Expanding the genetic code: a non-natural amino acid story

Athena Zitti and **Dafydd Jones** (Cardiff University, UK)

From enzymes to hormones, proteins are the most versatile macromolecules that serve a vital function throughout all biological systems. In nature, organisms are restricted to a 20 amino acid repertoire, which in turn limits the chemistry. Nature evolves and adapts but currently does so under a "chemistry-limited" circumstance. This in turn has limited the aspirations of protein designers and engineers that aim to expand the functional and structural properties of proteins into new and exciting realms. One way nature has circumvented this problem is to recruit non-proteinaceous cofactors. Another way is to change one of the most fundamental concepts underlying biology: the genetic code. Pyrrolysine and selenocysteine are the two main genetically encodable proteinogenic "21st" amino acids; stop codons are recruited to encode the incorporation during ribosomal polypeptide synthesis. This ability of nature to go beyond the standard 20 amino acid repertoire inspired researchers to engineer the fundamental elements of ribosomal polypeptide synthesis and allow full genetic encoding of non-natural amino acids. This in turn helped advance synthetic biology and protein engineering in ways that were not possible by using the standard 20 amino acids. To this day, a vast repertoire of non-natural amino acids is available and is continuously expanding with increasing scientific needs.

Natural VS non-natural amino acids (nnAA)

There are 20 common amino acids encoded as codons by the genetic code. To most bioscientists, this may seem like enough, given the impressive array of functions that proteins can undertake, but to organic chemists, the chemistry is very limited. Expanding the chemistry available to proteins has long been an aim of biochemists. Chemical labelling (e.g., via thiol or primary amines) is one such route that is still used routinely today. Another is auxotrophic approaches whereby synthetic analogues of natural amino acids (e.g., selenomethionine in place of methionine to facilitate structural studies) are "fed" to bacteria and incorporated in their place during ribosomal protein synthesis (RPS). The problem with both these approaches is precision – in terms of target (the protein) and position (the residue). To a protein engineer, precision is important as we want to elicit defined changes directly into the protein as the "sequence to structure to function" flow reigns supreme. So, in order to retain precision embedded within the genetic code but still allow new chemically useful amino acids to be incorporated, we reprogramme the genetic code.

Hijacking cellular protein synthesis – nnAA incorporation

The main method we, in the Jones lab at Cardiff University and others use to incorporate these new amino acids is by "hijacking" an existing codon. The one most commonly reprogrammed is the amber stop codon (UAG), the least used codon in the genome and naturally encodes the incorporation of pyrrolysine in a few selected organisms. To reprogramme the amber stop codon, extensive modification of some of the fundamental components of the RPS was required, including: (1) Changes to the base sequence of tRNA (the anticodon binding loop) to recognise the UAG codon; (2) a tRNA/aminoacyl-tRNA synthetase pair specific only to the non-natural amino acid and not endogenous amino acids; and (3) the ability of the new nnAA to enter the cell. The final step is introducing TAG mutations via site-directed mutagenesis at required positions to allow precise incorporation of the nnAAs during RPS. This tour de force in genetic/protein engineering and directed evolution led by Peter Shultz's group generated these new components, which allowed incorporation of nnAAs in response to the UAG codon (Figure 1). In recent years, this has been added to through genome engineering to generate strains of *E. coli* with all the amber stop codons removed and repurposing additional codons for incorporating multiple different nnAAs into proteins.

From new enzymatic catalytic mechanisms to non-native post-translational modification approaches, to improved biophysical analysis approaches, nnAA incorporation has already had a major impact on how we understand and engineer biological systems. We are currently using nnAAs to improve how we integrate proteins with non-biological systems, to generate new bionanohybrid systems using primarily phenyl azide

