Structural insights into hydrolytic defluorination of difluoroacetate by microbial fluoroacetate dehalogenases

Khusnutdinova, Anna N.; Batyrova, Khorcheska; Brown, Greg; Fedorchuck, Tatiana; Chai, Yao Sheng; Skarina, Tatiana; Flick, Robert; Petit, Alain-Pierre; Savchenko, Alexei; Stogios, Peter; Yakunin, Alexander

Febs Journal

DOI:
10.1111/febs.16903

Published: 01/10/2023

Peer reviewed version

Structural insights into hydrolytic defluorination of difluoroacetate by microbial fluoroacetate dehalogenases

Anna N. Khusnutdinova[1,2,3], Khorcheska A. Batyrova[1,2], Greg Brown[1], Tatiana Fedorchuk[1,2], Yao Sheng Chai[1], Tatiana Skarina[1], Robert Flick[1], Alain-Pierre Petit[1,3], Alexei Savchenko[1,4], Peter Stogios[1], and Alexander F. Yakunin[1,5]

[1] Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, M5S 3E5, Canada

[2] Institute of Basic Biological Problems, Russian Academy of Sciences, Pushchino, 142290, Russia

[3] Biological Chemistry and Drug Discovery Division, School of Life Sciences, University of Dundee, Dundee, DD1 5EH, UK

[4] Department of Microbiology, Immunology & Infectious Diseases, University of Calgary, Health Research Innovation Centre, Calgary, Alberta, T2N 4N1, Canada

[5] Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK

Correspondence

A.F. Yakunin, Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK

Tel: +44(0) 1248 388399

E-mail: a.iakounine@bangor.ac.uk

Running title

Bacterial difluoroacetate dehalogenases

Abbreviations

ABH, α/β hydrolase; DFA, difluoroacetic acid; DFPA, 2,2-difluoropropionic acid; DHA, dihydroxyacetic acid; FA, fluoroacetic acid; FAD, fluoroacetate dehalogenase; FCA, α-fluorocarboxylic acid; FHA, 2-fluoro,2-hydroxyacetate; GA, glycolic acid; HAD, haloacid dehalogenase; PFOA, perfluorooctanoic acid; TFA, trifluoroacetic acid; TFPA, 5,5,5-trifluoropentanoic acid.

Key words

difluoroacetate; hydrolytic defluorination; fluoroacetate dehalogenase; crystal structure; *Dechloromonas aromatica*

Conflict of interest

The authors declare that they have no conflict of interest with the contents of this article.

Abstract

Fluorine forms the strongest single bond to carbon with the highest bond dissociation energy among natural products. However, fluoroacetate dehalogenases have been shown to hydrolyze this bond in fluoroacetate under mild reaction conditions. Furthermore, two recent studies demonstrated that the fluoroacetate dehalogenase RPA1163 from *Rhodopseudomonas palustris* can also accept bulkier substrates. In this study, we explored the substrate promiscuity of microbial fluoroacetate dehalogenases and their ability to defluorinate polyfluorinated organic acids. Enzymatic screening of eight purified dehalogenases with reported fluoroacetate defluorination activity revealed significant hydrolytic activity against difluoroacetate in three proteins. Product analysis using liquid chromatography-mass spectrometry identified glyoxylic acid as the final product of enzymatic DFA defluorination. The crystal structures of DAR3835 from *Dechloromonas aromatica* and NOS0089 from *Nostoc* sp. were determined in the apo-state along with the DAR3835 H274N glycolyl intermediate. Structure-based site-directed mutagenesis of DAR3835 demonstrated a key role for the catalytic triad and other active site residues in the defluorination of both fluoroacetate and difluoroacetate. Computational analysis of the dimer structures of DAR3835, NOS0089, and RPA1163 indicated the presence of one substrate access tunnel in each protomer. Moreover, protein-ligand docking simulations suggested similar catalytic mechanisms for the defluorination of both fluoroacetate and difluoroacetate, with difluoroacetate being defluorinated via two consecutive defluorination reactions producing glyoxylate as the final product. Thus, our findings provide molecular insights into substrate promiscuity and catalytic mechanism of fluoroacetate dehalogenases, which are promising biocatalysts for applications in synthetic chemistry and bioremediation of fluorochemicals.

Introduction

The carbon-fluorine (C-F) bond is the strongest single bond to carbon with the bond dissociation energy up to 120 kcal/mol, the highest among all natural products [1, 2]. This supreme stability arises because fluorine is the most electronegative element reinforcing ionic forces through strong polarization of the C-F bond. The reactivity of this bond is further lowered by the poor accessibility of the valence electrons of the bonded fluorine atom [3]. Therefore, fluorinated organic compounds (organofluorines) are renowned for their inertness, thermostability, hydrophobicity, and lipophobicity [4, 5]. They are extensively used in a vast number of industrial applications including pharmaceuticals, agrochemicals, polymers, surfactants, fire retardants, and surface treatment [1, 6-8]. Currently, over 9,000 fluorinated chemicals have been synthesized for commercial applications with organofluorines comprising 30% of all pharmaceuticals and up to 30% of agrochemicals [9-12].

Among organofluorines, α-fluorocarboxylic acids (FCAs) have attracted considerable attention due to their inherent bioactivities and utility as building blocks in the construction of complex pharmaceutical molecules [13, 14]. The growing demand for FCAs including poly- and per-fluoroalkyl acids (e.g. perfluorooctanoic acid, PFOA) calls for the fast development of synthetic technologies. Although they can be produced using chemical catalysis, enantioselective introduction of fluorine into the α-position of FCAs remain challenging, and these protocols involve toxic chemicals and solvents [15, 16]. In addition to the key role played in synthetic chemistry, organofluorines (FCAs, PFOAs, per- and poly-fluoroalkyl acids) are well known as toxic, persistent and bioaccumulative pollutants widely distributed in various environments including water, soil, animals, and humans [17-20]. Although physical and chemical technologies have been proposed for degrading fluorinated pollutants, they are performed under harsh conditions and require large energy inputs. In nature, fluoroacetate (FA) and 5′-fluoro-5′-deoxyadenosine have been identified as natural products, and previous studies have also identified microbial strains involved in defluorination of these compounds [1, 21]. Furthermore, microbial biotransformations of trifluoroacetate (TFA), difluoroacetate (DFA), as well as fluorobenzoate, fluorophenol, and perfluorooctanoic acid (PFOA) under aerobic or anaerobic conditions have been reported [1, 22-28].

In recent years, biocatalytic technologies for organofluorine synthesis and decomposition are emerging as promising alternatives, because they are eco-friendly and can be performed under mild conditions with low energy use and high enantio-, regio-, and chemo-selectivity [29-34]. The global shift to sustainable industrial processes for fluorinated building blocks increases the demand for robust biocatalysts catalyzing the formation of the C-F bond or its cleavage. Recent studies identified several microbial enzymes, which can readily form or break the C-F bond under mild physiological conditions (fluorinases and defluorinases) [11, 32, 35-37]. From the three types of known dehalogenases (hydrolytic, reductive, and redox metalloenzymes), the hydrolytic dehalogenases are particularly attractive because they do not require cofactors and external energy inputs [38]. These enzymes belong to two large protein superfamilies: α/β hydrolases (ABHs, PF00561) and haloacid dehalogenase-like hydrolases (HADs, PF00702). Collectively, the ABH and HAD dehalogenases transform a broad variety of chlorinated, brominated, and iodinated substrates making them promising biocatalysts for applications in biocatalysis and bioremediation [39]. Interestingly, despite the absence of sequence similarity between the ABH and HAD dehalogenases, these enzymes employ a similar two-step mechanism of substrate dehalogenation involving an aspartate nucleophile

and a covalent ester intermediate [36, 40]. However, defluorination activity was found only in a few FA dehalogenases (ABHs) and L-2-haloacid dehalogenases (HADs) [41-46].

The first biochemically and structurally characterized defluorinating enzymes were FA dehalogenases (FADs) hydrolyzing FA with the formation of glycolic acid and fluoride ion as products [36, 42, 43, 45, 47, 48]. Although the dissociation energy of the C-F bond is much higher than those of other C-halogen bonds (96.6, 66.4, and 56.0 kcal/mol for C-F, C-Cl, and C-Br bonds, respectively), FADs exhibit the highest activity against FA, whereas for HADs defluorination is significantly slower than dechlorination of related substrates [36, 41, 42, 46, 48]. Earlier studies suggested that the catalytic mechanism of FADs involves a classical S_N2 reaction, in which an aspartate nucleophile directly ejects the fluoride from the FA C2 with inversion of stereochemistry [42, 43, 47, 48]. In a prior study on the biochemical and structural characterization of the FA dehalogenase RPA1163 from *Rhodopseudomonas palustris*, the catalytic mechanism of this defluorinase was characterized by mapping the reaction coordinates and capturing structural snapshots along the defluorination reaction including the free enzyme, enzyme-FA Michaelis complex, glycolyl-enzyme covalent intermediate, and enzyme-product complex harboring glycolic acid [36] (Fig. 1). Further structural studies revealed a subtle asymmetry in the RPA1163 apo-dimer, which plays a role in the phenomenon of "half-of-the-sites" reactivity with only one of the two protomers poised for substrate binding [49, 50]. After substrate binding, the dimer asymmetry becomes more pronounced with the empty protomer playing a key role in sampling successive reaction states and compensating for entropy penalties to binding [49].

The RPA1163 structures revealed that this enzyme employs the conserved catalytic triad Asp110-His280-Asp134 with Asp110 acting as a nucleophile, which directly displaces the substrate fluoride anion in an S_N2 attack at C2 (Fig. 1). The cleavage of the C-F bond produces the glycolyl-enzyme intermediate, which is then hydrolyzed by a water molecule activated by the His280 base (Fig. 1) [36]. Another structural study with the *Burkholderia* FA dehalogenase revealed that FA can be bound in the enzyme active site in an alternative orientation with the fluorine atom making contacts with the guanidino group of Arg108 (Arg114 in RPA1163) and a water molecule [43]. Additional studies with RPA1163 confirmed the proposed catalytic mechanism and revealed that this enzyme also accepts bulky substrates including 2-fluoro-2-phenylacetic and 2-fluoro-2-phenylpropionic acids [30, 51]. Furthermore, structural and biochemical studies of RPA1163 identified the active site residues critical for catalytic activity (Arg110, Arg111, Arg114, Arg134, His155, Trp156, Tyr219, and His280), as well as potential hot-spots (Trp185) for engineering this enzyme for improved activity toward bulky α-carboxylic acids for kinetic resolution of these substrates on a gram scale [36, 51]. Using *in silico* calculations and semi-rational mutagenesis, two mutant RPA1163 proteins (W185N and W185T) were generated, which showed improved activity against these substrates [51]. Very recently, mass spectrometry-based assays also showed that wild type RPA1163 appears to catalyze the defluorination of DFA to glyoxylate [52].

Overall, most of the previous studies on enzymatic defluorination were carried out using a single enzyme, the fluoroacetate dehalogenase RPA1163 from *Rhodopseudomonas palustris*, and FA as substrate [30, 36, 50, 51, 53]. However, polyfluorinated acids including DFA, TFA, and PFOA have wide industrial use and attracted increasing attention due to environmental persistence and aquatic toxicity [54]. For example, DFA is used for the

regioselective hydrodifluoromethylation of alkenes in organic synthesis of pharmaceutical building blocks [55]. In addition, these fluorochemicals can be formed in the environment during abiotic and biological degradation of more complex fluorocarbons (surfactants, refrigerants, pesticides) [54, 56, 57]. Numerous environmental and microbiological studies demonstrated the biodegradation of polyfluorinated acids under aerobic and anaerobic conditions suggesting that there are enzymes capable of defluorination of these compounds [19, 21, 26-28, 58, 59]. In this study, eight purified hydrolytic dehalogenases were screened for defluorination activity against DFA and other substrates revealing three enzymes with significant activity against DFA. The organic reaction product of DFA defluorination by the fluoroacetate dehalogenase DAR3835 from *Dechloromonas aromatica* was identified as glyoxylic acid indicating that this enzyme catalyzes the complete defluorination of DFA. Crystal structures of DAR3835 and NOS0089 from *Nostoc* sp. were determined including the structure of the DAR3835-glycolyl intermediate. Structure-based site-directed mutagenesis of DAR3835 revealed the active site residues essential for DFA defluorination and together with ligand docking and tunnel modeling suggested the potential catalytic mechanism of DFA defluorination.

