

Designing A Bioinspired Degradation System for Forever Chemicals in Water Using Molecular Simulations

Danny Nguyen^{1*} Jesús Valdiviezo²

1. Julia R. Masterman Laboratory and Demonstration School, 1699 Spring Garden St, Philadelphia, Pennsylvania, 19130, United States

2. Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts, 02115, United States

Abstract

Per- and polyfluoroalkyl substances (PFAS), often termed “forever chemicals”, are resistant to natural biodegradation processes. PFAS poses many health risks and environmental challenges and is correlated with causing certain types of cancer, immune system disorders, and pollution. As a result, many approaches have been attempted to degrade PFAS and more similar substances but have been inefficient in degradation because of high cost, producing toxic byproducts, and high energy intensity. This research proposes a method of utilizing the fatty acid photodecarboxylase (5NCC), a photo enzyme that requires light to catalyze its reactions obtained from the Protein Data Bank, to degrade PFAS. Since the molecular characteristics of PFAS are similar to the protein’s natural ligand, palmitic acid (PLM), we suggest that PFAS is also degradable by the fatty acid photodecarboxylase (5NCC) protein. We produced the highest occupied molecular orbitals (HOMO), the lowest occupied molecular orbitals (LUMO), and molecular dynamic (MD) simulations of the 5NCC enzyme with a carboxylate PFAS ligand to assess binding affinity and examine if it can emulate the PLM ligand. These experimentations investigates the efficiency of carboxylate versions of a novel PFAS degradation system and represents a step forward in achieving sustainable solutions for the PFAS contamination global issue.

Keywords: Chemistry; Computational Chemistry; PFAS degradation; remediation; Fatty acid Photodecarboxylase

Introduction

Per- and polyfluoroalkyl substances (PFAS) are anthropogenic substances containing multiple Carbon-Fluorine bonds with a hydrophobic polar headgroup and tail. PFAS are a diverse group of synthetic chemicals that have been produced for over 90 years and can be grouped into 2 categories: short-chain and long-chain.¹ PFAS are highly nondegradable substances that serve in a variety of sectors like pharmaceuticals, consumer goods, electronics, and even in environmental services for water treatment because of their strong C-F bonds which gives them thermal and chemical stability.^{1–3}

However, the same unique properties of PFAS also contribute to bio-accumulation and are toxic to both humans and the environment.⁴ It has persistent global use because of its practical applications but as a result, PFAS contamination is pervasive and affects drinking water, agriculture, and livestock around the world.^{4–5} According to some scientific studies, exposure to certain levels of PFAS can lead to: increased rates of prostate, kidney, and testicular cancers, weakened immune system, increased

cholesterol and obesity, reproductive effects, developmental effects in children, and interference with the human body's natural hormones.^{3,6,7}

To understand degradation methods of PFAS, we first must understand its physical properties and chemical composition to understand why certain methods are efficient or inefficient. Structurally, PFAS has a carbon chain, carbon-fluorine (C-F) bonds, a functional group, and an amphiphilic structure.^{8,9} Most notably, the hazardous characteristics of PFAS are mainly due to its strong C-F bonds. The strong bonds are caused by properties such as high electronegativity and low polarizability of fluorine which leads to weak intermolecular interactions. This contributes to their mutually hydro- and lipophobic (stain-resistant) and surfactant properties and make them thermally and chemically stable.¹⁰ Because of such unique amphiphilic properties, this makes PFAS more capable of bioaccumulation and sorption,^{11–12} thereby posing toxicity to organisms.

The carbon chain is a backbone of 2 to 13+ carbon atoms that can be grouped into short and longer carbon chains connecting the “tail” of the molecule and the “head of the molecules, which is a charged function group. It is also known that longer chains, water-attracting and polar, are more bioaccumulative and have a stronger affinity for solid-phase organic carbon while shorter chains, water-repelling and nonpolar, are more mobile in water.

The functional groups are commonly carboxyl groups and are attached at one end of the chain. PFAS with carboxylic acid or sulfonamide functional groups have surfactant-like properties because of its amphiphilic structure. This makes the structure more complex than traditional surfactants where PFAS are able to accumulate at fluid-fluid interfaces and interact with both polar and nonpolar regions within living organisms. Some PFAS functional groups (ionic PFAS) can dissociate into anions or cations in aqueous solution under appropriate pH conditions.⁸

Traditional methods have been used to degrade PFAS: thermal destruction, chemical redox, or electrochemical oxidation.¹³ However, there are many drawbacks because these methods are chain length-specific and functional group-specific which leads to inefficacy in combating the wide range of PFAS present in the environment. There is also concern over high energy costs, the ability to perform these methods on a larger scale, and the incomplete degradation creating short-chain PFAS.¹⁴

Research in PFAS degradation using computational methods poses a more reliable alternative solution than traditional methods because of cost and time efficiency. Computational chemistry can be used to enhance these processes where it is involved in selecting and optimizing absorbents that efficiently remove PFAS from contaminated environments.¹⁵ Molecular dynamics (MD) simulations can enhance a deeper understanding of contaminant-mineral interactions and the energetics of adsorption.¹⁶

Protein engineering like De novo and in silico design as a whole also hold potential in advancing research in PFAS degradation. De novo design utilizes prior molecular structures to create novel molecules or proteins to have a strong binding affinity to a specific protein target.^{17–18} This allows for exploration of new structures beyond naturally-occurring molecular structures which can lead to improved therapeutics and functions.^{18–19} In silico protein design modifies existing protein structures to achieve new functions, properties, and to improve binding affinity.²⁰ Protein modification can range from minor adjustments of an existing protein or a substantial redesign to make a novel structure. For instance, an approach is rational mutagenesis where specific residues, amino acids, or active sites are mutated to have a different function²¹.

The drawbacks of protein engineering as there are many amino acid sequences which makes it difficult to explore specific optimal solutions and experimental validation is required in order to evaluate

the new structure and its functionality.^{22–23} However, de novo and in silico still offer several advantages because enzymes can be tested virtually and replicated many times for more results without the risk of PFAS exposure making it cost effective, safe, and rapid.

