

# SCIENTIFIC REPORTS



OPEN

## A structural and mechanistic study of $\pi$ -clamp-mediated cysteine perfluoroarylation

Peng Dai, Jonathan K. Williams, Chi Zhang, Matthew Welborn, James J. Shepherd, Tianyu Zhu, Troy Van Voorhis, Mei Hong & Bradley L. Pentelute

Natural enzymes use local environments to tune the reactivity of amino acid side chains. In searching for small peptides with similar properties, we discovered a four-residue  $\pi$ -clamp motif (Phe-Cys-Pro-Phe) for regio- and chemoselective arylation of cysteine in ribosomally produced proteins. Here we report mutational, computational, and structural findings directed toward elucidating the molecular factors that drive  $\pi$ -clamp-mediated arylation. We show the significance of a *trans* conformation prolyl amide bond for the  $\pi$ -clamp reactivity. The  $\pi$ -clamp cysteine arylation reaction enthalpy of activation ( $\Delta H^\ddagger$ ) is significantly lower than a non- $\pi$ -clamp cysteine. Solid-state NMR chemical shifts indicate the prolyl amide bond in the  $\pi$ -clamp motif adopts a 1:1 ratio of the *cis* and *trans* conformation, while in the reaction product Pro3 was exclusively in *trans*. In two structural models of the perfluoroarylated product, distinct interactions at 4.7 Å between Phe1 side chain and perfluoroaryl electrophile moiety are observed. Further, solution  $^{19}\text{F}$  NMR and isothermal titration calorimetry measurements suggest interactions between hydrophobic side chains in a  $\pi$ -clamp mutant and the perfluoroaryl probe. These studies led us to design a  $\pi$ -clamp mutant with an 85-fold rate enhancement. These findings will guide us toward the discovery of small reactive peptides to facilitate abiotic chemistry in water.

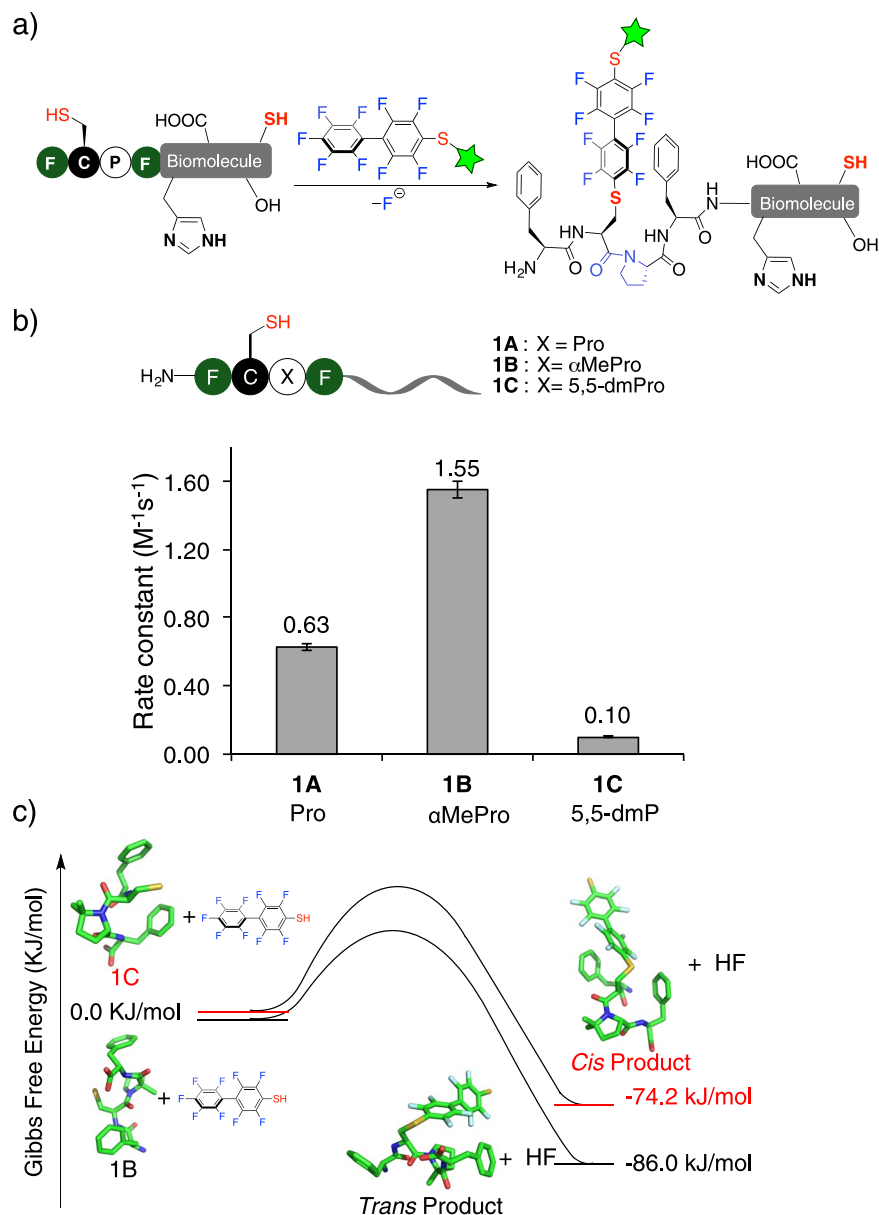
Chemo- and regioselective modification of proteins is a chemical challenge because polypeptide chains contain numerous similarly reactive functional groups<sup>1,2</sup>. Often the polymer, fluorophore, or small-molecule drug is rendered electrophilic for conjugation chemistry at nucleophilic protein side-chains. Many undesired side reactions occur and purification often fails. Major advances to overcome these chemistry challenges have been made, including the use of non-natural clickable amino acids within proteins for bio-orthogonal chemistry<sup>3,4</sup>, engineered enzymatic systems<sup>5</sup> such as sortase A<sup>6</sup>, and non-enzymatic routes that leverage selective chemistry between a matched reaction pair<sup>7–11</sup> such as the reaction of a N-terminal cysteine with a thioester<sup>11</sup>.

Recently we reported an enzyme-free, one-step, regio- and chemoselective protein-modification method based on cysteine arylation<sup>12</sup>. We discovered that the  $\pi$ -clamp tetrapeptide (Phe-Cys-Pro-Phe) possessed unique reactivity to achieve self-labeling with a perfluoroaryl (PFA) electrophile (Fig. 1a). The reaction even proceeded selectively in complex glycosylated IgG proteins. Each residue within the  $\pi$ -clamp is important as mutations often diminished the reaction efficiency. Further, the  $\pi$ -clamp-mediated arylation reaction rate is tunable by over four orders of magnitude with different salts<sup>13</sup>. Ammonium sulfate and other structure-stabilizing salts accelerated the reaction, while the addition of a denaturing salt such as guanidinium chloride impeded the reaction. Taken together, these results prompted us to carry out a systematic investigation of the  $\pi$ -clamp-promoted reaction. Here we employ kinetic studies, solution and solid-state NMR, and density functional theory (DFT) calculations to determine the structure-reactivity relationships that underscore this  $\pi$ -clamp mediated reaction. Our findings suggest several structural and chemical features contribute to the unique reactivity of  $\pi$ -clamp including the *trans* prolyl amide bond, lowered reaction  $\Delta H^\ddagger$ , reduced cysteine  $pK_a$  and side chain-perfluoroaryl electrophile interactions.

### Results

***Trans* prolyl amide accelerates  $\pi$ -clamp-mediated arylation reaction.** We first investigated if the prolyl amide conformation being either *cis* or *trans* in the  $\pi$ -clamp was important. Proline is conformationally unique among amino acids as the backbone phi ( $\phi$ ) torsion angle is restricted by the side chain ring structure.

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, United States. Correspondence and requests for materials should be addressed to B.L.P. (email: [blp@mit.edu](mailto:blp@mit.edu))



**Figure 1.** *Trans* conformation of the Cys-Pro amide bond is critical for  $\pi$ -clamp-mediated arylation. (a) Scheme of  $\pi$ -clamp-mediated cysteine arylation. (b) Reaction rate of  $\pi$ -clamp-mediated arylation is influenced by proline mutation. Peptides **1A–1C** were reacted with probe **2**. Reaction conditions: 200 mM phosphate, 20 mM TCEP, pH 8.0, 37°C. (c) Calculated Gibbs free energy changes of the perfluoroarylation reaction involving *trans* Pro  $\pi$ -clamp mutant (**1B**, black) and *cis* Pro  $\pi$ -clamp mutant (**1C**, red). Representative computed geometries of the  $\pi$ -clamp mutants (**1B** and **1C**) and the perfluoroarylated products used in the DFT calculation are also shown.

