Article

Investigating Substrate Scope and Enantioselectivity of a Defluorinase by a Stereochemical Probe

Jian-bo Wang, Adriana Ilie, SHUGUANG YUAN, and Manfred T. Reetz

J. Am. Chem. Soc., **Just Accepted Manuscript** • Publication Date (Web): 25 Jul 2017

Investigating Substrate Scope and Enantioselectivity of a Defluorinase by a Stereochemical Probe

Jian-bo Wang,[†,F] Adriana Ilie, [†,F] Shuguang Yuan*,[III] and Manfred T. Reetz*,[†,F]

[†]Department of Chemistry, Philipps-University, Hans-Meerwein-Strasse 4, 35032 Marburg, Germany,

[F]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Muelheim, Germany,

[III]Laboratory of Physical Chemistry of Polymers and Membranes, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH B3 495 (Bâtiment CH) Station 6, CH-1015 Lausanne, Switzerland

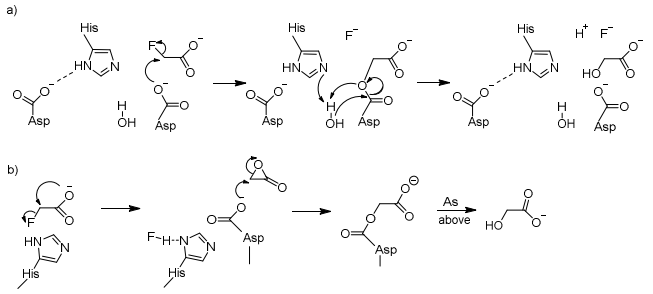
ABSTRACT: The possibility of a double Walden inversion mechanism of the fluoracetate dehalogenase FAcD (RPA1163) has been studied by subjecting *rac*-2-fluoro-2-phenyl acetic acid to the defluorination process. This stereochemical probe led to inversion of configuration in a kinetic resolution with an extremely high selectivity factor (E >500), showing that the classical mechanism involving S_N2 reaction by Asp110 pertains. The high preference for the (*S*)-substrate is of synthetic value. Wide substrate scope of RPA1163 in such hydrolytic kinetic resolutions can be expected because the reaction of the even more sterically demanding *rac*-2-fluoro-2-benzyl acetic acid proceeded similarly. Substrate acceptance and stereoselectivity were explained by extensive molecular modelling (MM) and molecular dynamics (MD) computations. These computations were also applied to fluoroacetic acid itself, leading to further insights.

INTRODUCTION

Fluorinated organic compounds are used in a vast number of industrial applications, ranging from polymers and plant-protecting agents to therapeutic drugs.[1] Not only the development of efficient synthetic techniques,[2] but also the invention of ecology-driven degradation processes[3] constitute prominent tasks. Due to the exceptionally high dissociation energy of C-F bonds, ranging up to 130 kcal/mol, the establishment of mild degradation methods is challenging. In all of these endeavors, synthetic reagents and man-made catalysts dominate current methodology, while biocatalysts are just beginning to play a role.[4] It can be expected that the study of substrate scope coupled with knowledge of the enzyme mechanism will help in paving the way to novel (industrial) applications.

Metabolism of organofluorine compounds such as F-containing therapeutic drugs is another important research area.[5] Enzymes called defluorinases are known to cleave C-F bonds under ambient conditions,[6] one of the first types being fluoroacetate dehalogenases (FAcD).[7] These enzymes catalyze the hydrolysis of the very stable fluoroacetic acid (FAc) with formation of glycolic acid. FAc is one of the most toxic compounds to mammals, yet it is used as a commercial rodenticide in a number of countries, including the USA.

Several early studies showed that the mechanism of a typical FAcD involves a classical S_N2 reaction in which an aspartate (Asp105) functions as the nucleophile with direct expulsion of fluoride,[8] analogous to homologous haloalkane dehalogenases,[9] which however do not accept FAc. Subsequently, in a study on the other defluorinase RPA1163, the details of this accepted mechanism were illuminated by mapping the reaction coordinates of defluorination using X-ray crystallography, thereby providing snapshots which picture the free enzyme, enzyme-FAc complex, enzyme-glycolyl covalent complex, and enzyme-product complex harboring glycolic acid (**Scheme 1a**).[10] A recent study provided further illuminating details, revealing a subtle asymmetry in the crystal structure of the apo-RPA1163 homodimer which plays a role in the observed "half-of-the sites" reactivity.[11] This means that at any instance only one of the two protomers participates in substrate binding.



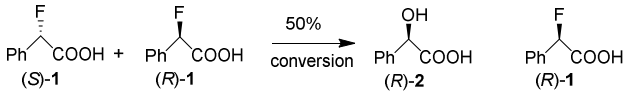
Scheme 1. a) Widely accepted mechanism of the defluorination of fluoroacetic acid (FAc) catalyzed by RPA1163 as the fluoroacetate dehalogenase (FAcD).[8,10] b) Possible enzymatic defluorination of FAc by an intramolecular S_N reaction followed by an S_N2 transformation.

While the intermediacy of the FAc-complexes and the recently proposed domain-specific dynamics are indisputable,[10,11] the possibility of a double Walden inversion mechanism[12,13] involving two successive nucleophilic substitution reactions with the intermediacy of an α-lactone playing a pivotal role cannot be rigorously excluded. Relevant is the observation that the solvolysis of chiral α-bromopropionic acid under basic conditions with formation of lactic acid occurs with retention of configuration.[14] In related chemistry, it has been demonstrated that the deamination of (*R*)-alanine by the action of HNO_2 with formation of (*R*)-lactic acid also proceeds

via an α-lactone followed by an S_N2 reaction with overall retention of configuration.[15a] Indeed, this is a general synthetic method, (*S*)-amino acids leading to the respective (*S*)-hydroxy carboxylic acids.[15b-c] α-Lactones are now accepted as reactive intermediates in these and other reactions,[16] and some have even been isolated.[17]

Following this line of thought, fluoroacetic acid could be activated in the enzyme's binding pocket by deprotonation of the carboxyl which then induces an intramolecular nucleophilic substitution reaction with formation of a reactive α-lactone (**Scheme 1b**). The usual fluoride anion stabilization can be expected in this scenario. In a second step, the intermediate α-lactone would undergo an S_N2 reaction by Asp110, followed by the usual hydrolysis of the covalent enzyme-ester intermediate.