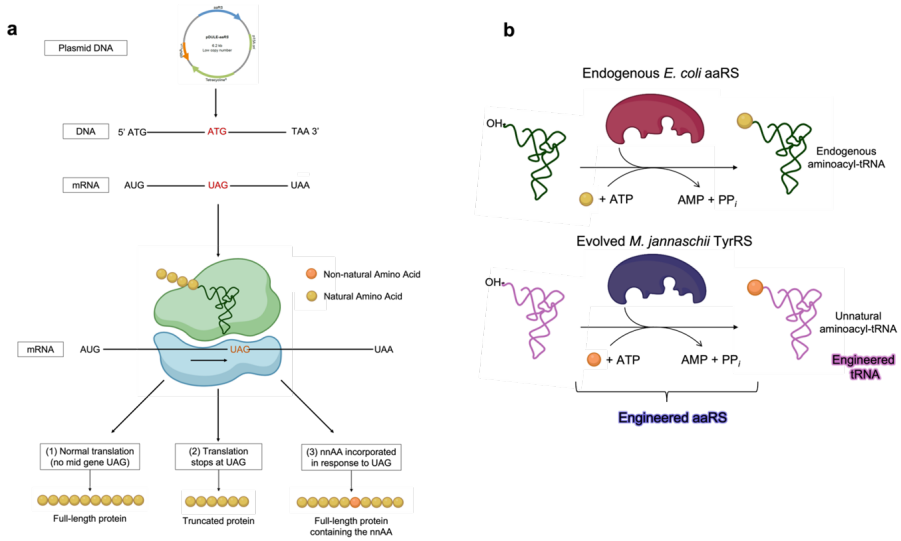

Figure 1. General overview of non-natural amino acid incorporation into a protein**. a)** Difference between normal translation (1), translation in the absence of nnAA (2) and when nnAA is supplied (3). **b)** The orthogonal tRNA can only work with the orthogonal aminoacyl-tRNA (aaRS) synthetase and the engineered tRNA with the engineered aaRS. In the presence of the UAG codon and supplied nnAA the end result is a modified peptide molecule with the non-natural incorporated.

chemistry. Phenyl azide chemistry has a long-forgotten role in biochemistry, as it was once conjugated to proteins where the inherent photochemistry was used to crosslink to near neighbours. Azide chemistry (Figure 2) has had a recent resurgence largely due to the advent of biorthogonal click chemistry for which Carolyn Bertozzi won the 2022 Nobel Prize for Chemistry.

Non-naturals revolutionizing protein engineering

From new enzymatic catalytic mechanisms to non-native post-translational modification approaches, to improved biophysical analysis approaches, nnAA incorporation has already had a major impact on how we understand and engineer biological systems. We are currently using nnAAs to improve how we integrate

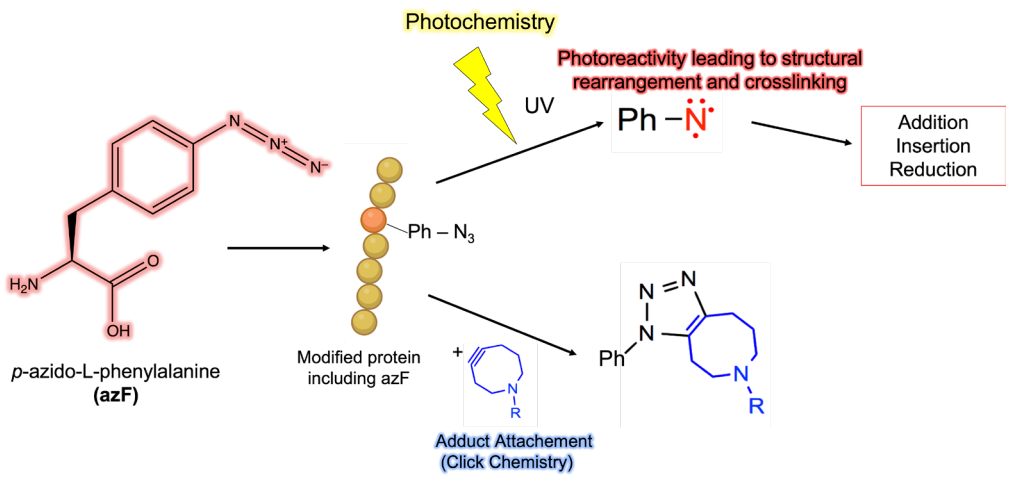

Figure 2. Post-translation functional control upon azF incorporation. azF can be used to modulate protein activity and generate proteins with novel properties that allow downstream procedures such as Click Chemistry.

proteins with non-biological systems, to generate new bionanohybrid systems using primarily phenyl azide chemistry. Phenyl azide chemistry has a long-forgotten role in biochemistry, as it was once conjugated to proteins where the inherent photochemistry was used to cross-link to near neighbours. Azide chemistry (Figure 2) has had a recent resurgence largely due to the advent of biorthogonal click chemistry for which Carolyn Bertozzi won the 2022 Nobel Prize for Chemistry.