Results and Discussion

Screening of purified dehalogenases for defluorination activity against difluoroacetate and polyfluorinated acids

To identify defluorinating dehalogenases active against DFA and other fluorinated acids, we selected eight bacterial enzymes, for which FA dehalogenase activity was reported in our previous study [41]. We hypothesized that defluorination activity of these enzymes might not be limited by FA, and some of them will also exhibit promiscuous hydrolytic activity with related substrates. These proteins include four α/β hydrolases (DAR3835 from *Dechloromonas aromatica*, NOS0089 from *Nostoc* sp., POL4478 from *Polaromonas* sp., and RPA1163 from *Rhodopseudomonas palustris*) and four haloacid dehalogenase-like (HAD) hydrolases (ADE3811 from *Anaeromyxobacter dehalogenans*, POL0530 and POL4516 from *Polaromonas* sp., and RJO0230 from *Rhodococcus jostii*) (Table S1, Fig. S1). These proteins share low to moderate sequence similarity to each other, from 17.0% to 53.6% sequence identity (Table S2, Fig. S2). Purified proteins were screened for defluorination activity using a pH-indicator assay with FA, DFA, as well as 2,2-difluoropropionic acid (DFPA), 5,5,5-trifluoropentanoic acid (TFPA), and perfluorooctanoic acid (PFOA) as substrates. As shown in Fig. 2, the screens confirmed the presence of FA defluorination activity in all purified proteins. Furthermore, significant DFA defluorination activity was observed in three α/β hydrolases (DAR3835, NOS0089, and POL4478) with the highest activity observed in DAR3835, which was comparable to its FA defluorination activity (Fig. 2). DFA defluorination activity of NOS0089 and POL4478 was slightly lower than that with FA (60-75%), whereas RPA1163 and HAD-like proteins (ADE3811, POL0530, POL4615, and RJO0230) exhibited low activity against DFA (Fig. 2). Finally, these screens revealed negligible or no defluorination activity with DFPA, TFPA, and PFOA in all tested proteins (Fig. 2).

Analysis of reaction products of enzymatic defluorination of FA and DFA

Liquid chromatography-mass spectrometry (LC-MS) analysis identified glycolic acid as the organic reaction product of FA defluorination by purified DAR3835 confirming that this enzyme is using the same defluorination mechanism as RPA1163 (Fig. 3). However, LC-MS analysis of reaction products of DFA defluorination by DAR3835 and RPA1163 revealed the formation of glyoxylic acid as the product (Fig. 3, Fig. S3A). This also suggests that both enzymes catalyze complete defluorination of DFA via two sequential defluorination reactions with the formation of 2-fluoro-2-hydroxyacetic acid (FHA) as the intermediate product. It is anticipated that the complete DFA defluorination by DAR3835 and RPA1163 will produce 2-dihydroxyacetate (a geminal diol) as the final product. In water solutions, 2-dihydroxyacetate exists in equilibrium with glyoxylic acid [60], but it is dehydrated to glyoxylic acid in organic solvents (during hydrophobic chromatography and MS analysis, see Materials and Methods). Recently, DFA defluorination activity of RPA1163 with the formation of glyoxylic acid as the final product was also reported by Yue et al. [52]. Purified DAR3835 exhibited comparable defluorination activities against both FA and DFA with K_M 3.8 ± 0.6 mM and 3.3 ± 0.5 mM, respectively (V_{max} 0.21-0.29 ± 0.02 μmoles/min per mg protein) (Fig. S3B). These values are in the same order of magnitude as those reported previously for FA defluorination by RPA1163 [41]. Since DAR3835, NOS0089, and POL4478 showed the highest DFA defluorination activity compared to other tested dehalogenases (Fig. 2), these proteins were selected for structural analysis.

Crystal structures of the DFA defluorinating dehalogenases: DAR3835 and NOS0089

Purified α/β hydrolases with high DFA defluorination activity (DAR3835, NOS0089, and POL4478) were submitted to crystallization trials, and crystal structures of DAR3835 (PDB code 8SDC) and NOS0089 (PDB code 3QYJ) were determined by molecular replacement at 1.86 Å and 1.78 Å resolution, respectively (see Materials and Methods for details) (Table S3). The structure of FA dehalogenase from *Burkholderia* sp. FA1 (PDB code 1Y37) was used as a model for NOS0089 (47% sequence identity), which was subsequently used (PDB code 3QYJ) as a model to solve the structure of DAR3835 (54% sequence identity). The crystal structure of the DAR3835 protomer revealed a classical α/β hydrolase fold with a core domain comprised of a slightly twisted central β-sheet containing eight parallel β-strands (except for the antiparallel β-2), which is surrounded by four α-helices on one side and seven α-helices on another side (Fig. 4). The DAR3835 core domain is covered by a U-shaped all-α cap domain (Thr142-Phe190) containing six short α-helices connected by flexible loops (Fig. 4). Similarly, the NOS0089 structure revealed an α/β hydrolase core domain with mostly parallel central β-sheet containing eight β-sheets surrounded by ten α-helices and covered with an all-α cap domain (Fig. S4).

A Dali search for structural homologues of DAR3835 in the PDB database identified NOS0089 as the top match: Z score 50.2 and root-mean-square deviation (rmsd) 0.8 Å (54% sequence identity), as well as the structures of FA dehalogenases from *Burkholderia* sp. FA1 (PDB code 3B12, Z score 47.2, rmsd 1.1 Å, 41% sequence identity) and RPA1163 (PDB code 6GXT, Z score 46.6, rmsd 1.2 Å, 42% sequence identity). This search also recognized hundreds of structurally homologous α/β hydrolases with low overall sequence similarity (10-40% identity) and annotated as putative hydrolases, epoxide hydrolases, esterases, and

peptidases. For both DAR3835 and NOS0089, the crystallographic asymmetric units showed the presence of two protomers connected via their cap domains suggesting that like most proteins [61], they exist as dimers in solution (Fig. 4, Fig. S4). This is supported by the results of the quaternary prediction server PISA [62] indicating that both DAR3835 and NOS0089 may form dimers in solution. Recent structural studies with RPA1163 demonstrated that the cap domains of this homodimeric enzyme mediate substrate-coupled allosteric interactions of two protomers resulting in half-of-the-sites reactivity with one subunit reacting at a time [49, 50]. Analysis of surface potential of the dimeric structures of DAR3835, NOS0089, and RPA1163 revealed a similar distribution of several patches of positively and negatively charged residues in these proteins with DAR3835 and RPA1163 showing higher surface negative charge compared to NOS0089 (Fig. 5). Analysis of surface hydrophobicity of the structures of protein dimers demonstrated the predominance of polar residues in these proteins with a few hydrophobic patches including a narrow active site opening between the core and cap domains (Tyr150, Trp153, Trp182, Met184, Phe247, Tyr252) (Fig. 5, 6A). Similarly, the NOS0089 structure showed the presence of mostly hydrophobic residues near the active site opening (Tyr147, Trp150, Trp179, Phe183, Ile243, Tyr248) (Fig. 6B).

The cap domains of both DAR3835 and NOS0089 include many hydrophobic residues establishing a hydrophobic surface descending down to the substrate access tunnel and active site (Fig. 6). In both enzymes, the active site is deeply buried inside the core domain and is connected to the exterior by the narrow substrate access tunnel (Fig. 6A, 6B). In DAR3835, the location of the active site is indicated by the bound Cl^- ion positioned between the side chains of conserved Asp107 (3.2 Å) and Arg111 (3.1 Å) (Fig. 6C). Asp107 is part of the DAR3835 catalytic triad including Asp107 (a nucleophile), His274 (a base, 2.8 Å to Asp107) and Asp131 (an acid, 2.6 Å to His274) (Fig. 6C). Similarly, the catalytic triad of NOS0089 comprises Asp104, His270 (2.8 Å to Asp104), and Asp128 (2.5 Å to His270) (Fig. 6D). Like in RPA1163 [36], the FA carboxylates appear to be coordinated by the side chains of conserved Arg108 and Arg111 located at the bottom of the DAR3835 active site (Arg105 and Arg108 in NOS0089), whereas the fluorine atom is H-bonded to His152, Trp153, and Tyr215 (His149, Trp150, and Tyr211 in NOS0089) (Fig. 6C, 6D). Previous studies with the FA dehalogenases from *Burkholderia* sp. FA1 and *R. palustris* (RPA1163) indicated that the cleavage of the C-F bond is greatly facilitated by hydrogen bond interactions between the fluorine atom and these three active site residues [36, 42, 63]. Thus, the characterized microbial defluorinases DAR3835, NOS0089, and RPA1163 have similar three-dimensional structures with conserved catalytic triads (Asp-His-Asp) and substrate binding residues involved in the coordination of the substrate fluorine and carboxylate groups.

Based on the FA defluorination mechanism proposed for RPA1163 [36], the cleavage of the C-F bond by DAR3835 is initiated by the side chain oxygen atom of Asp107 attacking the C2 carbon of FA and expelling the fluoride product with the formation of the D107-glycolyl ester intermediate (Fig. 1). Next, the D107-glycolyl intermediate is hydrolyzed using the catalytic water molecule activated by the His274 base and Asp131 acid and producing glycolate as the second product. This catalytic mechanism is supported by the crystal structure of the D107-glycolyl intermediate determined using the catalytically inactive DAR3835 H274N mutant protein co-crystallized with FA (Fig. 6E, 6F). This structure shows the glycolyl moiety covalently attached to the Asp107 side chain with the glycolyl

carboxylate oxygens coordinated by the side chain nitrogens of Arg108 (2.6 Å) and Arg111 (2.9 Å) (Fig. 6E, 6F). In addition, this structure revealed the presence of the Cl^- ion, mimicking the position of the bound F^- product and located near the side chains of Arg111 (2.9 Å) and Trp153 (4.3 Å) (Fig. 6E, 6F).

The proposed catalytic mechanism of FA defluorination by DAR3835 is in line with the results of site-directed mutagenesis of this protein, which revealed a strong negative effect of mutations of the catalytic triad residues (D107N, H274N), as well as of the residues involved in the coordination of the FA carboxylate (R108A) and fluorine (W153A) (Fig. 7). Since the D107A mutant protein was found to be insoluble, the D107N protein was generated and found to retain detectable FA defluorination activity (~15% of wild type activity) indicating that the Asn side chain can also support FA defluorination (Fig. 7). In addition, FA defluorination activity was found to be reduced in the DAR3835 mutant proteins Y37A and H106A (3.5-3.6 Å to the catalytic Asp107), F154A and I167A (part of the cap domain), and F275A (3.7 Å to the catalytic His274) (Fig. S5). In contrast, alanine replacement mutagenesis of the active site residues Met166 and Cys178 produced mutant proteins with the wild-type FA defluorination activity (Fig. 7). Interestingly, DFA defluorination activity of DAR3835 was negatively affected by most of these mutations suggesting that this activity is more sensitive to structural perturbations in the active site than FA defluorination (Fig. 7). For example, the C178A and F275A mutant proteins showed the wild-type FA defluorination activity, but were inactive toward DFA implying that these residues might be contributing to DFA coordination in the DAR3835 active site (Fig. S5). Similarly, site-directed mutagenesis of RPA1163 also showed that mutations of the active site residues had a stronger negative impact on DFA defluorination activity compared to FA defluorination (Fig. S3).

Structural analysis of the active sites of DAR3835, NOS0089, and RPA1163

To gain additional insight into DFA defluorination activity of purified defluorinases, we performed a comparative analysis of the active sites of enzymes with high (DAR3835 and NOS0089) and low (RPA1163) DFA defluorination activity. In many enzymes including hydrolases, the active sites are deeply buried inside the core domains and can be reached by substrates via one or several access tunnels [64-66]. In haloalkane dehalogenases, the active sites are buried between the core and cap domains, and they are accessible via a single tunnel in the *Xanthobacter autotrophicus* DhlA, two tunnels in the *Rhodococcus* sp. DhaA, or three tunnels in the *Sphingomonas paucimobilis* LinB [67]. It has been shown that the substrate access tunnels connecting active sites with bulk solvent influence ligand binding and substrate specificity in enzymes with buried active sites [68-70].