The Xaa-Pro amide bond can undergo *cis/trans* isomerization while other amide bonds are found in the *trans* conformation for most proteins<sup>14,15</sup>. Pro is thought to be a key player in the  $\pi$ -clamp, as mutation to D-Pro abrogates the reaction rate<sup>12</sup>. To understand if the conformation of the Cys-Pro amide bond perturbs the reaction we used mutation studies. Proline analogues containing certain functional groups can constrain the Cys-Pro conformation; analogues  $\alpha$ -methylproline ( $\alpha$ MePro)<sup>16,17</sup> and 5,5-dimethylproline (5,5-dmP)<sup>18,19</sup> strongly promote the *trans* and *cis* conformations, respectively. Three peptides incorporating proline (the  $\pi$ -clamp, peptide **1A**),  $\alpha$ -methylproline (peptide **1B**) and 5,5-dimethylproline (peptide **1C**) at residue 3 were synthesized. The rate constants of their reactions with the PFA probe **2** were measured (Figs S1–S3). We found the  $\alpha$ -methylproline mutation (**1B**) increased the rate constant by a factor of 2.5 relative to  $\pi$ -clamp **1A**, whereas the 5,5-dimethylproline mutation (**1C**) led to a significant loss of reactivity (Fig. 1b), indicating that *trans* Pro promoted the  $\pi$ -clamp-mediated conjugation.

Computational tools were employed to calculate the Gibbs free energy change ( $\Delta G$ ) of  $\pi$ -clamp-mediated arylation reaction, to further study whether a *trans* proline analogue provided any energetic advantage. We

2M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	$\Delta H^\ddagger$ (KJ·mol <sup>-1</sup> )	$-T\Delta S^\ddagger$ (KJ·mol <sup>-1</sup> )	$\Delta G^\ddagger$ (KJ·mol <sup>-1</sup> )	k (M <sup>-1</sup> ·s <sup>-1</sup> )
–	89 ± 6	–11 ± 6	78 ± 8	0.63
+	49 ± 3	18 ± 3	67 ± 4	43.3

**Table 1.** Influence of ammonium sulfate on thermodynamic parameters (310 K) for the arylation reaction between peptide **1 A** and probe **2**.

previously investigated the reaction energy pathway for *cis* Pro  $\pi$ -clamp<sup>12</sup>. To calculate  $\Delta G$  of the arylation reaction between peptide **1 B** and PFA probe using density functional theory (DFT), we first extracted three snapshots from molecular dynamics (MD) simulations of the *trans* proline  $\pi$ -clamp and then manually added a methyl group to form  $\alpha$ MePro. Similarly, for starting structures of peptide **1 C**, we extracted four snapshots from MD simulations of the *cis* proline  $\pi$ -clamp and manually added two methyl groups to form 5,5-dmP. For the product's starting structure, we manually connected the PFA group to the peptide cysteine. After performing geometry optimizations on the starting structures of the peptide and product (Fig. S4),  $\Delta G$  of the arylation reaction was calculated using DFT (see Section 3 in supporting information for details). We found that *trans* proline provided a thermodynamic advantage (lower Gibbs free energy of reaction) over *cis* proline by 11.8 kJ/mol for  $\pi$ -clamp-mediated arylation (Fig. 1c and Table S3). Taken together, both mutation and computational studies show a *trans* prolyl amide bond is preferred in the  $\pi$ -clamp-mediated arylation reaction.

**Enthalpy of activation is decreased in  $\pi$ -clamp-mediated arylation.** The reactivity of  $\pi$ -clamp cysteine was higher than that of other cysteines, leading to the site-selectivity of  $\pi$ -clamp. Compared to the reaction between PFA probe **2** and the  $\pi$ -clamp double-glycine mutant peptide **3** (GCPG, secondary rate constant  $k = 0.00065 \pm 0.00005 \text{ M}^{-1} \text{ s}^{-1}$ , Fig. S5), the rate constant for  $\pi$ -clamp peptide **1 A** ( $0.63 \pm 0.02 \text{ M}^{-1} \text{ s}^{-1}$ ) was enhanced 1000-fold. Ammonium sulfate further accelerated  $\pi$ -clamp mediated arylation without affecting the regioselectivity<sup>13</sup>.

We investigated whether and how the  $\pi$ -clamp cysteine arylation reaction activation parameters were different from other cysteine using Eyring equation. We chose to compare  $\pi$ -clamp peptide **1 A**, double glycine mutant **3**,  $\pi$ -clamp proline analogue mutants **1 B–1 C** and  $\pi$ -clamp mutants with phenylalanine mutated to leucine (**1 D**), cyclohexylalanine (**1 E**) and pyrenylalanine (**1 F**). PFA probe **2** was reacted with these peptides and the reaction rate constants ( $k$ ) were measured at different temperatures ( $T$ ) (Table S4, Figs S6–S36) to calculate the enthalpy of activation ( $\Delta H^\ddagger$ ), the entropy of activation ( $\Delta S^\ddagger$ ), and the Gibbs energy of activation ( $\Delta G^\ddagger$ ) (see Methods Section for details). As summarized in Table S5, the  $\Delta H^\ddagger$  for  $\pi$ -clamp and mutants (**1 A–1 F**) were lowered by 37–84 kJ/mol compared to that of the double glycine mutant **3**, contributing to the significantly enhanced reaction rate. Similarly, to shed light on whether ammonium sulfate changed activation parameters to accelerate the  $\pi$ -clamp-mediated arylation, the rate constant for the reaction between **1 A** and **2** in the presence of two molar (2 M) ammonium sulfate was measured at different temperatures (Figs S38–S42). We found both  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  were changed by ammonium sulfate (Table 1), but the change of  $\Delta H^\ddagger$  is more significant and is responsible for the reaction acceleration.

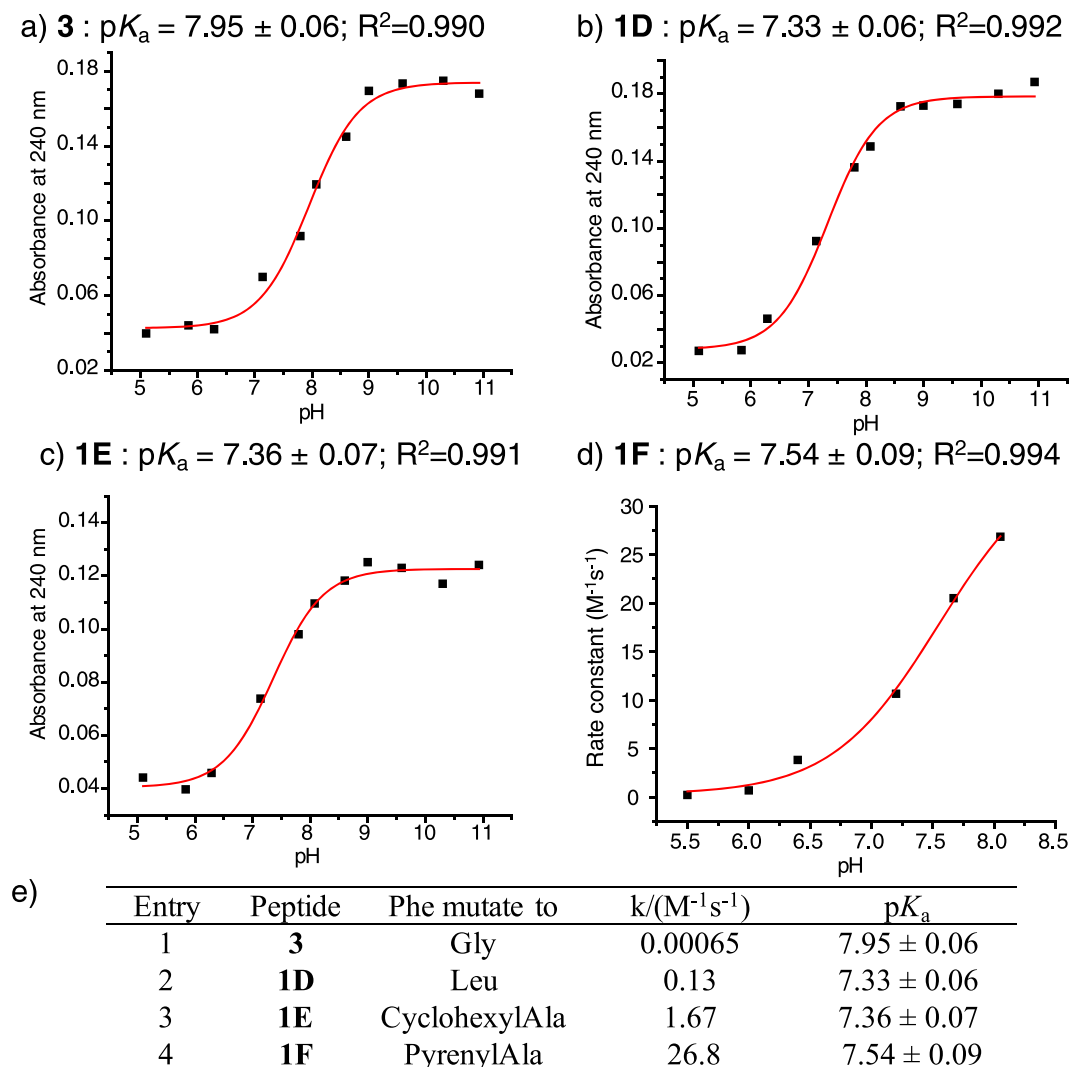
**$\pi$ -clamp cysteine  $pK_a$  is reduced even for Phe mutants.** We have shown that the cysteine  $pK_a$  of  $\pi$ -clamp is lowered by 0.6 pH unit compared to that of the double glycine mutant<sup>12</sup>. To test whether the cysteine  $pK_a$  values in  $\pi$ -clamp mutants are also perturbed and how important the cysteine  $pK_a$  value is for enhancing its reactivity, we measured the cysteine  $pK_a$  in **1 D**, **1 E**, **1 F** and **3**.

