Assembling a novel photosynthetic system that 'fills the green gap'

Author: Amelia Claxton

Abstract

Phototrophic organisms, including green plants and some bacteria, exploit solar energy to power electron transport chains involved in organic compound synthesis. Increasing demand for renewable energy alternatives has created unprecedented focus on the development and manipulation of natural photosynthetic systems to employ solar energy as a fossil fuel alternative. SpyTag/SpyCatcher was successfully deployed to spontaneously fuse a green plant light harvesting system (LHC) to a bacterial reaction centre (RC), forming a novel photosystem – dubbed a 'chimera'. Inspired to expand this chimera, the focus of this project was to evaluate suitable ligation systems: transglutaminase, sortase A, split inteins and SnoopTag/SnoopCatcher, for incorporation of a phycocyanin monomer into the photosynthetic network to 'fill the green gap' and utilize 550-650 nm solar energy.

Main body

Shifting global energy consumption from fossil fuels to renewable alternatives has grown in urgency over recent decades. CO_2 emissions have exceeded 400 ppm for the first time in 800,000 years as a result of mankind's dependency on fossil fuels; coal, oil and gas combustion contributed to 89% global CO_2 emissions in 2018[1,2]. One such renewable alternative is solar energy – an inexhaustible energy source of which 50,000 EJ is predicted to be easily harvestable[3]. Phototrophs, including green plants, algae and bacteria, naturally combine solar energy with electron transfer across photosynthetic systems to synthesise organic compounds[4]. Development and manipulation of these photosynthetic systems has become a paramount and environmentally attractive solution to addressing modern energy demands and cleaner energy consumption[5,6,7,8].

During photosynthesis, green plants, algae and some bacteria store solar energy within chemical bonds to be released during respiration. Systems composed of antenna complexes funnel photon energy to the reaction centre (RC) *via* resonance energy transfer across progressively red-shifted pigments. These antenna networks are comprised of photosensitive pigments; chlorophylls, bacteriochlorophylls, carotenoids and phycobiliproteins, which function to expand the energy scope a reaction centre can employ for charge separation. Within the RC, photon energy triggers a charge separation that initiates transmembrane electron transfer across the RC's cofactors, polarising the membrane into positive and negative poles. The electron holes are reset via oxidation of a donor substrate; in green plants

this donor is water. Bacteria and cyanobacteria differ from green plants in their photosynthetic systems, e.g. bacteria utilise bacteriochlorophyll RCs, although the principle of transmembrane electron transfer is analogous. The quantum efficiency (the efficiency of a one photon to one electron conversion) of natural photosynthesis is close to 100% - a consequence of this highly efficient electron transport system. Therefore, there are great interests in adopting photosynthetic proteins (direct use of entire photosynthetic systems is too challenging given their complexity) to solve our energy problems.

Utilizing the SpyTag/SpyCatcher system, Liu et al.[9] attached a green plant light harvesting complex (LHC) from *Arabidopsis thaliana* to a bacterial RC from *Rhodobacter sphaeroides*. This produced a unique polychromatic solar energy harvesting chimera, bridging chlorophyll (Chl) and bacteriochlorophyll (BChl) pigments, creating a photosynthetic system not seen in nature[9]. Within these chimeras, RC charge separation, usually dependent on 870 nm absorbance, was initiated by 650 nm excitation because of downhill resonance energy transfer from plant 650 nm absorbing LHC chlorophylls. Extending the spectral range of the energy harvested by the RC created a novel and more responsive photosynthetic system than those that nature provides. Inspired by the apparent ease of LHC-RC ligation, the focus is to expand this chimera to 'fill the green gap'[10] by incorporating a c-phycocyanin a-subunit (CpcA) into a heterotrimeric system. This project supported this venture by evaluating the most suitable protein ligation systems to aid future experimental designs.

The key consideration to forming this heterotrimeric system, using proteins that have no propensity to associate, is how the proteins will be covalently attached together so that they are close enough to warrant electron transfer but not too close to produce steric hinderance. SpyTag/SpyCatcher has proved successful at robustly ligating the LHC to the RC, but to introduce the c-phycocyanin to this chimera a different system must be used to prevent cross-reactivity. *In vitro* protein ligation techniques in the current literature guided this research, with 4 key methods proving the most suitable: transglutaminase, sortase A, split inteins and SnoopTag/SnoopCatcher.

