-Clamp-mediated cysteine conjugation
π-Clamp Mediated Cysteine Conjugation

Chi Zhang1, Matthew Welborn1, Tianyu Zhu1, Nicole J. Yang2, Michael S. Santos2, Troy Van Voorhis1, and Bradley L. Pentelute1,*

1Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
2Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Abstract

Site-selective functionalization of complex molecules is a grand challenge in chemistry. Protecting groups or catalysts must be used to selectively modify one site among many that are similarly reactive. General strategies are rare such the local chemical environment around the target site is tuned for selective transformation. Here we show a four amino acid sequence (Phe-Cys-Pro-Phe), which we call the “π-clamp”, tunes the reactivity of its cysteine thiol for the site-selective conjugation with perfluoroaromatic reagents. We used the π-clamp to selectively modify one cysteine site in proteins containing multiple endogenous cysteine residues (e.g. antibodies and cysteine-based enzymes), which was impossible with prior cysteine modification methods. The modified π-clamp antibodies retained binding affinity to their targets, enabling the synthesis of site-specific antibody-drug conjugates (ADCs) for selective killing of HER2-positive breast cancer cells. The π-clamp is an unexpected approach for site-selective chemistry and provides opportunities to modify biomolecules for research and therapeutics.

Site-selective chemistry1–5 is essential for creating homogeneously modified biologics6,7, studying protein structure and function8, generating materials with defined composition9, and on-demand modification of complex small molecules10,11. Existing approaches for site-selective chemistry utilize either reaction pairs that are orthogonal to other functional groups or catalysts that mediate selective reactions

Supplementary Information

Materials and Methods
Supplementary Text
Supplementary References and Notes
Tables S1 – S4
Figures S1 – S37
at a particular site among many competing ones (Fig. 1a, strategy 2).\textsuperscript{14-19} These strategies have been widely used in protein modification and have led to the development of multiple bio-orthogonal handles\textsuperscript{20-25} and enzyme-tag pairs\textsuperscript{26-31}.

Natural proteins precisely control selective reactions and interactions by building large three-dimensional structures from polypeptides usually much greater than 100 residues.\textsuperscript{32} For example, enzymes have folded structures where particular amino acids are placed in a specialized active-site environment.\textsuperscript{33} Inspired by this, we envisioned a new strategy for site-selective chemistry on proteins by fine-tuning the local environment around an amino acid residue in a small peptide sequence (Fig. 1b). This is challenging because peptides are highly dynamic and unstructured thereby presenting a formidable challenge to build defined environments for selective chemical transformations.

Our design efforts leveraged cysteine because Nature has shown its robust catalytic role in enzymes,\textsuperscript{34,35} and prior efforts indicate the reactivity of a cysteine residue can vary in different protein environments.\textsuperscript{36} Further, cysteine is the first choice in bioconjugation to modify proteins often via maleimide ligation or alkylation.\textsuperscript{37,38} However, these traditional cysteine-based bioconjugations are significantly limited because they are not site-specific. When these methods are applied to protein targets with multiple cysteine residues, a mixture of heterogeneous products are generated, as exemplified by recent efforts to conjugate small molecule drugs to antibodies through cysteine-based reactions.\textsuperscript{39}

Small peptide tags that contain multiple cysteine residues have been used for bioconjugation. Tsien and co-workers have developed biarsenic reagents that selectively react with tetra-cysteine motifs in peptides and proteins.\textsuperscript{40,41} More recently, organic arsenics have been used to modify two cysteine residues generated from reducing a disulfide bond.\textsuperscript{42} These methods can present challenges with thiol selectivity\textsuperscript{43} and none report the site-specific modification of one cysteine residue in the presence of many as enzymes or multiple chemical steps must be used to accomplish this feat.\textsuperscript{44,45} An enzyme-free and one-step method for site-selective cysteine conjugation has yet to be developed.

We have previously described a perfluoroaryl-cysteine S\textsubscript{N}Ar approach for peptide and protein modifications.\textsuperscript{46-49} The reactions between perfluoroaryl groups and cysteine residues are fast in organic solvent but extremely sluggish in water unless an enzyme is used.\textsuperscript{47,48} This observation inspired us to develop small peptides to promote the S\textsubscript{N}Ar reaction in an analogous fashion to enzymes.

