Bioorthogonal Chemistry: Enabling nCAA Protein Labelling

A. Manos-Turvey,
Wipf Group: Frontiers of Chemistry
Jan. 17th, 2015
From the Genome to Proteins

- **Pre-genomic era:** ended with the completion of the human genome project in 2003
- **Genomic era:** resulted in the rapid collection of immense databases of genomic sequences
- **Post-genomic era:** decipher the structural and functional information about biomolecules encoded by the genomic data
  - 3-4 D structure
  - Dynamics
  - Function

The ability to incorporate non-canonical amino acids (nCAAs) into proteins, allowing chemical probes to be synthesised, is aiding this search.

Methods of nCAA incorporation in proteins are well documented:

- Solid phase synthesis
- Native chemical ligation
- *In vitro* translation

Proteins within the cells of interest require “bioorthogonal chemistry” to allow for labelling in the cellular environment.

To be labelled a bioorthogonal reaction, a reaction must meet several criteria – many only meet some of these prescriptions

- Incorporated functionality + probe react selectively with one another
- Reactions yield stable covalent linkages with no by-products
- Reactants must be kinetically, thermodynamically and metabolically stable, and non-toxic, prior to reaction

Many chemoselective reactions have been found, but few are truly bioorthogonal

- Many remain restricted to cell surface labelling or in vitro systems
- Kinetics of the reaction are very important

R.K. Lim, Q. Lin, Chem. Commun., 2010, 46, 1589-1600
Bioorthogonal Chemistry Kinetics

- Biological processes are very rapid, bioorthogonal reactions need to be able to compete with these

- Bioorthogonal reactions often follow second-order kinetics and the rate of reaction is dependant upon:
  - [biomolecule]
  - [labelling reagent]
  - Second-order rate constant

- Difficulties:
  - Target biomolecules are usually in low abundance in their native environment
  - Higher conc. of labelling reagents can increase insolubility/chances of off-target effects
  - Need reactions which have higher intrinsic rate constants

- Faster chemoselective reactions are needed to increase utility of labelling
Bioorthogonal Chemical Reactions

- The predominant transformations in the “bioorthogonal toolkit” are:

  ![Diagram showing polar reactions and cycloadditions]

  - Polar reactions
    - $E \rightarrow E \cdot \cdot \cdot N u$ (transformation)
  - Cycloadditions
    - $[4+2]$ (reaction)
    - $[3+2]$ (reaction)

Ketone/Aldehyde-Hydrazine/Alkoxylamine Reactions

- Aldehyde and ketone reactions were amongst the first bioorthogonal reactions to be identified
  - Ketones are preferable due to lower general activity under physiological conditions
- Under acidic conditions the carbonyl group is protonated and reacts with amines to form reversible Schiff base

![Chemical Reaction Diagram]

- Difficulties:
  - Commonly requires acidic pH
  - Slow kinetics ($10^{-4}$-$10^{-3}$ M$^-1$s$^-1$)
  - Requires high [reagent] (mM)
  - Competition from naturally occurring aldehyde/ketone metabolites

Aniline activation:

- Can be run at pH 7
- Accelerates the reaction > 40 fold (> 400 at pH 4.5)
- Formation of the aniline Schiff base becomes the rate determining step
- Live cell surface labelling has proven successful

Ketone/Aldehyde-Hydrazine/Alkoxylamine Reactions

Aniline activation:

- Hydrazones and oximes can be hydrolytically cleaved over extended periods of time

Ketone/Aldehyde-Hydrazine/Alkoxylamine Reactions

**Pictet-Spengler Reaction**

\[
\text{Tryptamine} + R\text{CHO} \rightarrow \text{tetrahydrocarboline}
\]

\[k_2 \sim 10^{-4} \text{ M}^{-1}\text{s}^{-1} \text{ at pH 4-5}\]

**Pictet-Spengler Ligation**

- Accelerates the reaction in aqueous media

\[
\text{Oxacarboline} 4-5 \text{ fold faster}
\]

\[= \text{labelling reagent} \quad \text{= biomolecule}\]

**Pictet-Spengler Ligation**

- Successfully used to label proteins and antibodies
- Shows heightened stability to its oxime counterparts
- Still requires acidic conditions
- Mainly *in vitro* utility

