**University of Bristol**

**School of Biochemistry**

**Methods for Producing a Heterooligomeric**

**Complex of Photosynthetic Proteins from Different Species: An Appraisal**

Fergus Teague

Research Project Dissertation

Dr Mike Jones

**Word count: 3468**

This project is all my own work except where indicated. All text, figures and data which are not my own work are indicated and the sources acknowledged.

**Signature: Fergus Teague Date: 17/04/2022**

# Abstract

**The climate crisis is one of the most pressing matters threatening humanity, with our reliance on fossil fuels pushing many natural systems to the brink of collapse. This concern has driven research into a variety of new technologies utilising organic components capable of harvesting sunlight and converting this to useful energy. In doing so, numerous teams have created novel biohybrid and bioinspired systems for use *in vitro*. In this paper we build on work by Liu *et al*. (2020) and aim to orthogonally link three photosynthetic proteins from different species, producing a singular heterooligomeric complex. Specifically, we aim to link light harvesting complex I and II from *Arabidopsis* (*A*.) *thaliana* with the reaction centre from the purple proteobacterium *Rhodobacter sphaeroides*. Several potential methods for doing so are discussed, including their respective strengths and limitations, before a research plan is outlined for linking photosynthetic proteins using the SpyCatcher/SpyTag and SnoopCatcher/SnoopTag peptide systems. Potential issues with this plan are detailed, as is the impact of this work on the synthetic biology and biotechnology fields.**

Table of Contents

[Abstract 2](#_Toc101780601)

[List of Abbreviations: 4](#_Toc101780602)

[Introduction 5](#_Toc101780603)

[Current Methods for Linking Proteins: A Review 6](#_Toc101780604)

[Split Inteins 6](#_Toc101780605)

[SpyTag Systems 7](#_Toc101780606)

[Coiled-Coil Driven Assembly 8](#_Toc101780607)

[Hypothesis and Objectives 9](#_Toc101780608)

[Research Proposal 9](#_Toc101780609)

[Expression and purification of the adapted LHCI, LHCII, and RC. 9](#_Toc101780610)

[Production of Complexes and Purification 9](#_Toc101780611)

[Evaluation of Products 10](#_Toc101780612)

[Transmission Electron Microscopy 10](#_Toc101780613)

[Spectroscopic Analysis 10](#_Toc101780614)

[Timeline 11](#_Toc101780615)

[Discussion 11](#_Toc101780616)

[Acknowledgements 12](#_Toc101780617)

[References: 13](#_Toc101780618)

# List of Abbreviations:

**SpyCatcher/SpyTag (SC/ST)**

**Protein-*trans* splicing (PTS)**

**DNA polymerase III (DnaE)**

***Synechocystis sp.* (Ssp)**

***Nostoc punctiforme* (Npu)**

***Rhodobacter sphaeroides* (Rbs)**

**Light harvesting complex I (LHCI)**

**Light harvesting complex II (LHCII)**

**Reaction centre (RC)**

**SnoopCatcher/SnoopTag (SnC/SnT)**

**Green fluorescent protein (GFP)**

**Wild type (WT)**

**Gel filtration chromatography (GFC)**

**Phycobiliproteins (PBP)**

# Introduction

Technological advances and an ever-growing population have forced the demand for energy beyond a sustainable point and brought on a climate crisis that has seen more extreme weather, global temperatures rising, and the threatening of countless species (1). Humanity is too dependent on finite fossil fuels; in 2021 it is estimated their burning released 36.4 billion tonnes of carbon (2). Not only do they contribute to global warming more than any other industry (3), but their use is expected to continue to increase during this century (4), directly opposing necessary climate action to keep temperatures below 1.5 degrees Celsius relative to pre-industrial levels (5). To prevent an irreversible climate disaster, human energy consumption must be shifted to renewable and sustainable sources as soon as possible (6).

Bio-inspired technology and synthetic biology will play a key role in developing new strategies to tackle many problems currently faced, for example, the environmental cost of increased agricultural output. Estimates for food production demand in 2050 outline a need for agricultural productivity to increase by upwards of 110% more than levels seen in 2005 (7). A desire for greater efficiency has led to a dependency on agricultural chemicals: pesticides, herbicides, and fertilisers. Several known groups of agrochemicals are no longer used in the industry due to their harmful effects (8) but legislation varies between countries, and it is difficult to uniformly ban their use (9). Discrepancies between countries has created the need for environmental monitoring systems to track levels of potentially harmful agrochemicals present in various ecosystems. New bio-inspired technology has shown to have a potential role in environmental monitoring (10,11). This specialist equipment has the potential to be powered by energy produced by biological components and have the quality of being easily disposable.