In recent years it has become of great interest how synthetic biology can be integrated with nanotechnology in order to generate bionanohybrid molecules. These new nano-systems are essentially a combination of biomolecules with other organic and inorganic molecular scaffolds. An example of such is the interfacing of protein with carbon nanotubes (CNTs). The reason why proteins are so useful is that they work on the nanoscale. In addition, using proteins has many benefits as they self-assemble, have structural and functional variety and are transferable across different organisms due to their genetic code; they are compatible with both bio and non-biomolecules and finally, as previously explained, we can engineer and program their function to fit our desired application.

Coming back to the usefulness of phenyl azide chemistry, we have used both its chemical properties (Figure 2) to link proteins in defined and precise ways to nano-carbon materials such as CNTs and graphene. Nano-carbon is an excellent base material for biosensors as their inherent electrical properties can be changed by, e.g., the electrostatic surface properties of a protein. Prior to the use of phenyl azide chemistry, protein attachment to such nanomaterials was essentially random (Figure 3). Given that the protein attachment site, and thus orientation, can have a major impact on the nanomaterial properties, this was far from ideal in generating sensitive bionanohybrids with defined properties. By using click chemistry, we can attach binding proteins in designed orientations onto nano-carbon leading to defined conductance changes on ligand–protein binding. This has recently been applied to generate an antibiotic resistance marker biosensor. BLIP2, a β-lactamase inhibitory protein which binds β-lactamases such as TEM-1, was engineered to contain *p*-azido-L-phenylalanine (Figure 2) at defined sites. The presence of the biorthogonal azide reaction handle allowed tethering of single-wall CNTs at defined orientations. The devices generated with the different BLIP2 variants were successful in detecting TEM-1 via changes in conductance, while the performance of the device is dependent on the defined BLIP2 attachment site and, in extent, orientation, including being able to the detect different electrostatic surface features of TEM-1 via the conductance profile.

Engineered GFP with *p*-azido-L-phenylalanine (Figure 2) incorporated at different residues have been utilized for the construction of nanoscale field effect transistors (FETs) – the key component of

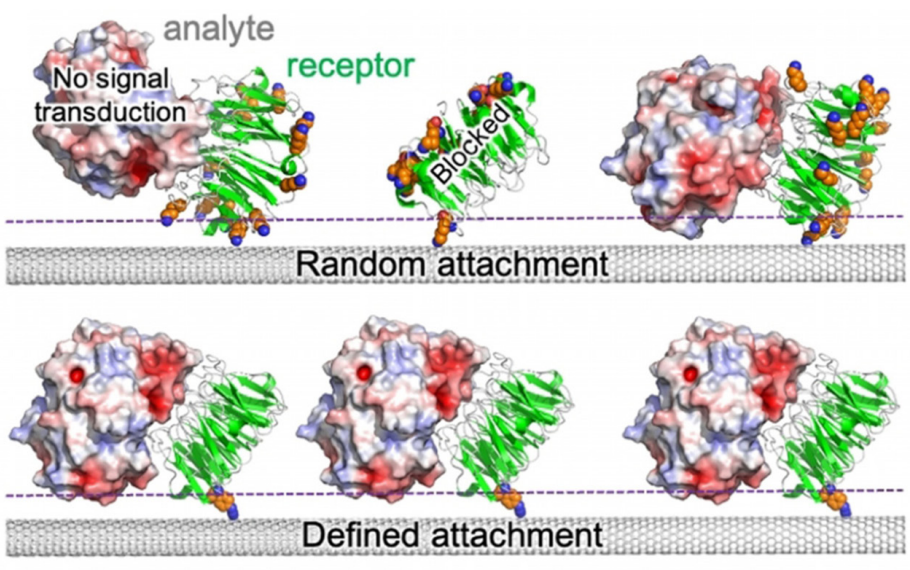

Figure 3. BLIP2 interfacing sites (green). Representation of the different BLIP2-CNT interfacing approaches when potential interface sites have standard amine attachment approaches (top) vs the attachment sites when biorthogonal nnAAs are incorporated instead (bottom). Shown at the top are different binding orientations that lead to random attachment where one has no communication with the CNT and the other blocks the binding of TEM-1 (red, white and blue shades). (Taken from Xinzhao Xu et al. 2021)

modern electronic devices. Photochemical attachment (Figure 2) was used to generate a direct and intimate link between the protein and nano-carbon. By sampling different attachment positions and thus orientations, we successfully generated a light responsive transistor or a memory system. The key to generating these two different functional affects was to attach the fluorescent protein at two different residues, something only possible using the expanded genetic approach.

