Chapter 7

Chemoselective Ligations via Pd(0) Chemistry on Unnatural Amino Acids Incorporated into Proteins
**Introduction**

Specific non-covalent associations between biomolecules are the central processes that drive physiology. The considerable structural diversity inherent in macromolecules allows for the recognition of any given structure, a principle central to the immune system [1]. Antibodies exhibit the ability to find and bind specifically to their targets in complex mixtures of molecules that have very similar properties, an attribute that has proven invaluable for many of clinical applications [2-8]. Efforts towards chemoselective ligation are in effect attempts to emulate such specificity utilizing small molecules that will form a stable covalent bond with the target [9]. In such cases specificity derives not from a cumulative effect of point contacts, but from a reactive group that will selectively modify a complementary chemical moiety. Ideally these reactions should be amenable to complex media and should run in completely aqueous environments at near neutral pH and ambient temperatures. Additionally, reactions should proceed without perturbation of biomolecular structure. There have been several successful examples of such chemistries (Figure 7.1)[10-14]. However, only native chemical ligation, entry f, uses existing protein functionality [15, 16]. The modified Staudinger ligation [17-19], entry g, has been shown to utilize introduced non-natural moieties [20-22]. Our ability to introduce unnatural amino acids allows many of these reactions to be performed on biosynthesized proteins [21]. Introduction of aryl halides, as in *para*-bromophenylalanine (pBrPhe) [23] and *para*-iodophenylalanine (pIPhe) [24], as well as terminally unsaturated unnatural amino acids, homoallylglycine (HAG) [25, 26], homopropargylglycine (HPG) and *para*-ethynylphenylalanine (pCCHPhe) [24],
Figure 7.1. Chemoselective ligation reactions (a-c) represent attack of a “hypernucleophile” on a ketone aldehyde (d) attack of a β-aminothiol on an aldehyde to form a thiozolidinone (e) formation of a thioester via attack of a thioacid on [\(-\)halocarbonyl (f) native chemical ligation (g) modified staudinger ligation (h) Cu(I) mediated 3+2 electrocyclization.
provide access to a wide regime of potentially chemoselective ligations, the family of reactions known as palladium catalyzed cross-couplings [27-30]. These chemistries allow the formation of carbon-carbon [31, 32], carbon-nitrogen and carbon-oxygen bonds between arylhalides and a wide variety of coupling partners [33]. Further, palladium catalyzed cyclizations, annulations and cascades can provide complex structures from simple components [34-38]. The creation carbon-carbon bonds, unusual for chemoselective reactions, not only provides excellent stability but also allows extension of conjugation, which can introduce interesting physical or photophysical properties.

We chose to focus specifically on the couplings of aryl halides with terminal acetylenes and terminal alkenes because we have the ability to introduce these functionalities into proteins. Terminal acetylenic-arylhalide couplings in the presence of a palladium source are generally termed Castro-Stevens or Sonagashira couplings; the latter is defined by the addition of a copper co-catalyst [30]. Oxidative insertion of palladium into the aryl halide bond provides an activated substrate for attack by the in situ formed copper acetylide (Figure 7.2), essentially a transmetallation for the terminal acetylene [39]. Reductive elimination of the palladium regenerates the catalytic Pd(0) species as well as providing the product. Heck reactions, terminal alkene-aryl halide couplings, proceed in similar fashion. Oxidative insertion proceeds by exactly the same mechanism. Coordination of the terminal alkene is followed by insertion of the aryl palladium species (Figure 7.2). Hydride elimination releases the catalytic species and installs the olefin into the product. Both of these reactions have been explored extensively in organic settings and, to a more
Figure 7.2. Simplified Sonagashira and Heck catalytic cycles

**Sonagahira coupling**

Pd$^0$ or Pd$^{II}$ precatalyst

Oxidative Insertion

Reductive Elimination

Transmetallation

**Heck coupling**

Pd$^{II}$ or Pd$^{II}$ precatalyst

Oxidative Insertion

Olefin Insertion

**Hydride Elimination**
limited extent, in aqueous media [40-43]. Aqueous conditions tend to accelerate these reactions but the response to salts and buffers is pronounced and poorly understood [39, 43]. These two reactions, typical of palladium cross couplings, can install and, perhaps more importantly, tolerate a wide range of chemical functionalities [44]. However, no studies have specifically explored either the Heck or Sonagashira couplings response to the full range of chemical moieties found in proteins. Additionally, there is little data on the effects of Pd(0) species on the functionality inherent in proteins.