**Results**

Here we describe the identification of the π-clamp sequence to mediate site-specific cysteine modification in water without an enzyme, which overcomes the selectivity challenge for cysteine bioconjugation (Fig. 1c). This offers a fundamentally new mode for site-specific chemistry by fine-tuning the microenvironment of a four-residue stretch within a complex protein or peptide. Through a library selection approach (Fig. S26 in the Supplementary Information), we find the sequence Phe-Cys-Pro-Trp within a polypeptide exhibits enhanced reactivity for a perfluoroaryl electrophilic probe (Fig. S1 in the Supplementary Information).

Nat Chem. Author manuscript; available in PMC 2016 May 09.
via nucleophilic aromatic substitution reaction. This observation is in stark contrast to our prior efforts\textsuperscript{47} which showed that cysteine residues and perfluoroaryl moieties do not react in water. Thus the Phe-Cys-Pro-Trp sequence appears to radically modify the reactivity of the cysteine thiol. Further mutating the Phe and Trp to Gly eliminated the reaction. Based on these findings and a molecular model of Phe-Cys-Pro-Trp, we hypothesize that the Phe and Trp side chains activate the cysteine thiol and interact with the incoming perfluoroaryl group, while the Pro serves to position the Cys, Phe, and Trp residues into a conformation that promotes the reaction. We refer to this distinctive amino acid sequence Xaa-Cys-Pro-Xaa (Xaa = electron-rich aromatic amino acids including Phe, Trp, or Tyr) as a $\pi$-clamp.

To investigate the $\pi$-clamp mediated conjugation, we mutated the aromatic residues. Each of 9 peptides (Xaa-Cys-Pro-Xaa-Gly-Leu-Leu-Lys-Asn-Lys, where Xaa was Phe, Trp, or Tyr) were tested for reaction with a perfluoroaryl-probe (2) in 0.2 M phosphate buffer at pH 8.0 and 37 °C with 20 mM TCEP added as the reducing agent. All 9 peptides reacted with probe 2 (rate constants = 0.05 to 0.73 M$^{-1}$ S$^{-1}$, see Table S2 in the Supplementary Information). In contrast, the double glycine mutant (1A) formed no product (Fig. 2a, entry 1). The Phe-Phe $\pi$-clamp peptide (1E) gave quantitative conversion in 30 minutes (rate constants = 0.73 M$^{-1}$ S$^{-1}$, Fig. 2a, entry 5). Single mutations of each Phe to Gly (1B and 1C, Fig. 2a, entries 2 and 3) or converting the L-Pro to D-Pro (1D, Fig. 2a, entry 4) significantly decreased the rate of the arylation reaction. These studies indicate that each amino acid in the $\pi$-clamp is essential for product formation.

$\pi$-clamp mediated conjugation is highly selective as indicated by our thiol competition experiments. The $\pi$-clamp peptide (1E) was found to undergo quantitative conversion with the perfluoroaryl probe (2) in the presence of a double glycine mutant peptide (1A) that served as the competing thiol species. Only the $\pi$-clamp peptide reacted quantitatively to form conjugated product in 30 minutes (Fig. 2b).

To further investigate the $\pi$-clamp mediated cysteine conjugation, we carried out additional studies to understand if location mattered and the substrate scope. We found that the $\pi$-clamp was efficiently modified irrespective of its position on the polypeptide chain (Fig. 3, Fig. S3 – S6 in the Supplementary Information). $\pi$-clamp at the N-terminus (1E), the C-terminus (1N), and the middle (1O) of the polypeptide chain were readily modified with a diverse set of perfluoroaryl-linked probes including peptide, biotin, fluorescein, alkyne, and polyethylene glycol (2 – 6).

