 different species, but only a limited number of newly identified *Pseudomonas* aldehyde dehydrogenases have been biochemically characterized (32).

In contrast to the better-studied aromatic aldehyde dehydrogenases identified from various *Pseudomonas* species (46–48), which have been studied because of their ability to catabolize and biodegrade a range of compounds, including napthalene-derived molecules, and metabolic aldehyde dehydrogenases linked to the osmoprotectant betaine and growth in biofilms (49–51), less is understood about the biochemistry of aliphatic aldehyde dehydrogenases in these microbes. Here we comprehensively examined and identified AldC from *Pto*DC3000 as a long-chain aliphatic aldehyde dehydrogenase using an integrated approach combining phylogenetic, crystallographic, and biochemical analyses.

The biochemical activity of AldC from *Pto*DC3000 is consistent with the traditional role of aldehyde dehydrogenases as metabolic clean-up enzymes that convert reactive aldehydes into less active carboxytes (11, 12); however, growing evidence suggests that these enzymes may have broader
contributions to the specialized environments that different *Pseudomonas* species occupy. For example, recent work identified AldA as an indole-3-acetaldehyde dehydrogenase involved in the synthesis of the phytohormone IAA in the plant pathogen *P. syringae* DC3000 and demonstrated a role for this protein in virulence of the pathogen in plant leaves (21). Given the growth environment of this pathogenic *Pseudomonas* species in the interior of plant leaves, it is possible that aliphatic molecules in the apoplast could contribute to energy sources for the microbe or to the synthesis of surfactants that help establish microbial colonization in the host plant (52–54). The physiological contribution of AldC to the life cycle of *P. syringae* DC3000 and related plant pathogenic or plant-associated species and strains remains to be evaluated but raises the question of how these microbes develop host relationships in the natural environment. Overall, this study provides molecular insight for understanding the evolution of the prokaryotic aldehyde dehydrogenase superfamily and for the potential development of inhibitors for the pathogenic *Pseudomonas* species, which could be useful for pathogen control in agriculture.

**Experimental procedures**

**Cloning and site-directed mutagenesis**


**Protein expression and purification**

Transformed *E. coli* BL21(DE3) cells containing either WT or mutant AldC construct were grown at 37 °C in Terrific broth with 50 μg ml⁻¹ kanamycin until A₆₀₀nm ≈ 0.6–0.9. After induction with 1 mM isopropyl-1-thio-β-D-galactopyranoside, the cells were grown at 18 °C overnight. Following centrifugation (5,000 × g for 30 min), the cell pellets were resuspended in 50 mM Tris, pH 8.0, 500 mM NaCl, 25 mM imidazole, 10% (v/v) glycerol, and 1% (v/v) Tween 20. Following lysis by sonication, the cell debris was removed by centrifugation (12,000 × g for 45 min), and the supernatant was loaded onto a Ni²⁺–nitrioloactic acid column (Qiagen). The column was washed with 50 mM Tris, pH 8.0, 500 mM NaCl, 25 mM imidazole, and 10% (v/v) glycerol to remove unbound proteins. Column-bound AldC protein was eluted using 50 mM Tris, pH 8.0, 500 mM NaCl, 25 mM imidazole, 10% glycerol, and 250 mM imidazole. AldC was further purified by size-exclusion chromatography with a Superdex 200 16/60 size-exclusion column (GE Healthcare) equilibrated in 25 mM Hepes (pH 7.5) and 100 mM NaCl. Fractions corresponding to the purified AldC protein were pooled and concentrated to ~9 mg ml⁻¹. Protein concentrations were determined using the Bradford method with BSA as a standard.

**Enzyme assays and steady-state kinetic analysis**

Enzymatic activity of WT and mutant AldC was measured by monitoring NADH formation ([ε₃₄₀nm = 6220 M⁻¹ cm⁻¹; 100-μl volume]) at A₃₄₀nm using an EPOCH2 microplate spectrophotometer (BioTek). The substrate and cofactor screening experiments were performed at 25 °C in a standard assay mix of 100 mM Tris HCl (pH 8.0) and 100 mM KCl with a fixed concentration of NAD⁺ (2 mM) and aldehyde (5 mM). The panel of aldehydes and their chemical structures used for substrate screening is shown in Table S2. Steady-state kinetic parameters of WT and mutant AldC were determined at 25 °C in a standard assay mix with either fixed NAD⁺ (2.0 mM) and varied aldehyde (*i.e.* valeraldehyde, hexanal, heptanal, octanal, nonanal, hydrocinnamaldehyde, or 4-pyridinecarboxaldehyde; 0.01–10 mM) or with fixed octanal (5 mM) and varied cofactor (NAD⁺ or NADP⁺; 0.05–5 mM). Protein amounts ranged from 0.1 μg for WT AldC to 100 μg for less active AldC mutants. The resulting initial velocity data were fit to the Michaelis–Menten equation, v = (kcat[S])/(Km + [S]), using SigmaPlot.

**Protein crystallography**

Protein crystals of the AldC C291A mutant were grown by the hanging-drop vapor-diffusion method at 4 °C. Crystals of the AldC C291A mutant (9 mg ml⁻¹) in complex with octanal and NAD⁺ formed grew in drops of a 1:1 mixture of proteins and crystallization buffer: 20% (w/v) PEG-1000, 100 mM Tris HCl (pH 7.0), 2 mM octanal, and 5 mM NAD⁺. Crystals were stabilized in cryoprotectant (mixture with either 30% (v/v) glycerol or 30% (v/v) ethylene glycol) before flash-freezing in liquid nitrogen for data collection at 100 K. Diffraction data were collected at Beamline 19-ID of the Advanced Photon Source at the Argonne National Laboratory HKL3000 was used to index, integrate, and scale the collected X-ray data (55). Molecular replacement was used to solve the X-ray crystal structure of AldC using PHASER (56) with the three-dimensional structure of the AldA indole-3-acetaldehyde dehydrogenase from *P. syringae*, which shares 39% amino acid identity with AldC, as a search model (Protein Data Bank code 3IUW) (21). COOT (57) and PHENIX (58) were used for iterative rounds of model building and refinement, respectively. Data collection and refinement statistics are summarized in Table S2. The final model included Leu⁶–Leu⁴⁹⁰ of chain A, Leu⁶–Asp⁴⁸⁹ of chain B, 1 octanal, and 1 NAD⁺ in each chain, and 197 waters. Atomic coordinates and structure factors for the AldC(C291A)–octanal–NAD⁺ complex were deposited in the RCSB Protein Data Bank (code 6X9L).
Analysis of AldC from *P. syringae* DC3000

**Data availability**

All data are contained in the article with structural information deposited in the Protein Data Bank (code 6X9L), as noted above.


**Funding and additional information**—This work was supported by National Science Foundation Grants IOS-1645908 (to B. N. K.) and MCB-1614539 (to J. M. J.) and a grant from the Howard Hughes Medical Institute Science Education Program (to J. M. J.). C.K.H. was supported by National Science Foundation Grants IOS-1645908 (to B. N. K.) and 1109-16602 (to J. M. J.).

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

**Abbreviations**—The abbreviations used are: IAA, indole-3-acetic acid; RMSD, root-mean-square deviation.

**References**


Analysis of AldC from *P. syringae DC3000*


