Chemoselective Chemical Ligations of Biological Relevance

"Bioconjugate Techniques"

Hermanson says that a 3rd edition is underway.
Some Recent Reviews


Why are We Interested?

Carboxylic Acid Ketoconjugation at Neutral pH:

The ketone-forming step can take place at room temperature and in water.
Ketoconjugations of $\text{RCO}_2\text{H}$

*From Small Molecules to Chemical Biology to Materials?*

\[ \text{R} = \text{ANYTHING} \]
- small molecules
- biologicals
- carbon allotropes (GO, C60)
- quantum things
- surfaces
- materials

\[
\begin{align*}
\text{O} & \quad \text{R} \quad \text{OH} \\
\text{R} & \quad \text{CO} \\
\text{O} & \quad \text{R} \quad \text{SR'}
\end{align*}
\]

Recyclable reagent in solution or attached to solid support

\[ \text{R'}-\text{B(OH)}_2 \]
- catalyst
- neutral pH

Ketoconjugation

**Trivalent Modular Conjugations**

\[ \text{ANYTHING} - \overset{\text{OH}}{\text{C}} - \overset{(\text{HO})_2\text{B}}{\text{O}} \rightarrow \text{ANYTHING} \]

- Selective C-C conjugation at neutral pH in $\text{H}_2\text{O}$

\[ \text{ANYTHING} \]

** MODULE 1 **

\[ \text{N} \]

** MODULE 2 **

** MODULE 1 **

** MODULE 1 **

- Hydrolysis/proteolysis stable carbon-carbon bonds

\[ \text{the ketone: orthogonal third valence} \]
Chemoselective Bioconjugation

- The chemoselective coupling of one reactive functional group specifically with another reactive functional group without side reactions:
  - Between biomolecules and other biomolecules, surfaces, or particles
  - Often in aqueous solutions or biological buffers and in the presence of biological material (proteins, DNA, RNA, complex carbohydrates).
  - A pair of specifically "tuned" chemoselective reactants is built into the design of cross-linkers, labeling reagents, modified biomolecules, and surfaces/particles.

- Requires the rapid and selective reaction of substrates
- Reactants must be stable near neutral pH and in aqueous environments

Bioorthogonal Chemistry

• **Bioorthogonal reactions** – a particular type of bioconjugation where one specific "abiotic" functional group reacts with another specific "abiotic" functional group without any potential for cross-reactivity with biomolecular functionality in vivo.

• Bioorthogonal reactions can be used to selectively functionalize biomolecules in complex biological milieu.

• Bioorthogonal reactions are performed:
  – Inside of cells (cytosol)
  – On the surface of cells
  – In complex cell lysates

---

*Obtaining total chemoselectivity and perfect "bioorthogonality" is an ideal that is seldom achieved. Some of the more recent chemoselective ligations, which possess varying degrees of bioorthogonality, will be discussed.*


"Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality",
Bioconjugation methods rely on a variety of specific functional group interactions to achieve selectivity. They include, but are not limited to:

- Carbon-sulfur conjugates
- Hydrazide-aldehyde and related -C=N- conjugates
- Phenylboronic acid (PBA)-salicylhydroxamic acid (SHA) conjugates
- 4+2 and 3+2 dipolar cycloadditions
- Staudinger reaction between an azide and a phosphine

Native bioconjugation methods use naturally-occurring biological processes, such as the formation of a peptide bond between a C-terminal thiol ester and an N-terminal cysteine.

Carbon—Sulfur Linkages

Traditional protein conjugation chemistries have exploited the reactivity of surface-exposed nucleophilic amino acids, such as cysteine (SH) or lysine (NH₂).

As a result, these methods typically result in heterogeneous mixtures of products.

Thiol linkages at Cys represent one of the oldest methods of bioconjugation.

- Technique relies on cysteine residues, which are the second least common amino acid in natural proteins (represent 1.7% of AA residues in natural proteins).
- Therefore, somewhat selective bioconjugation using cysteine residues can sometimes be achieved.
- Common ligating functional groups for thiols are A α-iodoamides (alkylation), B maleimides (conjugate addition), and C disulfides (disulfide exchange).

Other non-selective linkages used for bioconjugation are based on the formation of C=N bonds from hydrazides and oximes reacting with aldehyde and ketone functional groups.