---

Azide-Phosphine (Staudinger) Reactions

- Azide groups are absent from biological systems
  - Small in size
  - Stable under physiological conditions
  - React with bioorthogonal triaryl phosphines

Staudinger Reduction

\[
\text{N}_3^- + \text{P} \equiv \text{Ph} \quad \text{H}_2\text{O} \quad \text{NH}_2 + \text{O} \equiv \text{P} \equiv \text{Ph}
\]

- Need to trap the reaction at the aza-ylide intermediate

Azide-Phosphine (Staudinger) Reactions

Staudinger Ligation

- Intermediate aza-ylide undergoes intramolecular formation to give an amide
  - A electrophilic trap was incorporated into the phosphine
  - Stable under physiological conditions
  - Reacts with bioorthogonal triaryl phosphines ($k_2 = 10^{-3} \text{ M}^{-1}\text{s}^{-1}$)

\[ \text{N}_3 + \begin{array}{c} \text{Ph}_2\text{P} \end{array} \xrightarrow{-\text{N}_2} \begin{array}{c} \text{Ph}_2\text{P} \end{array} + \text{O} \xrightarrow{-\text{MeOH}} \begin{array}{c} \text{Ph}_2\text{P} \end{array} + \text{H}_2\text{O} \longrightarrow \begin{array}{c} \text{Ph}_2\text{P} \end{array} + \text{N} \xrightarrow{\text{H}_2\text{O}} \begin{array}{c} \text{Ph}_2\text{P} \end{array} + \text{O} \]

- = biomolecule
- = labelling reagent

Azide-Phosphine (Staudinger) Reactions

Staudinger Ligation

(1) Inject Ac$_4$ManNAz solution for 7 days
(2) Isolate splenocytes

(1) Staudinger ligation

~O$_2$C – KDDEDDKYD – N

Phos-Flag

(2) FITC–anti-Flag

Flag

Detection by flow cytometry

Azide-Phosphine (Staudinger) Reactions

Traceless Staudinger Ligation
- Final amide-linked product does not contain a phosphine oxide group
- Useful in the synthesis of peptides
  - Does not require a cysteine residue (unlike NCL)

\[
\text{N}_3^- + \begin{array}{c} \text{Ph}_2\text{P}\text{O} \end{array} \xrightarrow{-\text{N}_2} \text{Aza-ylide} \xrightarrow{\text{H}_2\text{O}} \begin{array}{c} \text{NH} \end{array} + \begin{array}{c} \text{OH} \end{array}
\]

General Difficulties
- Slow kinetics \((10^{-3} \text{ M}^{-1}\text{s}^{-1})\)
  - Requires high [labelling reagent]
- Increasing nucleophilicity of phosphines leads to oxidation

Azide-Alkyne Reactions ([3+2] Cycloadditions)

- Azide groups are absent from biological systems
  - Small in size
  - Stable under physiological conditions
  - React with physiologically stable alkynes

**Copper Catalysed Azide-Alkyne 1,3-Dipolar Cycloaddition (CuACC)**

- Requires sufficient Cu(I) to maintain rate of reaction
  - *E. coli* stop growing after 16 h of 100 µM CuBr exposure
  - Mammalian cells tolerate < 500 µM of Cu(I) for ~ 1 h

H.C. Kolb, K.B. Sharpless, *Drug Discovery Today*, 2003, 8, 1128-1137
Azide-Alkyne Reactions ([3+2] Cycloadditions)

Copper Catalysed Azide-Alkyne 1,3-Dipolar Cycloaddition (CuACC)

- **Water soluble ligands**
  - Reduces cytotoxicity by acting as reductants
  - Has led to faster reaction times, improved kinetics ($k_2 = 10$-200 $M^{-1}s^{-1}$)
  - Effective for cell surface labelling

![Diagrams of BTTES and BTAA]

- Reaction conditions: propargyl alcohol (50 $\mu$M), 3-azido-7-hydroxycoumarin (100 $\mu$M), CuSO$_4$ (50 $\mu$M) ([ligand]/[CuSO$_4$]=6:1), potassium phosphate buffer (0.1 M, pH 7.0)/DMSO=95:5, sodium ascorbate (2.5 mM), rt