Several methods have been developed to measure the cysteine  $pK_a$  in biomolecules. Spectrophotometric titration<sup>20–22</sup>, where UV absorbance of the cysteine containing molecule is plotted against different pH values, is a convenient method. This method is based on the fact that ionization of cysteine thiol to thiolate results in a large increase in its molar absorption coefficient at 240 nm<sup>20</sup>. To calculate the cysteine  $pK_a$  of **1 D**, **1 E** and **3**, we measured the absorbance at 240 nm (A240) of the peptides dissolved in buffers at different pHs. The absorbance of the buffer was used as a blank. Then the  $pK_a$  was determined (Fig. 2a–c) by fitting the UV absorbance and pH values with the following equation:

$$y = \frac{A \times 10^{pH} + B \times 10^{pK_a}}{10^{pK_a} + 10^{pH}}$$

where  $y$  is the absorbance of the peptide at 240 nm,  $A$  is the upper plateau of absorbance at high pH, and  $B$  is the lower plateau at low pH. The cysteine  $pK_a$  in the double glycine mutant **3** was measured to be 7.95, while the cysteine  $pK_a$  of both **1 D** and **1 E** were 0.6 pH unit lower than that of **3**.

The A240 method could not be applied to **1 F** because of the strong absorbance of the pyrenyl group at 240 nm. Instead we measured  $pK_a$  based on the reactivity of cysteine at different pHs. The  $pK_a$  values of cysteine in proteins have been measured by determining their reaction rates with a reagent that is more reactive with thiolate than thiol<sup>22,23</sup>. Here we used the rate constant of the  $S_NAr$  reaction between **1 F** and **2** to fit for  $pK_a$ , since thiolate is more reactive than thiol in the perfluoroarylation reaction. The rate constants were plotted against pH and fit with the same equation as in the A240 method, where  $y$  is the reaction rate constant, and  $A$  and  $B$  are the upper and lower plateaus for the reaction rates (Fig. 2d). The cysteine  $pK_a$  of **1 F** was determined to be 0.4 pH unit lower than **3**. In Fig. 2e, we summarize the cysteine  $pK_a$  values and the rate constants for the peptides studied. We observed a more than 10-fold rate constant difference between **1 D** and **1 E**, even though their  $pK_a$  values were similar. In



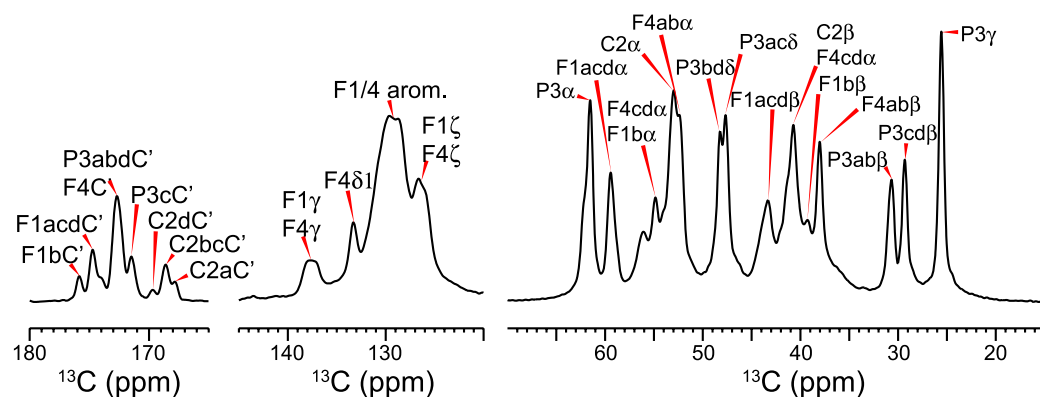
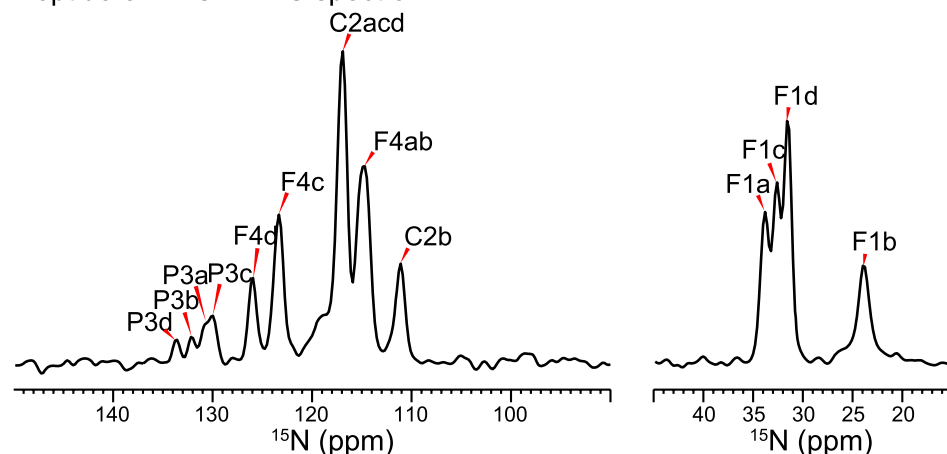
**Figure 2.** The cysteine  $pK_a$  in  $\pi$ -clamp mutants are reduced. Absorbance at 240 nm across different pHs were measured to determine the cysteine  $pK_a$  in peptide **3** (a), **1D** (b) and **1E** (c). (d) Rate constants for the reaction between **1F** and **2** were measured at different pHs to determine cysteine  $pK_a$  in peptide **1F**. (e) Summary of  $pK_a$  and the reaction rates of  $\pi$ -clamp mutants.

addition, peptide **1F** was more reactive than peptides **1D** and **1E**, but the cysteine  $pK_a$  of **1F** was higher. Although all three  $\pi$ -clamp mutants had a lower cysteine  $pK_a$  compared to that of the double glycine mutant **3** which facilitated ionization of the thiol to the more reactive thiolate, this  $pK_a$  effect was not the only factor to account for the enhanced reactivity of the  $\pi$ -clamp.