Many modern pieces of technology, classically smartphones, have become dependent on precious metals, a factor that is detrimental to the environment for two reason: their mining has a direct negative impact on the environment (12), and the recycling of these components is difficult and environmentally damaging (13). Advancements in photovoltaics have led to the development of photoelectrochemical cells capable of producing a current and voltage from a nearly completely biological system (14). Strides have also been made in the field of biomimetics. The use of biological or bio-inspired components in batteries is becoming a realistic way to bypass the limits of their crystal structure and constituting components (15).

In Liu et al. (2020) (16) a protein system composed of a bacterial reaction centre with either plant light harvesting complex I or II was created. This opened the door to creating more complex systems combining several other photosynthetic proteins. For example, utilising phycobiliproteins will complement the use of plant light harvesting systems and cover a greater range of the visible spectrum. To incorporate additional proteins requires a new system for linking proteins. Several attractive methods can be used to produce multi-protein systems in a genetically encodable fashion. The advantage of these systems over chemical ligation is that no additional reagents are needed, the proteins retain a more natural conformation, and they will spontaneously assemble when mixed. In this paper three techniques will be discussed, highlighting their uses and drawbacks, finally outlining a method used to create a novel heterotrimeric photosystem.

# Current Methods for Linking Proteins: A Review

## Split Inteins

Diagram

Description automatically generatedInteins and their partner domains, exteins, are the protein analogues of introns and exons. They possess self-splicing properties capable of removing the intein section through lysing the two peptide bonds flanking it and joining together the extein portion of the protein with a new peptide bond (17). This technique is natures example of a classic chemical synthesis process, native chemical ligation (18). Nature has found use for this mechanism in every domain (19) but was first discovered in *Saccharomyces cerevisiae* – bakers yeast(20). While inteins initially found use in expressed protein ligation – recombinant proteins with a C-terminal thioester that could undergo native chemical ligation (21) – it was the discovery of split inteins, catalysing protein-*trans* splicing (PTS), which opened possible avenues of discovery (22). A DNA polymerase III (DnaE) from *Synechocystis sp.* (Ssp) PCC6803 was seen to consist of two parts and is thought to be the first example of a naturally split intein. The mechanisms by which inteins and split inteins work can be seen in fig 1.

Figure 1: mechanistic overview of intein self-splicing, formation of a new peptide bond and scarless joining of the two extein regions. Split inteins first form a heterodimer by recombining before following the same pathway as inteins. Created with BioRender.com.

Under the control of genetics, the system could be manipulated easily, leading to several functions developing. Cyclisation of proteins created enzymes that are resistant to cellular metabolism and recycling (23); using DnaB of *Ssp*, recombinantly produced proteins that could be purified with tag affinity, and crucially have the tags removed via their self-splicing properties (24); protein semisynthesis can be harnessed to produce proteins with a lipid anchor attached (25), capable of embedding recombinant proteins in membranes. Unavoidable drawbacks have prevented the widespread use of split inteins, however. The flanking extein is often required to be identical to their native sequences to function at a kinetically useful rate (18). While a more promiscuous version of a DnaE homologue from *Nostoc punctiforme* (Npu) capable of withstanding changes to the extein sequences was produced, using directed mutagenesis to alter each desired split inteins to retain splicing activity is not a viable option. Further, the conserved nature of DnaE meant cross reactivity issues plagued attempted simultaneous uses of split inteins. For example, the same DnaE homologue from Npu was seen to have a higher catalytic rate (26), but cross-reactivity issues meant it could not be used with other homologues (27). A paper by Pinto *et al*. (2020) identified ten inteins that showed PTS activity without cross-reactivity; however, this was not without issues as well. The extein sequence was shown to still play a role in effecting the reactivity and catalytic potential, and the team also noted that the solubility of the target chimeric protein will affect the activity of the intein too (28).

## SpyTag Systems

SC/ST is a system derived from FbaB, a fibronectin binding protein from *Streptococcus pyogenes* – specifically, it comes from the CnaB2 domain. CnaB2 contains a crucial isopeptide bond, an amide bond formed via side chain carboxyl and amine interactions, that conveys high levels of stability in the protein (29). The bond forms between the amine group of lysine-31 and carboxyl group of aspartate-117, with glutamate-77 acting as a catalyst for the nucleophilic attack (figure 2). Through artificial splitting of this domain at the bond, and rational optimisation of the two components, a system capable of reconstitution was created (30).