Combining de novo design and in silico design with computational tools can provide insights that we are otherwise not able to have with traditional de novo and in silico protein design alone. Computational de novo and in silico protein design have allowed for rational, physics-based design of protein structures and sequences which enables the creation of novel protein folds and functions not found in nature. Moreover, computational methods can enhance structural precision and improve functional design which leads to rapid iteration that can validate experimental testing and optimization. Some computational protein engineering methods are coarse-grained molecular dynamic (MD) simulations, quantum mechanical methods like Density Function Theory (DFT), and machine learning approaches. Computational approaches examine proteins more extensively compared to experimental methods alone and allows for more targeted and efficient design processes.¹⁷

Computational means such as the usage and study of structures from the Protein Data Bank have contributed to the success of protein design applications.¹⁷ The usage of computational protein design have led to advancements like the observation of the new alpha-beta protein which had not been previously observed in nature.¹⁷ This molecule was assembled from fragments from the Protein Data Bank after analysis of the natural protein structure which led to the creation of its artificial structure.¹⁷ Furthermore, there have been more recent successes in computational protein designs such as novel folds, novel enzymes, vaccines, antibodies, novel protein assemblies, ligand-binding proteins, and membrane proteins.²⁴ Furthermore, Enzymes have been developed and optimized because of computational methods and hold a promise of continuing to enhance protein design.¹⁷

In the context of PFAS degradation, de novo and in silico protein design are approaches that make up protein engineering and has promising applications in developing enzymes that can degrade PFAS as they can aim to identify and optimize proteins that are capable of breaking down PFAS through advanced computational modeling and molecular design.

Another prior research performed MD simulations on the *Delftia acidovorans* bacterium, which were collected samples from PFAS-contaminated sites and identified as a promising microorganism capable of degrading PFAS.²⁵ The enzymes produced by *D. acidovorans*, dehalogenases, were engineered for expression in *Escherichia coli* to enable rapid production and purification and found potential for enzymatic defluorination of PFOA/S.²⁵ This study visually identified potential pockets and performed binding affinity modeling with a docking software called MCule-1CD and showed that PFAS was bounded to the dehalogenases. When cultured with PFOA during experimentation, *D. acidovorans* have shown an increase in Fluoride ion release in PFOA. However, this research also poses limitations. The models of some of these halogenases, DeHa I and II, were based on homology rather than direct structural data which could impact the reliability of binding studies and functional predictions as it lacks published crystal structures for DeHa I and II.

This past research introduced the use of bacterial enzymes as a potential solution in degrading PFAS but the drawbacks is the accuracy and precision when discovering enzymes with strong bind affinity to PFAS ligands as the 3D structures of dehalogenases enzymes used are uncertain. Therefore, a better alternative solution is to continue with bacterial enzymes but utilize structures that are more established and validated.

We propose the usage of the 5NCC protein, a structure of fatty acid Photodecarboxylase in complex with Flavin-Adenine Dinucleotide (FAD) and Palmitic Acid (PLM) from the Protein Data Bank, which provides the detailed 3D structures of the enzyme classified as an oxidoreductase. This enzyme is

derived from the green alga *Chlorella variabilis* and plays a pivotal role hydrocarbon production and the photodecarboxylation of fatty acids—a process that involves the removal of a carboxyl group from fatty acids using light energy.²⁶

In this current study, we propose that the fatty acid photodecarboxylase enzyme is able to remediate and/or degrade PFAS because of its ability to decarboxylate fatty acids, synthesize hydrocarbons from carboxylic acids, and use light to do these processes. We performed electronic structure experimentation, molecular dynamics simulations, and utilized RMSD and RMSF calculations to investigate the practicality of the 5NCC protein to degrade PFAS. By proposing a novel method of degradation, this current study holds potential to advance current research in PFAS capturing methods and help mitigate the spread and contamination of PFAS (carboxylate form) through an eco-friendly and green method.

Methods

We generated and analyzed the molecular orbitals of fatty acid carboxylates using the PySCF program, an open-source quantum chemistry library intended for electronic structure calculations. We installed many Python packages: RDKit for molecular structure editing, matplotlib and seaborn for plotting, numpy and pandas for numerical and data analysis, PySCF and py3Dmol for molecular visualization, and ForteCubeView for orbital plotting.^{27,28,29} The molecular orbitals of four molecules were created and studied using RDKIT: two PFAS molecules, a carboxylic and carboxylate type, two fatty acids, a carboxylic type Palmitic Acid (PLM) and a carboxylate type of polylactic acid (PLA). Hartree-Fock (HF) and post-HF algorithms were also successfully executed with PySCF. After the libraries and packages were imported, the 2D structures of the molecules were generated from the Smiles strings, a format used to represent chemical structures, using the Chem.MolFromSmiles() function. For instance, the function used Palmitic Acid's (PLM) SMILES string “CCCCCCCCCC(=O)O”, to produce its 2D structure.

The 3D structures were also obtained and prepared with many techniques. The xyz coordinates of the structure were collected using a self-defined function get_xyz and extracted by the Chem.MolToXYZBlock(mol). The molecules were embedded with Hydrogen atoms to ensure proper valence and bonding configurations and were optimized by the MMFF force field if necessary. Then, the 3D structure was visualized using the py3Dmol.view() function with inputs of a height and width of 300 (in pixels), a stick representation, and a sphere with a scale of 0.3.

After producing the 3D structures, the function run_calculation() calculates the electronic structure and other molecular properties of the molecule. The “sto-3g” dataset, used to approximate molecular orbitals, was utilized in the function. The command gto.M(...) used the Gaussian Type Orbital (GTO) basis sets to initialize the molecule, the unit of the atomic coordinates were set in Angstroms, molecular symmetry were enabled, and the mol.build() function built the molecules with the given parameters. Afterwards, the Restricted Hartree-Fock (RHF) calculation was run on the molecule, calculating the ground state energy of the molecule, assuming that all electrons are paired. The result of the RHF, which contained the molecular orbital coefficients and energy, and molecule was returned.

The molecular orbital (MO) energy levels and their occupancies were organized by a data table produced by the Pandas pd.DataFrame(...) for data analysis with inputs of Mf.mo_energy and mf.mo_occ. Mf.mo_energy contained the energy levels of the molecular orbitals calculated from Hartree-Fock and the mf.mo_occ command contained the occupancies of molecular orbitals running from 0-2, 0 being empty orbitals, 1 being partially occupied orbitals, and 2 being fully occupied orbitals. The canonical orbitals, directly computed by the HF calculation, the intrinsic atomic orbitals (IAOs), intrinsic bonding orbitals

(IBOs) and localized intrinsic valence virtual orbitals (LIVVOs) were collected. The Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO) from the molecular orbital (MO) data in the mean-field object (`mf`) is identified. Files were produced and plotted to visualize all different types of molecular orbitals and displayed regions where 85% of the total electron density or orbital probability is contained.

The preparation process for the molecular dynamics simulations included importing/installing condacolab, RDKit, OpenMM Force Fields, numpy, MDTraj, pdbfixer, OpenMM, OpenFF Toolkit, and GAFF Template Generator for Force Field (GAFF). The pdb file of the 5NCC protein and its ligand were obtained from the Protein Data Bank.