The presence of substrate access tunnels in the apo-structures of DAR3835, NOS0089, and RPA1163 was analyzed using the MOLE 2.5 software, which also allows to compute cavities (pockets) and their physicochemical properties [71, 72]. This analysis identified the presence of single elongated tunnels in these proteins (one tunnel per protomer) connecting the catalytic Asp in their active sites with the protein exterior (Fig. 8). Like in many α/β hydrolases with cap domains [66], the exits of defluorinase tunnels are located between the core and cap domains with mostly hydrophobic protein surface (Fig. 5, 6, 8). The presence of one elongated substrate-access tunnel in the crystal structure of RPA1163 has also been reported previously [36, 50]. In terms of shape, both protomers of DAR3835 and NOS0089

showed the presence of similar elongated tunnels, whereas different tunnels were observed in the RPA1163 apo-dimer with an elongated tunnel in the protomer A and a two-branched tunnel found in the protomer B (Fig. 8). The lengths of the calculated defluorinase tunnels span from 15 Å to 21 Å with DAR3835 and NOS0089 showing longer tunnels compared to RPA1163, whereas their physicochemical parameters (charge, hydrophobicity, polarity) were similar (Table S4). Furthermore, the two protomers of DAR3835 and NOS0089 revealed similar volumes of active site pockets (724.9/733.6 $Å^3$ and 909.3/1043.3 $Å^3$, respectively), whereas analysis of the RPA1163 subunits showed highly dissimilar pocket volumes (668.3 $Å^3$ and 1062.3 $Å^3$) (Table S5). Similar results were obtained after calculating the active site volumes of the glycolyl-enzyme intermediates of DAR3835 (996.3 $Å^3$ and 1072.5 $Å^3$) and RPA1163 (624.0 $Å^3$ and 1234.0 $Å^3$) (data not shown). The results of our analysis of active site volumes in the RPA1163 protomers are in line with the recent structural studies of this protein, which demonstrated "half-of-the-sites reactivity" based on dimer asymmetry and protomer dynamics in this enzyme [49, 50]. These studies revealed the presence of an allosteric pocket above the RPA1163 active site with a second FA molecule directly coordinated by the side chains of Tyr141 and Lys152 [49, 50]. In RPA1163, the side chain of Lys152 provides a key hydrogen bond stabilizing the second FA molecule bound in the allosteric site and resulting in enzyme substrate inhibition, which was abolished in the K152I mutant protein [50]. Although the structures of DAR3835 and NOS0089 demonstrated the presence of long substrate access tunnels leading to the active site (Fig. 8), the Lys and Tyr residues corresponding to the RPA1163 Lys152 and Tyr141 are not conserved in these proteins and other FA dehalogenases (Fig. S2). Thus, in contrast to RPA1163, structural analysis of the active sites of DAR3835 and NOS0089 revealed no apparent dimer asymmetry.

Since co-crystallization of purified wild-type and mutant DAR3835 proteins with FA and DFA produced no diffracting crystals, we performed ligand docking simulations to determine energetically favorable orientations of these substrates in the DAR3835 active site using the Autodock program [73]. Docking calculations for FA and DFA binding in the RPA1163 active site produced the substrate binding models similar to that observed in the crystal structure of the RPA1163-FA complex (PDB code 3R3V). The top-ranked conformations with minimal binding energies show that the substrate carboxylate group is coordinated by the conserved Arg111 and Arg114 and the fluorine atom is bound close to the halide-binding pocket (His155, Trp156, and Tyr219) (Fig. 9, Table S6). Similarly, the docking results for binding of FA and DFA in the DAR3835 active site suggest that Arg108 and Arg111 act as carboxyl-stabilizing residues, whereas His152, Trp153, and Tyr215 function as fluorine-stabilizing residues (Fig. 9, Table S6). In addition, ligand docking in the DAR3835 active site suggested that in this protein substrate binding might also be facilitated by the side chain of Tyr150, which is conserved in FA dehalogenases with high DFA activity (Tyr147 in NOS0089 and Tyr160 in POL4478), but replaced by Ile153 in RPA1163 (Fig. S2). Overall, the ligand docking results suggest that enzymatic defluorination of FA and DFA proceed via the same catalytic mechanism.

Proposed catalytic mechanism of DFA defluorination by DAR3835

Similar to other members of the α/β hydrolase superfamily [66], the catalytic machinery of defluorinases involves a catalytic triad composed of a nucleophile, a base, and an acid (in DAR3835: Asp107, His274, and Asp131, respectively). Previous structural and computational studies using the crystal structures of FA dehalogenases from *Burkholderia* sp. FA1 and RPA1163 suggested that enzymatic defluorination of FA involves the formation and hydrolysis of the covalent glycolyl-enzyme adduct, which is expected to proceed via a tetrahedral intermediate stabilized by an oxyanion hole (the main chain amides of Phe40 and Arg111 in RPA1163) [36, 52, 63]. Since DAR3835 produces glyoxylic acid as the final product of DFA defluorination, we propose that enzymatic defluorination of DFA is achieved via two successive defluorination reactions with the formation of two covalent acyl-enzyme intermediates and two tetrahedral intermediates (Fig. 10).

In the DAR3835/DFA (Michaelis) complex, the bound DFA is anchored in the active site by hydrogen bonds with its carboxylate oxygens coordinated by conserved Arg108 and Arg111 and the F_1 fluorine atom interacting with the three fluoride-stabilizing residues: His152, Trp153, and Tyr215 (Fig. 10). Next, the S_N2 reaction is initiated by nucleophilic attack of the Oδ2 atom of Asp107 at C2 expelling the fluoride-1 with inversion of configuration at C2 and producing a covalent fluoro-glycolyl-enzyme intermediate (acyl intermediate-1) (Fig. 10). The hydrolysis of the acyl intermediate-1 is catalyzed by the nucleophilic attack of the water molecule (activated by a proton transfer to His274) producing the intermediate product 2-fluoro-2-hydroxy-acetate (FHA) via the tetrahedral intermediate-1 stabilized by the oxyanion hole formed by the main chain NH groups of Tyr37 and Arg108 (Fig. 10). The second nucleophilic (S_N2) attack by Asp107 at C2 produces the second fluoride product (F_2) and 2-hydroxy-glycolyl-enzyme intermediate (acyl intermediate-2), which is hydrolyzed by an activated water molecule via the tetrahedral intermediate-2 (Fig. 10). This will generate the final product of DFA defluorination, gem-diol 2,2-dihydroxyacetate (DHA), which is dehydrated to glyoxylate in organic solvents. Furthermore, recent molecular dynamics simulations and QM/MM calculations using the RPA1163 structure suggested that DFA defluorination by this enzyme can be accomplished by the S_N2 mechanism without the formation of a tetrahedral intermediate [52]. According to this model, enzymatic defluorination of DFA can proceed via the direct cleavage of the acyl-enzyme C^2-O^1 bond without forming a tetrahedral intermediate (Fig. S6).

Conclusions

Recently, the FA dehalogenase RPA1163 from *Rhodopseudomonas palustris* attracted attention due to its ability to accept bulkier substrates (e.g. 2-fluoro-2-phenyl acetate) suggesting the potential for applications in greener chemical synthesis and environmental bioremediation [30, 51]. In the present study, we have demonstrated the ability of microbial FA dehalogenases to catalyze the complete defluorination of difluoroacetate to glyoxylic acid, thus extending the substrate scope of these enzymes to difluorinated carboxylic acids. These results indicate that microbial FA dehalogenases exhibit considerable substrate promiscuity and can also accept polyfluorinated substrates. Crystal structures of the DFA defluorinases DAR3835 and NOS0089 along with the DAR3835 glycolyl intermediate provided structural insights into the active site residues and potential catalytic mechanisms of DFA defluorination. Computational analysis of the DAR3835 and NOS0089 structures

showed the presence of similar active site pockets and tunnels in both protomers and suggested that DFA defluorination proceeds via two consecutive defluorination reactions. It appears that the utility of fluorinated chemicals will ensure their continued use in the foreseeable future, stressing the need for the development of efficient biocatalytic technologies for their bioremediation and chemical synthesis. Thus, screening and protein engineering of microbial FA dehalogenases have the potential to generate robust defluorinases with different substrate scopes for industrial applications.

Materials and methods

Reagents

All chemicals and substrates used in this study were of the highest purity (analytical grade) and were purchased from Sigma/Aldrich/Merck including difluoroacetic acid (DFA, 98%, catalog # 142859), fluoroacetic acid (FA, 95%, # 796875), 2,2-difluoropropionic acid (DFPA, 97%, # 684686), 5,5,5-trifluoropentanoic acid (TFPA, # CDS002681), and perfluorooctanoic acid (PFOA, 95%, #171468).

Protein purification and site-directed mutagenesis

The coding sequences of selected dehalogenase genes were PCR amplified using genomic DNA and cloned into a modified pET15b (Novagen) expression vector encoding an N-terminal 6His-tag. For protein expression, the plasmids were transformed into *Escherichia coli* BL21(DE3) Codon-Plus strain (Stratagene) as the expression host. The proteins were expressed and affinity purified using metal-chelate chromatography on Ni-NTA (Qiagen) as described previously [41]. Site-directed mutagenesis (alanine scanning) was performed based on the QuikChange® site-directed mutagenesis kit (Stratagene), and all mutations were verified by DNA sequencing.

Enzyme activity assays and reaction product analysis using LC-MS

Defluorination activity of purified proteins was measured using a pH indicator dye-based spectrophotometric assay developed for measuring dehalogenase activity quantitatively in microplates [74]. This assay is based on a decrease in the pH of a weakly buffered reaction mixture by hydrohalic acid (hydrofluoric acid) produced by dehalogenating enzymes, and the color change of the pH indicator phenol red was monitored spectrophotometrically at 540 nm [36, 41, 74]. The reaction mixture contained (final volume 200 µl) 1.0 mM HEPES-K buffer (pH 7.5), 2 µg/ml of phenol red (a pH indicator), purified enzyme (10 µg or 20 µg, as indicated), and fluorinated substrates (5 mM for substrate screening or 10 mM for product analysis or assays with mutant proteins). After incubation at 30ºC (1-5 h or overnight), the absorbance at 540 nm was recorded. All enzyme assays were performed in triplicate. For determination of kinetic parameters, defluorinase activity was determined over a range of substrate concentrations (0-20 mM), and kinetic parameters were calculated by non-linear regression analysis of raw data fit to the Michaelis-Menten function using GraphPad Prism software (version 5.00 for Windows). Reaction mixtures for defluorination product analysis using mass spectrometry (MS) included (final volume 200 µl) 5 mM DFA (or FA), 10 mM HEPES-K (pH 7.5), and 50 µg of purified protein. After 2 h of incubation at 30ºC, the

samples were filtered through centrifugal filter devices (Mw cut-off 10 kDa), and the flow-through fractions were used for MS product analysis.

Liquid chromatography-mass spectrometry (LC-MS) analysis of fluoroacetate (m/z 77.0044), difluoroacetic acid (m/z 94.995), glycolic acid (75.0088), and glyoxylic acid (m/z 72.9931) was performed using a Dionex Ultimate 3000 UHPLC systems and a Q-Exactive mass spectrometer equipped with a HESI-II probe (all from Thermo Scientific) and controlled by Thermo XCalibur 4.1 software. LC fractionation of chemicals was conducted on a Hypersil Gold C18 column (50 mm x 2.1 mm, 1.9 µ particle size, Thermo Scientific) equipped with a guard column (column temperature 40ºC). Solvent A was 5 mM ammonium acetate in water, solvent B was 5 mM ammonium acetate in methanol (flow rate 0.3 ml/min). Autosampler temperature was maintained at 10ºC, and the injection volume was 10 µl. The elution gradient was: 0-1 min, 5% B; 1-7 min, linear gradient to 100% B; 7-10 min, 100% B; 10-10.5 min, linear gradient to 0% B; 10.5-15.0 min, 0% B. Data collection was done in negative ionization mode with a scan range *m/z* 70-300, resolution 140 000 at 1 Hz, AGC target of 3e6 and a maximum injection time of 200 ms. Standard solutions of glyoxylic acid and difluoroacetic acid were used for validation of retention time and *m/z.*

Protein crystallization, data collection, and structure determination

Purified proteins DAR3835 and NOS0089 were crystallized using the sitting-drop vapour diffusion method. Crystals of the wild-type DAR3835 and NOS0089 were obtained by mixing 0.6 µl of protein solution (1 mM $MgCl_2$, 5 mM FA, and 20 mg/ml of DAR3835 or 25 mg/ml of NOS0089) with 0.6 µl of reservoir solution (DAR3835: 25% (w/v) PEG3350, 0.2 M $MgCl_2$, 0.1 M Bis-Tris, pH 5.5; NOS0089: 25% PEG3350, 0.2 M NaCl, 0.1 M Na-citrate buffer, pH 5.6). For the DAR3835 H274N protein (D107-glyoxylate intermediate), the reservoir solution contained 25% (w/v) PEG3350, 0.2 M NaCl, and 0.1 M Bis-Tris, pH 5.5. Prior to data collection, crystals were cryoprotected either by transferring into the reservoir solution containing 2% (v/v) PEG200 followed by paratone or using only paratone oil. Diffraction data was collected at 100 K on a Rigaku HF-007 rotating anode with R-AXIS IV detector or a Rigaku FR-E+ Superbright rotating anode with R-AXIS IV++ detector, processed using HKL-3000 [75] or XDS [76] and CCP4/Aimless [77], and structures were solved by Molecular Replacement using Phenix phaser [78]. The structure of NOS0089 (ALR0039) was determined using a model constructed by the Phyre2 server [79] onto the structure of FA dehalogenase from *Burkholderia* sp. FA1 (PDB code 1Y37, 47% sequence identity to NOS0089). The structures of DAR3835 were solved using a Phyre2 model constructed based on the structure of NOS0089 (54% sequence identity to DAR3835). For DAR3835, difference (Fo-Fc) densities for chlorine atoms in the crystals of the wild type protein and the D107-glycolyl intermediate (H274N protein) were unambiguous. Refinement of NOS0089 was performed with Refmac5 [80], whereas refinement of the DAR3835 structures was executed using Phenix.refine and included TLS parameterization. B-factors for all structures were refined as isotropic, and geometry was verified using Phenix.molprobity and the wwPDB server.