#### ssNMR indicates $\pi$ -clamp adopts 1:1 *cis/trans* Pro3 while the perfluoroarylated product adopts only *trans*.

To characterize the backbone conformation of the  $\pi$ -clamp and the S-perfluoroarylated product (PFA-tagged  $\pi$ -clamp), **1D** and **2D**  $^{13}C$  magic-angle-spinning (MAS) NMR spectra were measured on  $\pi$ -clamp peptide **5** (Fig. S43) and  $\pi$ -clamp PFA-tagged product **6** (Fig. S44), in which the  $\pi$ -clamp motif Phe-Cys-Pro-Phe were uniformly  $^{13}C$  and  $^{15}N$  labeled. The  $^{13}C$  cross-polarization (CP) spectrum of peptide **6** (Fig. 3a) shows sharp resonances, but with more peaks than expected for four labeled residues, indicating the presence of multiple resolvable peptide conformations. The Pro3 conformational heterogeneity is manifested as two resolved  $C\beta$  peaks around 30 ppm, and two partially resolved  $C\delta$  peaks around 48 ppm. In the carbonyl region, structural polymorphism is also readily observed, especially among the multiple Cys2 CO peaks, ~169 ppm, which are resolved from the carbonyl signals of the other labeled residues. The  $^{15}N$  CP spectrum (Fig. 3b) further highlights this conformational polymorphism: the N-terminal Phe1 amino group shows four resolved  $^{15}N$  signals at 22–35 ppm, indicating four distinct conformations. Likewise, Pro3 exhibits four resolved  $^{15}N$  peaks at 130–135 ppm.

To further resolve and assign the  $^{13}C$  and  $^{15}N$  chemical shifts, we measured 2D  $^{13}C$ - $^{13}C$  correlation spectra (Fig. 4) of peptides **5** and **6**, and 2D  $^{15}N$ - $^{13}C$  correlation spectra (Fig. S45) of peptide **6**. The 2D  $^{13}C$ - $^{13}C$  correlation spectrum of peptide **5** (Fig. 4a) shows larger  $^{13}C$  line widths than the PFA-tagged peptide **6** (Fig. 4b) due to the

a) Peptide 6  $^{13}\text{C}$  CP-MAS spectrab) Peptide 6  $^{15}\text{N}$  CP-MAS spectra

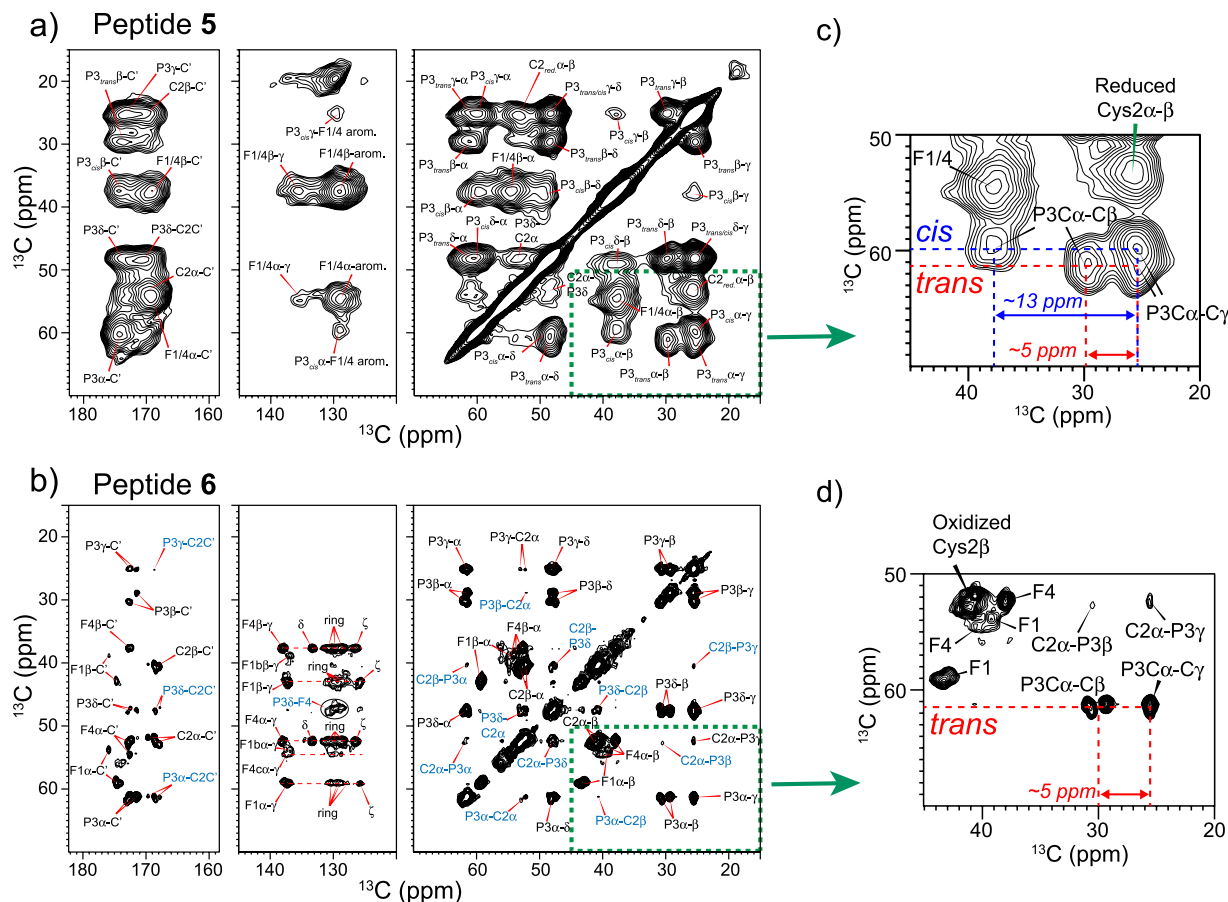
**Figure 3.**  $^{13}\text{C}$  (a) and  $^{15}\text{N}$  (b) cross-polarization magic-angle spinning (CP-MAS) spectra of the PFA-tagged  $\pi$ -clamp peptide 6 show four sets (denoted as (a,b,c,d) after the residue number) of well resolved peaks, indicating conformational polymorphism.

conformational distribution and the dry state of the peptide. Measurement of the dried, unreacted peptide 5 by solid-state NMR was necessary due to the unfavorable dynamics of this very soluble peptide in a hydrated state. The aliphatic linewidths of peptide 5 range from 2.0 ppm to 3.0 ppm, while the PFA-tagged peptide 6 exhibits  $^{13}\text{C}$  linewidths of 0.7–1.4 ppm. Within this linewidth range, peptide 5 shows a single set of chemical shifts for each residue, with the exception of Pro3, which exhibits chemical shifts for both *cis* and *trans* isomers (*vide infra*). The Phe1 and Phe4 chemical shifts are not well resolved: their  $\text{C}\alpha$ – $\text{C}\beta$  cross peaks partially overlap, and a single set of  $\text{C}\beta$  cross peaks with the aromatic side chain carbons are observed (Fig. 4a), indicating that the two Phe residues have similar conformations. Cys2 displays a  $\text{C}\beta$  chemical shift of ~25 ppm (Fig. 4a and c), which corresponds to the reduced state, confirming that the unreacted peptide 5 does not form disulfide bonds.

Upon reaction of peptide 5 with the PFA probe to form product 6, all  $^{13}\text{C}$  and  $^{15}\text{N}$  peaks sharpen (Figs 4b,d and S45), allowing us to resolve four sets of chemical shifts for each residue. In peptide 6, the inter-residue Cys2-Pro3 cross peaks allow the multiple sets of chemical shifts to be separated into four different molecules. No cross peaks between the different forms are observed at a  $^{13}\text{C}$  spin diffusion mixing time of 50 ms. Phe1 and Phe4 exhibit resolved  $\text{C}\alpha$  and  $\text{C}\beta$  chemical shifts that differ by ~5 ppm and ~3 ppm respectively (Fig. 4b). Multiple sets of aromatic  $\text{C}\gamma$ ,  $\text{C}\delta$ ,  $\text{C}\epsilon$  and  $\text{C}\zeta$  chemical shifts are also resolved for Phe1 and Phe4, indicating that Phe1 and Phe4 experience different chemical environments, likely due to the PFA tag. Cys2 shows a  $\text{C}\beta$  chemical shift of ~40 ppm, characteristic of oxidized Cys2, as expected for PFA-modified cysteine.