A function was used from the `pdbfixer` library to prepare the protein. This addressed any missing residues or atoms where it was identified. Nonstandard residues were replaced by standard ones and added any missing atom, and removed any heteroatoms from the protein structure. It also managed the protonation state based on the pH level where any missing hydrogen atoms were added to meet the specified pH level. The ligands were individually handled by splitting into separate components based on residues, hydrogen atoms were removed and added to prepare for protonation, and bond orders were assigned to the ligand based on a reference molecule template. A visual comparison depicting the ligand before and after were produced for analysis and ensuring the success of the preparation. The 5NCCA protein's other natural ligand, FAD, was also produced from its smiles string.

The RDKIT molecule was converted to an OpenFF molecule then into a topology in order to be used in OpenMM. The protein and ligands were merged along with the FAD ligand into a single topology and atomic positions suitable for a simulation. The protein force field used was "amber14-all.xml" and the solvent force field was "amber14/tip3pfb.xml". Solvent molecules were added to the system with an ionic strength of 0.15 M and a 1 nanometer buffer zone was added to ensure the presence of solvents around the system. The solvation box was the TIP3 water model.

Setting up the simulation, the target temperature for the simulation was 300 K and the friction coefficient (γ) was 1 ps^{-1} , and the time step for the integration was 2 fs. The simulation system undergoes energy minimization and then stores a topology and atomic positions into a PDB file. The simulation ran in two different ways. If it was on google colab, the simulation ran for 100 picoseconds (ps), the trajectory file was written every 10 picoseconds(ps), and the process was logged every 5 picoseconds (ps). If run locally, the simulation ran for 20 femtoseconds (fs), trajectory file writing every step (2 fs), progress being logged every step 2 femtoseconds, with a temperature of 300 Kelvin (K). The resulting simulation was stored in a pdb and trajectory file titled "topology.pdb" and "trajectory.xtc" in the XTC format and titled, trajectory.xtc.

After molecular simulations were complete, the trajectory and topology files of the 2 cases, the PFAS ligand bind with the enzyme and the PLM ligand bind with the enzyme, were imputed into a python script to compute the Root Mean Square Deviation (RMSD) calculations of the ligand and backbone of the protein. The script utilized the MDTraj library to perform trajectory analysis. The analysis was performed after energy minimization and equilibration to observe the structural stability of both ligands under the binding of the enzymes and assess if both enzymes shared similar dynamic characteristics. 4 graphs were produced, 2 for the PLM ligand and 2 for the PFAS ligand.

Results and Discussion



Figure 1: This is a visualization of the fatty acid photodecarboxylate enzyme (green) in its natural state collected from the Protein Data Bank and obtained by being expressed in *E. coli*.²⁶ Along with the enzyme, it also depicts its ligands PLM (blue) and FAD (red). This enzyme converts fatty acids to hydrocarbons in response to blue light initiated by the light-capturing flavin adenine dinucleotide (FAD) cofactor with a quantum yield >80%. The protein's hydrocarbon production activity was optimum around pH 8.5 and decreased sharply when temperature rose above 35°C³⁰

We first evaluated the structure of the PLM ligand to the PFAS ligand to see if there are any structural similarities between PFAS and PLM and can thus consider PFAS as potentially degradable by the 5NCC enzyme. Figure 2 shows Palmitic acid having 16-carbon aliphatic chain and a carboxylate anion (-COO⁻) group at one end. PFAS was also shown to have 16-carbon aliphatic chain with 33 Fluorine atoms and a carboxylate anion (-COO⁻) group at one end. As illustrated, both PFAS and PLM showed structural similarities in carbon chain and a hydrophobic tail end which suggests that there may be similar chemical reactivity and binding affinity, assuming that the enzyme can't differentiate between the two ligands. A significant difference is that PLM consists of a hydrocarbon backbone and is not entirely fluorinated compared to carbon-fluorine bonds and their more fluorinated backbone which accounts for how PFAS is less biodegradable than PLM.

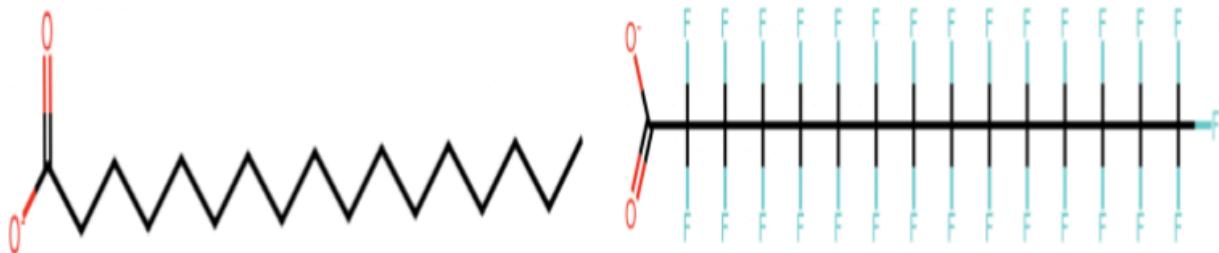


Figure 2: 2D structure representation of carboxylate palmitic acid (PLM) and a carboxylate version of PFAS produced by . In this depiction, palmitic acid consists of a 16-carbon aliphatic chain and a carboxylate anion (-COO⁻) group at one end. PFAS also has a 16-carbon aliphatic chain with 33 Fluorine atoms and a carboxylate anion (-COO⁻) group at one end. The

representations were produced by the Chem.MolFromSmiles command from the RDkit library which imputed the molecules' SMILES string code to convert into a 2D figure.³¹

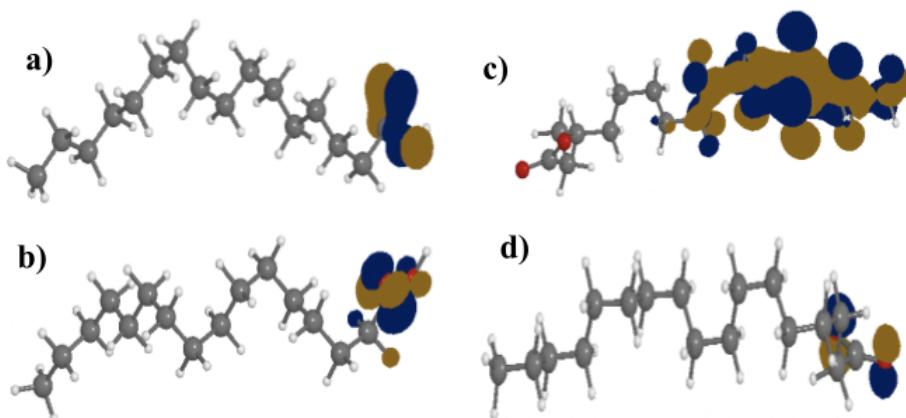


Figure 3: Frontier molecular orbitals (HOMO and LUMO) of palmitic acid (PLM) ligand of both carboxylate and carboxylic versions of the palmitic acid ligand associated with the Fatty acid Photodecarboxylase (5NCC) protein and obtained from the protein data bank and generated with PySCF and fortetcubeview.²⁹ These structures were produced with its smiles string code. a) Depicts the HOMO orbital of carboxylic PLM. b) Depicts the LUMO orbital of carboxylic PLM c) Depicts the HOMO orbital of carboxylate PLM d) Depicts the LUMO orbital of carboxylate PLM.