In this study, we report the characterization of both the Sonagashira and Heck couplings as routes to modify proteins bearing unnatural amino acids with aryl halides or terminally unsaturated moieties. We develop model systems to optimize these couplings under conditions amenable to proteins. We also explore the individual effect of the natural amino acids on these reactions as well as look for any unwanted side reactions with their side chain functionality. Further, we demonstrate the utility of this reaction to modify murine dihydrofolate reductase (mDHFR) with a version of the FLAG epitope. Finally, we incorporate pIPhe into Barstar for the purpose fluorescent labeling of this protein. Importantly, we attempt to show selective modification of the exposed pIPhe, indicating that this chemistry does not perturb structure.
Materials and Methods

Triphenylphosphine trisulfonate (TPPTS) and triphenylphosphine monosulfonate (TPPMS) were purchased from Strem and stored under argon. para-Iodophenylalanine was purchased from Chem-Impex. All other chemicals were purchased from Aldrich. All chemicals were used as received without further purification.

Model Reactions

Model Heck reactions were carried out with acrylic acid in the presence of N-acetyl-para-bromophenylalanine under a variety of conditions to effect crude optimization. Sonagashira reactions were pursued under the same conditions, but with propargylalcohol in place of acrylic acid. In general reactions were run from 30-80°C, in either water or mixed acetonitrile/water. Palladium acetate was used as a palladium source and either TPPTS or TPPMS or triphenylphosphine served as primary ligands. Triethylamine, potassium carbonate and potassium acetate were used as general bases. All reaction mixtures were deoxygenated by the freeze pump thaw method and stirred under argon. Quenching was accomplished by adding excess acetic acid. The resulting mixture was extracted with ethyl acetate (2x), which was pooled and dried over sodium sulfate. The resultant liquor was dried and was composed of exclusively the N-acetylated starting material and product (recovery typically >90%), allowing facile evaluation by $^1$H NMR spectroscopy.

Heck product:
**1H NMR (D$_2$O):**
$\delta$ 1.74 (s, 3H, NH-CO-CH$_3$), 2.76 (dd, J = 9.1, 13.8, 1H, CH-CHH-aryl), 3.05 (dd, J = 4.7, 14, 1H, CH-CHH-aryl), 4.28 (dd, J = 4.7, 9, 1H, CH-CHH-aryl), 6.33 (d, J = 16, 1H, aryl-CH-CH-CO$_2$H), 7.13 (d, J = 8.2, 2H, aryl-H$_A$), 7.21 (d, J = 16.1, 1H, aryl-CH-CH-CO$_2$H), 7.38 (d, J = 8.2, 2H, aryl-H$_B$).

ESI (negative mode) 276 (C$_{12}$H$_{13}$NO$_4$ -H$^+$ requires 275.9)

**Sonagashira product:**
$\delta$ 1.62 (s, 3H, NH-CO-CH$_3$), 2.60 (dd, J = 9.0, 13.5, 1H, CH-CHH-aryl), 2.89 (dd, J = 4.8, 13.7, 1H, CH-CHH-aryl), 3.98 (s, 2H, HO-CH$_2$-C-C-aryl),

4.16 (dd, J = 4.7, 9.1, 1H, aryl-CH-CH-CO$_2$H), 6.88 (d, J = 8.1, 2H, aryl-H$_A$), 7.14 (d, J = 8.1, aryl-H$_B$).