Transglutaminase

Transglutaminase (TG) catalyses isopeptide bonds between glutamine y-carboxamide groups and lysine E-amino groups to synthesise macromolecular protein complexes[11]. Microbial TGs (mTG) are more preferred than mammalian TGs for protein ligation reactions because of their independency from Ca^{2+} and GTP, high robustness and cheaper purification[12]. The most prevalent mTGs are from *Streptomyces* and *Bacilli* bacterial strains due to their considerable

stability at a range of pH and temperatures[13]. Importantly, mTG is compatible in n-dodecyl *B*-D-maltoside (DDM) detergent[14]. mTGs have been exploited in various protein ligations ranging from antibody-drug conjugate synthesis[15] and poly(ethylene glycol) ligation[16] to chitosan-tissue binding *in vivo*[17].

mTG is fairly promiscuous when choosing Gln and Lys substrate residues. mTG requires Gln donor residues to reside in a consensus sequence within an easily accessible flexible loop region, else the efficiency and yield can be greatly effected[12,18]. Caporale et al. characterised short peptide sequences that support efficient mTG catalysis through synthetic libraries; of which the sequences LQSP (acyl donor) and KAYA-NH2 (acyl acceptor) revealed a 75% product formation in 60 minutes[19]. mTG Gln specificity is highly influenced by local secondary structure[20]. mTG was used to PEGylate apomyoglobin, which revealed the specific ligation of PEG to Gln91 despite 6 Gln residues being available - Gln91 corresponded to a locally unfolded region of protein[21]. Moreover, Caporale et al. realised successful mTG ligation only after the LQSP Gln donor was flanked by a flexible β-alanine linker[19]. Therefore, the Gln donor sequence will require synthesis into a highly flexible linker – mere exposure to the solvent, i.e. at the C-term, is not sufficient to support high reactivity (Figure 1).

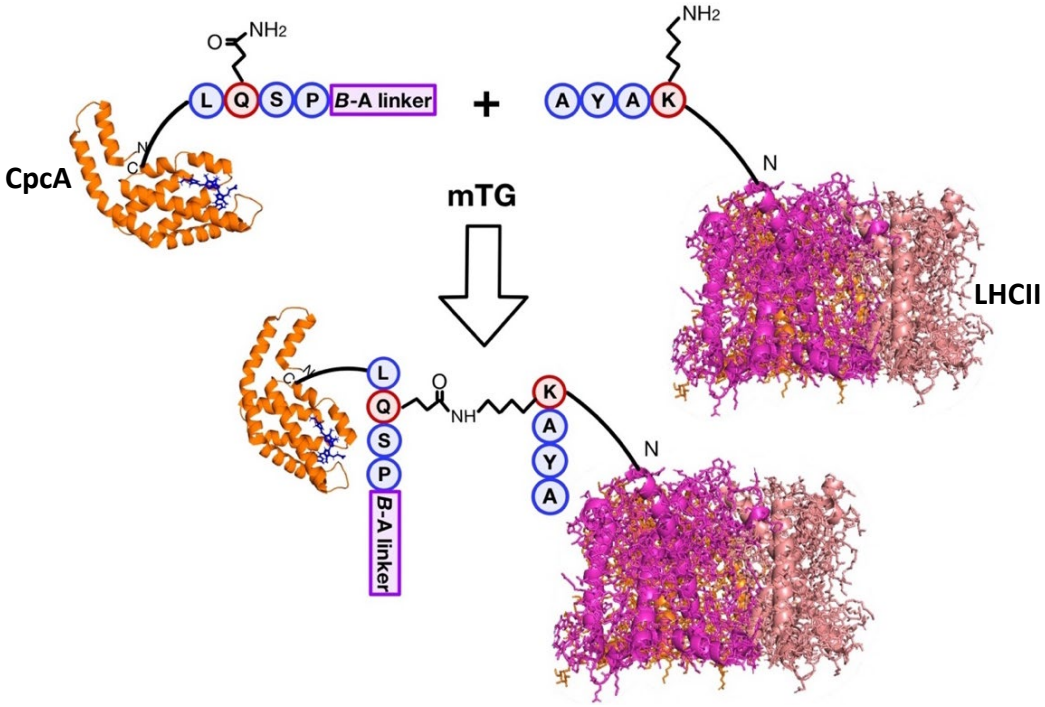

Figure 1: mTG mediated ligation of CpcA and LHCII through catalysis of covalent Gln-Lys bridges. *Specified by LQSP acyl donor and KAYA acyl acceptor tags, distinguished Gln y-carboxamide and Lys E-amino groups are favoured mTG substrates. (B-A linker = β-alanine linker).*

However, site-specific ligation between CpcA and LHC would not be guaranteed because of its promiscuity for native Gln residues residing in flexible regions. During antibody-drug conjugation, mTG mediated undesired Gln295 antibody ligation, even though the LLQGA Gln tag had been expressed on other Gln residues[15]. Additionally, mTG reveals very low substrate specificity regarding the Lys residue – any Lys exposed to the mTG active site will be used for ligation[18]. This unpredictable off-site ligation has led to few biochemical studies opting for mTG as a protein-based ligation method[22]. Whilst mTG low substrate specificity makes the enzyme versatile, for CpcA-LHC ligation this attribute is rather unfavourable.