Since FAc is not chiral, such a reaction mode cannot be proven or disproven on the basis of the presently available mechanistic data. In order to test this mechanistic alternative, we designed a stereochemical probe by resorting to the chiral substrate 2-fluoro-2-phenyl acetic acid (**1**) (**Scheme 2**). Since it is sterically more demanding than FAc, the question of substrate acceptance by RPA1163 had to be addressed first. In principle both (*R*)- and (*S*)-enantiomers could be tested in separate experiments. However, should this substrate be accepted by wildtype RPA1163, then only one of the two enantiomers might react in a stereoselective hydrolytic kinetic resolution, which would provide the necessary stereochemical evidence for mechanistic conclusions. High enantioselectivity at 50% conversion in this particular case and in similar transformations of related racemates would also be of preparative significance in synthetic organic chemistry, enabling asymmetric biocatalytic access to chiral fluoro compounds. To date only the sterically small unsubstituted FAc has been used as a substrate in RPA1163 catalysis.[8,10,11]



Scheme 2. Hydrolytic kinetic resolution of *rac*-1 catalyzed by the fluoroacetate dehalogenase RPA1163.

In the present study, we report that the fluoracetate dehalogenase FAcD (RPA1163) is indeed capable of not only accepting compound **1**, but also a sterically more demanding substrate. This sets the stage for hydrolytic kinetic resolution. In these reactions we discovered an unusually high degree of stereoselectivity favoring the transformation of (*S*)-**1**, which reveals the synthetic potential of RPA1163. Also serving as a stereochemical probe, the reaction demonstrates that α-lactones are not involved as short-lived intermediates. We also performed molecular dynamics (MD) computations with the monomer and the dimer, which shed light on the origin of stereoselectivity.

Results and discussion

Substrate acceptance and kinetic resolution of *rac*-1. Upon subjecting *rac*-**1** to RPA1163-catalysis, we discovered that the reaction occurs rapidly, showing for the first time that RPA1163 is capable of accepting bulky substrates, not just the sterically small FAc. The transformation stops completely at 50% conversion, the only new product being (*R*)-**2**, leaving behind unreacted (*R*)-**1** in a ratio of 1:1 **(Scheme 2)**. Using the Sih-formula,[18] the selectivity factor, reflecting the relative rate of reaction of (*S*)- versus (*R*)-**1**, proved to be exceptionally large, E = >500. The stereochemical outcome proves that (*S*)-**1** reacts with inversion of configuration leading to the sole formation of (*R*)-**2**. These results allow for three conclusions:

- The fluoroacetate dehalogenase RPA1163 is capable of accepting a rather bulky substrate.
- The mechanism in which an aspartate expels fluoride directly according to **Scheme 1a** operates, and not the alternative overall Walden retention process by way of two inversion events (**Scheme 1b**).
- Hydrolytic kinetic resolution is possible with an unusually high degree of enantioselectivity, which is of interest in synthetic organic chemistry.

Kinetics. Using *rac*-**1** as substrate and RPA1163 as the fluoroacetate defluorinase, the following kinetic parameters were measured:

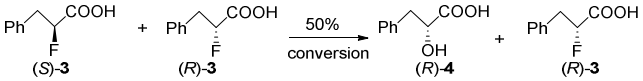
k_{cat} = 61.3±0.5 min^{-1} ; K_m = 1.7±0.2 mM; k_{cat}/K_m = 601 sec^{-1}M^{-1}

These results can be compared to the kinetic parameters of the natural substrate FAc. Due to the difficulty in determining the concentrations of FAc and glycolic acid in the aqueous medium, we were unable to obtain kinetic data as in the case of substrate **1**. Fortunately, in a previous study steady state kinetics were performed by using a microcalorimetric method, which provided the following data[10]:

k_{cat} = 6.7±0.6 min^{-1} ; K_m = 3.3±0.2 mM; k_{cat}/K_M = 33 sec^{-1}M^{-1}

While the results of the two determinations may not be precisely comparable, they indicate that the sterically demanding non-natural substrate (*S*)-**1** is actually more active, a surprising finding. These results suggest that fluoroacetate may not be the natural substrate, but some as yet unknown compound.

Another Chiral Substrate. In order to test whether wildtype RPA1163 also accepts other even more sterically demanding substrates, *rac*-**3** was subjected to hydrolytic kinetic resolution under comparable reaction conditions (Scheme 3). As before, a similar reaction occurred, showing the same mode of stereoselectivity likewise with a very high selectivity factor of E > 500. This finding suggests once more that RPA1163 is a fluroacetate dehalogenase with potentially broad substrate acceptance.



Scheme 3. Hydrolytic kinetic resolution of *rac*-3 catalyzed by the fluoroacetate dehalogenase RPA1163.

Molecular Dynamics Simulations. In order to gain insight into the origin of enantioselectivity, we performed extensive molecular modelling (MM) and molecular dynamics (MD) simulations starting from the RPA1163 crystal structures. The overall RPA1163 structure (pdb: 5SWN)[10] comprises eleven α-helices and eight β-strands, which fold into a compact catalytic pocket (**Figure S1 and S2**). It is a dimer. The catalytic machinery RPA1163 is composed of D110, D134, and H280, all three residues being located in loops: D110 is situated between β5 and helix3, D134 is situated between β6 and helix4, and H280 is situated between β8 and helix11 (**Figure S1B**). Loop B (residues

G249-T259) is unresolved in most solved crystal structures (**Figure S3**), which indicates high flexibility of this region. Moreover, we find that loop A (residues M145-P166) is flexible in the monomer, although it is less so in the dimer (**Figure S1 and S2**). We computationally "repaired" this region and used the repaired model for all computational tasks.