Azide-Alkyne Reactions ([3+2] Cycloadditions)

Copper Catalysed Azide-Alkyne 1,3-Dipolar Cycloaddition (CuACC)

- **Copper-chelating organic azides**
  - Raises the effective [Cu(I)] at the reaction site through proximal pyridine N
    - Reaction rate at 10 µM exceeds that of non-chelating variants at 100 µM
  - Can be used in concert with water soluble ligands ([Cu(I)] = 10 µM)
  - Effective for labelling proteins in live cells

![Chemical Reaction Diagram](image)

Azide-Alkyne Reactions ([3+2] Cycloadditions)

Strain-Promoted Alkyne-Azide Cycloaddition (SPAAC)

- Copper Free
  - Uses cyclooctynes as reagent
  - Shows no observable cytotoxicity
  - Some are now commercially available

- Original cyclooctyne bioorthogonal reactions are slower than Staudinger ligations
- Improvements in derivatives has only led to rates of ~0.1 to 1 M⁻¹s⁻¹

Nitrone-Cyclooctyne Reactions ([3+2] Cycloadditions)

Strain-Promoted Alkyne-Nitrone Cycloaddition (SPANC)

- Uses more reactive 1,3-dipole nitrones in place of azides
  - Rate constants up to 60 M\(^{-1}\)s\(^{-1}\) in model reactions
  - 60 x faster than SPAAC
- Lower [reagent]
- Cyclic nitrones are more stable than acyclic counterparts
  - Susceptibility of nitrones to hydrolysis can be overcome

Alkene-Tetrazole Reactions (Photoclick Cycloaddition)

- Tetrazoles activated with light will generate an imine in situ (1967)
  - An aqueous variant was first reported in 2008

- Originally UV activated (4 min), now near IR excited (2 h)
  - Reduced/removal of phototoxicity to cells ($k_2 \sim 10 \text{ M}^{-1}\text{s}^{-1}$)

Pyrazoline products are fluorogenic
Has been used to label in vitro and in living cells

Difficulties
- Can be quenched by chloride ions/acidic conditions
- Internal cis-alkenes are endogenous in biological systems, could compete

1,2,4,5-Tetrazines are reacted with electron-rich dienophiles (alkenes) = iEDDA

- Highly strained alkenes (such as *trans*-cyclooctene) react very rapidly in organic solvent
- Previously described tetrazines were not stable in H₂O
Strained Alkene/Alkyne-Tetrazine Reactions (Inverse-Electron-Demand Diels-Alder Cycloadditions)

- Governed by the $\text{HOMO}_{\text{dienophile}} - \text{LUMO}_{\text{diene}}$ gap
  - EWG groups at the 3- and 6-positions of the tetrazine lower the LUMO = Faster iEDDA
  - Dienophiles with EDG substituents raise the HOMO = Faster iEDDA
  - High ring strain in the dienophile helps reduce the activation energy

iEDDA was first shown to be bioorthogonal in 2008
- 3,6-Diaryl-s-tetrazines were found to be H₂O stable
- Can be run in cell media and cell lysate with > 80% yield (k₂ = 2000 M⁻¹s⁻¹)
- Successfully used to label proteins in vitro and in vivo

![Chemical Reaction Diagram]

Tetrazines can be conjugated to fluorophores to create fluorogenic products

- The tetrazine quenches fluorescence until activation by iEDDA


A range of strained alkenes/alkynes (a) and tetrazines (b) have now been investigated.

Difficulties

- Tetrazine synthetic availability
  - Potentially explosive nature of synthetic precursors to tetrazines
Incorporation of nCAAs for Bioorthogonal Reactions

- Proteins typically consist of 20 naturally occurring amino acids
- Syntheses of nCAAs bearing bioorthogonal groups are now well documented
- Highjacking protein translation can lead to successful nCAA incorporation
  - High fidelity of protein translation
  - Yields can be increased by encoding repressed translational promoters

Residue-specific
- 1950s – replace one of the 20 amino acids with a similar synthetic variant

Site-specific
- 1980s – a nCAA can be incorporated at a specific site within a protein, in the presence of the natural amino acids