Sequence and structural analyses

Multiple sequence alignments of defluorinating α/β hydrolases and HAD-like enzymes were constructed using the MAFFT online service (version 7) (https://mafft.cbrc.jp/alignment/server/) [81]. The PyMOL Molecular Graphics System

(version 1.7.2.1, Schrodinger, LLC, http://www.pymol.org/pymol) was used for the analysis of crystal structures of defluorinases and preparation of structural figures. Surface electrostatic potential of defluorinases was calculated using the PDB2PQR and APBS servers (https://server.poissonboltzmann.org/) [82, 83], whereas the surface hydrophobicity was displayed using Chimera v1.15 and ChimeraX1.4 [84]. Molecular docking simulation of ligand binding (FA and DFA) to the DAR3835 and RPA1163 dimers was performed using the AutoDock suite of automated docking tools (AutoDock 4.2.6 / Autodock tools 1.5.6 with ten minimal energy binding positions calculated for each ligand and protomer [73]. For the grid coordinates, the following active site residues were used for RPA1163 (Asp110, Asp134, Tyr141, Trp156, and His280) and DAR3835 (Asp107, Asp131, Trp153, and His274). Protein secondary structure analysis was carried out using the protein secondary structure prediction server JPred4 [85]. The universal toolkit MOLE 2.5 was used for the identification and analysis of tunnels and active sites in the dimeric structures of DAR3835 (PDB code 8SDC), NOS0089 (PDB code 3QYJ), and RPA1163 (PDB code 5K3D) [71]. The tunnels were calculated using the catalytic Asp107 (DAR3835), Asp104 (NOS0089), or Asp110 (RPA1163) as starting points and the tunnel bottleneck threshold 1 Å (cavity parameters: probe radius 3 Å, interior threshold 1.25 Å, minimal depth 5 Å). The Voronoi scale and Length+Radius weight functions were used for building protein tunnels and calculating standard deviation (SD) (https://webchem.ncbr.muni.cz/Platform/App/Mole) [72]. Polarity of amino acids lining the cavity was calculated as an average of amino acid polarities assigned according to Zimmerman et al., 1968 [86].

Author contributions

ANK, AS and AFY designed the project and experiments. ANK, KAB, TF, YSC, GB, and RF conducted biochemical experiments and MS analyses. TS, APP, PS, and AS carried out crystallographic studies and structure determination. The manuscript was written by ANK, PS, and AFY with the contribution of all other coauthors.

Acknowledgements

The authors thank all members of the BioZone Centre for Applied Science and Bioengineering and BioZone MS facility for help in conduction the experiments. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) Strategic Network grant IBN and the Samsung (SEC) GRO-2018 grant “Structural studies and protein engineering of microbial defluorinases”. Diffraction data were collected at the Structural Biology Center (Advanced Photon Source) supported by US Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

Data availability statement

The atomic coordinates and structure factors of the models described in this manuscript were deposited into the Protein Data Bank with accession codes: 8SDC (wild type DAR3835, apo-structure), 8SDD (DAR3835 H274N, D107- glycolyl-intermediate), and 3QYJ (wild type NOS0089).

References

1. Goldman, P. (1969) The carbon-fluorine bond in compounds of biological interest, *Science.* **164**, 1123-30.
2. Lemal, D. M. (2004) Perspective on fluorocarbon chemistry, *J Org Chem.* **69**, 1-11.
3. Douvris, C. & Ozerov, O. V. (2008) Hydrodefluorination of perfluoroalkyl groups using silylium-carborane catalysts, *Science.* **321**, 1188-90.
4. Smart, B. E. (2001) Fluorine substituent effects (on bioactivity). *J Fluor Chem.* **109**, 3-11.
5. Ritter, S. K. (2010) Fluorochemicals go short., *Chem Eng News.* **88**, 12-17.
6. Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J. & Meanwell, N. A. (2015) Applications of Fluorine in Medicinal Chemistry, *J Med Chem.* **58**, 8315-59.
7. Zhou, Y., Wang, J., Gu, Z., Wang, S., Zhu, W., Acena, J. L., Soloshonok, V. A., Izawa, K. & Liu, H. (2016) Next Generation of Fluorine-Containing Pharmaceuticals, Compounds Currently in Phase II-III Clinical Trials of Major Pharmaceutical Companies: New Structural Trends and Therapeutic Areas, *Chem Rev.* **116**, 422-518.
8. Fujiwara, T., and O'Hagan, D. (2014) Successful fluorine-containing herbicide agrochemicals., *J Fluor Chem.* **167**, 16-29.
9. Muller, K., Faeh, C. & Diederich, F. (2007) Fluorine in pharmaceuticals: looking beyond intuition, *Science.* **317**, 1881-6.
10. Key, B. D., Howell, R.D., and Criddle, C.S. (1997) Fluorinated organics in the biosphere., *Environ Sci Technol.* **31**, 2445-2454.
11. Wang, Y. & Liu, A. (2020) Carbon-fluorine bond cleavage mediated by metalloenzymes, *Chem Soc Rev.* **49**, 4906-4925.
12. Wackett, L. P. (2022) Nothing lasts forever: understanding microbial biodegradation of polyfluorinated compounds and perfluorinated alkyl substances, *Microb Biotechnol.* **15**, 773-792.
13. Huang, H., Meegalla, S. K., Lanter, J. C., Winters, M. P., Zhao, S., Littrell, J., Qi, J., Rady, B., Lee, P. S., Liu, J., Martin, T., Lam, W. W., Xu, F., Lim, H. K., Wilde, T., Silva, J., Otieno, M., Pocai, A. & Player, M. R. (2019) Discovery of a GPR40 Superagonist: The Impact of Aryl Propionic Acid alpha-Fluorination, *ACS Med Chem Lett.* **10**, 16-21.
14. Zhang, Y. K., Plattner, J. J., Freund, Y. R., Easom, E. E., Zhou, Y., Ye, L., Zhou, H., Waterson, D., Gamo, F. J., Sanz, L. M., Ge, M., Li, Z., Li, L., Wang, H. & Cui, H. (2012) Benzoxaborole antimalarial agents. Part 2: Discovery of fluoro-substituted 7-(2-carboxyethyl)-1,3-dihydro-1-hydroxy-2,1-benzoxaboroles, *Bioorg Med Chem Lett.* **22**, 1299-307.
15. Liang, T., Neumann, C. N. & Ritter, T. (2013) Introduction of fluorine and fluorine-containing functional groups, *Angew Chem Int Ed Engl.* **52**, 8214-64.
16. Gray, E. E., Nielsen, M. K., Choquette, K. A., Kalow, J. A., Graham, T. J. & Doyle, A. G. (2016) Nucleophilic (Radio)Fluorination of alpha-Diazocarbonyl Compounds Enabled by Copper-Catalyzed H-F Insertion, *J Am Chem Soc.* **138**, 10802-5.
17. Armitage, J. M., Macleod, M. & Cousins, I. T. (2009) Comparative assessment of the global fate and transport pathways of long-chain perfluorocarboxylic acids (PFCAs) and perfluorocarboxylates (PFCs) emitted from direct sources, *Environ Sci Technol.* **43**, 5830-6.
18. D'Eon, J. C. & Mabury, S. A. (2007) Production of perfluorinated carboxylic acids (PFCAs) from the biotransformation of polyfluoroalkyl phosphate surfactants (PAPS): exploring routes of human contamination, *Environ Sci Technol.* **41**, 4799-805.
19. Yu, X., Nishimura, F. & Hidaka, T. (2018) Effects of microbial activity on perfluorinated carboxylic acids (PFCAs) generation during aerobic biotransformation of fluorotelomer alcohols in activated sludge, *Sci Total Environ.* **610-611**, 776-785.

20. Post, G. B., Gleason, J. A. & Cooper, K. R. (2017) Key scientific issues in developing drinking water guidelines for perfluoroalkyl acids: Contaminants of emerging concern, *PLoS Biol.* **15**, e2002855.
21. Seong, H. J., Kwon, S.W., Seo, D.-C., Kim, J.-H., and Jiang, Y.-S. (2019) Enzymatic defluorination of fluorinated compounds., *Appl Biol Chem.* **62**, 62.
22. Luo, Q., Wang, Z., Feng, M., Chiang, D., Woodward, D., Liang, S., Lu, J. & Huang, Q. (2017) Factors controlling the rate of perfluorooctanoic acid degradation in laccase-mediator systems: The impact of metal ions, *Environ Pollut.* **224**, 649-657.
23. Davis, C. K., Webb, R. I., Sly, L. I., Denman, S. E. & McSweeney, C. S. (2012) Isolation and survey of novel fluoroacetate-degrading bacteria belonging to the phylum Synergistetes, *FEMS Microbiol Ecol.* **80**, 671-84.
24. Oltmanns, R. H., Muller, R., Otto, M. K. & Lingens, F. (1989) Evidence for a new pathway in the bacterial degradation of 4-fluorobenzoate, *Appl Environ Microbiol.* **55**, 2499-504.
25. Ferreira, M. I., Iida, T., Hasan, S. A., Nakamura, K., Fraaije, M. W., Janssen, D. B. & Kudo, T. (2009) Analysis of two gene clusters involved in the degradation of 4-fluorophenol by Arthrobacter sp. strain IF1, *Appl Environ Microbiol.* **75**, 7767-73.
26. Sunantha, G. & Vasudevan, N. (2020) A method for detecting perfluorooctanoic acid and perfluorooctane sulfonate in water samples using genetically engineered bacterial biosensor, *Sci Total Environ*, 143544.
27. Visscher, P. T., Culbertson, C.W., and Oremland, R.S. (1994) Degradation of trifluoroacetate in oxic and anoxic sediments., *Nature.* **369**, 729-731.
28. Alexandrino, D. A. M., Ribeiro, I., Pinto, L. M., Cambra, R., Oliveira, R. S., Pereira, F. & Carvalho, M. F. (2018) Biodegradation of mono-, di- and trifluoroacetate by microbial cultures with different origins, *N Biotechnol.* **43**, 23-29.
29. Butt, C. M., Muir, D. C. & Mabury, S. A. (2014) Biotransformation pathways of fluorotelomer-based polyfluoroalkyl substances: a review, *Environ Toxicol Chem.* **33**, 243-67.
30. Wang, J. B., Ilie, A., Yuan, S. & Reetz, M. T. (2017) Investigating Substrate Scope and Enantioselectivity of a Defluorinase by a Stereochemical Probe, *J Am Chem Soc.* **139**, 11241-11247.
31. Luo, Q., Yan, X., Lu, J. & Huang, Q. (2018) Perfluorooctanesulfonate Degrades in a Laccase-Mediator System, *Environ Sci Technol.* **52**, 10617-10626.
32. Murphy, C. D. (2010) Biodegradation and biotransformation of organofluorine compounds, *Biotechnol Lett.* **32**, 351-9.
33. Haas, R. & Nikel, P. I. (2023) Challenges and opportunities in bringing nonbiological atoms to life with synthetic metabolism, *Trends Biotechnol.* **41**, 27-45.
34. Calero, P., Volke, D. C., Lowe, P. T., Gotfredsen, C. H., O'Hagan, D. & Nikel, P. I. (2020) A fluoride-responsive genetic circuit enables in vivo biofluorination in engineered Pseudomonas putida, *Nat Commun.* **11**, 5045.
35. Natarajan, R., Azerad, R., Badet, B., and Copin, E. (2005) Microbial cleavage of C-F bond., *J Fluorine Chem.* **126**, 425-436.
36. Chan, P. W., Yakunin, A. F., Edwards, E. A. & Pai, E. F. (2011) Mapping the reaction coordinates of enzymatic defluorination, *J Am Chem Soc.* **133**, 7461-8.
37. Pardo, I., Bednar, D., Calero, P., Volke, D. C., Damborsky, J. & Nikel, P. I. (2022) A Nonconventional Archaeal Fluorinase Identified by In Silico Mining for Enhanced Fluorine Biocatalysis, *ACS Catal.* **12**, 6570-6577.
38. Fetzner, S. & Lingens, F. (1994) Bacterial dehalogenases: biochemistry, genetics, and biotechnological applications, *Microbiol Rev.* **58**, 641-85.