We next investigated the regioselectivity on a 55-kDa protein substrate (Fig. 4a). Model protein 7 was designed to contain an N-terminal cysteine and a $\pi$-terminal $\pi$-clamp. A protease cleavage site was positioned upstream of the $\pi$-clamp thereby allowing for the unequivocal verification of the regioselectivity. Upon reacting the protein (7) with probe 2 for 2 hours, we observed > 95% formation of the mono-labeled product (7A). The N-terminal free cysteine was subsequently labeled with fluorescein-5-maleimide producing the dual-labeled product (7B). Upon protease cleavage, only two products were generated: a protein with maleimide-labeled N-terminal cysteine (7C) and a $\pi$-clamp arylated species, confirming the absolute regioselectivity endowed by the $\pi$-clamp.
Next, we site-specifically modified a cysteine-containing transpeptidase Sortase A (SrtA)\textsuperscript{50} (Fig. 4b). An N-terminal π-clamp SrtA variant (8) reacted with probe 2 to produce > 95% mono-labeled product (8A). The modified variant displayed full catalytic activity (Fig. S10 in the Supplementary Information). No reaction took place with SrtA without the π-clamp (9). In sharp contrast, when the π-clamp-Sortase (8) was reacted with bromoacetamide, a mixture of products was produced with labeling of both cysteine residues (Fig. S9 in the Supplementary Information).

IgG molecules modified with small molecule drugs (antibody-drug conjugates, ADCs) are currently used as therapeutic agents.\textsuperscript{51} However, attaching small molecule agents site-specifically to cysteines in IgGs is as of yet impossible, and thus commercial ADCs are heterogeneous mixtures of conjugates.\textsuperscript{51} Approaches to engineer cysteine substitutions in antibodies produce mixed disulfides with cysteine or glutathione, thus a fine-tuned reduction-oxidation protocol must be used to afford the free cysteine thiols for selective drug conjugation in the presence of disulfide bonds.\textsuperscript{52,53}

We anticipated that the π-clamp IgG could be used to overcome this specificity problem in ADC synthesis, which is notably challenging because IgGs harbor 32 native cysteine residues. The π-clamp mediated modification on antibodies will be a single-step and site-specific antibody-drug conjugation technology that does not require significant antibody engineering or extra chemical steps\textsuperscript{52,53}. To this end, we inserted the Phe-Cys-Pro-Phe sequence into the C-termini of the heavy chains of trastuzumab\textsuperscript{54}. Reacting the π-clamp trastuzumab (protein 10) with either a biotin-perfluoroaryl probe (11-Biotin) or a drug-perfluoroaryl probe (11-MMAF) under reducing conditions, we observed facile formation of the heavy chain mono-labeled products (10-Biotin or 10-MMAF) by LC-MS analysis (Fig. 5a). Antibodies without the π-clamp showed no desired modification under the same conditions (Supplementary Fig. S27), highlighting the specificity of the conjugation. Moreover, this selective conjugation reaction works with other antibodies, reacting a π-clamp C225 antibody\textsuperscript{55,56} with 11-Biotin resulted in only the selective modifications on the π-clamp cysteine residues (Fig. S28), suggesting that the π-clamp could be a general strategy for site-selective antibody modification.

Under the developed reaction conditions (0.2 M phosphate, 20 mM TCEP, pH 8.0, at 37 °C), only the inter-chain disulfides and the π-clamp cysteine residues are reduced (Supplementary Fig. S38), and the modified antibodies retained binding affinity to their targets. Biotin modified π-clamp trastuzumab (10-Biotin) showed similar binding affinity to HER2 (K_D = 0.2 ± 0.2 nM) compared to native trastuzumab non-selectively modified with a (PEG)\textsubscript{4}-Biotin (trastuzumab-(PEG)\textsubscript{4}-Biotin, K_D = 0.3 ± 0.1 nM) (Fig. 5b and Supplementary Fig. S31). In addition, both proteins 10 and 10-Biotin readily bound to BT474 cells (HER2-positive) (Fig. 5c and Supplementary Fig. S32 and S33). As another antibody test case, biotin modified C225 antibody (12-Biotin) showed similar binding to A431 cells (EGFR-positive) compared to the native C225 antibody (Supplementary Fig. S34 and S35). Collectively, insertion of the π-clamp into the heavy chains of antibodies and subsequent modification with drugs or probes did not significantly alter the binding properties.
Using the π-clamp mediated cysteine conjugation, we synthesized a site-specific antibody drug conjugate using π-clamp trastuzumab (protein 10) and a monomethyl auristatin F (MMAF) linked to a perfluoroaryl group (11-MMAF, see Supplementary Information for synthesis). LC-MS analysis of the conjugation reaction showed selective labeling of the heavy chain π-clamp cysteine residues (Fig. 5a). The prepared ADC selectively killed BT474 cells (HER2 positive) but was not effective for CHO cells (HER2 negative), indicating that the observed toxicity is receptor-dependent.