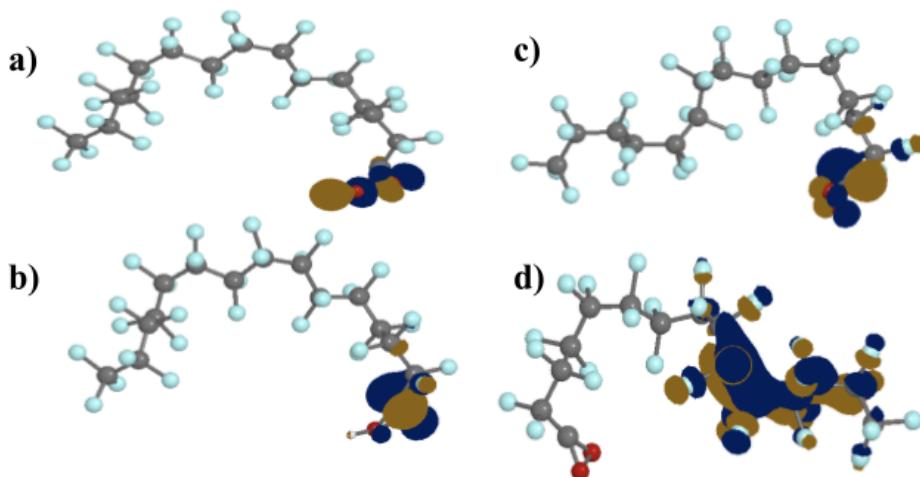


Figure 4: Frontier molecular orbitals (HOMO and LUMO) of both carboxylate and carboxylic versions of the PFAS ligand generated with PySCF and fortetcubeview.²⁹ These structures were produced with its smiles string code. a) Depicts the HOMO orbital of carboxylic PFAS. b) Depicts the LUMO orbital of carboxylic PFAS c) Depicts the HOMO orbital of carboxylate PFAS d) Depicts the LUMO orbital of carboxylate PFAS.

System	E_HOMO	E_LUMO
--------	--------	--------

PLM (carboxylic)	1.23 eV	16.66 eV
PLM (carboxylate)	2.95 eV	16.67 eV
PFAS (carboxylic)	-0.40 eV	11.18 eV
PFAS (carboxylate)	2.61 eV	10.48 eV

Figure 5: Energy values (electron volts) of molecular orbitals (HOMO and LUMO) of PLM and PFAS (both carboxylate and carboxylic versions) produced with PySCF.³²

Molecular Orbitals of PFAS and PLM (both carboxylate and carboxylic versions) were also obtained to acquire insight into the reactivity of the molecules in order to understand its interactions with the enzyme. Significantly, we observed that the HOMO values of the carboxylate versions of PFAS (2.61 eV) and PLM (2.95) had the lowest energy level gap (0.34 eV) which suggests that the molecules have comparable reactivity. Although the LUMO values of carboxylate PFAS (10.48) and PLM (16.67) have a bigger gap, it's relatively still comparable with a gap of 6.19 eV.

Residue Name and #	Type	Residue Name and #	Type	Residue Name and #	Type
ALA'128/CA	Hydrophobic	GLY'455/C	Hydrophobic	GLU'550/CB	Polar (Charged)
THR'131/CA	Polar	MET'456/SD	Polar (Uncharged)	LEU'551/CG	Hydrophobic
PHE'134/CD2	Hydrophobic	ALA'457/CA	Hydrophobic	PHE'552/C	Hydrophobic
LEU'386/CG	Hydrophobic	LEU'458/CA	Hydrophobic	PRO'553/N	Polar (Uncharged)
ILE'398/CG2	Hydrophobic	GLY'462/C	Hydrophobic	GLY'554/C	Hydrophobic
ALA'399/N	Hydrophobic	VAL'463/CG2	Hydrophobic	SER'555/OG	Polar (Uncharged)
ASP'402/O	Polar (Charged)	THR'465/O	Polar (Uncharged)	GLY'556/C	Hydrophobic
SER'429/O	Polar (Uncharged)	TYR'466/CD1	Hydrophobic	VAL'557/N	Hydrophobic
THR'430/O	Polar (Uncharged)	PHE'469/CZ	Hydrophobic	HIS'572/O	Polar (Uncharged)
GLY'431/C	Hydrophobic	TRP'479/CD2	Hydrophobic	SER'573/N	Polar (Uncharged)
CYS'432/CA	Polar (Uncharged)	THR'484/CB	Polar (Uncharged)	SER'574/OG	Polar (Uncharged)
ARG'451/CB	Polar (Charged)	GLN'486/CG	Polar (Uncharged)	ASN'575/CB	Polar (Uncharged)
VAL'453/CG1	Hydrophobic	ARG'535/CZ	Polar (Charged)	PLM'702/CC	Hydrophobic

Figure 6: Residues in close proximity to enzyme (within 5 Angstroms) categorized by its abbreviated name, sequence number within the protein, and it's type (hydrophobic or polar). Total of 39 residues detected. This analysis provides potential enzymatic degradation pathways for PFAS.