39. Swanson, P. E. (1999) Dehalogenases applied to industrial-scale biocatalysis, *Curr Opin Biotechnol.* **10**, 365-9.
40. Verschueren, K. H., Seljee, F., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1993) Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase, *Nature.* **363**, 693-8.
41. Chan, W. Y., Wong, M., Guthrie, J., Savchenko, A. V., Yakunin, A. F., Pai, E. F. & Edwards, E. A. (2010) Sequence- and activity-based screening of microbial genomes for novel dehalogenases, *Microb Biotechnol.* **3**, 107-20.
42. Jitsumori, K., Omi, R., Kurihara, T., Kurata, A., Mihara, H., Miyahara, I., Hirotsu, K. & Esaki, N. (2009) X-Ray crystallographic and mutational studies of fluoroacetate dehalogenase from Burkholderia sp. strain FA1, *J Bacteriol.* **191**, 2630-7.
43. Nakayama, T., Kamachi, T., Jitsumori, K., Omi, R., Hirotsu, K., Esaki, N., Kurihara, T. & Yoshizawa, K. (2012) Substrate specificity of fluoroacetate dehalogenase: an insight from crystallographic analysis, fluorescence spectroscopy, and theoretical computations, *Chemistry.* **18**, 8392-402.
44. Kurihara, T., Yamauchi, T., Ichiyama, S., Takahata, H., and Esaki, N. (2003) Purification, characterization, and gene cloning of a novel fluoroacetate dehalogenase from Burkholderia sp. FA1., *J Mol Catal B - Enzym.* **23**, 347-355.
45. Donnelly, C. & Murphy, C. D. (2009) Purification and properties of fluoroacetate dehalogenase from Pseudomonas fluorescens DSM 8341, *Biotechnol Lett.* **31**, 245-50.
46. Chan, P. W. Y., Chakrabarti, N., Ing, C., Halgas, O., To, T. K. W., Walti, M., Petit, A. P., Tran, C., Savchenko, A., Yakunin, A. F., Edwards, E. A., Pomes, R. & Pai, E. F. (2022) Defluorination Capability of l-2-Haloacid Dehalogenases in the HAD-Like Hydrolase Superfamily Correlates with Active Site Compactness, *Chembiochem.* **23**, e202100414.
47. Kurihara, T. & Esaki, N. (2008) Bacterial hydrolytic dehalogenases and related enzymes: occurrences, reaction mechanisms, and applications, *Chem Rec.* **8**, 67-74.
48. Liu, J. Q., Kurihara, T., Ichiyama, S., Miyagi, M., Tsunasawa, S., Kawasaki, H., Soda, K. & Esaki, N. (1998) Reaction mechanism of fluoroacetate dehalogenase from Moraxella sp. B, *J Biol Chem.* **273**, 30897-902.
49. Kim, T. H., Mehrabi, P., Ren, Z., Sljoka, A., Ing, C., Bezginov, A., Ye, L., Pomes, R., Prosser, R. S. & Pai, E. F. (2017) The role of dimer asymmetry and protomer dynamics in enzyme catalysis, *Science.* **355**.
50. Mehrabi, P., Di Pietrantonio, C., Kim, T. H., Sljoka, A., Taverner, K., Ing, C., Kruglyak, N., Pomes, R., Pai, E. F. & Prosser, R. S. (2019) Substrate-Based Allosteric Regulation of a Homodimeric Enzyme, *J Am Chem Soc.* **141**, 11540-11556.
51. Zhang, H., Tian, S., Yue, Y., Li, M., Tong, W., Xu, G., Chen, B., Ma, M., Li, Y., and Wang, J. (2020) Semirational design of fluoroacetate dehalogenase RPA1163 for kinetic resolution of a-fluorocarboxylic acids on a gram scale, *ACS Catalysis.* **10**, 3143-3151.
52. Yue, Y., Fan, J., Xin, G., Huang, Q., Wang, J. B., Li, Y., Zhang, Q. & Wang, W. (2021) Comprehensive Understanding of Fluoroacetate Dehalogenase-Catalyzed Degradation of Fluorocarboxylic Acids: A QM/MM Approach, *Environ Sci Technol.* **55**, 9817-9825.
53. Li, Y., Yue, Y., Zhang, H., Yang, Z., Wang, H., Tian, S., Wang, J. B., Zhang, Q. & Wang, W. (2019) Harnessing fluoroacetate dehalogenase for defluorination of fluorocarboxylic acids: in silico and in vitro approach, *Environ Int.* **131**, 104999.
54. McCulloch, A. (2003) Fluorocarbons in the global environment: a review of the important interactions with atmospheric chemistry and physics., *J Fluor Chem.* **123**, 21-29.
55. Meyer, C. F., Hell, S. M., Misale, A., Trabanco, A. A. & Gouverneur, V. (2019) Hydrodifluoromethylation of Alkenes with Difluoroacetic Acid, *Angew Chem Int Ed Engl.* **58**, 8829-8833.

56. Sun, M., Cui, J., Guo, J., Zhai, Z., Zuo, P. & Zhang, J. (2020) Fluorochemicals biodegradation as a potential source of trifluoroacetic acid (TFA) to the environment, *Chemosphere.* **254**, 126894.
57. Stuke, S., Bemboom, P., Wirkner, H., Smith, W. & Lock, S. J. (2021) Residue analytical method for the determination of trifluoroacetic acid and difluoroacetic acid in plant matrices by capillary electrophoresis tandem mass spectrometry (CE-MS/MS), *J Chromatogr A.* **1646**, 462096.
58. Kim, B. R., Suidan, M.T., Wallington, T.J., and Du, X. (2000) Biodegradability of trifluoroacetic acid., *Environ Eng Sci.* **17**, 337-342.
59. Huang, S. & Jaffe, P. R. (2019) Defluorination of Perfluorooctanoic Acid (PFOA) and Perfluorooctane Sulfonate (PFOS) by Acidimicrobium sp. Strain A6, *Environ Sci Technol.* **53**, 11410-11419.
60. Sorensen, P. E., Bruhn, K., and Lindelov, F. (1974) Kinetics and equilibria for the revesible hydration of the aldehyde group in glyoxylic acid., *Acta Chem Scand.* **28**, 162-168.
61. Goodsell, D. S. & Olson, A. J. (2000) Structural symmetry and protein function, *Annu Rev Biophys Biomol Struct.* **29**, 105-53.
62. Krissinel, E. & Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol.* **372**, 774-97.
63. Kamachi, T., Nakayama, T., Shitamichi, O., Jitsumori, K., Kurihara, T., Esaki, N. & Yoshizawa, K. (2009) The catalytic mechanism of fluoroacetate dehalogenase: a computational exploration of biological dehalogenation, *Chemistry.* **15**, 7394-403.
64. Raushel, F. M., Thoden, J. B. & Holden, H. M. (2003) Enzymes with molecular tunnels, *Acc Chem Res.* **36**, 539-48.
65. Huang, X., Holden, H. M. & Raushel, F. M. (2001) Channeling of substrates and intermediates in enzyme-catalyzed reactions, *Annu Rev Biochem.* **70**, 149-80.
66. Bauer, T. L., Buchholz, P. C. F. & Pleiss, J. (2020) The modular structure of alpha/beta-hydrolases, *FEBS J.* **287**, 1035-1053.
67. Otyepka, M. & Damborsky, J. (2002) Functionally relevant motions of haloalkane dehalogenases occur in the specificity-modulating cap domains, *Protein Sci.* **11**, 1206-17.
68. Prokop, Z., Gora, A., Brezovsky, J., Chaloupkova, R., Stepankova, V., and Damborsky, J. (2012) Engineering of protein tunnels: keyhole-lock-key model for catalysis by enzymes with buried active sites., *Protein Engineering Handbook (Lutz S and Bornscheuer U, eds), Wiley.* **3**, 421-464.
69. Kingsley, L. J. & Lill, M. A. (2015) Substrate tunnels in enzymes: structure-function relationships and computational methodology, *Proteins.* **83**, 599-611.
70. Brezovsky, J., Babkova, P., Degtjatik, O., Fortova, A., Gora, A., Iermak, I., Rezacova, P., Dvorak, P., Smatanova, I.K., Prokop, Z., Chaloupkova, R., and Damborsky, J. (2016) Engineering a de novo transport channel., *ACS Catalysis.* **6**, 7597-7610.
71. Berka, K., Hanak, O., Sehnal, D., Banas, P., Navratilova, V., Jaiswal, D., Ionescu, C. M., Svobodova Varekova, R., Koca, J. & Otyepka, M. (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels, *Nucleic Acids Res.* **40**, W222-7.
72. Sehnal, D., Svobodova Varekova, R., Berka, K., Pravda, L., Navratilova, V., Banas, P., Ionescu, C. M., Otyepka, M. & Koca, J. (2013) MOLE 2.0: advanced approach for analysis of biomacromolecular channels, *J Cheminform.* **5**, 39.
73. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J Comput Chem.* **30**, 2785-91.
74. Holloway, P., Trevors, J.T., and Lee, H. (1998) A colorimetric assay for detecting haloalkane dehalogenase activity., *J Microbiol Methods.* **32**, 31-36.

75. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes, *Acta Crystallogr D Biol Crystallogr.* **62**, 859-66.
76. Kabsch, W. (2010) Xds, *Acta Crystallogr D Biol Crystallogr.* **66**, 125-32.
77. Evans, P. R. & Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta Crystallogr D Biol Crystallogr.* **69**, 1204-14.
78. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr.* **66**, 213-21.
79. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nat Protoc.* **10**, 845-58.
80. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr D Biol Crystallogr.* **67**, 355-67.
81. Katoh, K., Rozewicki, J. & Yamada, K. D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, *Brief Bioinform.* **20**, 1160-1166.
82. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, *Nucleic Acids Res.* **32**, W665-7.
83. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc Natl Acad Sci U S A.* **98**, 10037-41.
84. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem.* **25**, 1605-12.
85. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. (2015) JPred4: a protein secondary structure prediction server, *Nucleic Acids Res.* **43**, W389-94.
86. Zimmerman, J. M., Eliezer, N. & Simha, R. (1968) The characterization of amino acid sequences in proteins by statistical methods, *J Theor Biol.* **21**, 170-201.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. SDS-PAGE analysis of purified FA dehalogenases used in this study.

Fig. S2. Multiple sequence alignment of hydrolytic dehalogenases used in this study.

Fig. S3. Defluorination activity of purified DAR3835 and RPA1163.

Fig. S4. Crystal structure of NOS0089: overall fold.

Fig. S5. DAR3835 active site: amino acid residues selected for site-directed mutagenesis.

Fig. S6. Potential catalytic mechanism of DFA defluorination by DAR3835 without tetrahedral intermediates.

Table S1. Bacterial dehalogenases purified and screened in this study.

Table S2. Amino acid sequence identity of bacterial dehalogenases used in this study.

Table S3. X-ray crystallographic statistics for the structures of DAR3835 and NOS0089.

Table S4. Structure-based computational analysis of protein tunnels in DAR3835, NOS0089, and RPA1163.

Table S5. Molecular parameters of the active site cavities of DAR3835, NOS0089, and RPA1163.

Table S6. Binding energies for ligand binding to DAR3835 and RPA1163.