To investigate the mechanism of the π-clamp mediated reaction, we first used molecular dynamics (MD) to sample the conformational arrangements of the π-clamp peptide (1E) (Fig. 6a). Simulations indicated that 1E adopts four primary conformations when a cis-Pro is present: a “π-clamp” (S1) with the phenyl rings of Phe-1 and Phe-4 interacting face-on with the Cys-2 thiol; a “half-clamp” (S2) where only the Phe-4 side chain interacts with the Cys-2 thiol; S3 in which the Phe-1 and Phe-4 side chains are stacked together, leaving the Cys-2 thiol exposed; and an open configuration (S4) where all side chains are too far apart to interact. MD simulation for π-clamp peptide (1E) with a trans-Pro indicated two “open” structures with the cysteine thiol not interacting a Phe residue and one structure with Phe-4 side chain interacting with Cys-2 thiol (see Supplementary Fig. S37).

With these MD structures in hand, we used density functional theory (DFT) to investigate the nucleophilic aromatic substitution energy pathway for structures with a cis-Pro. We found that the half-clamp structure S2 stabilized the arylation product by approximately 5 kcal/mol compared to the double glycine mutant, indicating the important role of Phe-4 in promoting the arylation reaction. This is consistent with our mutation studies showing that Phe-4 alone can partially mediate the arylation reaction (Fig. 2a, Entry 3). The product generated from the open structure (S4) has similar free energy compared to that of the double glycine mutant, further substantiating the hypothesis that the two phenylalanine side chains are important for the arylation reaction with the perfluoroaryl groups.

The most stable product was observed with the “π-clamp” structure (S1) of which the free energy was approximately 7 kcal/mol lower than that of the double glycine mutant. We further found that the activation energy for the formation of the transition state57 (III in Fig. 6b) was decreased by approximately 3 kcal/mol when the π-clamp (S1) was present (see further discussion in the Supplementary Information), presumably because of the phenyl rings recognizing the perfluoroaryl group and activating cysteine sulfur before conjugation. Collectively, these DFT calculations indicated that the π-clamp offers both a kinetic advantage (lower activation energy) and a thermodynamic advantage (lower free energy) over the double glycine mutant for the selective reaction with the perfluoroaryl reagents.

**Discussion and Conclusion**

Here we describe the discovery of a π-clamp to mediate site-selective cysteine conjugation. The π-clamp is composed of natural amino acids and shares some essential features of large enzymes, yet it mediates a purely abiotic cysteine perfluoroarylation reaction. The π-clamp tunes the reactivity of a cysteine thiol in its “active-site”, recognizes the perfluoroaromatic reaction partner, and decreases the activation energy for the reaction. In addition, the π-
clamp has practical applications in protein labeling.\(^4\) The reported reaction is site-specific, operational under physiologically relevant conditions, enzyme-free, and as efficient as the commonly used azide-alkyne click chemistry\(^{58,59}\) (\(\pi\)-clamp rate constant: 0.73 M\(^{-1}\)•S\(^{-1}\)).

Compared to existing bioconjugation techniques\(^{38}\), the advantages of the \(\pi\)-clamp include its (1) small size that offers minimal structural perturbation to the target protein; (2) genetic encodability for straightforward incorporation; (3) ability to perform protecting-group-free dual cysteine modification; (4) and reaction mode that tunes the kinetic parameters to favor the cysteine perfluoroarylation reaction. This mode of reaction is distinct when compared to other advanced cysteine bioconjugations that use entropy to favor conjugation.\(^{40-42}\)

The unexpected mode of site-specificity provided by the \(\pi\)-clamp requires further mention. In all existing conjugation methods\(^{38}\), selectivity results from the judicious choice of certain functional groups so that each reaction pair undergoes conjugation in the presence of many other potentially reactive groups. For example, the unnatural handles used for click reactions are orthogonal to other functional groups on the target of interest.\(^{12}\) In contrast, selectivity in the \(\pi\)-clamp mediated conjugation is achieved by fine-tuning the local chemical environment and reactivity as proteins do. This provides a complementary strategy to non-natural amino acid-mediated bioconjugation.\(^{60}\) By fine-tuning the peptide microenvironment to allow for selective modification, the \(\pi\)-clamp significantly expands the chemistry available for selectively tailoring biomolecules.