It is crucial to understand the impact the incorporation of nnAAs, especially in this context, and how utilizing this defined and homogeneous attachment allowed the sampling of those distinctive conductance profiles. Applications such as these not only provide important information we couldn't get before but also show the significant potential of generating new clinical diagnostic devices and even potentially eco-friendly replacements for classically semiconductor-based electronic components.

The future of non-naturals

We are still exploring what we can achieve with nnAAs, and tools are still being developed that will facilitate and expand their use. The ones coming online at the moment that hold the greatest promise are the genome-edited bacteria that have certain codons fully removed from their genome, allowing more efficient incorporation of multiple nnAAs. The nnAAs also hold great promise in allowing proteins to break away from their biological context into new realms through new catalytic mechanism, generating useful bionanodevices and materials to define drug conjugates. They will also allow us to explore biological systems in new ways through defining post-translational modification states (e.g., phosphorylation and methylation straight off the ribosome) to regulating biological processes. ■

Further Reading

- Liu, C.C., and Schultz, P.G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem*. **79**, 413–444. 10.1146/annurev.biochem.052308.105824
- Xu, X., Bowen, B.J., Gwyther, R. E., et al. (2021). Tuning electrostatic gating of semiconducting carbon nanotubes by controlling protein orientation in biosensing devices. *Angew. Chem. Int. Ed*. **60**, 20184–20189. 10.1002/anie.202104044
- Xie, J., and Schultz, P.G. (2006). A chemical toolkit for proteins — an expanded genetic code. *Nat. Rev. Mol. Cell Biol*. **7**, 775–782. 10.1038/nrm2005
- Worthy, H.L., Auhim, H.S., Jamieson, W.D., et al. (2019). Positive functional synergy of structurally integrated artificial protein dimers assembled by Click Chemistry. *Commun. Chemistry* **2**. 10.1038/s42004-019-0185-5
- Haimovich, A.D., Muir, P., and Isaacs, F.J. (2015) Genomes by design. *Nat. Rev. Genet*., *16*(9), 501–516. 10.1038/nrg3956
- Baskin, J., and Bertozzi, C. (2007). Bioorthogonal click chemistry: Covalent labeling in living systems. *QSAR Comb. Sci*. **26**, 1211–1219. 10.1002/qsar.200740086
- Dien, V.T., Morris, S.E., Karadeema, R.J., and Romesberg, F.E. (2018). Expansion of the genetic code via expansion of the genetic alphabet. *Curr. Opin. Chem. Biol*. **46**, 196–202. 10.1016/j.cbpa.2018.08.009
- Hohsaka, T., and Sisido, M. (2002). Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol*. **6**, 809–815. 10.1016/s1367-5931(02)00376-9
- Kim, C.H., Axup, J.Y., and Schultz, P.G. (2013). Protein conjugation with genetically encoded unnatural amino acids. *Curr. Opin. Chem. Biol*. **17**, 412–419. 10.1016/j.cbpa.2013.04.017
- Reddington, S.C., Baldwin, A.J., Thompson, R., et al. (2015). Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chem Sci*. **6**, 1159–1166. 10.1039/c4sc02827a
- Reddington, S., Watson, P., Rizkallah, P., et al. (2013). Genetically encoding phenyl azide chemistry: New uses and ideas for classical biochemistry. *Biochem. Soc. Trans*. **41**, 1177–1182. 10.1042/bst20130094

Athena Zitt has finished her integrated master's degree in biochemistry at Cardiff University. Her research interest has always been on genetic and protein engineering with a focus on fluorescent proteins and their applications. She is currently working towards obtaining her PhD at Cardiff University under the supervision of professors Dafydd Jones, Paola Borri and Wolfgang Langbein and Dr Pete Watson. Her project focuses on developing and applying genetically encoded proteins, as pre-resonant coherent Raman scattering tags for next-generation imaging. Email: zittia@cardiff.ac.uk

Dafydd Jones is currently a professor of biochemistry in the School of Biosciences at Cardiff University. His group's research includes the construction of new protein components and scaffolds, including bionanohybrid systems. His group have developed approaches to engineer proteins to contain new chemistry that allows their precise interfacing with other biomolecules and nanomaterials such as nanocarbon. Email: jonesdd@cardiff.ac.uk