Figures

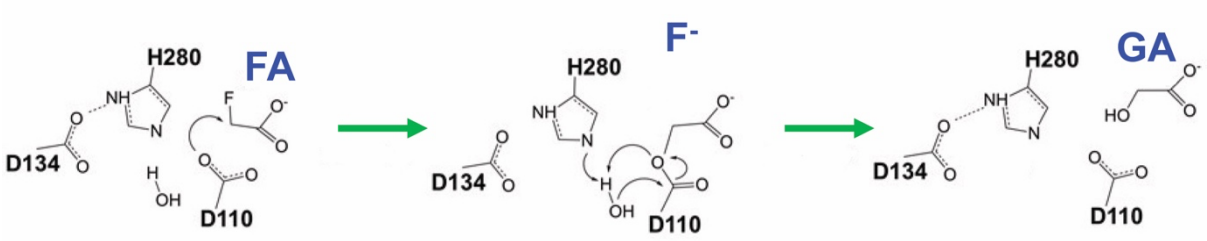


Fig. 1. The accepted two-step mechanism of fluoroacetate (FA) dehalogenation catalyzed by RPA1163 from *Rhodopseudomonas palustris*. The reaction involves the conserved catalytic triad Asp110-His280-Asp134 catalyzing a classical S_N2 reaction, in which Asp110 functions as the nucleophile with direct expulsion of fluoride and formation of the covalent enzyme-glycolyl intermediate. The resulting covalent intermediate is then hydrolyzed by a water molecule activated by the His280 base producing glycolic acid (GA) [30, 36].

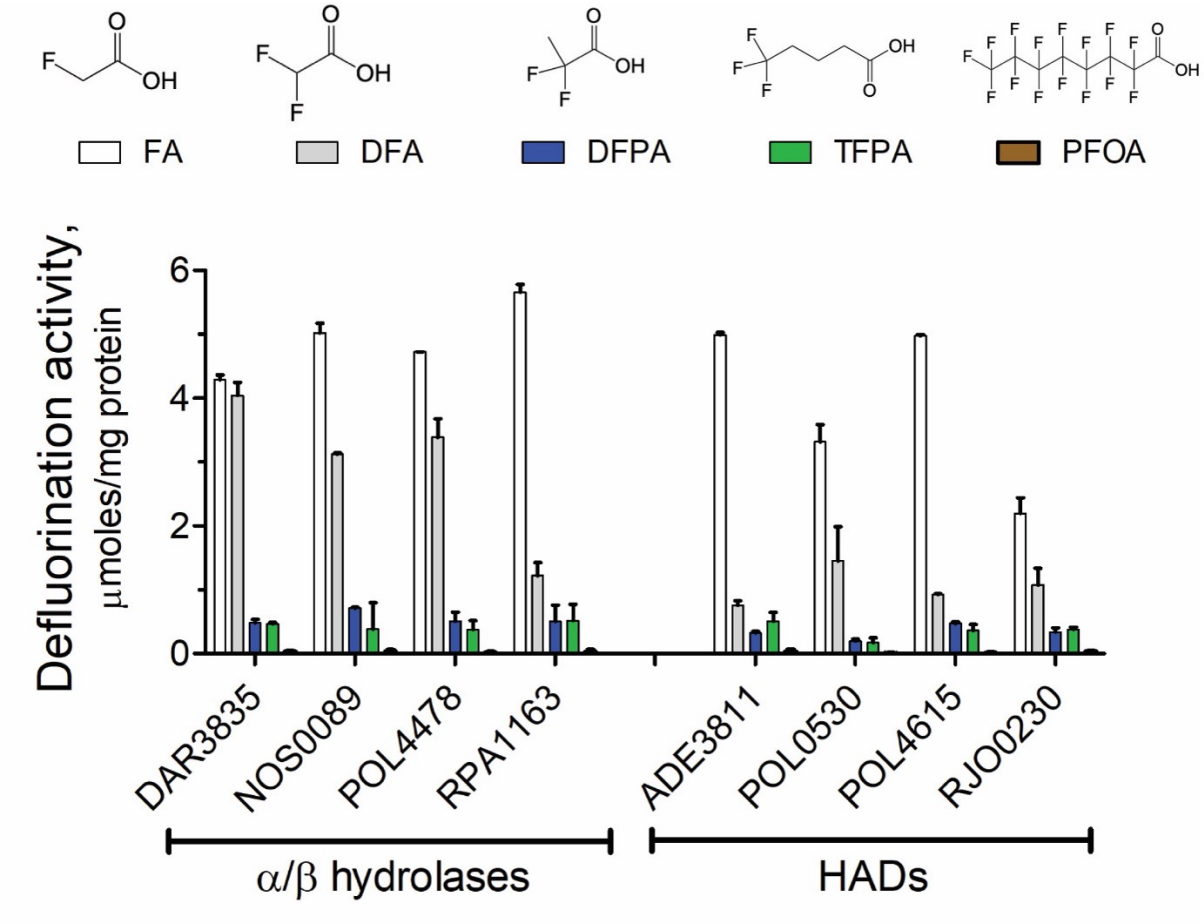


Fig. 2. Screening of purified bacterial dehalogenases for defluorination activity against four substrates: fluoroacetate (FA), difluoroacetate (DFA), 2,2-difluoropropionic acid (DFPA), 5,5,5-trifluoropentanoic acid (TFPA), and perfluorooctanoic acid (PFOA). Defluorination activity was determined using a pH-shift assay with phenol red (2 μg/ml), 5 mM substrate, and purified protein (20 μg) (2 h incubation at 30ºC). All assays were carried out in triplicate, and results are means ± SD from at least two independent determinations.

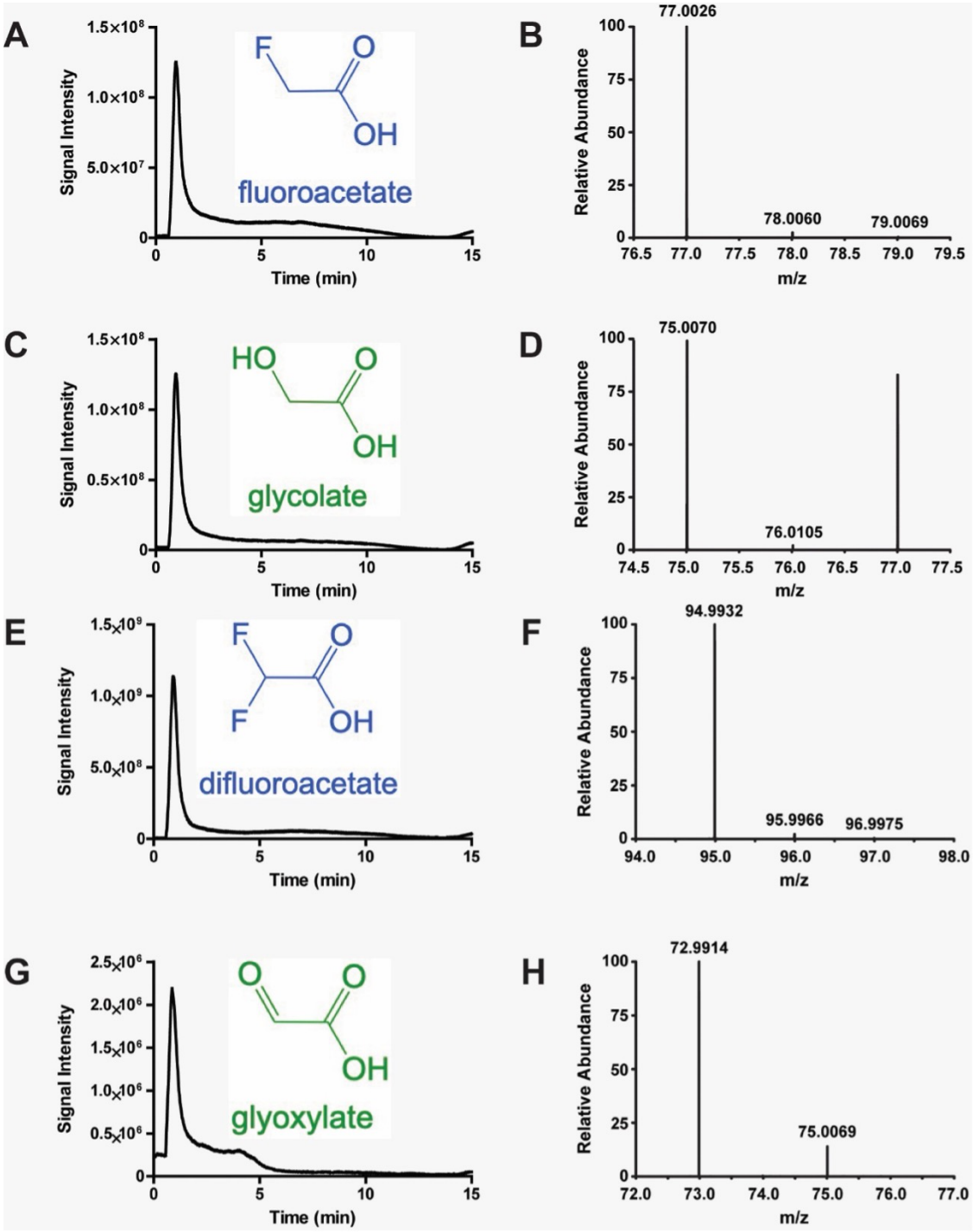


Fig. 3. Defluorination activity of DAR3835: LC-MS analysis of reaction products. (A, C, E, G), Extracted ion chromatograms of FA (A), glycolic acid (C), DFA (E), and glyoxylic acid (G) isolated via C18 column under negative ionization. (B, D, F, H), MS spectra of the C18 column peaks of FA (B), glycolic acid (D), DFA (F), and glyoxylic acid (H). M/Z values for the hydrogen adducts of the products under negative ionization are shown.

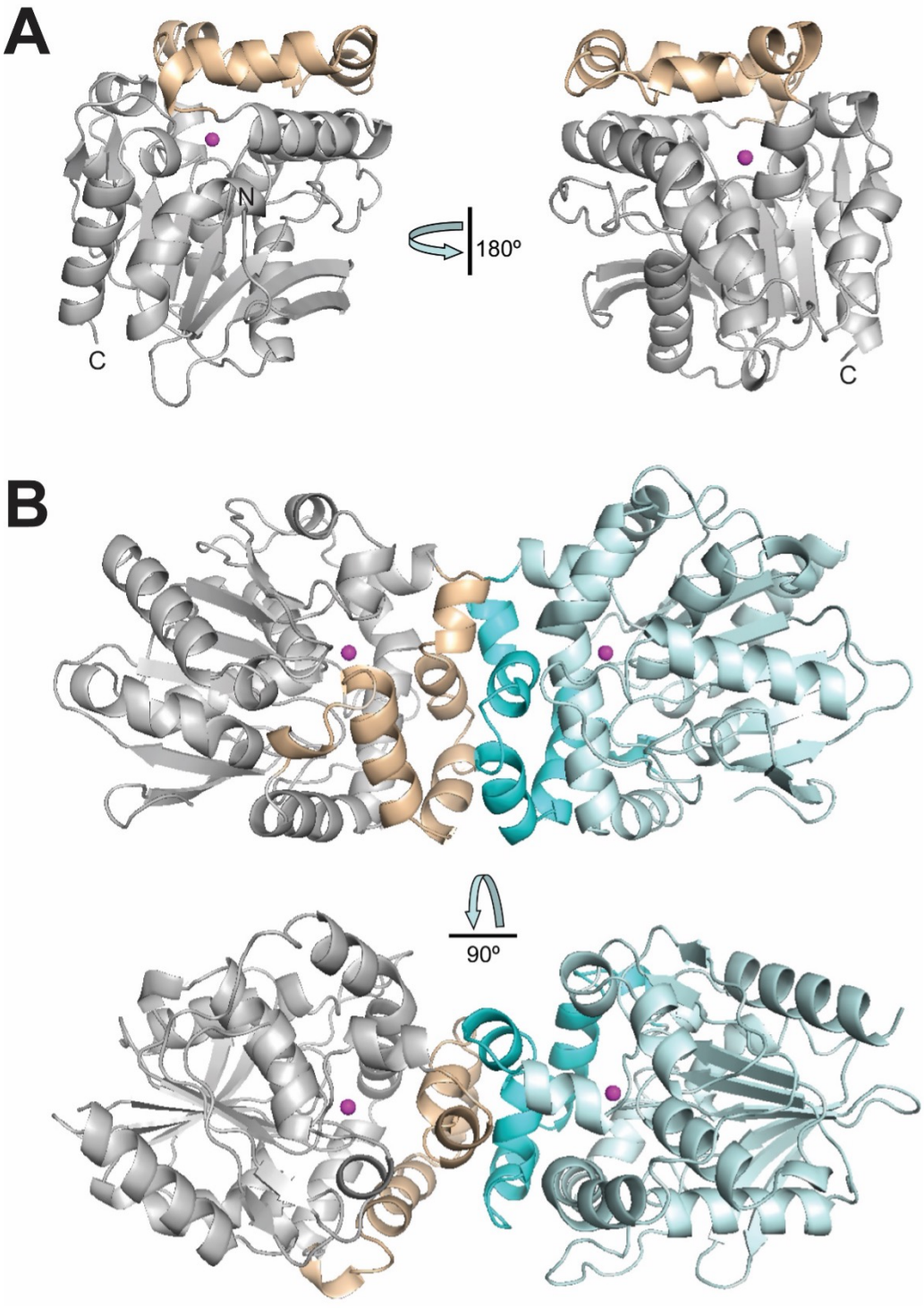


Fig. 4. Overall crystal structure of DAR3835. Two views each of the DAR3835 protomer (A) and dimer (B) related by 180º and 90º rotations, respectively. (A), the DAR3835 protomer: the protein core and cap domains are shown as ribbon diagrams colored grey and light orange, respectively, whereas the bound Cl^- ion is designated by the magenta-colored sphere. (B), the DAR3835 dimer: the subunits are colored using different colors for the protein core and cap domains.