**Interference experiments**

Interference experiments were run under optimized conditions for each both the Heck and the Sonagashira type reactions. For the Heck reaction N-acetyl-para-bromophenylalanine (50 mg, 0.175 mmol), acrylic acid (18 mL, 19 mg, 0.262 mmol), potassium carbonate (71 mg, 0.525 mmol), palladium acetate (1 mg, 4.37 μmol), TPPTS (10 mg, 17.5 μmol) and one of the N-acetyl amino acids (0.175 mmol) was added to 350 mL of water deoxygenated by freeze pump thaw method and allowed to stir under argon at 50°C overnight. For the Sonagashira reaction N-acetyl-para-bromophenylalanine (50 mg, 0.175 mmol), propargyl alcohol (16 mL, 15 mg, 0.262mmol), potassium carbonate (71 mg, 0.525 mmol), palladium acetate (1 mg, 4.37 μmol), TPPTS (10 mg, 17.5 μmol), copper iodide (2 mg, 10 μmol) and one of
the N-acetyl amino acids (0.175 mmol) was added to 350 μL of water deoxygenated by freeze pump thaw method and allowed to stir under argon at 50°C overnight. Workups were identical to that above. The amino acids used for testing the sensitivity of these chemistries excluded the amino acids with purely hydrocarbon side chains. Reactions were run in the presence of arginine, asparagine, cysteine, glutamine, histidine, lysine, serine, tryptophan and tyrosine individually. Quenching was achieved by adding excess acetic acid. Extraction of the subsequent mixture with ethyl acetate (3x), followed by drying over sodium sulfate and removal of solvent provided clean starting material and/or product, depending upon the success of the reaction. Evaluation of extent of reaction was performed by $^1$H in CD$_3$OD.

**Production of DHFR-pCCHPhe, HAG or HPG**

Expression experiments in 10 ml cultures were performed to produce murine dihydrofolate reductase containing *para*-ethynylphenylalanine (pCCHPhe), homoallylglycine (HAG) and homopropargylglycine (HPG). AF-IQ[pQE15-A294G], as described in Chapter 2, was utilized for incorporation of pCCHPhe. Cultures of CAG18491/pREP4/pQE15 [26], were used for the incorporation of HAG and HPG. Briefly, M9 minimal medium (50 mL) supplemented with 0.2 % glucose, 1 mg/L thiamine, 1 mM MgSO$_4$, 0.1 mM CaCl$_2$, 19 amino acids (at 20 mg/L), antibiotics (ampicillin 200 mg/L, chloramphenicol 35 mg/L) and phenylalanine (in the case of pCCHPhe) or methionine (in the case of HAG and HPG) (at 20 mg/L) was inoculated with 1 mL of an overnight culture of the expression strain. When the optical density at 600 nm reached 0.8-1.0, a medium shift was performed. Cells were sedimented by
centrifugation for 15 min at 3100g at 4 °C, the supernatant was removed and the cell pellets were washed twice with 0.9% NaCl. Cells were resuspended in supplemented M9 medium containing 250 mg/L of the analog of choice. Protein expression was induced 10 min after the medium shift by addition of isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 1 mM. Cells were cultured for 4 hours post-induction and protein expression was monitored by SDS polyacrylamide gel electrophoresis (PAGE, 12%). Proteins were purified by nickel chelation columns (promega) according to the manufacturers protocol, subsequently dialyzed and lyophilized.

**Modification of DHFR (pCCHPhe, HAG or HPG) with pIF-FLAG tag under denaturing conditions**

Purified and lyophilized samples of mDHFR-pCCHPhe, mDHFR-HPG and mDHFR-HAG were dissolved at a concentration of 1.5 mg/mL in coupling buffer (8M urea; 0.1M NaH₂PO₄; 0.01M Tris•Cl; pH 9.7). Each these solutions (16 µL) was mixed with a pIF-FLAG solution in coupling buffer (8.1 mg/mL, 16 µL). To each solution 1 µL of preformed catalyst solution (see below) was then added and the mixture was deoxygenated by three freeze pump thaw cycles. Catalyst solution was generated by freeze pump thaw deoxygenation of a 10 mM palladium acetate and 40 mM TPPTS solution in deionized water. Subsequent to deoxygenation the coupling solutions, buffer, protein, catalyst and pIF-FLAG, were mixed thoroughly by vortex and allowed to incubate at 50°C overnight.
Gel electrophoresis and western blot analysis