Sortase A

Sortase enzyme (SrtA) originating from *Staphylococcus aureus* has received attention for its site-specific protein modifications and catalysis of peptide bonds[24]. In nature, SrtA joins surface proteins to peptidoglycan in bacterial cell walls, through recognition of an LPXTG sorting motif on the C-terminal[23,25]. SrtA cleaves the LPXTG motif between the threonine and glycine residues, prior to ligating onto an N-terminal oligoglycine sequence. For site-specific protein ligations, the LPXTG motif must be sequenced within a polypeptide linker that features at the end of the protein[26]. Examples of using SrtA for protein ligation include fluorescent tagging *in vivo*, chromatography-free protein purification and antibody-drug conjugations[27,28,29]. In particular, Chen et al. showcased successful SrtA mediated ligation of two proteins onto E2 nanocages, whilst upholding protein functionality and solubility[30].

SrtA mutant, R8, has shown good tolerance to DDM with 40-50% yield[31,32]. This would be the SrtA variant recommended for CpcA-LHC binding due to DDM compatibility and fair reaction yield. Compared to mTG mediated ligation, SrtA reacts only to LPXTG motifs, making the enzyme much more site-specific than mTG[23]. This is important to ensure that the CpcA-LHC is specifically ligated across the linker and not randomly. However, mTG ligation benefits from being non-reversible meaning that the reaction will reach completion rather than equilibrium, producing a higher yield. Although, SrtA reversibility can be countered by removing product as it forms or by introducing depsipeptide substrates. Therefore, between the two enzymatic ligation systems, sortase would be more appropriate at ensuring ligation is site-specific - paramount to successful resonance energy transfer from CpcA PCB fluorophores to LHCII chlorophylls. Addtionally, SrtA protein ligation offers certain benefits over the SnoopTag/SnoopCatcher and split intein systems. SrtA ligation would ligate CpcA to LHC with a smaller linker (the LPXTG motif) than the SnoopTag/SnoopCatcher system which is ~143 amino acids long[33]. The smaller the distance between CpcA and LHC, the more efficient the Förster resonance energy transfer (FRET) across the pigments[34]. Additionally, SrtA ligation

would require the fusion of smaller polypeptide sequences to CpcA and LHC compared to split inteins, which could lead to better expressions of recombinant constructs.

Whilst SrtA mediated protein ligation is site specific and introduces a small distance between CpcA and LHC, the low reaction efficiency of SrtA may outweigh these advantages. At 1:1 molar ratio of reactants, the fusion yield ranges from 30-65% - an 85% yield has shown to require a 20-fold excess of reactant[35]. The low yield of SrtA reactions is because of the reaction's reversibility; LPTX- ligation to an oligoglycine nucleophile reforms the original LPTXG- substrate, meaning the reaction reaches equilibrium rather than completion (Figure 2).