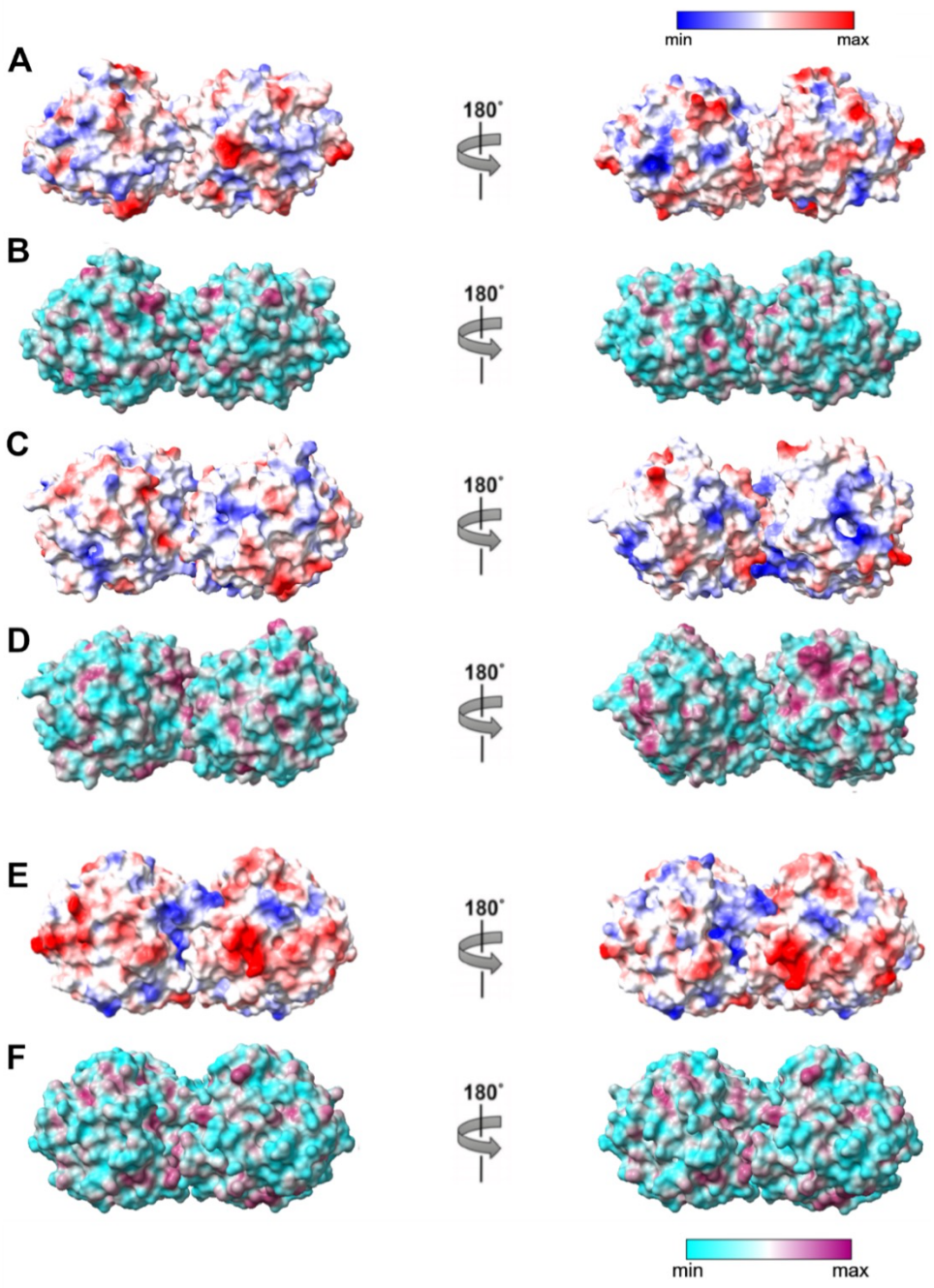


Fig. 5. Structural analysis of microbial defluorinases: surface potential (Coulombic charge) and hydrophobicity of protein dimers. Surface presentations of DAR3835 (A, B), NOS0089 (C, D), and RPA1163 (E, F) showing electrostatic potential (A, C, D) and hydrophobicity (B, D, F) of solvent accessible protein residues. Distribution of surface potential (A, C, E) has been shown as a blue (negatively charged) to red (positively charged) gradient, whereas distribution of surface hydrophobicity (B, D, F) is shown as a cyan (polar) to magenta (hydrophobic) gradient. The scale bars show coloring schemes of surface-exposed residues.

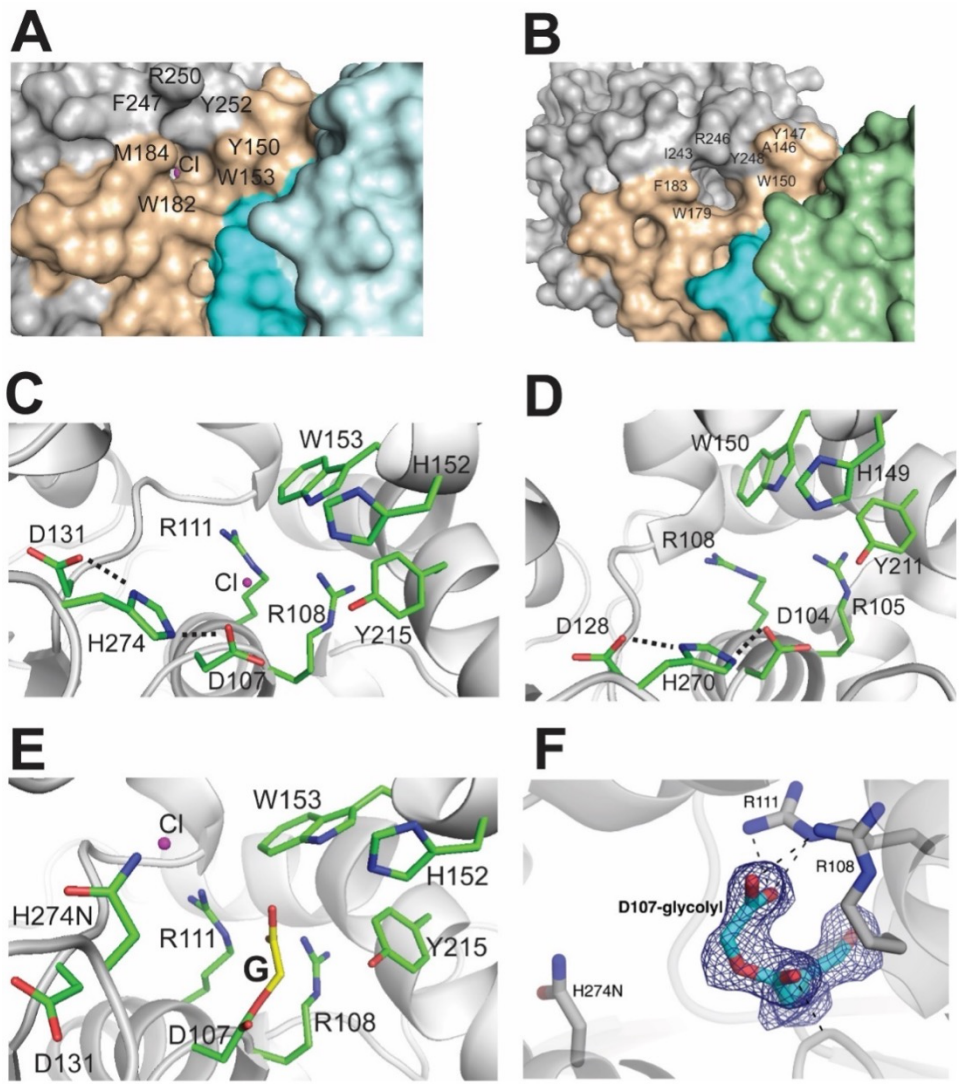


Fig. 6. Crystal structures of DFA defluorinating dehalogenases: active site openings and close-up view of active sites. (A, B), The surface presentation of the DAR3835 (A) and NOS0089 (B) dimers showing the opening of the substrate access tunnel leading to the active site and showing the bound Cl^- ion (surface presentation). The color schemes used were as in Fig. 3B (with the bound Cl^- ion shown as the magenta-colored sphere) and in Fig. S5B. Positions of surface residues near the active site opening are indicated by residue numbers. (C, D), Close-up view of the active sites of DAR3835 (C) and NOS0089 (D) (apo structures). (E, F), Close-up view of the DAR3835 H274N active site showing the covalent D107-glycolyl intermediate (E) and electron density for the D107-glycolyl intermediate (F). The density shown is a simulated annealing omit map (D107-glycolyl omitted) contoured at 3σ. In C, D, and E: the protein core domain is colored grey, the active site residues are shown as sticks with green carbons, the bound Cl^- ion is designated as the magenta-colored sphere, and the glycolyl moiety (attached to Asp107 in the DAR3835 H274N mutant protein) is shown as sticks with yellow carbons (labelled as G). The catalytic triad residues (in C and D) are indicated by dashed lines (Asp107-His274-Asp131 in DAR3835 and Asp104-His270-Asp128 in NOS0089). In F: the side chains of Arg108, Arg111, and His274Asn are shown as sticks with grey carbons, whereas the D107-glycolyl moiety is shown as sticks with cyan-colored carbons, and dashes indicate polar contacts.

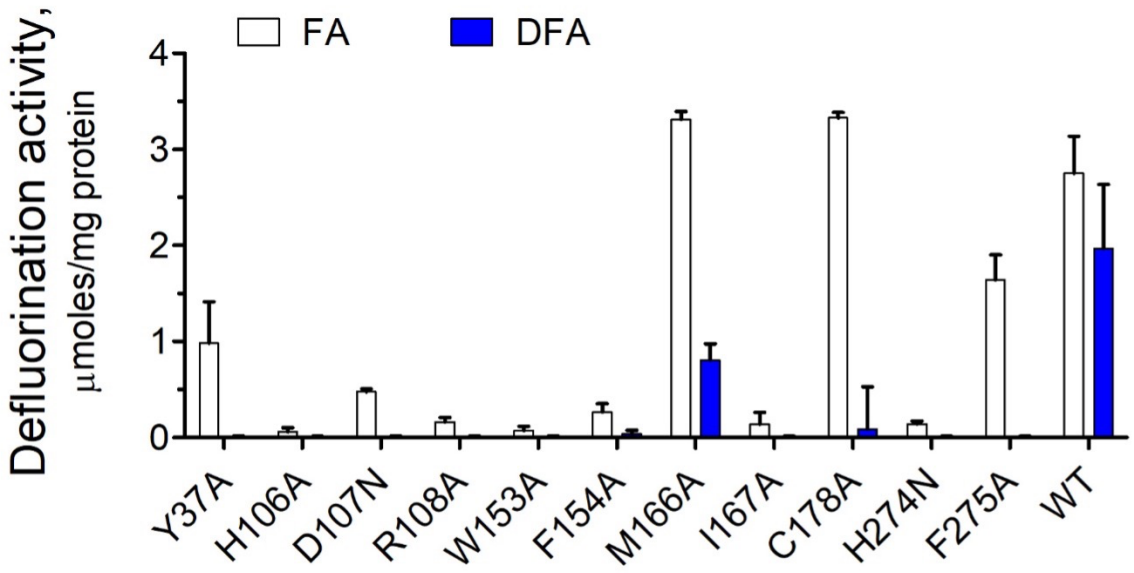


Fig. 7. Site-directed mutagenesis of DAR3835: defluorination activity of purified proteins. Enzymatic activity of purified wild type (WT) and mutant proteins was measured using FA or DFA as substrates. The D131A mutant protein was found to be insoluble, and therefore it was not analyzed in these experiments. The reaction mixtures contained 10 mM substrate, phenol red (2 µg/ml), and purified enzyme (20 µg/assay). All assays were carried out in triplicate, and results are means ± SD from at least two independent determinations.