Tricine gels were run directly on the reaction solutions using standard protocol. Western blotting was performed using a two antibody system. Tricine gels were put into TBST (0.15M NaCl, 0.05M Tris, 0.05% Tween) for about 30 minutes before doing the transfer to nitrocellulose. The membranes were then blocked with a solution of 5% dry non-fat milk in TBST for approximately one hour. The membrane was then placed in a solution of anti-FLAG M2 monoclonal antibody (Sigma, 1:10000) and 5% milk in TBST and gently agitated for one hour. The membranes were then washed with 5% milk in TBST and subsequently added to a solution of HRP-Sheep anti MouseIg (Amersham, 1:10000) in 5% milk TBST for one hour. Subsequently the membranes were washed and developed using ECL Western detection kit (Amersham-Pharmacia).

Lissamine rhodamine propargyl sulfonamide

An excess of propargyl amine (100 μL, 119 mg, 2.2 mmol) is added to a solution of lissamine rhodamine sulfonyl chloride (50 mg, 86 μmol) in dry pyridine (0.5 mL). The solution was allowed to sit at room temperature for 30 minutes at which point it was added dropwise to a stirring solution of diethyl ether. The resulting red precipitate was filtered from the mixture, dissolved in methanol and reprecipitated into diethyl ether (2x). The resulting dark red precipitate was dried under vacuum (46 mg, 92%)
Production of Barstar-pIF

An expression strain containing a plasmid encoding PheRS* (Chapter 1) and the target protein Barstar, termed AF-IQ[pQE60B*-A294G], was obtained from Kent Kishenbaum. Expression of Barstar and Barstar-pIF was accomplished following the same protocol established for the production of mDHFR-pCCH. Purification was performed as described [45]. Incorporation of pIF was established by tryptic digest coupled MALDI-TOF (data not shown) and purity was assessed by tricine gel electrophoresis.

Fluorescent modification of Barstar-pIF

A solution of Barstar (0.86 mM), lissamine rhodamine propargylsulfonamide (4 mM) and catalyst solution (0.7 mM) (same as above) in sodium phosphate buffer (20mM, pH=8) was deoxygenated by freeze pump thaw method and allowed to react overnight under argon at 37°C. Control reactions performed on denatured barstar were done in the presence of 4M urea. Modification was analyzed by 15% tris/tricine PAGE, followed by visualization using a UV lightbox. Proteolytic digests of the modified barstar were used to show a difference in labeling between the folded and denatured conditions used. Chymotryptic digests were performed by adding 10 μL of reaction solution to 90 μL of 50mM ammonium bicarbonate with 2 μL of a 0.1 ng/μL
solution of chymotrypsin (Promega). Lys-C digestion was performed by adding 10 μL of reaction solution to 90 μL of NH₄HCO₃ and 0.5 μL of Lys-C (0.1ng/μL in 0.1mM HOAc). The proteolytic fragments were visualized using a 20% tris/tricine PAGE gel.

**Results and Discussion**

*Optimization of Heck and Sonagashira couplings in aqueous conditions*

Preliminary optimizations of both the Heck and the Sonagashira type reactions were carried out using *para*-bromophenylalanine in the presence of acrylic acid or propargyl alcohol, respectively (Figure 7.3). Solvent system, base, temperature and time were the principal variables investigated. Interestingly both reaction types seemed to go best in pure water with an excess of potassium carbonate. This possibly reflects acceleration in these types of reactions in purely aqueous environments as reported [40-42]. The dramatic effect of potassium carbonate is not surprising as salt effects are known to be exceedingly important in palladium catalyzed crosslinking [39]. The conditions arrived upon are milder than those previously reported in under like conditions [41-43] and proceeded to completion as evaluated by ¹H NMR. While these conditions were far from fully optimized they were significantly better than previously reported palladium cross couplings in water involving arylbromides and were deemed acceptable to use with protein systems.
Figure 7.3. Partially optimizing conditions for Heck and Sonagashira couplings. In both cases N-acetyl-para-bromophenylalanine was used as aryl halide coupling partner. All reactions were deoxygenated exhaustively by the freeze-pump-thaw method. Reactions were quenched with AcOH, purified by extraction and analyzed by $^1$H NMR.
Demonstration of tolerance to protein functionality