To understand how substrate and product interact with the native RPA1163, we computationally constructed three complex models in the dimer: RPA1163-(*S*)-substrate complex (RPA1163-(*S*)-**1**), RPA1163-(*R*)-substrate complex (RPA1163-(*R*)-**1**) and the RPA1163-(*R*)-product complex (RPA1163-(*R*)-**2**), all of which were obtained using protein-ligand docking. Two dominant binding modes for both RPA1163-(*S*)-**1** and RPA1163-(*R*)-**1** occur during docking procedures (**Figure 1**). In the first binding mode (**Figure 1A**), which is more favorable than the other according to the docking scoring function, the F-atom points to the opposite position of base D109 and is close to H155. This special position leads to (*R*)-**2** after the S_N2 reaction. In the other binding mode, the F-atom faces D109 directly (**Figure 1B**). This second binding mode is identical to the first binding mode of RPA1163-(*S*)-**1** (**Figure 1D**). We then applied 4x20 ns MD simulations to each binding mode for Poisson–Boltzmann and surface area continuum solvation (MM/PBSA).[19] The results of relative energy calculations, widely used for ligand relative binding affinity evaluations, are shown in **Figure 1**.

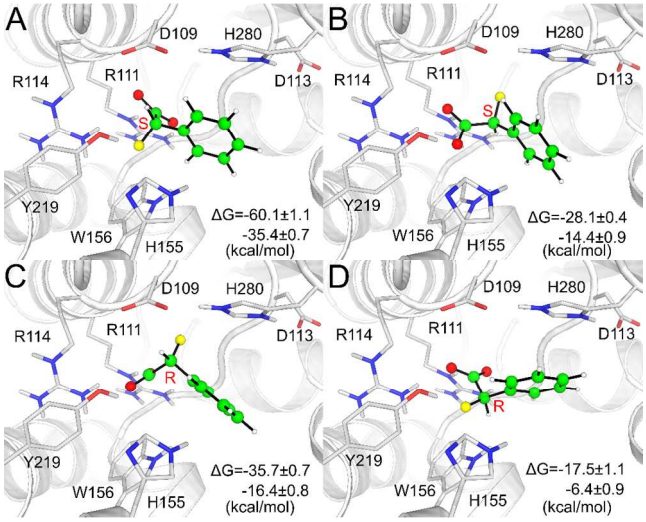


Figure 1. Two dominant binding modes of RPA1163 substrates and their relative ligand binding energies. The substrates are shown in green ball-and-sticks. The F-atoms are colored in yellow. (A) The binding mode 1 of (*S*)-**1**, with ΔG^{S1}=-60.1±1.1 (-35.4±0.7, monomer) kcal/mol. (B) The binding mode 2 of (*S*)-**1**, with ΔG^{S2}= -28.1±0.4 (-14.4±0.9, monomer) kcal/mol. (C) The binding mode 1 of (*R*)-**1**, with ΔG^{R1}=-35.7±0.7 (-16.4±0.8, monomer) kcal/mol. (D) The binding mode 2 of (*R*)-**1**, with ΔG^{S2}= -17.5±1.1 (-6.4±0.9, monomer) kcal/mol.

Accordingly, the binding energy of the (*S*)-**1** binding mode 1 (**Figure 1A**), in a pose leading to (*R*)-**2**, is clearly more favorable (ΔG_{S1}=-60.1±1.1 kcal/mol) than that of either (*S*)-**1** in binding mode 2 (ΔG_{S2}=-28.1±0.4 kcal/mol) or bound (*R*)-**1** (ΔG_{R1}=-35.7±0.7 kcal/mol, ΔG_{R2}=--17.5±1.1 kcal/mol). We also calculated the MM/PBSA for RPA1163 monomer which is in line with that of the dimer system. Interestingly, the calculated ΔG is higher than its corresponding dimer value. This is probably because loop A is much less flexible in dimer than that in monomer, stabilizing the substrate interactions (**Figure S1C, S2B and S4**). These computational findings are consistent with the stereochemical results. In order to further confirm the favorable binding modes of both RPA1163-(*S*)-**1** and RPA1163-(*R*)-**2** obtained from docking, we performed a 200 ns molecular dynamics (MD) simulation for each complex model in both dimer and monomer states. The interactions in dimer and monomer states proved to be identical. These computations revealed H-bond interactions, salt bridges, π-π stacking and hydrophobic contacts (**Figure S4)** which are all in line with the initial model derived only from docking computations. The interactions in the dimer systems are stronger than that in the monomer state, as indicated by interaction fingerprinting (**Figure S4)**.

Based on the experimental and computational findings, we propose the catalytic mechanism of RPA1163 as follows (**Figure 2**): (A) In the APO RPA1163, two loops (loop A and loop B) next to the catalytic pocket are very flexible and the dynamics allows the size of the pocket to vary extensively. H280, which stacks between D110 and D134, is protonated (calculated pKa = 8.6±0.2). (B) The motion of the substrate into RPA1163 leads to the shift of loop A and loop B, thereby producing large space in the binding pocket. This explains why the sterically demanding substrate **1** can readily enter and bind, necessary for catalysis. (C) After (*S*)-**1** adopts an energetically favored pose at the catalytic site, D110 acts as the nucleophile, attacking the F-bearing C-atom with inversion of configuration and expulsion of fluoride which is stabilized by H155. (D) An intermediate covalent enzyme ester complex forms, which becomes protonated by H280. (E) Hydrolytic ester cleavage takes place by the attack of nucleophilic water, which is activated by H155, leading to product (*R*)-**2**. In this process D110 becomes a free protonated residue, while (*R*)-**2** adopts a new pose on the way to departing from the binding pocket. (F) Finally, with the fluctuations of loop A and loop B, the product departs.