**Acknowledgments**

We thank Prof. Steve Buchwald (MIT) and Prof. K. Dane Wittrup (MIT) for comments and suggestions on the manuscript. This work was generously supported by an MIT start-up fund, NIH (R01GM110535) and the Sontag Foundation Distinguished Scientist Award for B.L.P. C.Z. is a recipient of the George Buchi Research Fellowship and the Koch Graduate Fellowship in Cancer Research of MIT. M.W. is an NSF Graduate Research Fellow. We thank the Biological Instrument Facility of MIT for providing the Octet BioLayer Interferometry System (NIF S10 OD016326), the MIT CEHS Instrument Facility for providing flow cytometer (P30-ES002109), and Prof. R. John Collier (Harvard) for contributing select laboratory equipment used in this study.

**References and Notes**


a. Existing strategies for site-selective chemistry

**Strategy 1: orthogonal chemistry**

\[ X \xrightarrow{Z} Y \]

**Strategy 2: catalyst recognition**

\[ X \xrightarrow{Y} XY \]

b. This work: site-selective chemistry by tuning the local environment

\[ X \xrightarrow{\pi\text{-clamp}} Y \]

c. \( \pi\)-Clamp

\[ \text{competing cys} \xrightarrow{\text{protein}} \text{SH} \]

\[ \text{\( \pi\)-clamp} \]

\[ \text{R} - \text{SH} \]

Figure 1. \( \pi\)-clamp mediated cysteine conjugation as a new strategy for site-selective chemistry

**a.** Existing strategies for site-selective chemistry. Strategy 1: selectivity arises from orthogonal chemistry between site Z and reagent Y. Strategy 2: catalyst mediates the reaction between a particular site X (highlighted in red) and reagent Y. **b.** This work demonstrates a new strategy for site-selective chemistry by fine-tuning the local chemical environment around the target site. A particular site X (highlighted in red) is tuned to react with reagent Y in the presence of other competing X sites. **c.** Cysteine residue inside the \( \pi\)-clamp selectively reacts with perfluoroaromatic probes in the presence of other competing cysteine residues and thiol species.
Figure 2. π-clamp mediated cysteine conjugation on peptides

a. Mutation studies show Phe-1, Pro-3, and Phe-4 are required for the observed reactivity. TCEP: tris(2-carboxylethyl) phosphine. Yields shown are from LC-MS analysis of the crude reactions at 30 minutes. b. Site-specific conjugation at the π-clamp in the presence of another competing cysteine peptide. Chromatograms shown are total ion currents (TIC) from LC-MS analysis of crude reaction mixtures at 0 minute (black) and 30 minutes (red). The mass spectrum of product 2E is shown as the inset.

Nat Chem. Author manuscript; available in PMC 2016 May 09.
Figure 3. π-clamp functions at distinct positions in polypeptides and is compatible with diverse perfluoroaryl-based probes

π-clamp at the N-terminus, the C-terminus, and the middle of peptides were readily reacted with perfluoroaryl probes bearing peptide molecule, affinity tag (biotin), fluorescent reporter (fluorescein isothiocyanate, FITC), click chemistry handle (alkyne), and polymer (polyethylene glycol, PEG). Yields shown are from LC-MS analysis of the crude reactions at 60 minutes. *Yields at 120 minutes. See Supplementary Information for LC-MS chromatograms.
Figure 4. π-clamp mediated site-specific conjugation on proteins with multiple cysteines