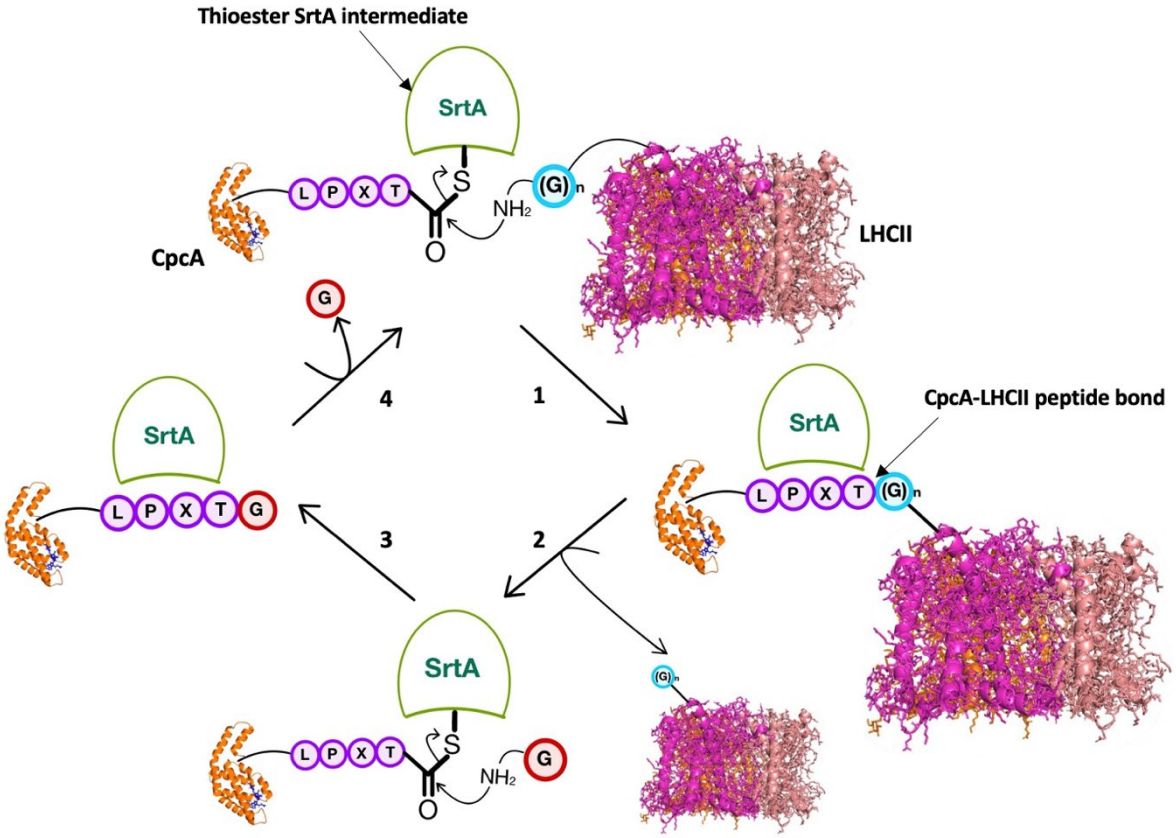

Figure 2: SrtA mediated ligation of CpcA and LHCII. *SrtA recognises the LPXTG motif on the C-term of CpcA and removes the glycine residue, producing a thioester SrtA intermediate. The polyglycine motif on LHCII undergoes nucleophilic attack on the intermediate; forming a peptide bond between CpcA and LCHII. The reformed LPXTG motif between CpcA and LHCII is subsequently recognised by SrtA, leading to cleavage of the desired product, reversing the reaction. PDB ID: CpcA - 4Z8K. LHCII - 1RWT.*

Approaches to improve yield include depsipeptide amino acid synthesis of the LPETG motif to prevent the reversible reaction. However, yield of depsipeptide substrates can be low and the protocol of depsipeptide synthesis would add considerable time and work to this study.

Moreover, CpcA monomers prove difficult to synthesise and purify due to their natural tendency to form hexameric structures[36]; excess phycocyanin monomers are difficult to obtain to counter low yield inefficiencies. Therefore, one-step ligation reactions like SnoopTag/SnoopCatcher and split inteins will prove more efficient and less time-consuming than SrtA mediated ligation.

Split Inteins

Inteins are protein introns, encoded by mobile intervening sequences. They autocatalytically excise from a polypeptide whilst covalently ligating the flanking sequences 'N-extein and C-extein' via a peptide bond[37]. Split inteins are more efficient than contiguous ones due to their lack of premature splicing, therefore only split inteins will be considered for binding CpcA to LHCII. Uses of split-intein mediated protein ligation include specific immunotoxin production, detecting protein-protein interactions and protein purification methods[38,39,40].

Nostoc punctiforme (Npu) DnaE split intein is the most utilised due to its fast ligation rates, 170-fold higher than *Synechocystis* sp. PCC6803 (Ssp) DnaE intein, and high efficiency (up to 90%, however, this is dependent on the extein sequence)[41]. Additionally, the Npu intein is compatible in DDM conditions[42]. Development of Npu split intein by *Stevens et. al* characterised the Cfa intein - a robust split intein that has a 2.5 fold faster splicing rate than Npu[43]. Therefore, the Cfa split intein would be the most recommended of the current split intein systems.

Split inteins are an efficient system for protein ligation as the added ligation motifs are excised. This is unlike SnoopTag/Catcher, SrtA and mTG ligation systems, which result in a short 'scar' at the excision site[37]. Therefore, the split intein system is the most attractive for conjoining CpcA and LHCII at the smallest distance to support efficient FRET[34]. Additionally, the Cfa intein is considerably more site-specific than mTG, and unlike SrtA is non-reversible. These qualities make the split intein system an effective choice for CpcA ligation to LHCII (Figure 3).