Finally, the question whether fluoroacetic acid (FAc) reacts similarly was addressed by introducing 200 ns MD simulations for RPA1163-fluoroacetate (RPA1163-FAc) and its corresponding product complex RPA1163-hydroxyacetate (RPA1163-HAc). The initial binding poses were first constructed by docking, and proved to be essentially identical to a recently solved RPA1163 mutant crystal structure (pdb: 5SWN, **Figure S5**).[11] Since FAc is sterically much smaller than substrate **1**, the interactions between RPA1163 and FAc are lower in number than those computed for the RPA1163-(*S*)-**1** complex (**Figure S6**). Only three interactions were found: A salt bridge with R111 as well as halogen interactions with H155 and W156. With the introducing of an additional hydroxyl group in the product, a few more interactions occur in the RPA1163-HAc complex, including less frequent H-bond interactions with K152, W185 and Y219. The frequency of the salt bridge interaction with R111 is lower than that in the RPA1163-FAc complex.

To illustrate the intrinsic property of RPA1163 and the catalytic mechanism, we performed an additional MD simulation for the native APO RPA1163. The root-mean-square deviation (RMSD) in the case of APO RPA1163 is notably higher than those of both RPA1163-(*S*)-**1** and RPA1163-(*R*)-**2** (**Figure S7 & S8**), indicating its pronounced flexibility. We then performed the principle component analysis (PCA), an efficient and reliable method to reveal the most important motions in proteins, specifically for revealing the RPA1163 MD simulation trajectory.[20] The PCA analysis indicates that loop A and loop B are capable of undergoing dramatic shifting, altering the size of catalytic site to a notable extent (**Figure S1C, S2B and movie 1,2**). The root-mean-square fluctuation (RMSF) and the B-factor values of each residue, calculated

from the three simulated systems (APO RPA1163, RPA1163-(*S*)-**1** and RPA1163-(*R*)-**2**), lend further support to this conclusion