We were especially interested in the prolyl amide bond conformation. Pro3 shows interesting chemical shift differences between the peptide 5 and the PFA-tagged peptide 6. In the unreacted state (peptide 5), two sets of Pro3  $\text{C}\beta$  and  $\text{C}\gamma$  chemical shifts are observed: one set shows a large  $\text{C}\beta$  and  $\text{C}\gamma$  chemical shift difference of ~13 ppm while the other set shows a smaller chemical shift difference of ~5 ppm (Fig. 4c). These  $\text{C}\beta$  and  $\text{C}\gamma$  chemical shift differences<sup>24</sup> are characteristic of the *cis* and *trans* isomers of Pro, and peak intensities indicate that the two forms have a *cis*:*trans* ratio of 1:1 in the unreacted peptide 5. Upon reaction with the PFA probe, the multiple resolved Pro3 chemical shifts all show similar  $\text{C}\beta$  and  $\text{C}\gamma$  chemical shift differences of ~5 ppm, indicating that the Pro3 in peptide 6 predominantly shifts to the *trans* conformer (Fig. 4d).

Overall, these  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts (Tables S6, S7) indicate there is no single well-defined secondary structure for the  $\pi$ -clamp motif in peptide 5 or 6. The unreacted  $\pi$ -clamp peptide 5 adopts a broadly distributed

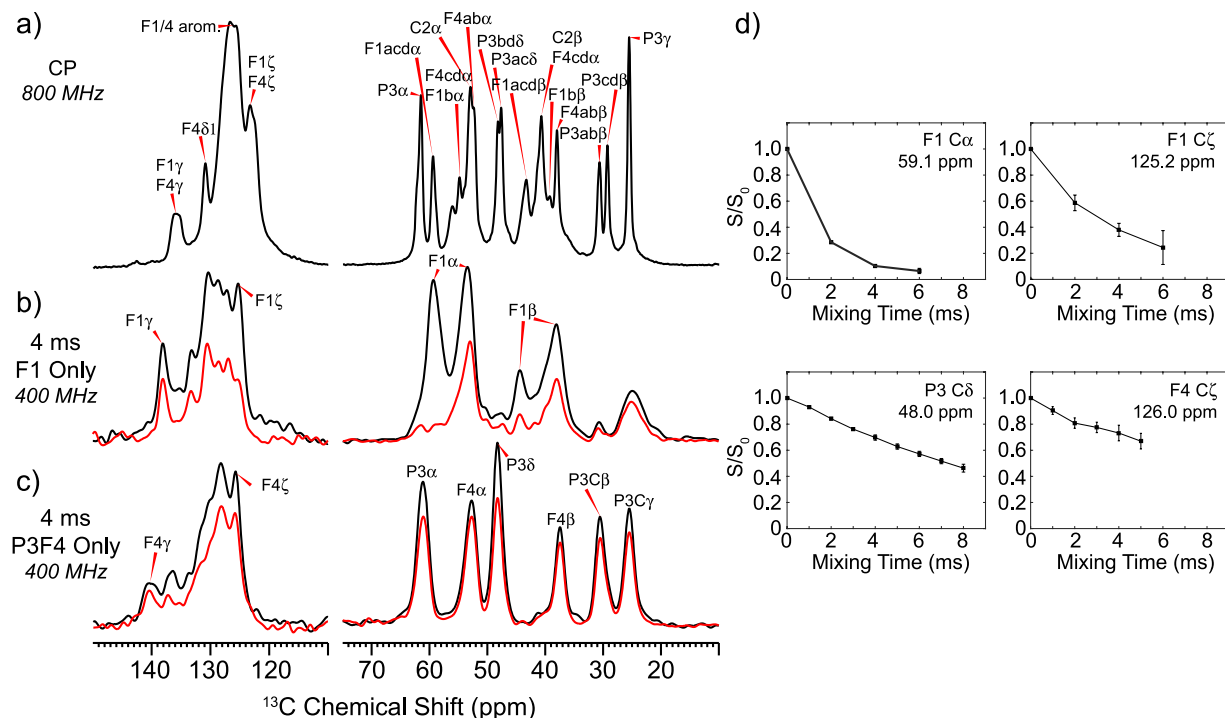


**Figure 4.** 2D  $^{13}\text{C}$ - $^{13}\text{C}$  correlation spectra of the (a) unreacted  $\pi$ -clamp peptide 5 and (b) PFA-tagged peptide 6. The spectra were acquired at 298 K. (c,d) Enlarged regions showing Pro3 and Cys2 cross peaks. The Cys2 C $\beta$  chemical shift indicates that Cys2 is in a reduced state ( $\sim 25$  ppm) in peptide 5 (c), and in an oxidized state ( $\sim 40$  ppm) in the PFA-tagged peptide 6 (d). The chemical shift difference between Pro C $\beta$  and C $\gamma$  indicates the isomer state. The Pro intensity in (c) indicates an about 1:1 ratio between the *cis* and *trans* isomers in the unreacted peptide 5. The Pro signal in (d) shows only the *trans* Pro conformation is adopted in PFA-tagged peptide 6.

conformation with an approximate 1:1 ratio of the *cis* and *trans* Pro3 isomer. Reacting with the PFA probe causes peptide 6 to adopt four distinct conformations, but only the *trans* isomer at Pro3. Using the compiled  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts of peptide 6, a TALOSN<sup>25</sup> prediction yielded similar backbone ( $\phi, \psi$ ) torsion angles for the four forms (Fig. S46), with the largest torsion angle difference occurring at Phe4. This difference is likely from prediction uncertainty due to the lack of chemical shift information of the residue C-terminal to Phe4. For the unreacted peptide 5, because of the large spectra linewidths, there is significant uncertainty in predicting these ( $\phi, \psi$ ) torsion angles. However, since all four distinct sets of peptide 6 chemical shifts fall within the broad linewidths of the unreacted peptide 5, the average backbone conformations may not differ dramatically between the two peptides.

**A clamp-like claw structure is one of the two reaction products from ssNMR and modeling studies.** To generate a structural model for the perfluoroarylated  $\pi$ -clamp product based on the backbone conformation predicted from chemical shifts,  $^{13}\text{C}$ - $^{19}\text{F}$  distances were measured using  $^{13}\text{C}$ - $^{19}\text{F}$  REDOR of two partially labeled  $\pi$ -clamp peptides: peptide 7 with uniform  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeling at only Phe1 (Fig. S47) and peptide 8 with uniform  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeling at Pro3 and Phe4 (Fig. S48). By examining these two peptides, we can determine the distance between the PFA and each Phe. Figure 5 shows representative REDOR  $S_0$  and  $S$  spectra at a mixing time of 4 ms.  $^{13}\text{C}$  spins that are close to the  $^{19}\text{F}$  spins exhibit larger intensity differences between the  $S_0$  and  $S$  spectra than  $^{13}\text{C}$  spins that are farther from  $^{19}\text{F}$  spins. Figure 5d plots the intensity ratio  $S/S_0$  as a function of the REDOR mixing time for a few representative peaks; the full set of REDOR curves used for modelling are summarized in Fig. S49. The Phe1 backbone C $\alpha$  peak has the largest dephasing among all peaks, reaching a value of  $0.07 \pm 0.02$  at 6 ms, while the Phe1 C $\zeta$  shows a more moderate dephasing trend. The Pro3 C $\delta$  and Phe4 C $\zeta$  show linear decay in contrast to the expected universal two-spin REDOR curve. We attribute these decay trajectories to the multi-spin nature of the system, with 8  $^{19}\text{F}$  spins at varying distances from each  $^{13}\text{C}$ .

Given the system complexity, we first analyzed these  $^{13}\text{C}$ - $^{19}\text{F}$  REDOR data semi-quantitatively. In contrast with the Phe1 backbone C $\alpha$  peak which has the largest degree of dephasing, the Phe4 C $\alpha$  and C $\beta$  backbone sites show



**Figure 5.**  $^{13}\text{C}$ - $^{19}\text{F}$  distances are measured using  $^{19}\text{F}$ -dephased and  $^{13}\text{C}$ -detected REDOR spectra of the PFA-tagged  $\pi$ -clamp peptides. Representative spectra were shown. (a)  $^{13}\text{C}$  CP-MAS spectra of the fully labeled peptide **6**, showing the aliphatic (right) and aromatic (left) region. Full intensity ( $S_0$ , black) and dephased ( $S$ , red) REDOR spectra at a REDOR mixing time of 4 ms for the F1-only labeled  $\pi$ -clamp peptide **7** (b) and P3F4-only labeled peptide **8** (c) are shown. (d) Experimental  $S/S_0$  dephasing values for four representative  $^{13}\text{C}$  resonances.