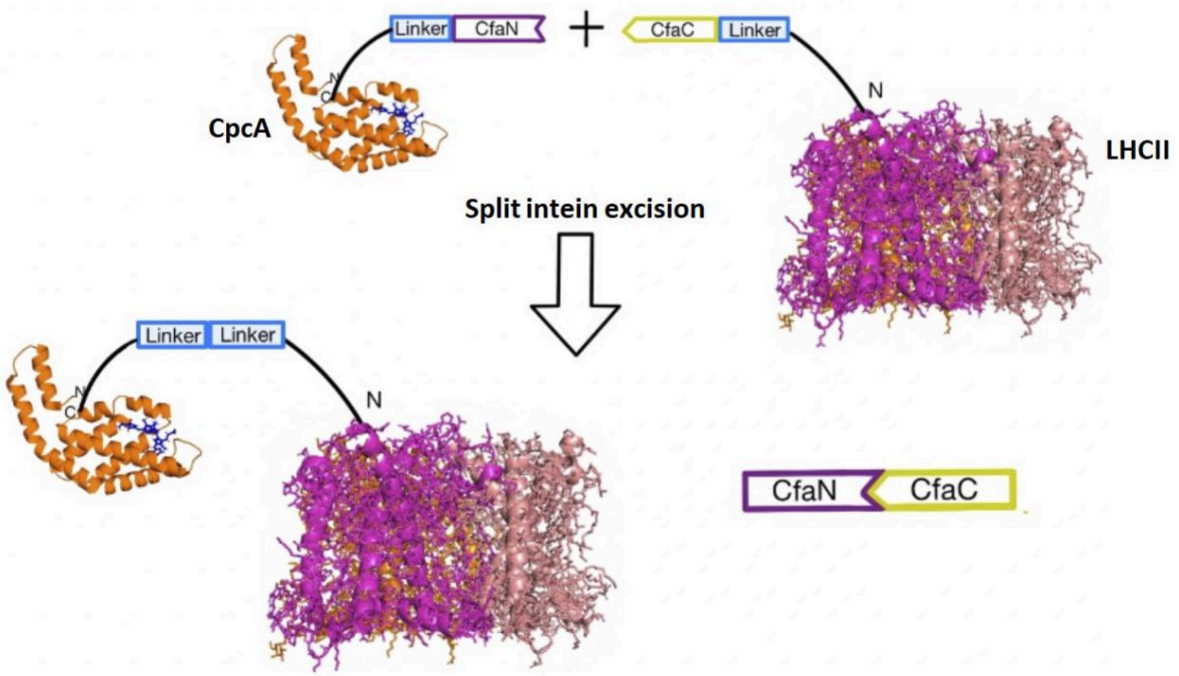

Figure 3: Cfa split intein catalysed ligation of CpcA-LHCII. *CfaN and CfaC autocatalytically excise causing covalent ligation of the extein sequences CpcA and LHCII. PDB ID: CpcA - 4Z8K. LHCII - 1RWT.*

However, the efficiency of split inteins has been revealed to decrease with increasing protein size. Despite the Npu intein system being one of the most robust, it's yield dramatically decreased upon ligating two Src homology 3 (SH3) domains[44]. LHCII monomers are ~70kDa, whilst the biggest SH3 domain used consisted of 91 amino acids, ~1/7th the kDa of LHCII[45]. Therefore, as the yield of Npu system suffered upon ligation of ~10kDa proteins, it is less likely that the Cfa intein system, derived from Npu, can support successful CpcA-LHCII ligation.

Overall, the split inteins are an attractive candidate for protein ligation due to their fast reaction rates, small 'scar' interference and high site-specificity[37]. However, their undocumented ligation of large proteins makes split inteins fairly unpredictable and breakdown of the intein system would cause by-product production rather than CpcA-LHCII.

SnoopTag/SnoopCatcher

Veggiani et al. composed SnoopTag/SnoopCatcher, an orthogonal SpyTag/SpyCatcher peptide-protein pair, through splitting *Streptoccus pneumoniae* (*S. pneumoniae*) RrgA adhesin complex into an 11 amino acid peptide (SnoopTag) and a 132 amino acid protein (SnoopCatcher)[33]. Upon incubation in TBS, 1.5M TMAO at pH 8.0, an isopeptide bond spontaneously forms between K742 (Tag) and N854 (Catcher), following a similar reaction

mechanism to SpyTag/SpyCatcher (Figure 4). Veggiani et al. showcased SnoopTag/SnoopCatcher autonomy from SpyTag/SpyCatcher by synthesising polyproteins using both systems[33]. Examples of SnoopTag/SnoopCatcher being used in site-specific protein ligations include synthesising polyprotein casettes, 2D nanosheets of SbsB and twin antigen VLPs[33,46,47].