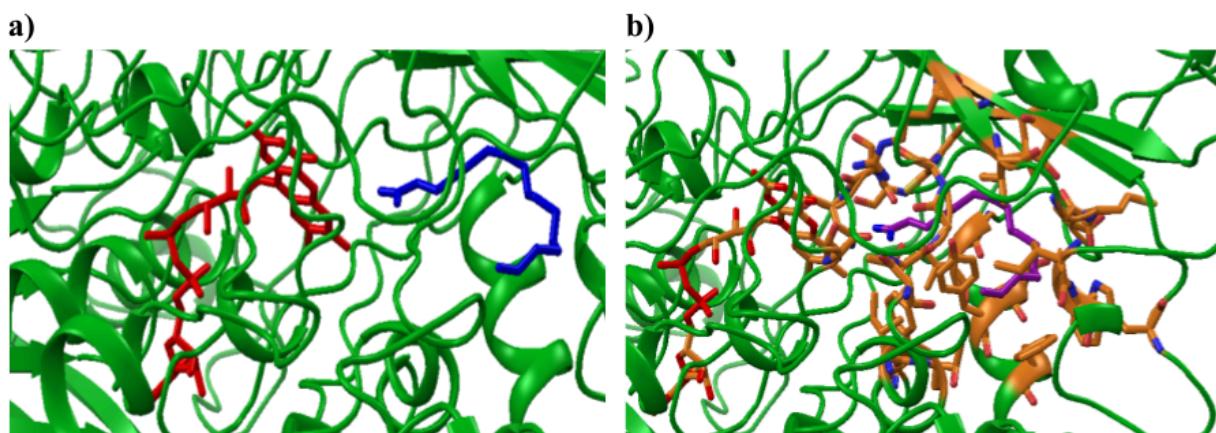


Figure 7: 3D structure representation (pdb file) of PLM (blue) and FAD (red) ligand with 5NCC protein depicting the protein's natural state and then depicting residues that are part of the binding site of the PLM ligand (orange) visualized using Pymol.³² a) Regular cartoon representation of the 5NCC protein b) Cartoon representation of the 5NCC depicting residues that are part of the binding site of the PLM ligand (orange).

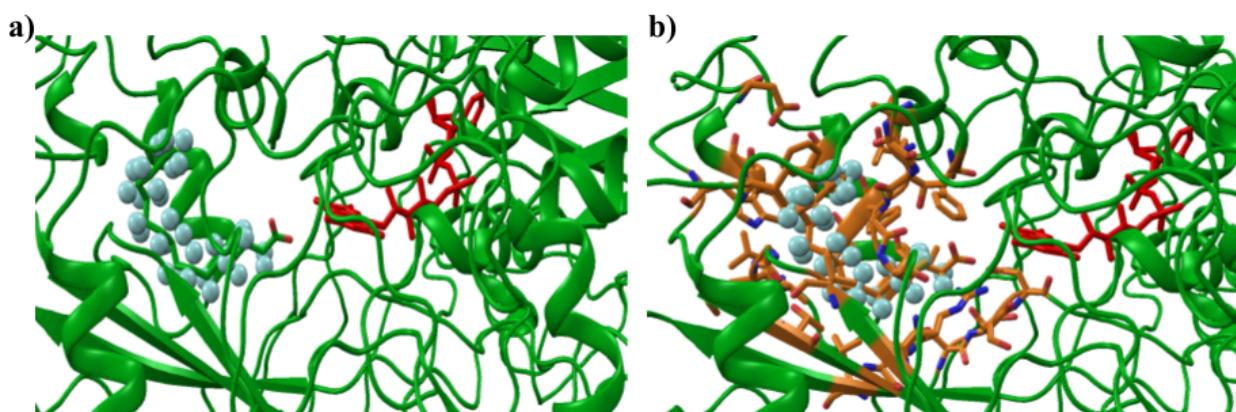


Figure 8: 3D structure representation (pdb file) of PFAS and FAD (red) ligand with 5NCC protein depicting the protein's natural state and then depicting residues that are part of the binding site of the PLM ligand (orange) visualized in Pymol. a) Regular cartoon representation of the 5NCC protein b) Cartoon representation of the 5NCC depicting residues that are part of the binding site of the PLM ligand (orange).

Figure 6 shows the 39 detected residues of both PFAS and PLM. When obtaining which residue reacted with the ligands, the PFAS ligand was directly substituted in for PLM and we observed that it was the same residue bonded to PLM, also bonded with PFAS. This indicates that they share similar binding modes within the enzyme's active site, also implying that the positive outcome of the enzyme potentially can't differentiate between the ligands because of structural similarity. Furthermore, this implies that the enzyme can also theoretically catalyze the same reaction as it does with PLM. This can also be supported by the diverse data (39 residues total) of residues that bind with these ligands of different interaction types, hydrophobic, polar (uncharged), and polar (charged), suggesting the idea that the enzyme's active site can accommodate various substrates. Most residues were hydrophobic (21 residues) which can provide a stable environment for PFAS to bind with the enzyme since a hydrophobic core might be formed. Moreover, with the presence of charged polar amino acids, it can lead to further stabilization of PFAS with the enzyme if the ligand has any positive charges or dipole moments. Essentially, the presence of ionic amino acids like Arginine (ARG) (2 residues total), highly reactive to negative-charged

groups, and Glutamate (GLU), interacts with positive-charged sites, in the binding site hints that it's involved in the catalytic activity by promoting tight substrate binding.

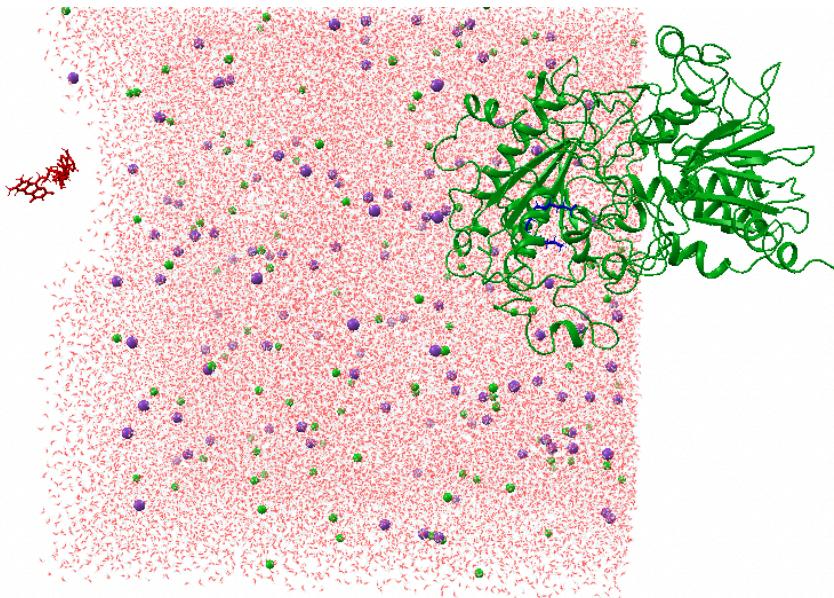


Figure 9: Representation of molecular dynamics simulations of PLM (blue) and FAD (red) ligand with 5NCC protein produced in Pymol. The system was solvated in water with oxygen (red), sodium (purple), and chloride (green) atoms added to neutralize ionic charges.