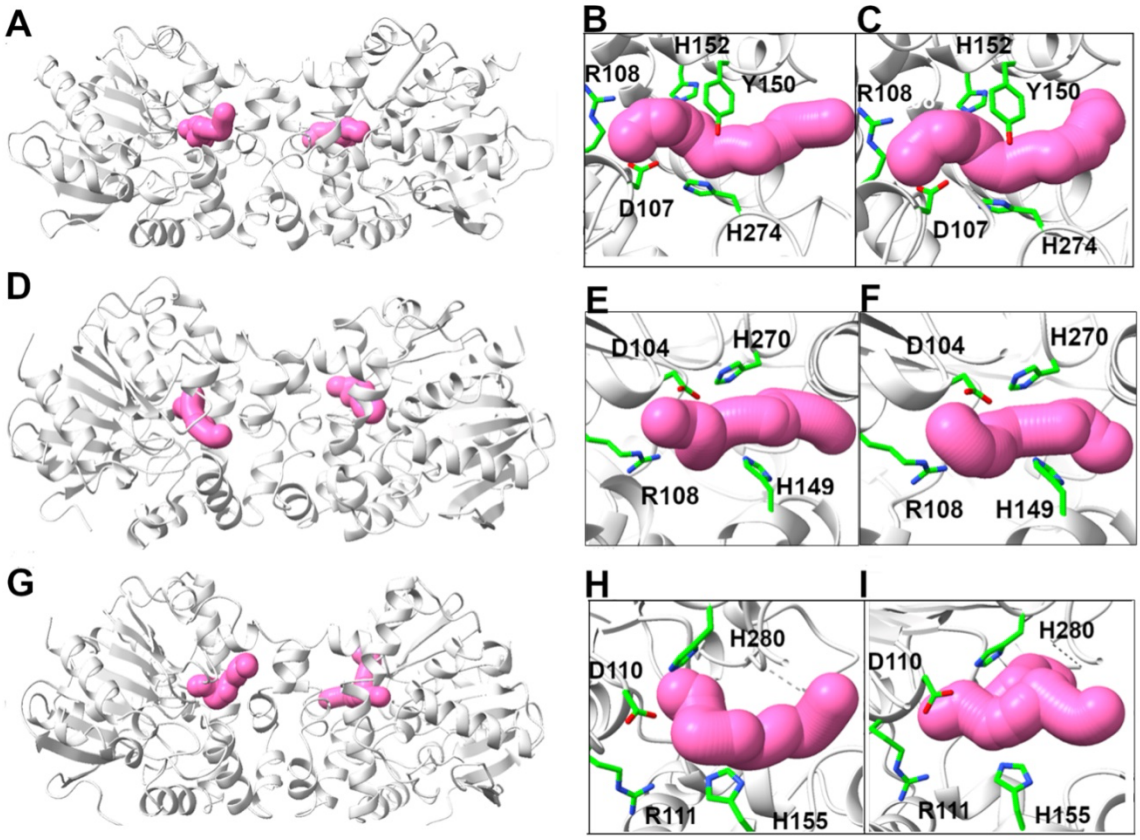


Fig. 8. Structural analysis of substrate access tunnels in microbial defluorinases: DAR3835 (A, B, C), NOS0089 (D, E, F), and RPA1163 (G, H, I). Overall view of the apo-dimers of DAR3835 (A), NOS0089 (D), and RPA1163 (G) showing the substrate access tunnels (presented as pink colored spheres). The right-side panels present close-up views of substrate access tunnels shown along their long axes: DAR3835 (B, C), NOS0089 (E, F), and RPA1163 (H, I). The substrate access tunnels connecting the catalytic Asp (Asp107 in DAR3835, Asp104 in NOS0089, and Asp110 in RPA1163) to the enzyme surface were calculated using the Asp side chains as the starting point and are visualized as a set of intersecting spheres (colored pink) within the protein subunits shown as grey ribbon diagrams with the active site residues presented as sticks with green carbons.

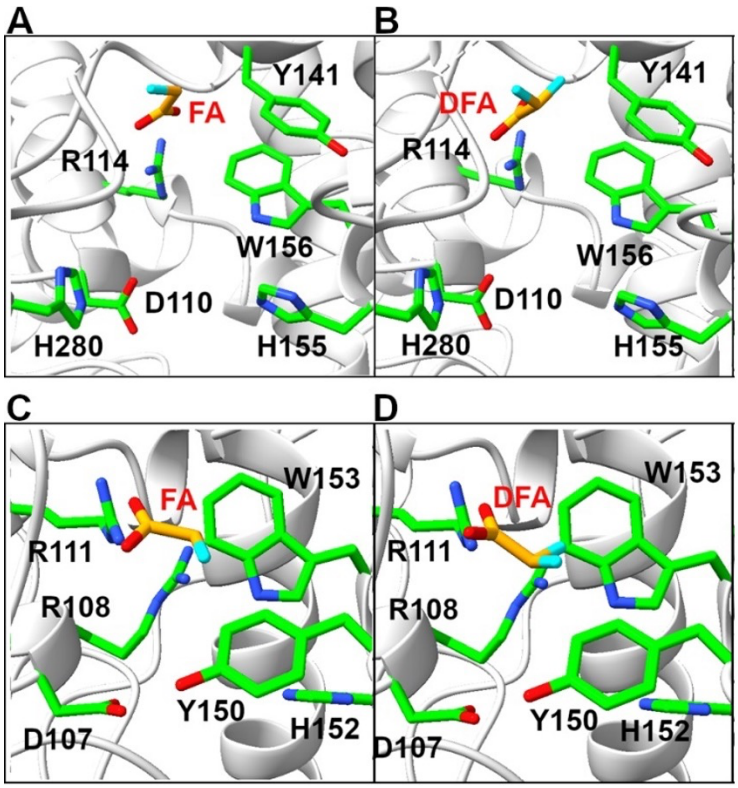


Fig. 9. Structural analysis of DFA defluorination by RPA1163 and DAR3835: ligand docking simulations. Binding of fluoroacetate (FA) or difluoroacetate (DFA) in the active sites of RPA1163 (A, B) and DAR3835 (C, D). The panels present the docking models with minimal binding energies shown in Table S6. Enzyme residues are shown as sticks with carbon atoms colored green, whereas the ligands (FA and DFA) are shown as sticks with carbons colored orange and fluorines colored cyan.

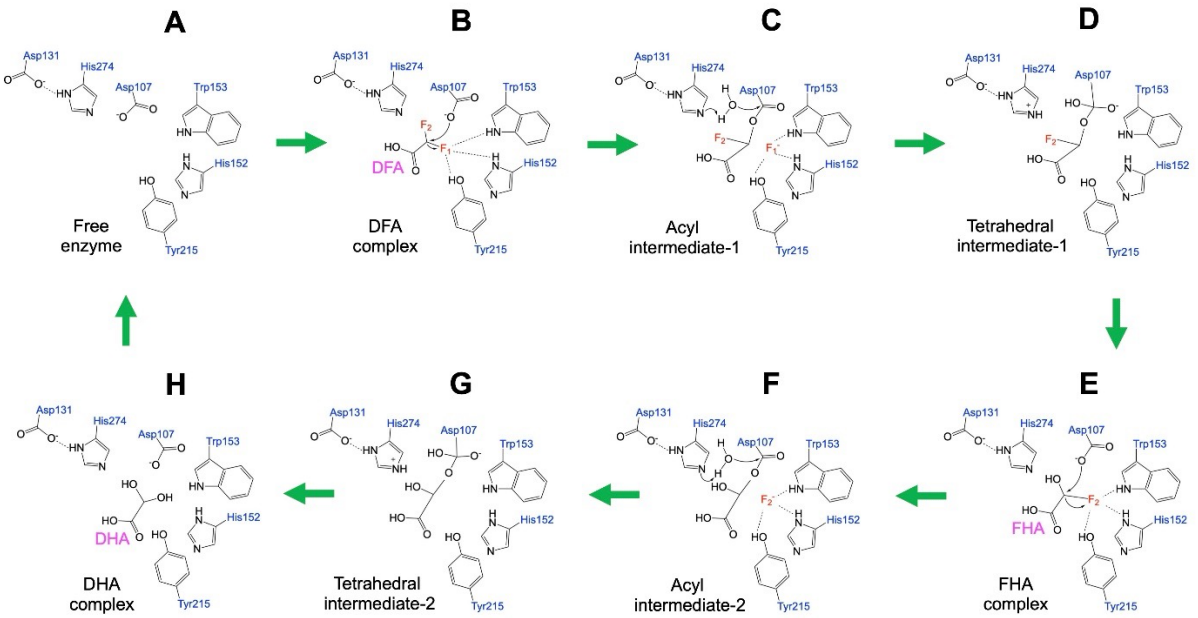
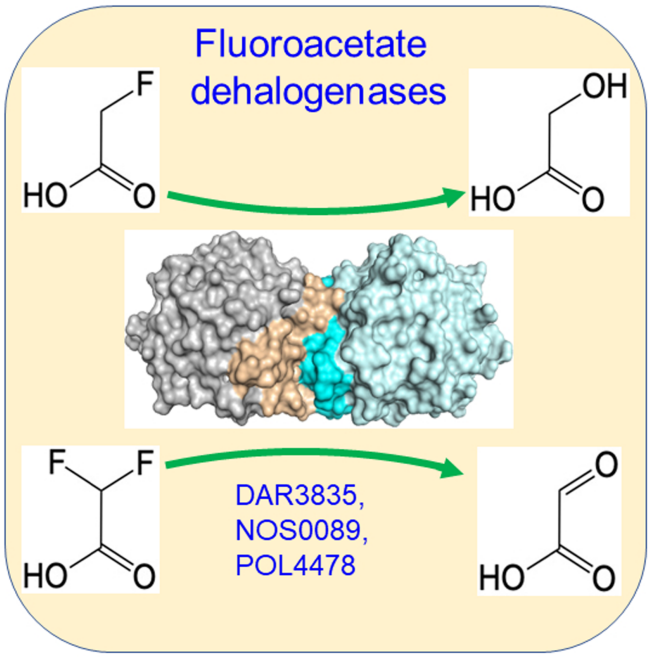


Fig. 10. Possible reaction mechanism of DFA defluorination by DAR3835. The reaction involves the conserved catalytic triad Asp107-His274-Asp131 and proceeds through the following steps: (A), free enzyme; (B), formation of the DAR3835-DFA (Michaelis) complex; (C), acyl intermediate-1 (2-fluoroglycolyl); (D), tetrahedral intermediate-1; (E), 2-fluoro-2-hydroxy-acetate (FHA) complex; (F), acyl intermediate-2 (2-hydroxyglycolyl); (G), tetrahedral intermediate-2; (H), dihydroxyacetate (DHA) complex followed by DHA release and producing free enzyme. In organic solvents, DHA is dehydrated to glyoxylic acid, which was detected using LC-MS.

Graphical abstract



Fluorine forms the strongest single bond to carbon, but microbial dehalogenases can hydrolyze it under mild reaction conditions. Here, we have demonstrated that several fluoroacetate dehalogenases also catalyze complete defluorination of difluoroacetate extending the substrate scope of these enzymes to polyfluorinated carboxylic acids. Crystal structures, product analysis, site-directed mutagenesis, and computational analyses provided insights into the active site and potential catalytic mechanism of difluoroacetate degrading defluorinases.