a, Protecting group-free one-pot dual labeling of a 55-kDa protein. The protein used was a fusion protein of the anthrax toxin lethal factor 1–263 (LFN) and diphtheria toxin domain A (DTA). Reaction conditions: (1) 50 μM 7, 1 mM 2, 0.2 M phosphate, 20 mM TCEP, 37 °C, 2 hours. (2) 50 μM 7A, 1 mM fluorescein-5-maleimide, 0.2 M phosphate pH 7.0, room temperature, 10 minutes. (3) 25 μM protein 7B, 0.1 mg/mL TEV protease, 50 mM Tris, 0.1 mM EDTA, 1 mM DTT, pH 8.0, room temperature, 15 hours. TEV: tobacco itch virus; EDTA: ethylenediaminetetraacetic acid; DTT: dithiothreitol; Tris: 2-amino-2-hydroxymethyl-propane-1, 3-diol. b, Left, quantitative and selective labeling of π-clamp SrtA (PDB entry: 1T2P); right, control shows no labeling of SrtA. Reactions conditions: 38 μM 8 or 9, 1 mM 2, 0.2 M phosphate, 20 mM TCEP, 37 °C, 6 hours.
Figure 5. \(\pi\)-clamp mediated site-specific antibody conjugation

a, site-specific conjugation of biotin or monomethyl auristatin F (MMAF) to \(\pi\)-clamp trastuzumab (protein 10). LC-MS analysis showed site-specific labeling of the \(\pi\)-clamp cysteine residues on the trastuzumab heavy chain. The antibodies were treated with PNGase F to remove the N-linked glycans before LC-MS analysis. Reaction conditions for biotin conjugation: 100 \(\mu\)M \(10\), 1 mM \(11\)-Biotin, 0.2 M phosphate, 20 mM TCEP, 37 °C, 4 hours. Reaction conditions for MMAF conjugation: 100 \(\mu\)M \(10\), 1 mM \(11\)-MMAF, 0.2 M phosphate, 20 mM TCEP, 5%DMSO, 37 °C, 16 hours.

b, the biotin-conjugated \(\pi\)-clamp trastuzumab (10-Biotin) binds to HER2 in the Octet binding assay (K\(_D\) = 0.2 ± 0.2 nM). 10-Biotin was immobilized on streptavidin tips and was sampled with serially diluted concentrations of recombinant HER2 (concentrations of HER2 in each experiment are shown next to the curve, and see Supplementary Information for data analysis and fitting).

c, 10-Biotin retained binding to BT474 cells (HER2 positive). Cells were treated with 10-Biotin or controls, washed with phosphate buffer saline (PBS) with 0.1% BSA, and then treated with streptavidin-AlexaFluor-647 before analyzed by flow cytometer. 10-(PEG)4-
Biotin and trastuzumab-(PEG)4-Biotin were prepared from reacting Biotin-(PEG)4-NHS with protein \textbf{10} or trastuzumab, respectively (see Supplementary Information for details). \textbf{d}, \textbf{10-MMAF} killed BT474 cells (HER2 positive) but was not effective for CHO cells (HER2 negative). EC$_{50}$ values for BT474 cells were 0.19 nM for \textbf{10-MMAF} and 41 nM for auristatin F. EC$_{50}$ value of auristatin F for CHO cells is 1.3 μM. Cells were seeded in 96-well white opaque plate at a density of 5 x 10$^3$/well (CHO) or 10 x 10$^3$/well (BT474). Cells were allowed to attach for 24 hours at 37 °C and 5% CO$_2$ in humidified atmosphere. Cells were then treated with serial dilutions of auristatin F, \textbf{10-MMAF}, or \textbf{10} for 96 hours (BT474) or 72 hours (CHO, treatment time was shorten to prevent overgrowth). Cell viability was quantified using CellTiter Glo assay and was normalized to cell only. Experiments were done in triplicate for each dose.
Figure 6. Structure and mechanism of the $\pi$-clamp

a. Four primary structures S1 – S4 were identified from MD simulation of $\pi$-clamp peptide 1E. The phenyl rings and cysteine thiol are shown as spheres; the rest of the peptide is drawn as sticks. b. Conjugation to the $\pi$-clamp is energetically favored compared to the double glycine mutant. Left, proposed nucleophilic aromatic substitution pathway for arylation at the $\pi$-clamp. Right, computed geometries and free energy surface of the nucleophilic aromatic substitution at the $\pi$-clamp (red). The free energy surface of the double glycine control is also shown (grey).