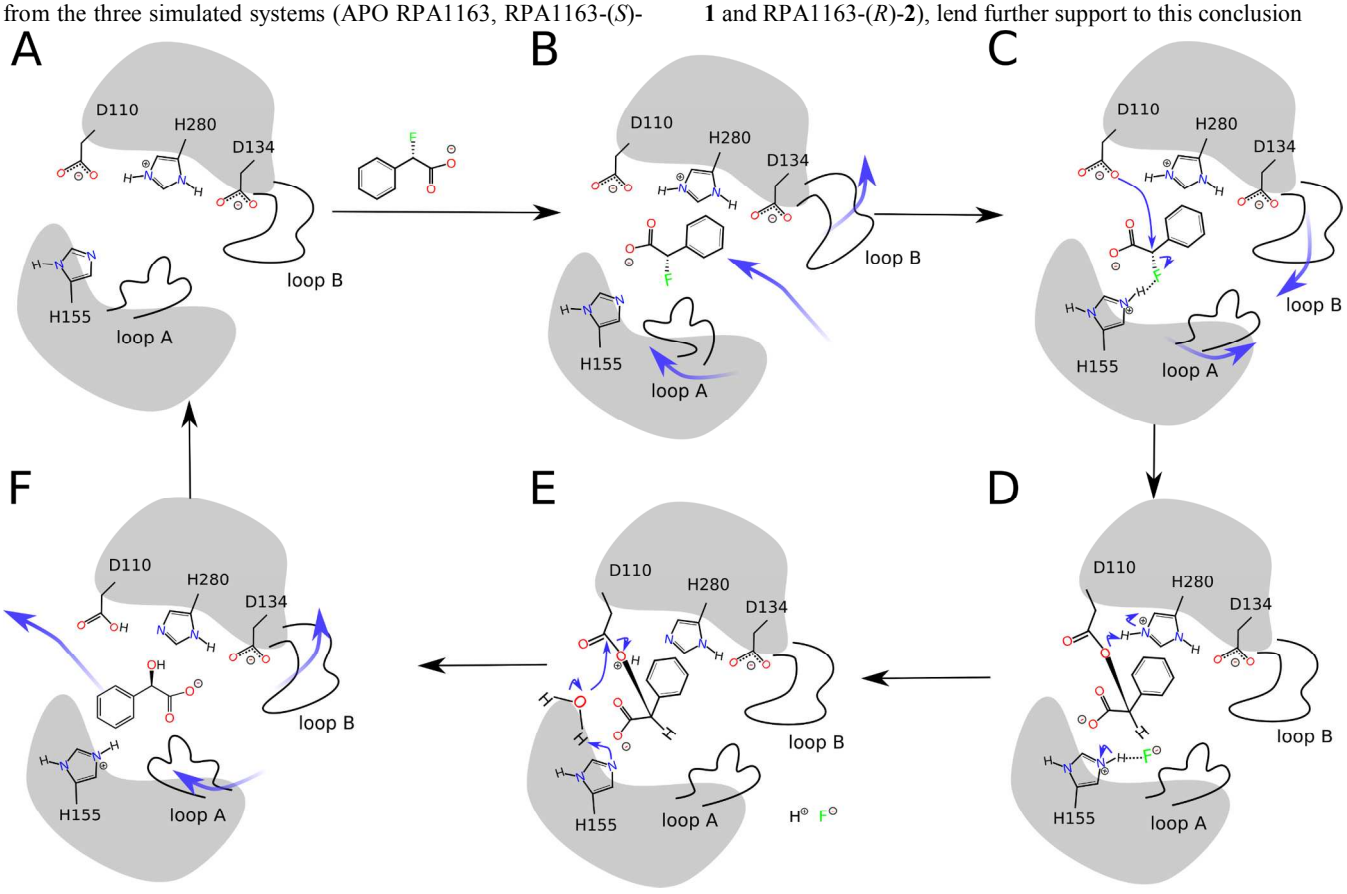


Figure 2. The catalytic mechanism of RPA1163. (A) In the APO RPA1163, two loops (loop A and loop B) next to the catalytic pocket are very flexible and the size of the pocket can be fluctuated to a large extent. (B) The entry of substrate into RPA1163 leads to the inward shift of loop A and outward shift of loop B, resulting in a large space in the binding pocket. This facilitates the substrate molecule entering the catalytic region. (C) After substrate situating at the catalytic site, D110 acts as a nucleophile, attacking the methenyl group connected with F atom. This leads to the leaving of F atom from the substrate. (D) A covalent intermediate complex forms (E) The covalent intermediate is hydrolyzed by a water molecule. D110 becomes a free deprotonated residue. The hydroxyl group from water is transferred to the product. (F) Finally, with the fluctuations of loop A and loop B, the product leaves catalytic pocket.

(**Figure S7B & S8B**). While loop C fluctuates noticeably more in APO RPA1163, loop A undergoes only mild shifting for all simulated systems.

Conclusion and Perspectives. In conclusion, we have designed and applied a stereochemical probe for gaining insight into the substrate scope and stereoselectivity of the fluoroacetate dehalogenase RPA1163. Using *rac*-2-fluoro-2-phenyl acetic acid in a kinetic resolution, inversion of configuration of the favored (*S*)-substrate was observed in a rapid reaction. This demonstrates for the first time that RPA1163 is capable of accepting bulky substrates other than the sterically small fluoroacetic acid, and that a synthetically useful level of stereoselectivity is possible ($E > 500$ in favor of the transformation of the (*S*)-enantiomer). Thus, this process constitutes a simple way to access enantiomerically pure (*R*)-2-fluoro-2-phenyl acetic acid. Surprisingly, the kinetics show that this non-natural compound is more reactive than the natural sterically small substrate FAc. A second even more bulky substrate, *rac*-2-fluoro-2-benzyl acetic acid, behaves similarly. The stereochemical results lend support to the previously postulated mechanism,[10,11] thereby excluding the possibility of an α-lactone as a reactive intermediate. Extensive MM and MD computations provide insight into the origin of stereoselectivity. Finally, our study also points to the potential application of RPA1163 as a catalyst in stereoselective transformations which allow access to enantiomerically pure α-fluoro carboxylic acids from racemic precursors.

MATERIALS AND METHODS

Bacterial strains, enzymes, DNA and chemicals. The expression strain was *E. coli* BL21 (DE3). All chemicals were purchased from TCI EUROPE N.V. Endonuclearase *Nde*I and *Sal*I were obtained from New England Biolabs. The gene of RPA1163 was synthesized by the company Lifetechnology and the DNA code was optimized for *E. coli* expression. It was used as template to amplify the target gene. The PCR products were digested by endonuclease (*Nde*I and *Sal*I) then were inserted into the *Nde*I/*Sal*I sites of pET28a to get the expressed plasmid (named DF/pET28a).

Protein expression, purification and reaction conditions. The expressed plasmid DF/pET28a was transformed into BL21 (DE3), single colonies were picked into the Lumbria Broth which contains 50 μg*mL-1 Kanamycin, then were incubated in 37℃ overnight. About 1% seeding cultures were transferred into the TB culture containing 50 μg*mL-1 Kanamycin, the expressed system was incubated in 37℃ for about 2-3 hours. After the OD600 nm had reached 0.6-0.8, 1 mM IPTG was added to induce

the protein expression, and the temperature was lowered to 25℃. After about 20 hours, the cells were gathered by centrifuging, and were then washed two times with Tris-H2SO4 buffer (pH=7.5). To purify the protein, the cells were lyzed by sonicating, then were purified by Ni-NTA. To use the whole cells to catalyze the reaction, the cells were dispersed in Tris-H_2SO_4 buffer, the dry weight of cells amounted to10-12 g*L-1, which were then tested with 5 mM, 10 mM, 15 mM or 20 mM substrate **1** and 10 mM, 15 mM or 20 mM substrate **3**. The total volume of the reaction system was 500μL, which was incubated at 30℃,1000 rpm for at least 2 hours.

Reaction protocol and kinetics measurement. Concentrated HCl was used to quench the reaction (1/10), then 700 μL ethyl acetate was added to extract the substrates and product. 500 μL organic layer was obtained after centrifuging. Then 300 μL methanol were added to the system. To derivate the substrates and the products, 30μL of 2M TMSCHN2 (trimethylsilyl diazomethane in diethyl ether) was added to the above mixture; 30℃, 1000 rpm for 30 min, then GC (Gas Chromatography) was used directly to analyze the samples. GC conditions: 50℃-220℃, 5℃/min, 220℃ for 5 minutes. Column: Hydrodex β-TBDAC, 25m*0.25mm ID. When the concentration of the substrates **1** was 5 mM, 10 mM, 15 mM and 20 mM, the conversion ratios were about 50%, 49%, 48%, 45%, the ee-values for unreacted substrates were >99%, 98%, 93%, 91%; for substrate **2**, the conversion ratios were about 50% when the concentration were 10 mM and 15 mM. However when the concentration of substrate **2** was 20 mM, the conversion ratio decreased to 35%. In all cases, only one product was produced. To measure the kinetics of substrate **1**, 1 μM purified RPA1163 was used, different concentrations of substrate were tested in 10 minutes, then concentrations of the products were calculated by the peaks' areas in the GC.