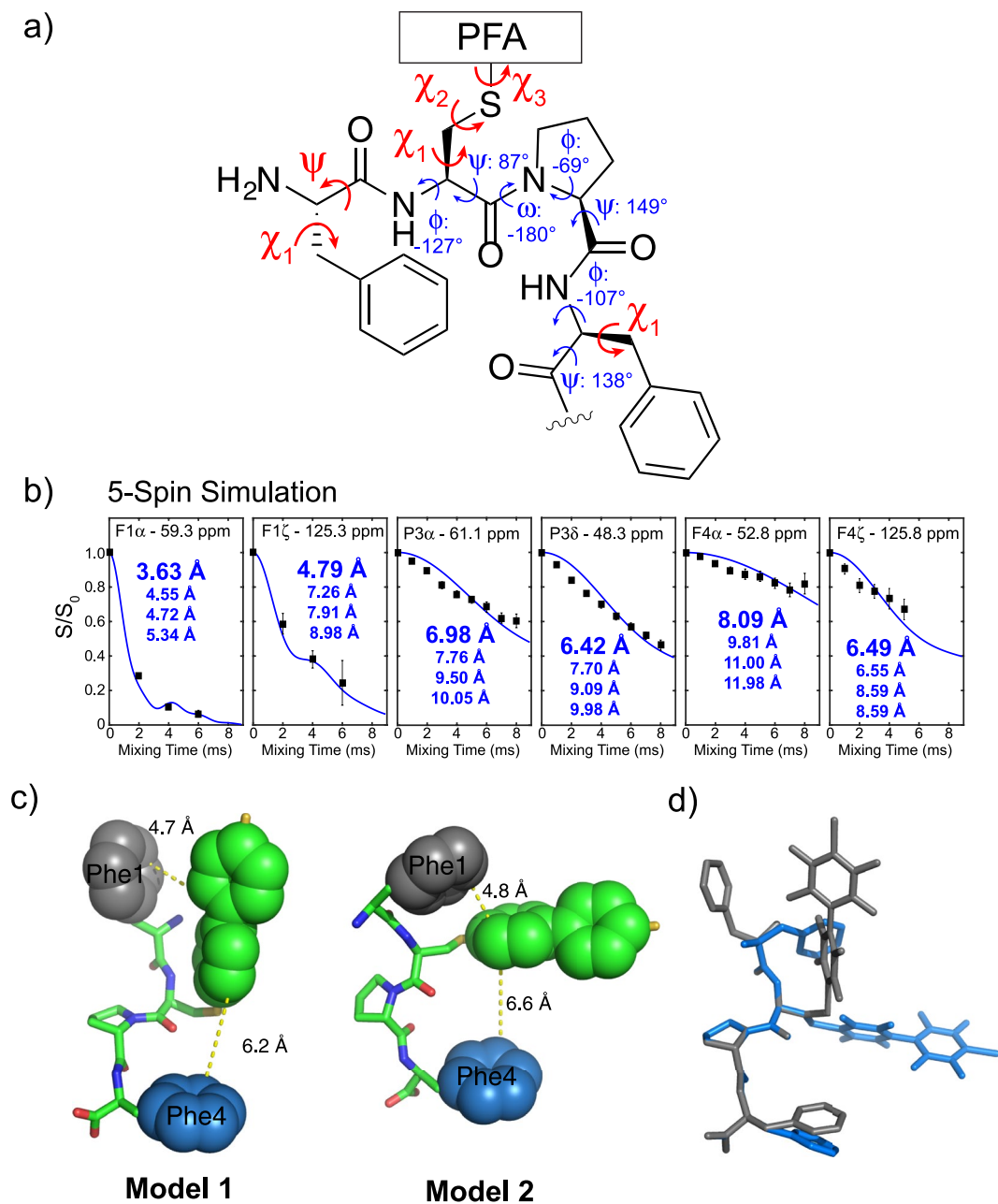
minimal dephasing, reaching only  $S/S_0$  values of  $\sim 0.8$  at 8 ms. This indicates that the PFA tag is closer to Phe1 than Phe4. Among the  $^{13}\text{C}$  signals from Pro3, C $\delta$  reaches the lowest dephasing at a  $S/S_0$  value of  $\sim 0.4$  at 8 ms, while Pro3 C $\alpha$ , C $\beta$  and C $\gamma$  each show minimum  $S/S_0$  of  $\sim 0.6$  at 8 ms, indicating Pro3 C $\delta$  is closer to the PFA tag.

More quantitative distance information was extracted by multi-spin simulations utilizing the SIMPSON program<sup>26</sup>. We compared the experimental data first to 2-spin simulations to obtain the apparent  $^{13}\text{C}$ - $^{19}\text{F}$  distance, and then to 5-spin system containing one  $^{13}\text{C}$  and four  $^{19}\text{F}$  spins in model-dependent simulations. In these simulations, the  $\psi$  and  $\chi_1$  angle of Phe1,  $\chi_1$ ,  $\chi_2$  and  $\chi_3$  of Cys2 and  $\chi_1$  of Phe4 were iteratively set to  $180^\circ$ ,  $60^\circ$  or  $-60^\circ$ . For each simulated curve, the root-mean-square deviation (RMSD) was calculated between each experimental REDOR data set, and the minimum among all calculated RMSD values corresponds to the best fit. The 2-spin simulations provided reasonable fits for each  $^{13}\text{C}$  site (Fig. S50). For all sites in Pro3, the distances range from 6.3–7.0 Å. Slightly longer distances for Phe4 C $\alpha$  and C $\beta$  (7.7 Å and 7.4 Å respectively) are observed, and the distance for Phe4 C $\zeta$  is 6.0 Å. For Phe1, the fast dephasing C $\alpha$  and C $\zeta$  fit relatively well to distances of 3.8 Å and 5.2 Å, respectively. The 5-spin model-dependent simulations unfortunately did not simultaneously fit well for all  $^{13}\text{C}$  sites with any set of dihedral angle combinations. However, taking the best-fit 5-spin REDOR curve for each  $^{13}\text{C}$  site (Figs 6b, S51), the shortest  $^{13}\text{C}$ - $^{19}\text{F}$  distance among the four nearest neighbor  $^{19}\text{F}$  agreed well with the best-fit distances determined from the 2-spin simulations.

These distances were taken as a set of REDOR derived, nearest-neighbor  $^{13}\text{C}$ - $^{19}\text{F}$  distances: 3.6 Å F1 $\alpha$ , 4.6 Å F1 $\beta$ , 6.0 Å F1 $\alpha'$ , 6.0 Å F1 $\beta'$ , 4.8 Å F1 $\zeta$ , 7.0 Å P3 $\alpha$ , 7.5 Å P3 $\beta$ , 7.7 Å P3 $\gamma$ , 6.4 Å P3 $\delta$ , 8.1 Å F4 $\alpha$ , 7.8 Å F4 $\beta$ , and 6.5 Å F4 $\zeta$ . Using a smaller interval of  $30^\circ$ , the dihedral angles were rotated iteratively again and the nearest neighbor  $^{13}\text{C}$ - $^{19}\text{F}$  distance was compared to the above set of distances. In the generated best-fit models, shown in Fig. 6c, model 1 (left) uses F1 C $\alpha$  and C $\beta$ , and model 2 (right, wide clamp-like) uses F1 C $\alpha'$  and C $\beta'$  for fitting. The  $^{13}\text{C}$ - $^{19}\text{F}$  distances from these models were used in 5-spin REDOR simulations, and plotted against the experimental REDOR  $S/S_0$  values (Figs S52, S53) for validation. Model 2 gave decent fits for all  $^{13}\text{C}$  sites while model 1 has worse fits, even though the shortest  $^{13}\text{C}$ - $^{19}\text{F}$  distances are similar in the two models. In both models, distinct non-covalent interactions between Phe1 side chain and perfluoroaryl electrophile at around 4.7 Å distance and similar backbone structures were observed (Fig. 6c and d).