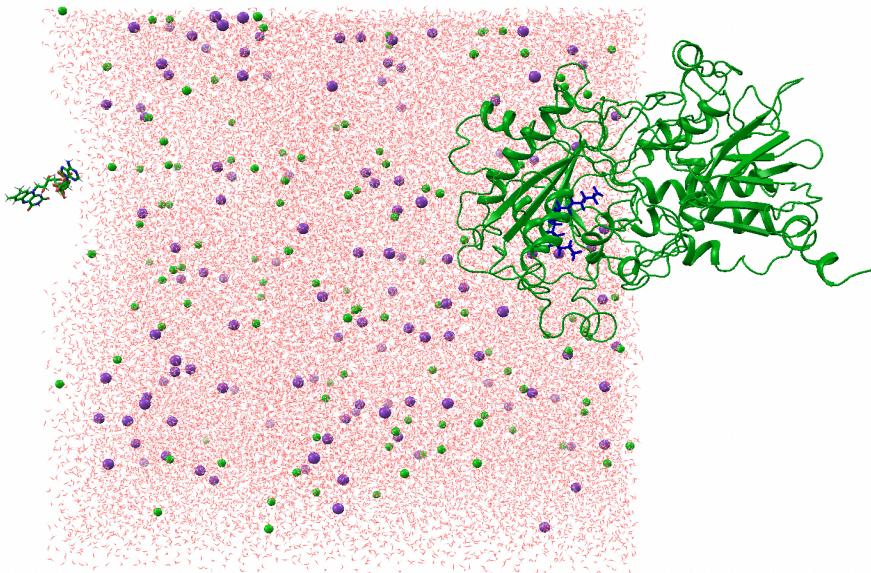


Figure 10: Representation of molecular dynamics simulations of PLM (blue) and FAD (red) ligand with 5NCC protein produced in Pymol. The system was solvated in water with oxygen (red), sodium (purple), and chloride (green) atoms added to neutralize ionic charges.

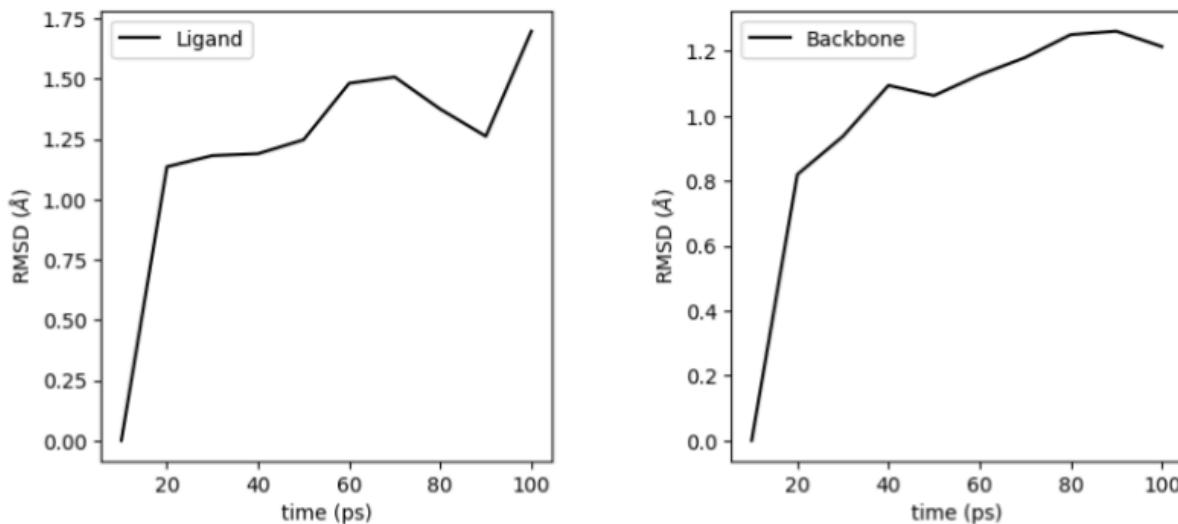


Figure 11: RMSD plots of the PLM ligand binded with the 5NCC protein. RMSD plot of the protein's PLM ligand (left) and RMSD plot of the protein's backbone (right). Both graphs were produced by a custom python script that utilizes Mdtraj library to perform trajectory analysis.³¹

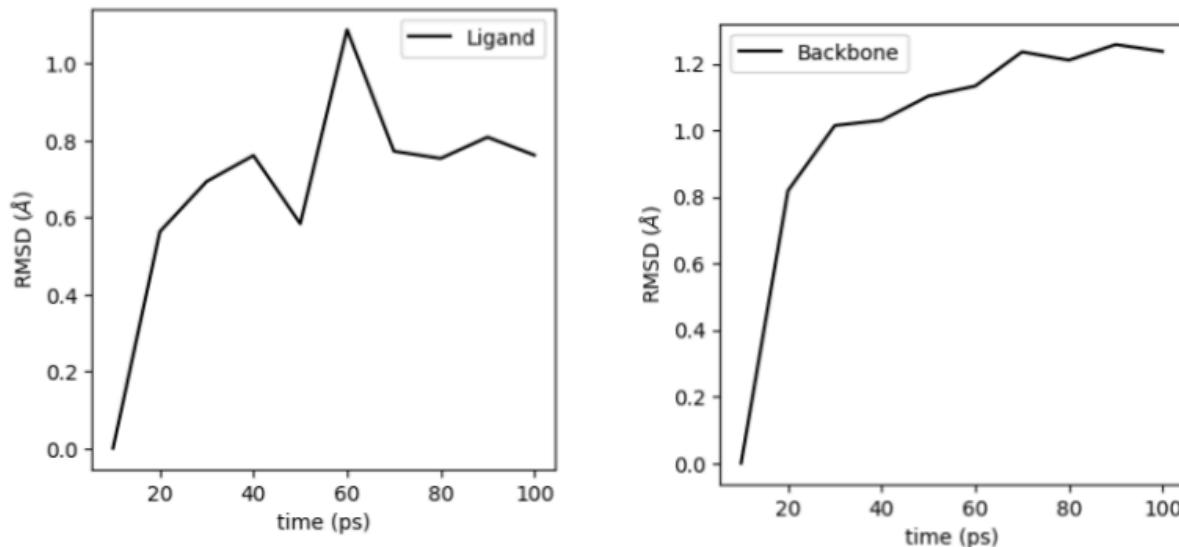


Figure 12: RMSD plots of the PFAS ligand coupled with the 5NCC protein. RMSD plot of the protein with the PFAS ligand (left) and RMSD plot of the protein's backbone (right) with PFAS present. Both graphs were produced by a python script that utilizes the Mdtraj library to perform trajectory analysis.