In effort to demonstrate Pd(0) cross couplings could be classified as chemoselective reactions a series of both Heck and Sonagashira reactions were carried out under the above conditions in the presence of each of the potentially interfering natural amino acids. These "interfering" amino acids, arginine, asparagine, cysteine, glutamine, histidine, lysine, serine, tryptophan and tyrosine, were also acylated to allow facile purification of the starting material, product and "interfering" species. This allowed facile examination of all three species and evaluation of the extent of reaction. The results, under Heck, conditions, demonstrated full conversion to product in the presence of all of the tested amino acids with the exception of cysteine in which case there was no conversion whatsoever. No evident alteration of NMR signature was seen for any of the natural amino acids that remained in the organic phase, indicating that this chemistry may be benign to natural protein functionality under these conditions. Sonagashira results were similar except that the reaction run in the presence of cysteine did not suffer any loss in yield. Inhibition of the reaction by cysteine is presumably the result of reaction between the free thiol and the catalytic palladium species. This issue may not prove problematic in many cases where the cysteines are buried, oxidized or not present at all.

Modification of mDHFR (pCCHPhe, HAG and HPG) with pIF-FLAG tag

In order to demonstrate the effectiveness of both the Heck and Sonagashira reactions on intact proteins, murine DHFR with either terminal alkynes or alkenes
was modified with a pIF containing FLAG epitope. mDHFR was produced with either a terminal alkene (in the form of a methionine analog, homoallylglycine) or terminal alkynes (HPG, a methionine analog, and pCCHPhe, a phenylalanine analog). All of the reactions demonstrated excellent selectivity by western blotting with anti-FLAG antibody (Figure 7.4). While reaction was evident in the presence of either mDHFR-HAG, mDHFR-HPG or mDHFR-pCCHF with the catalyst, no product was formed without catalyst. Further, in the presence of catalyst there was no reaction with mDHFR not armed with terminal unsaturation. Differences in reactivity between the introduced analogs is readily apparent from the comassie and western detection. mDHFR-HAG produced a few distinct bands above the parent band indicating 2-3 modifications per protein. mDHFR-HPG produced a slightly higher smear, consistent with a significant increase in number of modifications, while mDHFR-pEF produced a much higher smear. Given that there are only 8 phenylalanine sites and that under identical conditions pCCHPhe was shown to replace phenylalanine at an average of 62% of the sites, the height of the band must indicate complete or near-complete reaction of the pCCHPhe incorporated into mDHFR. It is notable that mDHFR does not contain any cysteines.

Selecteive fluorescent modification of Barstar-pIF with lissamine rhodamine propargylsulfonamide

In an effort to examine the specificity of palladium cross couplings, specifically the Sonagashira reaction, a new target protein was chosen. Barstar demonstrates high expression levels, is an easily purified protein with simple
Figure 7.4. Pd (0) mediated modification of DHFR with pIF-FLAG. DHFR-pCCHPhe, DHFR-HAG and DHFR-HPG correspond to mDHFR incorporating para-ethynylphenylalanine, homoallylglycine and homopropargylglycine, respectively. All reactions were run at 37°C for 8 hours with or without catalyst. Reactions run with 62.5 μM protein, 5.4 mM peptide, and 270 μM catalyst (if added) in 8M urea; 0.1M NaH₂PO₄; 0.01M TrisCl; at pH 9.7.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>DHFR-pCCHPhe</th>
<th>DHFR-HAG</th>
<th>DHFR-HPG</th>
<th>DHFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pd cat</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>CuI</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>pIF-FLAG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Coomassie