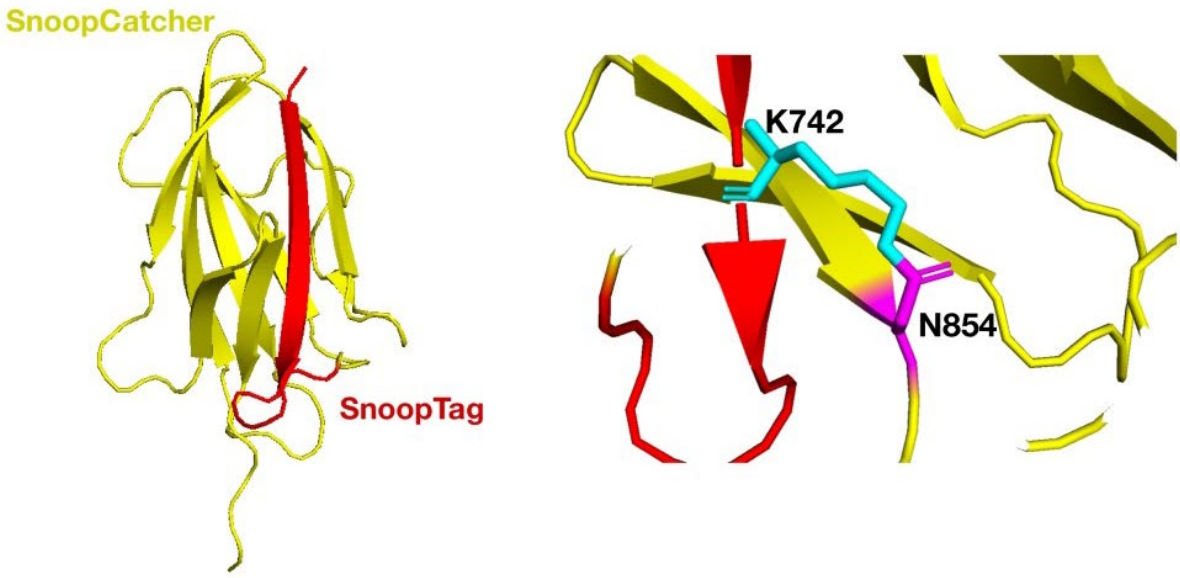

Figure 4: *X-ray crystal structure of SnoopTag (red) and SnoopCatcher (yellow) peptide-protein complex from the C-terminal domain of RrgA adhesin of S. pneumoniae. SnoopTag K742 residue (blue) and SnoopCatcher N854 (pink) undergo spontaneous isopeptide bond formation upon incubation. PDB ID: 2WW8.*

A 1:1 ratio of SnoopTag/SnoopCatcher substrates leads to 80% yield of the isopeptide product in 120 minutes[33]. Yield increases to 100% in 2:1 ratio and a 4:1 ratio reaches completion in just 40 minutes[33]. Whilst these rates are slower than SpyTag/SpyCatcher mediated ligation (half-time of ~1 minute[48]), the reaction is compatible with a range of pH (6-9), temperatures (4-37°), buffers and detergents[33].

Compared to SrtA and mTG ligation systems, the SnoopTag/SnoopCatcher offers a higher site-specificity and reproducibility. SnoopTag/SnoopCatcher mediates a one-step isopeptide bond formation, the substrates are highly specific to each other and the mechanism doesn't involve intermediate species - unlike the enzyme catalysed ligation systems. Moreover, split inteins may be unable to ligate the large LHCII construct to CpcA without producing considerable aggregates. Phycocyanin monomers prove difficult to synthesise and purify due to their natural tendency to form hexameric structures[36]; further supporting the requirement for a system that supports high yields of CpcA-LHCII and not by-products.

However, an important drawback to SnoopTag/SnoopCatcher is the sizeable distance that would separate CpcA and LHCII. When fused, the SnoopTag/SnoopCatcher construct would introduce ~143 residue long distance, as well as the linkers[33]. This is an important consideration as energy transfer from CpcA to LHCII is via FRET – the efficiency of which is notoriously dependent on distance[34]. In contrast, SrtA leaves ~5 amino acid gap, whilst split inteins introduce only three. Therefore, out of the considered ligation systems, SnoopTag/SnoopCatcher may form chimeras with the lowest FRET efficiency. In comparison, SnoopTag/SnoopCatcher is 27 residues longer than SpyTagΔ/SpyCatcherΔ, as used by Liu et al.[9]. The considerable spectral overlap of PCB emission with LHCII and RC absorbance may counteract this 27 residue increased distance[34].

Conclusions

To conclude, the most appropriate system currently available to produce a novel photosynthetic system based on protein ligations is SnoopTag/SnoopCatcher. SpyTag/SpyCatcher is already employed in LHCII-RC ligation and therefore a different system must be used to prevent cross-reactivity. The mTG protein ligation system reveals the lowest substrate selectivity and therefore the least reliable at producing site-specific and reproduceable constructs. SrtA is more specific than mTG but catalyses a reversible reaction, making high yield difficult to control. In contrast, split inteins are highly specific and catalyse a one-step reaction. However, split inteins prove unreliable in ligating larger proteins – considering that LHCII is a large membrane protein, the split intein system is fairly unpredictable. Therefore, the SnoopTag/SnoopCatcher is the most attractive system for synthesising site-specific constructs at reliably high yields. The compromised FRET functionality is an important consideration, although the sizeable PCB emission overlap with LHCII and RC will likely counteract this.

Potential problems do still exist when implementing this system. Even though SnoopTag/SnoopCatcher has proved functional in a range of detergents, pH and temperatures, its lack of considerable publications could mean it proves incompatible with certain ligations, such as LHCII, unlike its orthogonal pair SpyTag/SpyCatcher. Another potential drawback is that PCB fails to transfer energy to LHCII-RC caused by the sizeable distance between PCB and ChlA fluorophores as a consequence of using the SnoopTag/SnoopCatcher ligation system. FRET inefficiencies could be overcome by mutating SnoopTag/SnoopCatcher into smaller constructs; however, if this doesn't prove possible then

a different ligation system may be more successful. As reviewed, the split intein system would prove more successful in this aspect.