PLM and PFAS also shared similar dynamics characteristics as the PFAS ligand stayed stable and alike to PLM and certain times in the simulation. Shown by the RMSD graphs in figures 11 and 12, the RMSD calculations of both system's ligands were similar up to approximately 20 picoseconds (ps) with values around 0.58 Angstroms for PFAS and 1.12 Angstroms for PLM . More precisely, the RMSD values of the backbone of both systems had slight deviations throughout the simulation but at 100 picoseconds (ps), the values were around 1.2 Anstrom for PFAS and 1.19 for PLM. This infers that the 5NCC enzyme can accommodate PFAS just like PLM for early stages during the simulation for the ligands and the more accurate RMSD values of PFAS to PLM shows that it experiences conformational

changes to PLM. Nonetheless, the general RMSD calculations suggest similar dynamics in early stages and backbone.

There poses a limitation in analyzing structural characteristics of both molecules as our experimentation and findings can't accurately conclude that PFAS is degradable by the enzyme solely based on structural similarities. Yet theoretically, the 5NCC protein can cross-react with PFAS, particularly long-chain ones, since PLM and PFAS have similar chain length, hydrophobicity, and functional groups. Our experimentation of only observing and considering close-distant residues to the enzyme lacks considering other realistic factors and making generalizations that can hinder findings. During experimentation, we only detected residues that were within 5 Angstroms of the ligand. By doing this, farther distanced residues that can potentially influence the interaction of the ligand are not considered. Classifying residues as either hydrophobic or polar and observing their charges only can also lead to a limitation in predicting behavior of real-world enzyme-ligand systems. The purpose of this experimentation was to make a simple conclusion on if PFAS are dynamically similar to PLM and we discovered that the enzyme can allow for various substrates like PFAS to bind with it by certain residues that enhances the binding strength and excludes any disfavorable interactions. For the molecular orbital calculations, there was a limited number of molecular orbitals, only HOMO and LUMO. This causes our findings to not other relevant electronic properties that could influence reactivity and interactions with the enzyme. The ligand RMSD calculations can be interpreted to show a slight falter in the dynamic similarity between PFAS and PLM and there's a limited scope of analysis since only the ligands and backbones values were calculated.

Conclusion

This study investigated the structural, electronic, and dynamic characteristics of the carboxylate version of PFAS, binded with the 5NCC protein that was obtained from the Protein Data Bank, to the palmitic acid (PLM) ligand to assess the capability of the 5NCC enzyme degrading per- and polyfluoroalkyl substances (PFAS). Our findings illustrated significant structural similarities between PFAS and PLM - the same 16 carbon chain and carboxylate anion group - which showed that in the beginning of our experimentation, PFAS could have equivalent chemical reactivity and binding affinity. With the analysis of close-distant binding residues, we noticed a high amount of hydrophobic residues and ionic residues like Arginine (ARG) and Glutamate (GLU) that stabilize the ligand-enzyme system, which further supported that PFAS could be tightly bound with the enzyme. The molecular orbital comparisons also showed similar electronic reactivities by the small gap of 0.34 eV between the HOMO values of the molecules' carboxylate version. More notably, the RMSD calculations also strengthened our hypothesis of similar dynamics. However, this study also comes with limitations as the generalization of certain experimentations could affect findings by overlooking critical interactions that affect binding behavior and not accounting fully for dynamic characteristics of realistic enzyme-ligand binding. Overall, the 5NCC protein's structural and chemical flexibility necessary to interact with PFAS reinforces the proposition that the protein has the capabilities to degrade PFAS. The barrier that prevents using this idea in realistic applications is the challenge in designing the carboxylate version of PFAS structure in order for it to fit within the protein and bind. Yet ultimately, this understanding contributes towards further research as scientists are able to explore this new method and discover potential enzymatic pathways of bioremediation for PFAS.

Acknowledgements

I want to thank my mentor and collaborator, Dr. Jesús Valdiviezo from Harvard Medical School, who closely supported my research for the past few months and provided guidance and essential skills for

me to excel. I want to also thank Dr. Preston B. Moore from St.Joseph's University College of Pharmacy who provided me with basic computational skills needed to perform my experimentation.

References

1. Ma, B. Computational Design of Sorbent Materials for Per- and Polyfluoroalkyl Substances (PFAS) Remediation; American Chemical Society: 2024.
<https://pubs.acs.org/doi/10.1021/accountsmr.4c00142>.
2. The Water Research Foundation. Per- and Polyfluoroalkyl Substances; The Water Research Foundation: 2024. <https://www.waterrf.org/serve-file/4949-PFAS.pdf>.
3. United States Environmental Protection Agency. Our Current Understanding of the Human Health and Environmental Risks of PFAS; United States Environmental Protection Agency: 2024.
<https://www.epa.gov/pfas/our-current-understanding-human-health-and-environmental-risks-pfa-s>.
4. Cao, H.; Peng, J.; Zhou, Z.; Sun, Y.; Wang, Y.; Liang, Y. Insight into the Defluorination Ability of Per- and Polyfluoroalkyl Substances Based on Machine Learning and Quantum Chemical Computations. *Sci. Total Environ.* 2022, 807.
<https://www.sciencedirect.com/science/article/abs/pii/S0048969721060964>.
5. Souza, B. B.; Hewage, S. A.; Kewalramani, J. A.; Duin, A. C.; Meegoda, J. N. A ReaxFF-Based Molecular Dynamics Study of the Destruction of PFAS Due to Ultrasound. *Environ. Pollut.* 2023, 333.
<https://www.sciencedirect.com/science/article/abs/pii/S026974912301028X>.
6. Verma, S.; Lee, T.; Sahle-Demessie, E.; Ateia, M.; Nadagouda, M. N. Recent Advances on PFAS Degradation via Thermal and Nonthermal Methods. *Curr. Opin. Green Sustain. Chem.* 2023, 13.
<https://www.sciencedirect.com/science/article/abs/pii/S2666821122001818>.
7. Darrow, L. A.; Stein, C. R.; Steenland, K. Serum Perfluoroctanoic Acid and Perfluorooctane Sulfonate Concentrations in Relation to Birth Outcomes in the Mid-Ohio Valley, 2005–2010. *Environ. Health Perspect.* 2013, 121 (10), 1207–1213.
<https://doi.org/10.1289/ehp.1206372>.
8. Interstate Technology & Regulatory Council. Chemistry, Terminology, and Acronyms. <https://pfas-1.itrcweb.org/2-2-chemistry-terminology-and-acronyms/> (accessed Dec 26, 2024).
9. Buck, R. C.; Franklin, J.; Berger, U.; Conder, J. M.; Cousins, I. T.; de Voogt, P.; Jensen, A. A.; Kannan, K.; Mabury, S. A.; van Leeuwen, S. P. Perfluoroalkyl and Polyfluoroalkyl

Substances in the Environment: Terminology, Classification, and Origins. *Integr. Environ. Assess. Manag.* 2011, 7 (4), 513–541. <https://doi.org/10.1002/ieam.258>.