Boronic Acid-Salicylhydroxamate Pairs

- Phenylboronic acids (PBA) react with salicylhydroxamic acids (SHA) to form stable complexes.
- Complex formation occurs in aqueous buffers between pH 5 to 9, tolerates high salt concentration (up to 1.5 M), as well as the presence of water miscible solvents, detergents, protein denaturants, etc.
- Phenyl diboronic acids react with molecules or surfaces possessing at least two near-neighbor salicylhydroxamic acids to form highly stable bis-conjugates.

Commercial PBA/SHA Conjugate Pairs

- Commercially available PBA/SHA reagents:
  - NHS ester reacts with amines, hydrazide with aldehydes, maleimide group with thiols
  - Also available: pyridyl disulfide for reversible disulfide linkage to thiols, and iodoacetyl functional group to create thio ether bonds with thiols

1,2-Diol containing carbohydrates interfere with bioconjugation because the PBA-SHA interaction is reversible. Therefore, PBA/SHA conjugation is not bioorthogonal.

However, PBA-SHA reactive groups are stable to hydrolysis in aqueous solution. They are useful in creating protein-protein conjugates: *J. Biol. Chem.* **2000**, 275(12), 8952-8958.
Linkages Formed by Cycloadditions

A number of cycloaddition reactions are useful for bioconjugation.

**Cu-Promoted Azide-Alkyne Cycloaddition**

$$R - N_3 + R' \rightarrow Cu^1, rt \rightarrow \text{product}$$

**Diels-Alder Cycloaddition**

$$R - \text{alke} + \text{olefin} \rightarrow \text{product, rt, 24 h}$$

**Strain Promoted Azide-Alkyne Cycloaddition**

$$R - N_3 + \text{cycloaddition} \rightarrow \text{product, rt}$$

**Inverse Electron Demand Diels-Alder**

$$R'' - \text{addition} \rightarrow \text{product, rt, 40 min}$$
Normal Diels-Alder Reactions in Water


Waldmann: The site-specific C-terminal modification of the Rab7 protein with fluorophores by means of a normal Diels-Alder reaction in water.

The C-terminal protein thioester was prepared (recombinant technology) and ligated with a short peptide containing a C-terminal hexadienyl ester to yield a Rab7 protein hexadienyl ester. A maleimide Diels Alder reaction in water provided the protein functionalized with fluorescent probes.


Ring Strain/Inverse Demand DA


Abiotic strained alkenes and alkynes are relatively stable in cellular environs and can be used for bioorthogonal reactivity. trans-Cyclooctenes, norbornenes, bicyclononynes, and cyclopropenes react rapidly with electron-deficient tetrazines via inverse-electron-demand Diels−Alder reactions. Ring strain allows rapid reactions at rt. Useful for various bioorthogonal imaging applications.

Avidin is a tetrameric protein found in avian egg whites. It binds to biotin (vitamin B<sub>7</sub>) with a dissociation constant $K_D \approx 10^{-15}$ M, making it one of the strongest known non-covalent bonds.

Avidin-APC: stain
The original Huisgen dipolar cycloaddition of alkynes and azides typically requires elevated temperature and pressure. 

\[ R-N_3 + R' \rightleftharpoons \text{Cu(I)} \rightarrow R-N=C=N-R' \]

Cat Cu(I) dramatically accelerates the kinetics (about 1-million-fold): high yields at room temp and ambient pressure.

Azide and alkyne functional groups are abiotic and bioorthogonal: compatible with incorporation into live cells.


The azide-alkyne dipolar cycloaddition has revolutionized applications in organic synthesis and bioconjugation. Alkyl azides are very stable in aqueous solutions in the presence of biological molecules. In contrast, aryl azides are unstable to UV light and reducing agents. Alkynes not bearing activating groups are also very stable in the presence of biomolecules.
Metal-Free Azide—Alkyne Cycloadditions

Cu is toxic to living cells. Therefore, Cu-catalyzed azide—alkyne cycloadditions are problematic for applications in living systems.


**Metal-free bioorthogonal cycloaddition reactions have been reviewed**

"Cu-free click cycloaddition reactions in chemical biology"
The Classical Staudinger Reaction

- The Staudinger reaction (1919): the reduction of azides to amines via an iminophosphorane.

- The intermediate iminophosphorane also reacts with a variety of electrophiles to provide amines, imines, and related compounds:

Bertozzi/Saxon and Nilsson/Raines/Kiessling simultaneously disclosed that electrophiles such as esters can be used to trap the Staudinger iminophosphorane when they exist in the same molecule, thus providing a potential new bioconjugation tool for amide bond formation.