Future perspectives

Reliable synthesis of a novel photosynthetic system that harvests blue, green and red light (300-900 nm) would provide a more responsive system for photodevices, potentially leading to better functionality. Review of available protein-based ligation systems prompted the conviction that SnoopTag/SnoopCatcher is the most appropriate for assured reliability, high yields and site-specificity. Employing SnoopTag/SnoopCatcher will be outlined in future experiments, likely incorporating synthetic biology techniques to express CpcA-SnoopTag and LHCII-SnoopCatcher in separate *Escherichia coli* populations. Validation of CpcA-LHCII ligation could be determined by reverse-phase high pressure liquid chromatography (RP-HPLC), sucrose density gradients and blue native-polyacrylamide gel electrophoresis, which will reveal a higher mass compound (CpcA-LCHII) if ligation is successful. CpcA-LHCII ligations could then be integrated into the existing LHCII-RC chimera as SnoopTag/SnoopCatcher is autonomous from SpyTag/SpyCatcher, creating a novel system with no 'green gap'.

Looking forward, *in vitro* synthesis of novel photosynthetic systems has led to the development of *in vivo* novel systems, notably the manufacture of bacterial solar cells[49,50]. Engineering recombinant microorganisms with highly responsive photosystems offers exciting prospects to tackling future energy demands.

References

1. Ritchie, H. & Roser, M. CO_2 and Greenhouse Gas Emissions. *Our World in Data* (2017).
2. Trends in global CO2 and total greenhouse gas emissions: 2019 Report. 70.
3. Kabir, E., Kumar, P., Kumar, S., Adelodun, A. A. & Kim, K.-H. Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews* **82**, 894–900 (2018).
4. Baker, L. A. & Habershon, S. Photosynthetic pigment-protein complexes as highly connected networks: implications for robust energy transport. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**, 20170112 (2017).
5. Kanygin, A. *et al.* Rewiring photosynthesis: a photosystem I-hydrogenase chimera that makes H2in vivo. *Energy Environ. Sci.* **13**, 2903–2914 (2020).
6. Batista-Silva, W. *et al.* Engineering Improved Photosynthesis in the Era of Synthetic Biology. *Plant Communications* **1**, 100032 (2020).

7. Ort, D. R. *et al.* Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *PNAS* **112**, 8529–8536 (2015).
8. Sokol, K. P. *et al.* Bias-free photoelectrochemical water splitting with photosystem II on a dye-sensitized photoanode wired to hydrogenase. *Nature Energy* **3**, 944–951 (2018).
9. Liu, J., Friebe, V. M., Frese, R. N. & Jones, M. R. Polychromatic solar energy conversion in pigment-protein chimeras that unite the two kingdoms of (bacterio)chlorophyll-based photosynthesis. *Nature Communications* **11**, 1542 (2020).
10. Croce, R. & van Amerongen, H. Natural strategies for photosynthetic light harvesting. *Nature Chemical Biology* **10**, 492–501 (2014).
11. Strop, P. Versatility of Microbial Transglutaminase. *Bioconjugate Chem.* **25**, 855–862 (2014).
12. Ando, H. *et al.* Purification and Characteristics of a Novel Transglutaminase Derived from Microorganisms. *Agricultural and Biological Chemistry* **53**, 2613–2617 (1989).
13. Schneider, H., Deweid, L., Avrutina, O. & Kolmar, H. Recent progress in transglutaminase-mediated assembly of antibody-drug conjugates. *Analytical Biochemistry* **595**, 113615 (2020).
14. Takahara, M. *et al.* Enzymatic Cell-Surface Decoration with Proteins using Amphiphilic Lipid-Fused Peptide Substrates. *Chemistry – A European Journal* **25**, 7315–7321 (2019).
15. Farias, S. E. *et al.* Mass Spectrometric Characterization of Transglutaminase Based Site-Specific Antibody–Drug Conjugates. *Bioconjugate Chem.* **25**, 240–250 (2014).
16. Mero, A., Spolaore, B., Veronese, F. M. & Fontana, A. Transglutaminase-Mediated PEGylation of Proteins: Direct Identification of the Sites of Protein Modification by Mass Spectrometry using a Novel Monodisperse PEG. *Bioconjugate Chem.* **20**, 384–389 (2009).
17. Fernandez, J. G. *et al.* Direct Bonding of Chitosan Biomaterials to Tissues Using Transglutaminase for Surgical Repair or Device Implantation. *Tissue Engineering Part A* **23**, 135–142 (2016).
18. Rachel, N. M. & Pelletier, J. N. Biotechnological Applications of Transglutaminases. *Biomolecules* **3**, 870–888 (2013).
19. Caporale, A. *et al.* The LQSP tetrapeptide is a new highly efficient substrate of microbial transglutaminase for the site-specific derivatization of peptides and proteins. *Biotechnol J* **10**, 154–161 (2015).
20. Spolaore, B. *et al.* Local Unfolding Is Required for the Site-Specific Protein Modification by Transglutaminase. *Biochemistry* **51**, 8679–8689 (2012).
21. Fontana, A., Spolaore, B., Mero, A. & Veronese, F. M. Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Advanced Drug Delivery Reviews* **60**, 13–28 (2008).