Western Blot (anti-FLAG)
reversible folding and only two phenylalanines. Importantly, one of the phenylalanines is buried while the other remains relatively exposed (Figure 7.5). Our objective was to introduce para-iodophenylalanine and demonstrate modification at both sites when denatured, but at only the exposed site under native conditions. Because the large palladium catalyst has to be able to insert into the arylhalide bond the congested nature of the internal site should preclude any reaction (Figure 7.6). A fluorescent probe was created by the simple reaction of propargyl amine with Lissamine rhodamine sulfonamide in pyridine. Upon exposure to the alkyne-probe and palladium catalyst under native and denaturing conditions the protein was fluorescently tagged (Figure 7.7). Again no reaction was seen in the case of the target protein without analog incorporation (Figure 7.7B; lane 3). Notably the native modification showed less fluorescence intensity than the denatured (Figures 7.7B; lane 1 vs. 2 and 7.8B; lane 1 vs. 2 and 3, respectively), consistent with selective modification. Chymotryptic digest of both barstar modified under native conditions and denaturing conditions provide very different results (Figure 7.8C), indicating a selectivity linked to the structure of the target. Lys-C digest and LC-ESI-MS\textsuperscript{a} sequencing of the resulting peptides does demonstrate correctly modified peptide (data not shown). Importantly these correctly modified peptide mass spectrum provide further indication that the catalyst is not participating in unwanted side reactions.
**Figure 7.5.** Crystal structure of Barstar demonstrated as stick model. The phenylalanines, Phe56 and Phe74, are displayed space filling blue. Nominally Phe56 is surface exposed, whereas Phe74 is buried.
Figure 7.6. Schematic depicting the selective modification of Barstar-pIF with lissamine rhodamine propargyl sulfonamide. Cartoon represents ideal situation in which both Phe sites are occupied by pIPhe. Modification only at external sites depends upon steric exclusion of catalyst from internal Phe site.
Figure 7.7. Modification of Barstar with lissamine rhodamine propargyl sulfonamide. Lane 1: Barstar-pIPhe (0.86 mM), fluorophore (4 mM) and catalyst solution (0.7 mM) in 20mM Phosphate pH=8 (Native conditions), 37°C overnight. Lane 2: Barstar-pIPhe (0.86 mM), fluorophore (4 mM) and catalyst solution (0.7 mM) in 20mM Phosphate pH=8, 4 M urea, (denaturing conditions), 37°C overnight. Lane 3: Barstar-phe under identical conditions to lane 1. Gel A was developed with coomassie stain, whereas Gel B was imaged using UV. Otherwise the gels are identical.
Figure 7.8. Selective modification of Barstar with lissamine rhodamine propargyl sulfonamide. Lane 1: Barstar-pIPhe (0.86 mM), fluorophore (4 mM) and catalyst solution (0.7 mM) in 20mM Phosphate pH=8. Lane 2: Barstar-pIPhe (0.86 mM), fluorophore (4 mM) and catalyst solution (0.7mM) in 20mM Phosphate pH=8, 4 M urea. Lane 3: Barstar-pIPhe (0.86 mM), fluorophore (4 mM) and catalyst solution (0.7 mM) in 20 mM Phosphate pH=14. Image A was obtained by coomasie staining. Image B, otherwise identical to A, was obtained by UV detection. Gel C is the product of a chymotrypsin digest of lanes 1-3 in A and B, visualized in UV.

Chymotrypsin Digest
Conclusion

We have demonstrated that palladium catalyzed cross couplings, particularly the Heck and Sonagashira couplings, are viable for modification of proteins bearing the appropriate chemical functionality. These reactions proceed to completion under mild, fully aqueous conditions. The required conditions do seem to adversely affect any of the natural 20 amino acids, nor is the reaction affected by them with the exception of free cysteine, which presumably coordinates and "kills" the catalyst. We have demonstrated that the chemistry can effectively conjugate a peptide to protein bearing terminally unsaturated amino acids with efficacy depending on the nature of the terminal unsaturation. Thus these chemistries can be harnessed for chemoselective ligations as well as epitope tagging. Further, we have demonstrated fluorescent tagging with the alkyne within the modifying reagent and the arylhalide within the protein, demonstrating the versatility of these reactions. Proteolytic digests indicate the chemistry is selective for exposed residues. Pd(0) chemistry represents a remarkably broad set of chemistries that can effect carbon-carbon bond formations, linkages formerly inaccessible to chemoselective ligations on protein side chains. Such bonds provide unsurpassed stability in addition to allowing extension of [-systems, particularly in the modification of phenylalanine analogs. Palladium catalyzed crosslinking thus represents an attractive complement to current protein modification methods.
References