Staudinger Ligation

The concept was used to functionalize cell surfaces, selectively. Sugar azide units were incorporated into the cell surface through incubation. The live cells were then treated with the newly devised Staudinger ligation reagent. The biotin residue was then bound to a synthetic avidin bearing a fluorescent probe.

Biotin – for conjugation to avidin

Traceless Staudinger Ligation

The original Staudinger Ligation leaves a phosphine oxide moiety covalently attached within the conjugation linkage. Traceless Staudinger Ligations were developed to overcome this limitation.

**Original:** phosphorus on the linkage  
**Traceless:** phosphorus on the leaving group.

\[ R'\text{ONH} + \text{Ph}_2\text{P} + \text{X} \rightarrow \text{Ph}_2\text{P} + \text{R'ON} \]

Ligations Using Native Functionality

Native Chemical Ligation
Expressed Chemical Ligation

Limitations of Solid Phase Peptide Synthesis

• Automated solid phase peptide synthesis (SPPS) allows the routine synthesis of peptides up to 50 amino acids in length, the size of small protein domains.

• Generating larger proteins is limited by the insolubility of protected peptide segments, by racemization that occurs in solution-phase synthesis, and by the statistical accumulation of error peptide byproducts in stepwise solid phase synthesis.

• This fundamental limitation in the length of synthetic peptides was overcome by the development of a new chemoselective ligation reactions where two or more fully unprotected peptides are linked together: Native Chemical Ligation.
Native Chemical Ligation

- Kent and coworkers developed native chemical ligation (NCL) as one solution to overcome SPPS limitations.
- In NCL, one unprotected peptide with an N-terminal cysteine reacts with a second unprotected peptide possessing a C-terminal thioester to form a peptide bond.
- Native Chemical Ligation proceeds via a reversible trans-thioesterification, which is followed by an irreversible nucleophilic attack of an internally positioned amine.
- **Critical feature of NCL**: the avoidance of any protecting groups on either fragment.

Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B.

Kent, S.; Sohma, Y.; Liu, S.; Bang, D.; Pentelute, B.; Mandal, K.
NCL: Selectivity w/o Protecting Groups

Reversible transthioesterification

Spontaneous S→N acyl transfer
NCL Particulars

- **NCL is carried out in buffered aqueous solutions at neutral pH.** This avoids strong basic conditions which would free residues such as Lys for reactions with the thioesters.
- Neutral pH also avoids hydrolysis of thioesters, which are unstable to aqueous basic conditions.
- Acidic conditions, on the other hand, do not favor NCL, because the reactivity of both the Cys thiol and the N-terminal amine of the C peptide is reduced when they are fully protonated.

![Diagram](https://via.placeholder.com/150)

Preparation of C-terminal thiol ester via SPPS is best using alkyl thiols to generate stable alkyl thiol esters. But the subsequent exchange reaction of the alkyl thioester with the N-terminal cysteine SH is slow.

Therefore, NCL is often catalyzed by the inclusion in the reaction of thiol additives such as thiophenol, benzyl mercaptan, and more recently (4-carboxymethyl)thiophenol. The added thiol carries out thiol ester exchange at the alkyl thioester, generating a more reactive thiol ester, and also helps prevent aerobic oxidation of the N-terminal cysteine SH group.

An Application of NCL

NCL for the total chemical synthesis of human lysozyme (130 amino acid residues). The synthetic peptide had full catalytic activity and was characterized by high-resolution X-ray structure.

NCL and Staudinger Ligation

Reactivity Issues with NCL


Side chains significantly affect the NCL rates of the 20 different amino acid thioesters.

**Fastest:** Gly thioester reacts quantitatively in less than two hours.

**Slowest:** Thioesters containing β-branched amino acids or proline did not result in a quantitative conversion after two days.

**Sterically hindered** thioesters however can be ligated by using more reactive thiols.
Influence of the Thioester on NCL


- Alkyl thioesters are less reactive than aryl thioesters.
- Alkyl thioesters are good for SPPS synthesis and handling, but NCL requires a rapid thiol–thioester exchange, which is much more efficient with aryl thioesters.
- Therefore, peptide thioesters are commonly synthesized as their alkyl derivatives, often with sodium-2-mercaptoethanesulfonate (MESNA), and are then converted into the corresponding aryl thioesters *in situ* by the addition of an excess of an aryl thiol, such as the water-soluble (4-carboxymethyl)thiophenol.
Beyond the NCL Cysteine Limit

NCL requires a cysteine residue at the N-terminus of one coupling partner.