### A pyrenylalanine $\pi$ -clamp mutant non-covalently interacts with the perfluoroaryl electrophile in solution.

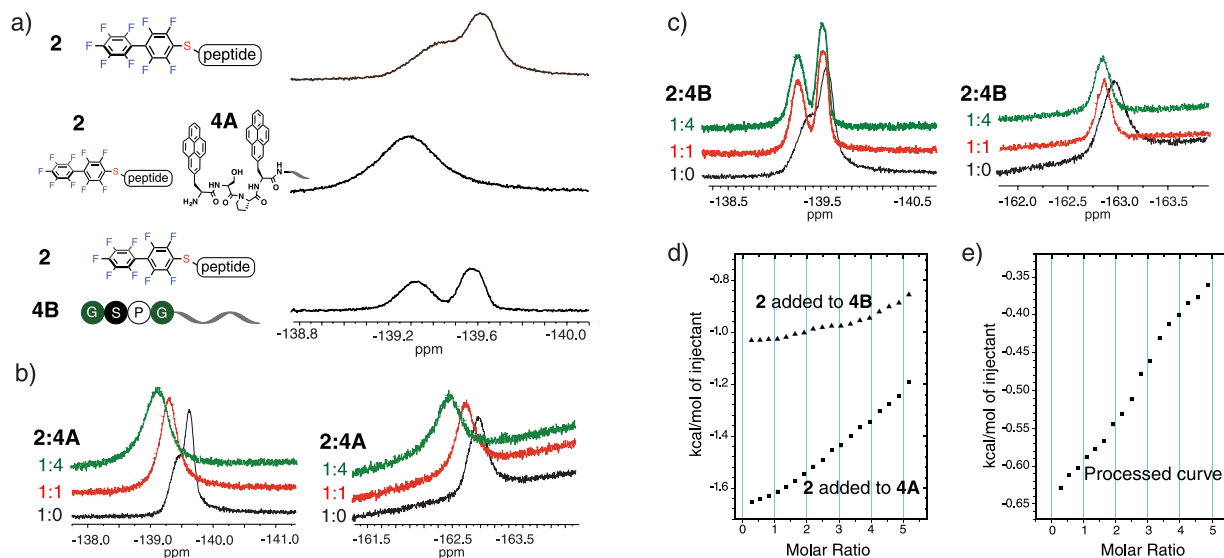
We hypothesized the phenylalanine side chains in  $\pi$ -clamp interact non-covalently with the PFA electrophile to facilitate the reaction.  $^{19}\text{F}$  solution NMR and isothermal titration calorimetry experiments were carried out to test this hypothesis. We started with an un-reactive  $\pi$ -clamp mutant (Phe-Ser-Pro-Phe), but were unable to detect interactions by NMR experiments. Prompted by our earlier experiments, we investigated a



**Figure 6.** Two structural models for PFA-tagged  $\pi$ -clamp show similar backbone structures but distinct Phe1-PFA interaction. (a) The average predicted backbone torsion angles (shown in blue) and the torsion angles varied for fitting (shown in red). (b) Representative model-dependent, best-fit 5-spin SIMPSON simulations of the experimental REDOR curves for the resolved  $^{13}\text{C}$  sites. For each simulation of torsion angle combinations, the four nearest  $^{19}\text{F}$  atoms to a  $^{13}\text{C}$  site were included in the simulation. (c) Structural model of the PFA-tagged  $\pi$ -clamp motif based on the predicted backbone torsion angles and REDOR distance restraints. Model 1 (left) uses Phe1 C $\alpha$  and C $\beta$  for fitting, giving torsion angles of Phe1  $\psi = 90^\circ$  and  $\chi_1 = -60^\circ$ , Cys2  $\chi_1 = 60^\circ$ ,  $\chi_2 = -60^\circ$ , and  $\chi_3 = -30^\circ$ , and Phe4  $\chi_1 = -90^\circ$ . Model 2 (right) uses Phe1 C $\alpha'$  and C $\beta'$  for fitting, giving torsion angles of Phe1  $\psi = -60^\circ$  and  $\chi_1 = 150^\circ$ , Cys2  $\chi_1 = -120^\circ$ ,  $\chi_2 = 180^\circ$ , and  $\chi_3 = 180^\circ$ , and Phe4  $\chi_1 = -120^\circ$ . The shortest carbon-to-carbon distance between Phe side chains and PFA were shown. (d) Superimposition of the two models around Pro3.

pyrenylalanine mutant which may interact with the PFA electrophile more strongly. Peptide **4A** (Py-Ser-Pro-Py, where Py is pyrenylalanine) and the control peptide double-glycine mutant **4B** (Gly-Ser-Pro-Gly) were prepared. We mixed **4A** with 1 equivalent of probe **2** and observed a change in  $^{19}\text{F}$  solution NMR chemical shift and peak shape when compared to the mixture of **4B** and probe **2**, or probe **2** alone (Fig. 7a). In addition, the  $^{19}\text{F}$  chemical shift was changed during titration when we added **4A** to a fixed amount of **2** (Fig. 7b). In contrast, this effect was not observed for the same titration of peptide **4B** (Fig. 7c). These results suggest the pyrenyl side chains are interacting with the PFA electrophile. To corroborate the  $^{19}\text{F}$  solution NMR studies, we investigated the interaction





**Figure 7.** F NMR and ITC indicate interactions between peptide 4A side chain and PFA probe 2. **(a)**  $^{19}\text{F}$  solution NMR spectra for probe 2 (top), 2 mixed with 1 equivalent 4A (middle), and 2 mixed with 1 equivalent 4B (bottom). Conditions: 2 mM peptide or probe, 20 mM phosphate, pH 7, 10%  $\text{D}_2\text{O}$ , 90%  $\text{H}_2\text{O}$ . **(b,c)**  $^{19}\text{F}$  solution NMR spectra for the titration experiment in which the equivalent of peptide 4A **(b)** or 4B **(c)** relative to 2 was increased. Concentration of 2 was fixed at 2 mM. **(d)** Isothermal titration calorimetry binding curves for 2 interacting with 4A and 4B. 5 mM probe 2 was added into 0.2 mM 4A or 4B in the titration experiment. **(e)** Processed ITC binding curve where the interaction between 2 and 4B was deducted as background from the interaction between 2 and 4A.

of 4A and 2 by isothermal titration calorimetry. Probe 2 was added to peptide 4A or 4B to detect non-covalent interaction (Fig. 7d). When compared to the non-pyrenylalanine case 4B as a control, weak binding was observed between 4A and 2. The binding curve (Fig. 7e) did not display simple two state interactions and could not be fit nor could we determine the binding stoichiometry.

## Discussion

The  $\pi$ -clamp accelerates abiotic perfluoroarylation chemistry in water. Here we carried out structural and mechanistic studies to understand this transformation. Several structural and chemical features were found to facilitate the reaction. We found that Pro3 played an important role in the  $\pi$ -clamp function. A  $\pi$ -clamp  $\alpha$ -methylproline mutant favoring the *trans* Pro conformation accelerated the rate of arylation by more than 15-fold, compared to the 5,5-dimethylproline mutant that favored a *cis* Pro conformation. *Trans* Pro provided a thermodynamic advantage over *cis* Pro for  $\pi$ -clamp-mediated arylation as shown by DFT calculation. The  $\Delta H^\ddagger$  for the  $\pi$ -clamp arylation reaction is decreased by 64 kJ/mol when compared to a non- $\pi$ -clamp cysteine, enabling 1000-fold rate enhancement and regioselective cysteine modification at the  $\pi$ -clamp. Ammonium sulfate salt lowers the  $\pi$ -clamp arylation reaction  $\Delta H^\ddagger$  by 40 kJ/mol relative to the reaction without the salt and thus promote the reaction. In addition, the  $\pi$ -clamp cysteine  $\text{pK}_a$  is reduced by 0.4–0.6 pH unit even for Phe mutants.

Structural and biophysical studies indicate certain conformations may promote the reaction, and side-chains in the  $\pi$ -clamp interact non-covalently with the PFA electrophile. Solid-state NMR indicated a mixture of *cis*/*trans* Pro in the unreacted  $\pi$ -clamp was fully converted to the *trans* conformation upon perfluoroarylation. Two ssNMR structural models of the reaction product highlight the interactions between the Phe1 side chain and perfluoroaryl electrophile moiety.  $^{19}\text{F}$  solution NMR titration experiments and the ITC data support the hypothesis that the  $\pi$ -clamp side chains may interact with the PFA electrophile.

The structural models from our ssNMR study are not fully consistent with the  $\pi$ -clamp  $\alpha$ -methylproline mutant product structures used in DFT calculations, although they both have *trans* Pro. As the comparison shown in Fig. S54, the largest torsion angles differences between the calculated structures and clamp-like NMR structure model 2 occurred at Phe1-psi and Pro3-psi. These two angles contribute to the calculated structures being more compact than the NMR structure. In addition, all of the calculated structures have Phe4 in close proximity to the PFA tag. This would theoretically cause the Phe4 C $\alpha$  and C $\beta$  simulated REDOR curves to decay rapidly, while the experimental data shows a much slower decay (Fig. S54). The inconsistency may come from the fact that we did not directly sample the PFA-labeled  $\pi$ -clamp product in MD simulations, due to a lack of parameters to describe the PFA in the reaction product. In addition, two slightly different molecules have been used in our computation and ssNMR studies. We think that the ssNMR models better describe the structural feature of the PFA labeled  $\pi$ -clamp product. However, our DFT calculations which indicate that *trans* Pro is thermodynamically favored over *cis* proline in the  $\pi$ -clamp perfluoroarylation reaction should be valid, considering that for all the geometrically-optimized calculated structures we sampled, larger or similar reaction  $\Delta G$  were observed for *trans* Pro compared to the *cis* Pro structures.