Confirming the enantioselectivity of the reaction. The reaction system was amplified to 1 mL, and 50 reactions were performed in parallel. The system included 15 mM of substrate **1**, 10 g*L-1 cells, incubated at 30℃, 1000 rpm for 2 h. Then 10% concentrated HCl was added for quenching. The reacted reaction mixtures were combined, and the total mixture extracted by adding 50 mL of ethyl acetate. After esterifying derivation by $TMSCHN_2$ leading to the respective methylesters, the sample was subjected to column chromatography (Et_2O: n-hexane: formic acid = 1 : 0.1: 0.01 v/v), leading to the isolation of (*R*)-mandelic acid, as shown by optical rotation measurement. The ee-value for the unreacted substrates is under these conditions 33%, [α]D22 = -38.8[20]. Both substrate and product were analyzed by NMR spectroscopy to confirm their structures (Figure S6 and S7). For substrate 1: ^{1}H NMR (300 MHz, CDCl3) δ 10.90 (s, ^{1}H), 7.55 – 7.37 (m, 5H), 5.83 (d, J = 47.4 Hz, 1H). ^{13}C NMR (75 MHz, CDCl3) δ 174.54 (d, J = 27.9 Hz), 133.57 (d, J = 20.5 Hz), 130.07 (d, J = 2.1 Hz), 129.04 (s), 126.81 (d, J = 6.1 Hz), 88.92 (d, J = 186.9 Hz). ^{19}F NMR (282 MHz, CDCl3) δ -181.50 (d, J = 15.0 Hz). NMR data is consistent with reported data.[21] NMR data of product (*R*)-**2** : ^{1}H NMR (300 MHz, acetone-d6) δ 7.72-7.48 (m, 2H), 7.42 – 7.26 (m, 3H), 5.21 (s, 1H). ^{13}C NMR (75 MHz, acetone-d6) δ 174.44 (s), 140.79 (s), 129.11 (s), 128.76 (s), 127.57 (s), 73.50 (s). NMR data is consistent with reported data.[22] [α]D22 = -123.24 (10 mg/mL in acetone), ee=99%.[23]

Loop filling and refinements. Since the loop B was missing in the complex structure, loop refinement protocol in Modeller[24] V9.10 was used to fulfill and refine this area. A total of 20,000 loops were generated for RPA1163 and a conformation with the lowest DOPE (Discrete Optimized Protein Energy) score was chosen for protein construction.

Protein structure preparations. All protein models were prepared in Schrödinger suite software under OPLS_2005 force field[25]. Hydrogen atoms were added to repaired crystal structures according to the physiological pH (7.5) with the PROPKA[26] tool in Protein Preparation tool in Maestro[27] to optimize the hydrogen bond network. Constrained energy minimizations were conducted on the full-atomic models, with heavy atom coverage to 0.4 Å.

Ligand structure preparations. All ligand structures were produced in Schrödinger Maestro software[27]. The LigPrep module in Schrodinger software was introduced for geometric optimization by using OPLS_2005 force field. The ionization states of ligands were calculated with Epik[28] tool employing Hammett and Taft methods in conjunction with ionization and tautomerization tools[28].

Protein-ligand docking. The docking of a ligand to the receptor was performed using Glide[29]. Cubic boxes centered on the ligand mass center with a radius 8 Å for all ligands defined the docking binding regions. Flexible ligand docking was executed for all structures. Twenty poses per ligand out of 20,000 were included in the post-docking energy minimization. The best scored pose for the ligand was chosen as the initial structure for MD simulations.

Molecular dynamics simulations. All unbiased MD simulations were performed in Gromacs 5.1.4[30]. All amino acid residues of the protein were modelled according to their protonation state at neutral pH. The protein was centered in a water box with a distance of 12 Å away from the protein. The total number of atoms was approximately 43, 000: 47 Na^+ and 41 Cl^- ions, and about 13,000 water molecules. Amber99SB*-ILDNP[31] force field was assigned to the protein, water and ions, while the ligands were treated by Amber GAFF2 force field[32] through ACPYPE[33] tool. The ligands were submitted to GAUSSIAN 09 program[34] for structure optimization at Hartree-Fock 6-31G* level prior to the generation of force field parameters. All bond lengths of hydrogen atoms in the system were constrained using M-SHAKE[35] Van der Waals and short-range electrostatic interactions were cut off at 10 Å. The whole system was heated linearly at constant volume (NVT ensemble, 1bar) from 0 to 310K over 400 ps. Ten nanoseconds equilibration was performed at constant pressure and temperature (NPT ensemble; 310 K, 1 bar) using the Nose-Hoover coupling scheme with two temperature groups. Long-range electrostatic interactions were computed by particle mesh Ewald (PME) summation. Finally, a 200 ns MD simulations with a time step of 2.0 fs were performed for APO RPA1163, RPA1163-SUB, RPA1163-PRO, RPA1163-FAc and RPA1163-HAc. The MD simulations results were analyzed in Gromacs[30] and VMD[36]. The PCA analysis have been done in VMD. Figures were prepared in PyMOL and Inkscape[37,38].

MM/PBSA relative binding energy calculation. 4x20 ns restrained MD simulations were performed in Gromacs 5.1.4[30] for each specific binding mode. Small force constant (1 kcal/mol/Å2) was added to each ligand to keep a specific binding pose. All generated trajectories (a total of 400 frames for each case) were used for the relative binding energy evaluation. The MM/PBSA calculations were finished using MMPBSA.py script in AmberTools 16[39].

Interaction fingerprint calculations. The interaction fingerprint between protein and ligand was done with PLIP tool[40]. We first extracted snapshots of the final 20ns MD simulation. We then used PLIP to convert protein−ligand coordinates into a bit-string fingerprint (TIFP) registering the corresponding molecular interaction pattern. TIFP fingerprints have been calculated for 500 protein−ligand complexes for each case, enabling a broad comparison of relationships between interaction pattern similarities and ligand or binding site pairwise similarities[41,19b]. In this work, we kept the default parameters of PLIP and focused on three types of interactions: H-bond interaction, π-stacking and hydrophobic contact.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website.