**Issues with the Cys criterion.**

1. NCL requires a Cys residue at the N-terminal ligation site, but Cys residues appear in peptide sequences to only 1.7%. This limits the number of retrosynthetic disconnections that can be considered when constructing proteins using NCL.

2. Cys residues are not "innocent" – they can undergo disulfide bond formation and elimination to reactive unsaturated amides.

**Avoiding the Cys limitations of NCL.**

Desulfurization and related approaches


However, desulfurization is incompatible with retaining cysteine residues elsewhere in a protein.
Danishefsky Auxiliary Approach

Chen, G. G.; Warren, J. D. J.; Chen, J. J.; Bin B Wu; Wan, Q. Q.; Danishefsky, S. J. S.

Expressed Chemical Ligation


Tom Muir conceived of fusing the NCL technique with known technology of recombinant protein production to construct large proteins through what he termed "expressed protein ligation" (EPL):

The Muir challenge: How to construct a specific *recombinant* protein C-terminal thioester for reaction (via NCL) with a SPPS-generated N-terminal cysteine peptide. Muir accomplished this second task using a genetically re-engineered mutant variant of a common biological process: **protein-splicing**


Protein Splicing: Inteins

In Nature, protein splicing takes place when a protein segment, termed an intein, is excised out of two flanking protein sequences, named exteins (the N-terminal extein is at the N-terminal end of the protein and the C-terminal extein is at the C-terminal end of the protein). The process generates a new protein by fusing together the two extein peptide segments, most commonly through a cysteine (and sometimes a serine) linkage.


The Protein Splicing Mechanism

The protein splicing process involves a series of acyl rearrangements and internal reactions catalyzed by the central intein protein domain. This process results in linkage of the two flanking polypeptides, the N- and C-exteins, via a normal peptide bond.

The S→N acyl transfer should be contrathermodynamic. However, the position of the equilibrium of the S→N acyl transfer of inteins is facilitated by ground-state destabilization of the cysteine linkage amide bond in the protein.
Asn → Ala Mutant Inteins

Mutant inteins (i.e., Asn → Ala) can not complete excision of the intein – it is trapped in the thiol ester state.

Mutant inteins can be treated with external thiols and the resulting C-terminal protein thiol ester purified.

How is this chemistry accomplished in actual practice?
Muir was the first to ligate a recombinant protein thioester with a synthetic N-terminal cysteine peptide. He termed the technique "expressed protein ligation" (EPL).


Commercial Intein Expression Kit: Expressed C-Terminal Thiol Esters

Commercial Protein Expression Kit: a plasmid is genetically engineered to express the target protein of interest, which is attached to the mutant intein, which is attached to a "chitin-binding-domain" protein.
Variations of EPL: Split Inteins

Muir's EPL technique is widely used, but the generation and isolation of the required recombinant protein α-thioesters remain challenging. To address this issue Muir has described a variant of EPL that uses "split inteins".

Split inteins are characterized by the fact that their primary sequence is cut into two polypeptides giving an N-terminal fragment (IntN) and a C-terminal fragment (IntC). Fragment complementation leads to reconstitution of the canonical intein fold, recovery of protein splicing activity, and ligation of the exteins.
Muir and coworkers devised a streamlined procedure for the purification and generation of recombinant protein α-thioesters using split inteins. The IntC portion, bound to a solid support, was mutated to prevent excision of the intein and allow build up of the protein of interest thiol ester. Kinetics of thiol release are much faster than with non-split inteins.
Muir Split Intein Approach

1. cell lysates or growth medium
2. wash, then RSH (POI thiol ester)
3. Cys-N-terminal peptide

modified protein
Application: Proteo-ketones

ECL Intein Technology

Target Protein

Intein

Chitin Binding Domain

solid support

Target Protein

O

S

FG

thiol cleavage site

HN

R

B(OH)_2

aerobic C-C bioconjugation
cat. Cu

Target Protein

ketoconjugated protein

Green Fluorescent Protein
with the Salaita Lab

Thiol pKa: 6.5 or above: Johnson and Kent, JACS, 2006, 128, 6640.

Ubiquitin
with the Wilkinson Lab