10. Kiss, E. Fluorinated Surfactants and Repellents, 2nd ed.; Marcel Dekker: 2001.
11. Shahsavari, E.; Rouch, D.; Khudur, L. S.; Thomas, D.; Aburto-Medina, A.; Ball, A. S. Challenges and Current Status of the Biological Treatment of PFAS-Contaminated Soils. *Front. Bioeng. Biotechnol.* 2021, 8, Article 602040. <https://doi.org/10.3389/fbioe.2020.602040>.
12. Wee, S. Y.; Aris, A. Z. Revisiting the “Forever Chemicals,” PFOA and PFOS Exposure in Drinking Water. *npj Clean Water* 2023, 6, Article 57. <https://doi.org/10.1038/s41545-023-00274-6>.
13. Marcelsky, M.; Aga, D. S.; Bradley, I. M.; Aich, N.; Ng, C. Mechanisms and Opportunities for Rational In Silico Design of Enzymes to Degrade Per- and Polyfluoroalkyl Substances (PFAS). *J. Chem. Inf. Model.* 2023, 63 (23), 7299–7319. <https://doi.org/10.1021/acs.jcim.3c01303>.
14. Heilesen, L.; Aarhus University. Bacterial Enzyme Traps and Breaks Down PFAS Molecules. *Phys.org*. <https://phys.org/news/2023-03-bacterial-enzyme-pfas-molecules.html> (accessed Dec 26, 2024).
15. Souza, B. B.; Meegoda, J. Insights into PFAS Environmental Fate through Computational Chemistry: A Review. *Sci. Total Environ.* 2024, 926. <https://doi.org/10.1016/j.scitotenv.2024.171738>.
16. Willemse, J. A. R.; Bourg, I. C. Molecular Dynamics Simulation of the Adsorption of Per- and Polyfluoroalkyl Substances (PFASs) on Smectite Clay. *J. Colloid Interface Sci.* 2021, 585, 337–346. <https://doi.org/10.1016/j.jcis.2020.11.071>.
17. Pan, X.; Kortemme, T. Recent Advances in de Novo Protein Design: Principles, Methods, and Applications. *J. Biol. Chem.* 2021, 296. <https://doi.org/10.1016/j.jbc.2021.100558>.
18. Pegg, S. C. H.; Haresco, J. J.; Kuntz, I. D. A Genetic Algorithm for Structure-Based de Novo Design. *J. Comput. Chem. Mol. Des.* 2001, 15 (9), 911–933. <https://doi.org/10.1023/A:1014389729000>.
19. Mouchlis, V. D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A. G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in de Novo Drug Design: From Conventional to Machine Learning Methods. *Int. J. Mol. Sci.* 2021, 22 (4), 1676. <https://doi.org/10.3390/ijms22041676>.
20. Watson, J. L.; Juergens, D.; Bennett, N. R.; et al. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* 2023, 620, 1089–1100. <https://doi.org/10.1038/s41586-023-06415-8>.

21. Dahiyat, B. I. In *Silico Protein Design*. In *Bioinformatics and Drug Discovery*; Larson, R. S., Ed.; Humana Press: 2006; Vol. 316, pp 359–374.
<https://doi.org/10.1385/1-59259-964-8:359>.
22. Cao, L.; Coventry, B.; Goreshnik, I.; et al. Design of Protein-Binding Proteins from the Target Structure Alone. *Nature* 2022, 605, 551–560.
<https://doi.org/10.1038/s41586-022-04654-9>.
23. Klepeis, J. L.; Floudas, C. A. In *Silico Protein Design: A Combinatorial and Global Optimization Approach*. *SIAM News* 2004, 37 (1). Retrieved from
<http://titan.princeton.edu/papers/john/protein.pdf>.
24. Wang, J.; Cao, H.; Zhang, J. Z. H.; Qi, Y. Computational Protein Design with Deep Learning Neural Networks. *Sci. Rep.* 2018, 8 (6349).
<https://www.nature.com/articles/s41598-018-24760-x>.
25. Harris, J. D.; Coon, C. M.; Doherty, M. E.; McHugh, E. A.; Warner, M. C.; Walters, C. L.; Orahood, O. M.; Loesch, A. E.; Hatfield, D. C.; Sitko, J. C.; Almand, E. A.; Steel, J. J. Engineering and Characterization of Dehalogenase Enzymes from *Delftia acidovorans* in Bioremediation of Perfluorinated Compounds. *Synth. Syst. Biotechnol.* 2022, 7 (2), 671–676.
<https://doi.org/10.1016/j.synbio.2022.02.005>.
26. RCSB Protein Data Bank. RCSB PDB - 5NCC: Fatty Acid Photodecarboxylase from *Chlorella variabilis* NC64A. RCSB PDB 2016. <https://www.rcsb.org/structure/5NCC>.
27. Landrum, G. A. RDKit: Open-Source Cheminformatics Software. *RDKit* 2006.
<https://www.rdkit.org>.
28. Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: The Python-Based Simulations of Chemistry Framework. *WIREs Comput. Mol. Sci.* 2018, 8, e1340.
<https://doi.org/10.1002/wcms.1340>.
29. OpenWeaver. fortetcubeview (Version 1.0.0) [Software]. PyPI 2023.
<https://pypi.org/project/fortetcubeview/>.
30. Sorigué, D.; Boulanger, C.; Pichon, A. An Algal Photoenzyme Converts Fatty Acids to Hydrocarbons. *Science* 2017, 357 (6349), 903–907.
<https://doi.org/10.1126/science.aan6349>.
31. MDTraj Development Team. MDTraj Documentation (Version 1.9.8.dev0). MDTraj 2023. <https://www.mdtraj.org/1.9.8.dev0/index.html>.
32. Schrödinger, LLC. The PyMOL Molecular Graphics System (Version 3.0). PyMOL 2024. <http://www.pymol.org>