Calculated models, movies and NMR data of related compounds.

AUTHOR INFORMATION

Corresponding Author

reetz@mpi-muelheim.mpg.de; shuguang.yuan@gmail.com

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

Support from the Max-Planck-Society is gratefully acknowledged.

REFERENCES

(1) (a) Banks, R. E.; Smart, B. E.; Tatlow, J. *Organofluorine Chemistry: Principles and Commercial Applications*, Springer Science & Business Media, **2013**. (b) Ojima, I. *Fluorine in Medicinal Chemistry and Chemical Biology*, John Wiley & Sons, **2009**. (c) Reddy, V. P. *Organofluorine Compounds in Biology and Medicine*, Elsevier, Amsterdam, **2015**. (d) Zhou, Y.; Wang, J.; Gu, Z.; Wang, S.; Zhu, W.; Aceña, J. L.; Soloshonok, V. A.; Izawa, K.; Liu, H. *Chem. Rev.* **2016**, *116*, 422. (e) Ni, C.; Hu, J. *Chem. Soc. Rev.* **2016**, *45*, 5441. (f) Fujiwara, T.; O'Hagan, D. *J. Fluorine Chem.* **2014**, *167*, 16. (g) Gouverneur. V. *Fluorine in Pharmaceutical and Medicinal Chemistry*, Imperial College Press, Singapore, **2012**.

(2) Reviews of synthetic methods for the production of organofluorine compounds: (a) O'Hagan, D. *Chem. Soc. Rev*. **2008**, *37*, 308**.** (b) Furuya, T.; Kamlet, A. S.; Ritter, T. *Nature* **2011**, *473*, 470. (c) Preshlock, S.; Tredwell, M.; Gouverneur, V. *Chem. Rev*. **2016**, *116*, 719. (d) Ma, J.-A.; Li, S. *Org. Chem. Frontiers* **2014**, *1*, 712. (e) Wu, J. *Tetrahedron Lett*. **2014**, *55*, 4289. (f) Lin, X.; Weng, Z. *Dalton Trans.* **2015**, *44*, 2021. (g) Shibata, N. *Bull. Chem. Soc. Jp.* **2016**, *89*, 1307. (h) Zard, S. Z. *Org. Biomol. Chem.* **2016**, *14*, 6891. (i) Schimler, S. D.; Cismesia, M. A.; Hanley, P. S.; Froese, R. D.; Jansma, M. J.; Bland, D. C.; Sanford, M. S. *J. Am. Chem. Soc*. **2017**, *139*, 1452.

(3) (a) Murphy, C. D. Appl. *Microbiol. Biotechnol.* **2016**, *100*, 2617. (b) Rotander, A.; Kärrman, A.; van Bavel, B.; Polder, A.; Rigét, F.; Auðunsson, G. A.; Víkingsson, G.; Gabrielsen, G. W.; Bloch, D.; Dam, M. *Chemosphere* **2012**, *86*, 278.

(4) Reviews of fluorinases: (a) O'Hagan, D.; Deng, H. *Chem. Rev*. **2014**, *115*, 634. (b) Thuronyi, B. W.; Chang, M. C. *Acc. Chem. Res*. **2015**, *48*, 584. (c) Senn, H. M. *Front. Chem*. **2014**, 2. (d) Eustáquio, A. S.; O'Hagan, D.; Moore, B. S. *J. Nat. Prod*. **2010**, *73*, 378.

(5) Murphy, C. D.; Sandford, G. *Expert Opin. Drug Metab. Toxicol.* **2015**, *11*, 589.

(6) (a) Au, K. G.; Walsh, C. T. *Bioorg. Chem.* **1984**, *12*, 197. (b) Kurihara, T.; Yamauchi, T.; Ichiyama, S.; Takahata, H.; Esaki, N. *J. Mol. Catal. B: Enzym.* **2003**, *23*, 347. (c) Natarajan, R.; Azerad, R.; Badet, B.; Copin, E. *J. Fluorine Chem*. **2005**, *126*, 424. (d). Murphy, C. D. *Biotechnol. Lett*. **2010**, *32*, 351.

(7) Goldman, P. *Science* **1969**, *164*, 1123.

(8) (a) Liu, J.-Q.; Kurihara, T.; Ichiyama, S.; Miyagi, M.; Tsunasawa, S.; Kawasaki, H.; Soda, K.; Esaki, N. *J. Biol. Chem.* **1998**, *273*, 30897. (b) Kurihara, T.; Esaki, N. *The Chemical Record* **2008**, *8*, 67.

(9) Reviews of dehalogenases: (a) van Hylckama, J. E. T.; Tang, L.; Spelberg, J. H. L.; Smilda, T.; Poelarends, G. J.; Bosma, T.; van Merode, A. E. J.; Fraaije, M. W.; Janssen, D. B. *J. Bacteriol*. **2001**, *183*, 5058. (b) de Jong, R. M.; Dijkstra, B. W. *Curr. Opin. Struct. Biol*. **2003**, *13*, 722. (c) Koudelakova, T.; Bidmanova, S.; Dvorak, P.; Pavelka, A.; Chaloupkova, R.; Prokop, Z.; Damborsky, J. *Biotech. J*. **2013**, *8*, 32. (d) Nagata, Y.; Ohtsubo, Y.; Tsuda, M. *Appl. Microbiol. Biotechnol*. **2015**, *99*, 9865.

(10) Chan, P. W.; Yakunin, A. F.; Edwards, E. A.; Pai, E. F. *J. Am. Chem. Soc*. **2011**, *133*, 7461.

(11) Kim, T. H.; Mehrabi, P.; Ren, Z.; Sljoka, A.; Ing, C.; Bezginov, A.; Ye, L.; Pomès, R.; Prosser, R. S.; Pai, E. F. *Science* **2017**, *355*, eaag2355.