Our studies here have led to the discovery of a  $\pi$ -clamp mutant **1 G** with an 85-fold rate enhancement. A convergent effect was observed when we combined  $\alpha$ -methylproline, which promoted the *trans* Pro conformation, and the large hydrophobic side chain in pyrenylalanine. Taken together, we generated  $\pi$ -clamp mutant **1 G** with a rate constant of  $53.3 \pm 2.3 \text{ M}^{-1} \text{ s}^{-1}$ , which is 85-fold higher  $\pi$ -clamp peptide **1 A** (Figs S55, S56). We anticipate our findings could direct us toward the design and screening of other small reactive peptides to facilitate selective abiotic chemistry enabled by local environment.

## Methods

**Kinetics Study.** The reactions were carried out with 200 mM phosphate, 20 mM TCEP at 37 °C unless otherwise noted. To measure the second order rate constants, reaction mixture was prepared on ice and divided into several 10- $\mu$ L aliquots. All aliquots were immediately put in 37 °C water bath unless otherwise noted. For reactions that takes more than 1 hour to monitor, all aliquots were heated in a PCR machine set at 37 °C to prevent solvent evaporation. Reactions were quenched by addition of 100  $\mu$ L 50% water: 50% acetonitrile: 0.5% TFA at different time points and then subjected to LC-MS analysis. The initial concentration of probe and substrate were known. The second-order rate constants were determined by fitting the following kinetics equation:

$$y = \frac{\ln \frac{[\text{peptide}]_0 [\text{probe}]_t}{[\text{peptide}]_t [\text{probe}]_0}}{[\text{probe}]_0 - [\text{peptide}]_0} = k_2 t$$

Error of reaction rate constant was obtained from the linear fitting of the kinetics curves for measuring the reaction rate constants.

**Determination of the standard enthalpy/entropy of activation.** The secondary rate constant ( $k$ ) for the reaction between  $\pi$ -clamp peptide and probe **2** was experimentally measured at different temperatures ( $T$ ). Then  $\ln(k/T)$  was plotted against  $1/T$ . The standard enthalpy of activation ( $\Delta H^\ddagger$ ) and the standard entropy of activation ( $\Delta S^\ddagger$ ) were calculated (Table S5) by fitting  $\ln(k/T)$  against  $1/T$  with the following Eyring equation:

$$\ln \frac{k}{T} = \ln \frac{\kappa k_B}{h} + \frac{\Delta S^\ddagger}{R} + \frac{-\Delta H^\ddagger}{R T}$$

where  $\kappa$  is transmission coefficient ( $\kappa = 1$ ),  $k_B$  is Boltzmann constant,  $h$  is Planck's constant,  $T$  is absolute temperature and  $R$  is gas constant. The errors for  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  were obtained from the linear fitting.  $\Delta G^\ddagger$  was calculated as  $\Delta H^\ddagger - T\Delta S^\ddagger$ , and the error for  $\Delta G^\ddagger$  was calculated from error propagation. The Eyring plots were summarized in Fig. S37.

## References

- Krall, N., Cruz, F. Pda, Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113 (2016).
- Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 4740 (2014).
- Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
- Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
- Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic labeling of proteins: Techniques and approaches. *Bioconjug. Chem.* **24**, 1277–1294 (2013).
- Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708 (2007).
- Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269 (1998).
- Halo, T. L., Appelbaum, J., Hobert, E. M., Balkin, D. M. & Schepartz, A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc.* **131**, 438–439 (2009).
- Kawakami, T., Ogawa, K., Goshima, N. & Natsume, T. DIVERSE System: *De Novo* Creation of Peptide Tags for Non-enzymatic Covalent Labeling by *in Vitro* Evolution for Protein Imaging Inside Living Cells. *Chem. Biol.* **22**, 1671–1679 (2015).
- Ramil, C. P., An, P., Yu, Z. & Lin, Q. Sequence-specific 2-cyanobenzothiazole ligation. *J. Am. Chem. Soc.* **138**, 5499–5502 (2016).
- Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
- Zhang, C. *et al.*  $\pi$ -Clamp-mediated cysteine conjugation. *Nat. Chem.* **8**, 120–128 (2016).
- Dai, P. *et al.* Salt Effect Accelerates Site-Selective Cysteine Bioconjugation. *ACS Cent. Sci.* **2**, 637–646 (2016).
- Detar, D. F. & Luthra, N. P. Conformations of Proline. *J. Am. Chem. Soc.* **76**, 1232–1244 (1977).
- Wedemeyer, W. J., Welker, E. & Scheraga, H. A. Proline Cis - Trans Isomerization and Protein Folding. *Biochemistry* **41**, 14637–14644 (2002).
- De Poli, M. *et al.* Is the backbone conformation of C $\alpha$ -methyl proline restricted to a single region? *Chem. - A Eur. J.* **15**, 8015–8025 (2009).
- Moretto, A. *et al.* C $\alpha$ -methyl proline: A unique example of split personality. *Biopolymers* **89**, 465–470 (2008).
- An, S. S. A. *et al.* Retention of the cis proline conformation in tripeptide fragments of bovine pancreatic ribonuclease A containing a non-natural proline analogue, 5,5-dimethylproline. *J. Am. Chem. Soc.* **121**, 11558–11566 (1999).
- Cerovský, V., Welker, E. & Scheraga, H. A. A convenient incorporation of conformationally constrained 5,5-dimethylproline into the ribonuclease A 89–124 sequence by condensation of synthetic peptide fragments. *J. Pept. Res.* **61**, 140–151 (2003).
- Donovan, J. W. The Spectrophotometric Titration of the Sulfhydryl and Phenolic Groups of Aldolase. *Biochemistry* **3**, 67–74 (1964).
- Polgár, L. Spectrophotometric determination of mercaptide ion, an activated form of SH-group in thiol enzymes. *FEBS Lett.* **38**, 187–190 (1974).
- Nelson, K. J., Parsonage, D., Hall, A., Karplus, P. A. & Poole, L. B. Cysteine pK Values for the Bacterial Peroxiredoxin AhpC Cysteine pK a Values for the Bacterial Peroxiredoxin AhpC †, ‡. *Biochemistry* **47**, 12860–12868 (2008).
- Lim, J. C. *et al.* A low pKa cysteine at the active site of mouse methionine sulfoxide reductase A. *J. Biol. Chem.* **287**, 25596–25601 (2012).

24. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on  $^{13}\text{C}$  chemical shift statistics. *J. Biomol. NMR* **24**, 149–154 (2002).
25. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
26. Bak, M., Rasmussen, J. T. & Nielsen, N. C. SIMPSON: a general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* **147**, 296–330 (2000).

## Acknowledgements

This work was supported by an MIT Startup fund, the National Institutes of Health (NIH; R01GM110535) and the Sontag Foundation Distinguished Scientist Award (to B.L.P.). C.Z. is a recipient of the George Büchi Research Fellowship, the MIT Koch Graduate Fellowship in Cancer Research, and the Bristol-Myers Squibb Graduate Fellowship in Synthetic Organic Chemistry. The solid-state NMR segment of this work was partially funded by NIH grant GM088204 to M.H. The theory work was funded by a grant from the NSF (CHE-1464804). J.J.S. would like to thank the Royal Commission for the Exhibition of 1851 for a Research Fellowship. The Biophysical Instrumentation Facility for the Study of Complex Macromolecular Systems (NSF-0070319) is gratefully acknowledged.

## Author Contributions

P.D., C.Z. and B.L.P. conceived the work and designed the experiments. P.D. performed the mechanistic and kinetics studies. J.K.W. and M.H. performed the solid-state NMR studies. M.W., J.J.S., T.Z. and T.V.H. performed the computational studies. P.D., J.K.W. and B.L.P. wrote the manuscript, with input from all other authors

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-08402-2](https://doi.org/10.1038/s41598-017-08402-2)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017