Amino Acid Transfer with tRNA Synthetases

1. A specific amino acid and ATP bind to the aminoacyl-tRNA synthetase.
2. The amino acid is activated by the covalent binding of AMP, and pyrophosphate is released.
3. The correct tRNA binds to the synthetase. The amino acid is covalently attached to the tRNA. AMP is released.
4. The charged tRNA is released.


http://biomoosnews.blogspot.com/ Sept 10th 2014 Blog
Residue-Specific Incorporation of nCAAs

- Produces globally modified proteins
- Simplest variants require no genetic alterations of the target organism
  - Targets entire proteome of the cell/organism

- Three key requirements:
  - Natural amino acid must be encoded at the genetic level
  - The unnatural amino acid must be a substrate for the natural variants tRNA synthetase
  - The desired protein must be expressed when the unnatural amino acid is in the cell

Residue-Specific Incorporation of nCAAs

= Met, or natural amino acids to be replaced
= one of other 19 amino acids
= unnatural amino acid
= codon for Met or natural amino acid to be replaced
= codons for other 19 natural amino acids

Residue-Specific Incorporation of nCAAs

- Greater alterations in nCAAs can be accommodated by mutating existing tRNA synthetase binding sites

Uses

- BONCAT & FUNCAT (Bioorthogonal/Fluorescent noncanonical amino acid tagging)
  - Determine kinetics of protein synthesis and separate newly synthesised proteins from pre-existing proteome
  - Analysis of localised synthesis of proteins critical to cell function (such as axons)
  - Labelling of newly synthesised proteins in multicellular organisms
    - Now progressed to cell selective protein labelling

Residue-Specific Incorporation of nCAAs

Difficulties

- Labels all proteins in an organism or cell unless specifically genetically altered
- Some nCAA incorporation can be toxic to the cells
  - Conc. and exposure time dependant
- Some nCAAs are poorly incorporated
  - Addressed by use of mutated tRNA synthetases that can be selectively expressed
- Limited by requirements of the cell, i.e. 20 natural amino acids

Site-Specific Incorporation of nCAAs

- Uses genetic code expansion
- Allows a single amino acid of a peptide to be altered in the presence of the all 20 naturally occurring amino acids

- Three key requirements:
  - An orthogonal amino-acyl-tRNA synthetase/tRNA pair
  - A blank codon to encode for a nCAA
    - Commonly the amber stop codon UAG (TAG) as 93% of E.coli genes end with TGA or TAA
  - Methods to ensure the transfer of the nCAA to the orthogonal tRNA occurs selectively

Site-Specific Incorporation of nCAAs

- Orthogonal tRNA synthetase: must not label any endogenous tRNA in a cell
- Orthogonal tRNA: must not be recognised by any endogenous tRNA synthetases
- Exploit the tRNA-synthetase/tRNA differences between organisms

Four tRNAs (with desirable synthetase partner) that recognise amber stop codons have been identified and used in site-specific manipulations of proteins

- tRNA may require mutation to recognise nCAAs
- tRNA does not need to be altered to recognise a manufactured blank codon

Methanococcus

Methanosarcinae

E.coli

MjTyrRS/tRNA$_{CUA}$

PyIRS/tRNA$_{CUA}$ orthogonal pair

EcTyrRS/tRNA$_{CUA}$

EcLeuRS/tRNA$_{CUA}$

BACTERIA

EUKARYOTES

Orthogonal translation pathway

- A new genetic code, evolved to include quadruplet codons
  - New ribosome synthesis
    - Will accommodate quadruplet codon tRNA
  - New mRNA
  - New tRNA-synthetase/tRNA pairs
    - Could non-α-amino acids be used?