22. Chan, S. K. & Lim, T. S. Bioengineering of microbial transglutaminase for biomedical applications. *Appl Microbiol Biotechnol* **103**, 2973–2984 (2019).
23. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *PNAS* **96**, 12424–12429 (1999).
24. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
25. Ton-That, H. & Schneewind, O. Anchor Structure of Staphylococcal Surface Proteins IV. INHIBITORS OF THE CELL WALL SORTING REACTION. *J. Biol. Chem.* **274**, 24316–24320 (1999).
26. Refaei, M. A. *et al.* Observing selected domains in multi-domain proteins via sortase-mediated ligation and NMR spectroscopy. *J Biomol NMR* **49**, 3–7 (2011).
27. Strijbis, K., Spooner, E. & Ploegh, H. L. Protein Ligation in Living Cells Using Sortase. *Traffic* **13**, 780–789 (2012).
28. Bellucci, J. J., Amiram, M., Bhattacharyya, J., McCafferty, D. & Chilkoti, A. “Three-in-one” chromatography-free purification, tag removal, and site-specific modification of recombinant fusion proteins using Sortase A and elastin-like polypeptides. *Angew Chem Int Ed Engl* **52**, 3703–3708 (2013).
29. Beerli, R. R., Hell, T., Merkel, A. S. & Grawunder, U. Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency. *PLOS ONE* **10**, e0131177 (2015).
30. Chen, Q. *et al.* Sortase A-mediated multi-functionalization of protein nanoparticles. *Chem. Commun.* **51**, 12107–12110 (2015).
31. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **108**, 11399–11404 (2011).
32. Li, J. *et al.* Optimization of sortase A ligation for flexible engineering of complex protein systems. *J Biol Chem* **295**, 2664–2675 (2020).
33. Veggiani, G. *et al.* Programmable polyproteams built using twin peptide superglues. *PNAS* **113**, 1202–1207 (2016).
34. Berney, C. & Danuser, G. FRET or No FRET: A Quantitative Comparison. *Biophys J* **84**, 3992–4010 (2003).
35. Levary, D. A., Parthasarathy, R., Boder, E. T. & Ackerman, M. E. Protein-Protein Fusion Catalyzed by Sortase A. *PLOS ONE* **6**, e18342 (2011).
36. Stadnichuk, I. N., Krasilnikov, P. M. & Zlenko, D. V. Cyanobacterial phycobilisomes and phycobiliproteins. *Microbiology* **84**, 101–111 (2015).
37. Shah, N. H. & Muir, T. W. Inteins: Nature’s Gift to Protein Chemists. *Chem Sci* **5**, 446–461 (2014).

38. Pirzer, T. *et al.* Generation of Potent Anti-HER1/2 Immunotoxins by Protein Ligation Using Split Inteins. *ACS Chem. Biol.* **13**, 2058–2066 (2018).
39. Yao, Z. *et al.* Split Intein-Mediated Protein Ligation for detecting protein-protein interactions and their inhibition. *Nature Communications* **11**, 2440 (2020).
40. Lu, W. *et al.* Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage. *J Chromatogr A* **1218**, 2553–2560 (2011).
41. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Letters* **583**, 909–914 (2009).
42. Mehler, M. *et al.* Assembling a Correctly Folded and Functional Heptahelical Membrane Protein by Protein Trans-splicing. *J Biol Chem* **290**, 27712–27722 (2015).
43. Stevens, A. J. *et al.* Design of a Split Intein with Exceptional Protein Splicing Activity. *J. Am. Chem. Soc.* **138**, 2162–2165 (2016).
44. Aranko, A. S., Züger, S., Buchinger, E. & Iwaï, H. In Vivo and In Vitro Protein Ligation by Naturally Occurring and Engineered Split DnaE Inteins. *PLOS ONE* **4**, e5185 (2009).
45. Pagliano, C., Barera, S., Chimirri, F., Saracco, G. & Barber, J. Comparison of the α and β isomeric forms of the detergent n-dodecyl-D-maltoside for solubilizing photosynthetic complexes from pea thylakoid membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**, 1506–1515 (2012).
46. Manea, F., Garda, V. G., Rad, B. & Ajo-Franklin, C. M. Programmable assembly of 2D crystalline protein arrays into covalently stacked 3D bionanomaterials. *Biotechnology and Bioengineering* **117**, 912–923 (2020).
47. Brune, K. D. *et al.* Dual Plug-and-Display Synthetic Assembly Using Orthogonal Reactive Proteins for Twin Antigen Immunization. *Bioconjugate Chem.* **28**, 1544–1551 (2017).
48. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *PNAS* **109**, E690–E697 (2012).
49. Paul, N. *et al.* Self-powered all weather sensory systems powered by Rhodobacter sphaeroides protein solar cells. *Biosensors and Bioelectronics* **165**, 112423 (2020).
50. J. McCormick, A. *et al.* Photosynthetic biofilms in pure culture harness solar energy in a mediatorless bio-photovoltaic cell (BPV) system. *Energy & Environmental Science* **4**, 4699–4709 (2011).