(12) (a) Walden, P. *Ber. Dtsch. Chem. Ges*. **1893**, *26*, 210. (b) Walden, P. *Ber. Dtsch. Chem. Ges.* **1886**, *29*,133. (c) Walden, *P. Ber. Dtsch. Chem. Ges*. **1899**, *32*, 1855.

(13) Hughes, E. *Trans. Faraday Soc.* **1938**, *34*, 202.

(14) (a) Rørdam, H. N. K. *J. Chem. Soc.* (Resumed) **1928**, 2447. (b) Winstein, S.; Grunwald, E.; Ingraham, L. *J. Am. Chem. Soc.* **1948**, *70*, 821.

(15) (a) Aitken, R. A.; Meehan, A.; Power, L. A. *Synthesis* **2015**, *47*, 1557. (b) Moumne, R.; Lavielle, S.; Karoyan, P. *J. Org. Chem*. **2006**, **71**, 3332. (c) Pitt, N.; Gani, D. *Tetrahedron Lett*. **1999**, *40*, 3811. (d) Jia, C.; Yang, K.-W.; Liu, C.-C.; Feng, L.; Xiao, J.-M.; Zhou, L.-S.; Zhang, Y.-L. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 482.

(16) (a) Rodriquez, C.; Williams, I. *J. Chem. Soc., Perkin Trans. 2* **1997**, 953. (b) Buchanan, J. G.; Charlton, M. H.; Mahon, M. F.; Robinson, J. J.; Ruggiero, G. D.; Williams, I. H. *J. Phys. Org. Chem*. **2002**, *15*, 642. (c) Buchanan, J. G.; Diggle, R. A.; Ruggiero, G. D.; Williams, I. H. *Chem. Commun.* **2006**, 1106. (d) Buchanan, J. G.; Ruggiero, G. D.; Williams, I. H. *Org. Biomol. Chem*. **2008**, *6*, 66.

(17) (a) Wheland, R.; Bartlett, P. D. *J. Am. Chem. Soc.* **1970**, *92*, 6057. (b) Adam, W.; Liu, J.-C.; Rodriguez, O. *J. Org. Chem*. **1973**, *38*, 2269. (c) Coe, P. L.; Sellars, A.; Tatlow, J. C.; Whittaker, G.; Fielding, H. C. *J. Chem. Soc., Chem. Commun*. **1982**, 362.

(18) Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. *J. Am. Chem. Soc*. **1982**, *104*, 7294.

(19) (a) Genheden, S.; Ryde, U. *Expert Opin. Drug Discov.* **2015**, *10*, 449. (b) Chan, H. S.; Filipek, S.; Yuan, S. *Sci. Rep.* **2016**, *6*, 34736.

(20) Effenberger, F.; Oßwald, S. *Tetrahedron: Asymmetry* **2001**, *12,* 279.

(21) Mizuta, S.; Stenhagen, I. S.; O'Duill, M.; Wolstenhulme, J.; Kirjavainen, A. K.; Forsback, S. J.; Tredwell, M.; Sandford, G.; Moore, P. R.; Huiban, M. *Org. lett.* **2013,** *15*, 2648.

(22) Zhu, S. F.; Chen, C.; Cai, Y.; Zhou, Q. L. *Angew. Chem., Inter. Ed.* **2008,** *47*, 932.

(23) Szeleczky, Z.; Bagi, P.; Pálovics, E.; Fogassy, E. *Tetrahedron: Asymmetry* **2014,** *25,* 1095.

(24) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. *Current protocols in protein science* **2007**, *Chapter 2*, Unit 29.

(25) Shivakumar, D.; Williams, J.; Wu, Y. J.; Damm, W.; Shelley, J.; Sherman, W. J. *Chem. Theory Comput.* **2010,** *6*, 1509.

(26) Sondergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. J. Chem. *Theory Comput.* **2011,** *7,* 2284.

(27) Schrödinger, L.; Schrödinger, LLC: **2015**.

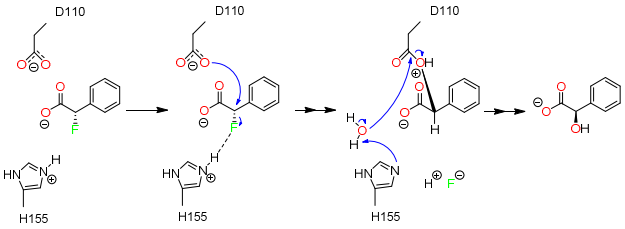
(28) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. *J. Comput. Aided Mol. Des.* **2010,** *24,* 591.

(29) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004,** *47,* 1739.

(30) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013,** *29,* 845.

(31) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. *Proteins* **2014,** *82,* 195.

(32) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004,** *25,* 1157.
(33) Sousa da Silva, A. W.; Vranken, W. F. *BMC Res. Notes* **2012,** *5,* 367.
(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V. M., B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.1; Gaussian, Inc.: Wallingford CT,* **2009.**
(35) Barth, E.; Kuczera, K.; Leimkuhler, B.; Skeel, R. D. *J. Comput. Chem.* **1995,** *16,* 1192.
(36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph Model* **1996,** *14,* 33.
(37) Yuan, S.; Chan, H. C.; Filipek, S.; Vogel, H. *Structure* **2016,** *24,* 2041.
(38) Yuan, S.; Chan, H. C. S.; Hu, Z. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017,** e1298.
(39) D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,; A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C.; Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I.; Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails,; R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao, P.A. Kollman; University of California: San Francisco, **2016**.
(40) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. *Nucleic Acids Res.* **2015,** *43,* W443.
(41) Marcou, G.; Rognan, D. *J. Chem. Inf. Model.* **2007,** *47,* 195.



For Table of Contents Only