Orthogonal translation pathway

- Used successfully to create a modified GST-calmodulin protein
  - Only synthesised if both nCAAs are incorporated
  - Has the potential to create new proteins and alter functions and structures

Site-Specific Incorporation of nCAAs

Uses

- More nCAAs have been successfully incorporated than in residue-specific experiments
- Help to investigate specific proteins within organisms and cells
- Being used to form antibody conjugates (such as ADC antibody-drug conjugates)
- Enables minimal disruption to normal protein folding
  - Can probe specific sites in the protein amenable to nCAA incorporation
- Can (and has) been adapted to DNA and RNA labelling (aptamers)
  - Could prove useful in early and specific cancer detection assays

Difficulties

- The number of orthogonal tRNA-synthetase/tRNA pairs is small
  - Limit of two nCAAs that can be incorporated in any cell
  - The process of mutation and selection can be painstaking

Conclusions and Scope

- Bioorthogonal chemistry has been developed
  - Chemoselective
  - Non-toxic/no byproducts

Conclusions and Scope

• Bioorthogonal chemistry has been developed
  o Reactions with faster kinetics are still needed
  o Need to investigate quantitative reaction specificity
  o Need to further optimise and analyse site-specific labelling in mammalian cells

• Some intracellular imaging difficulties need to be addressed
  o Selective imaging/washing out labelling probe
  o More easily adapted fluorescent probes are also needed to couple with this work

• As new reactions/reagents are discovered, the scope and utility of this field will increase

Acknowledgements

- Dr Peter Wipf
- Dr Jeff Brodsky
- Wipf and Brodsky Group Members (past and present)
Azide-Alkyne Reactions ([3+2] Cycloadditions)

Copper Catalysed Azide-Alkyne 1,3-Dipolar Cycloaddition (CuACC)

B.T. Worrell, J.A. Malik, V.V. Fokin, Science, 2013, 457-460
**Water soluble ligands mechanism**

*Figure 7.* Proposed mono- and binuclear speciation in the CuAAC process. $X =$ anionic donor such as $\alpha$-acetylde, halide, hydroxide, or triazolide; $S =$ neutral intermolecular donor such as DMSO, other solvent, $\pi$-alkyne, or organic azide. Productive complexes are not formed when only one heterocyclic donor arm (L) exists.

Azide-Alkyne Reactions ([3+2] Cycloadditions)

- Calculated and X-ray strain of BARAC strained cyclooctyne

Nitrone-Cyclooctyne Reactions ([3+2] Cycloadditions)

- Synthesis of nitrone bound to biomolecules

**Scheme 1.** One-pot N-terminal conjugation of a hexapeptide by SPANC: a) 1. NaIO₄, NH₄OAc buffer, pH 6.8, room temperature, 1 h; 2. p-MeOC₆H₄SH, room temperature, 1 h; then p-MeOC₆H₄NH₂, MeHNOH-HCl, room temperature, 20 min; b) 2, room temperature, 1 h.

Nitrone-Cyclooctyne Reactions ([3+2] Cycloadditions)

- Synthesis of nitrone bound to biomolecules

Scheme 2. One-pot N-terminal functionalization of IL-8 by SPANC: a) 1. NaO₄, NH₄OAc buffer, pH 6.9, room temperature, 1 h; 2. p-MeOC₆H₄SH, room temperature, 2 h; b) p-MeOC₆H₄NH₂, MeNHOH.HCl, room temperature, 20 min; c) cyclooctynol 2 or PEG-cyclooctyn 16, room temperature, 20 h.


Alex Manos-Turvey @ Wipf Group
Page 50 of 53
1/17/2015
Nitrone-Cyclooctyne Reactions ([3+2] Cycloadditions)

- Synthesis of nitrone bound to labelling tags

Fig. 1  In situ labeling of EGF-EGFR interactions via SPANC in MDA-MB-468 cells. (a) Cyclic nitrone modified EGF-1c bound to EGFR was labeled by SPANC with 2b-biotin for 30 min prior to streptavidin-FITC fluorescent labeling. (b) Fluorescence (top) and
Alkene-Tetrazole Reactions (Photoclick Cycloaddition)

- Mechanism of nitrile imine interconversion

Protein Translation 101 (Prokaryotes)

16S rRNA (30S subunit)

UCCUCC
AGGAGG

Shine-Delgarno sequence

(usually 7 bp long)

anticodon

fMet

Amino Acid Attachment Site

5' mRNA

Aminoacyl-tRNA

"A" Binding Site

Ext "E" Binding Site

Peptidyl-tRNA "P" Binding Site

Large Subunit

Small Subunit


https://wikispaces.psu.edu/display/Biol230WCE/Protein+Translation