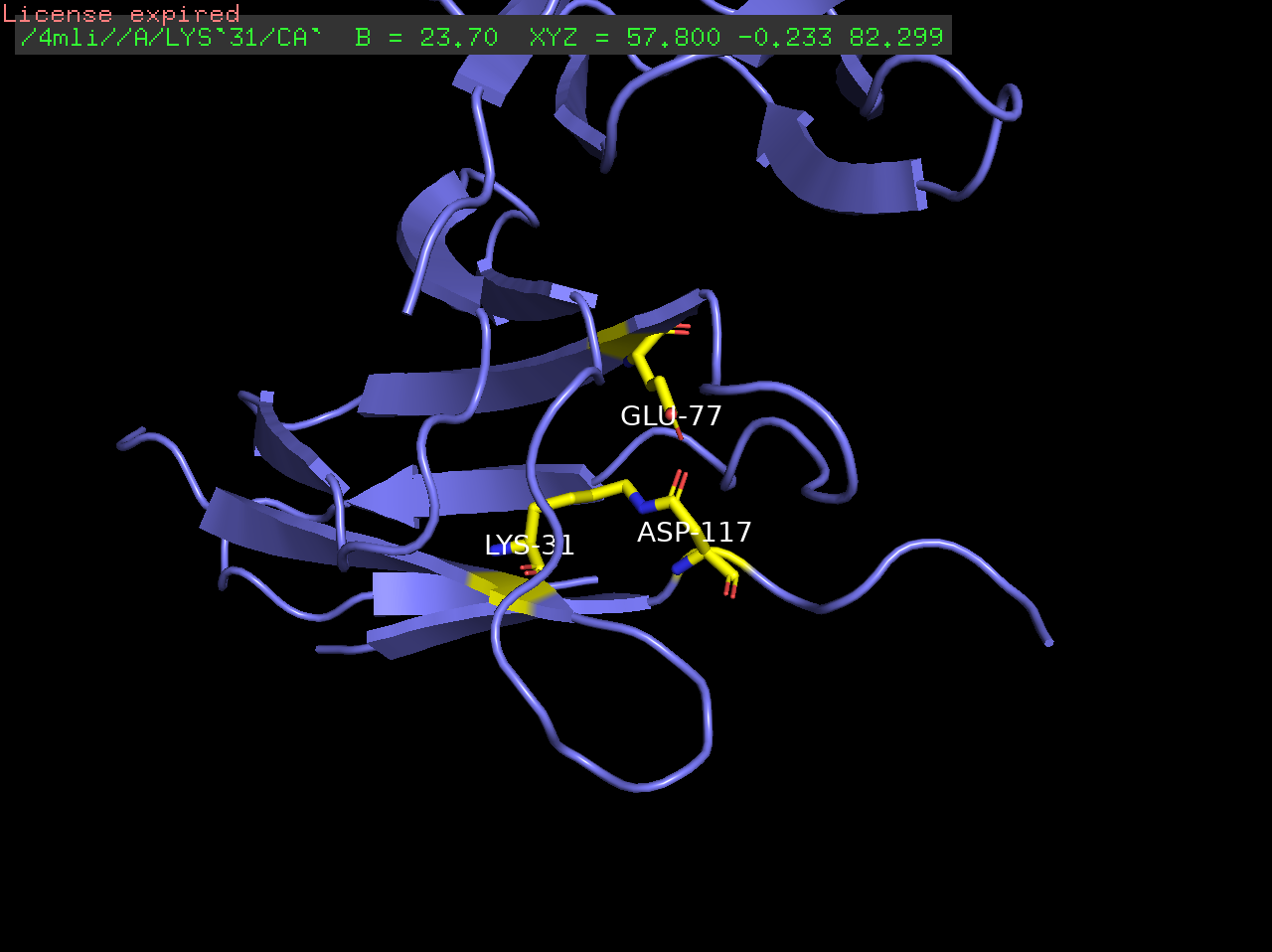


Figure 2: PyMol representation of the two reactive residues Lys-31 and Asp-117 and their isopeptide bond, as well as the catalytic Glu-77 (left). Highlighted in yellow is SpyTag with the Asp-117, while SpyCatcher is in blue with the other two residues (right).

The system has been utilised for several reasons; chief among them in this scenario is the assembly of an oligomeric system (31,32). The short SpyTag peptide can added to the terminal domains of proteins, as well as internally, and an adapted shorter SpyCatcher (33) protein being added to another, either directly or through linker regions if flexibility in the final product is necessary. SC/ST was the system used in the original Liu et al. (2020) paper (34) and was shown to be effective in producing chimeric proteins combining the purple bacterial reaction centre of *Rhodobacter sphaeroides* (Rbs) and light harvesting complexes I and II (LHCI and LHCII) from *Arabidopsis* (*A*.) *thaliana.* While promising, the system is limited to only joining two proteins together if performed in a one-pot reaction.

Instead, a combinatorial approach using SnoopCatcher/SnoopTag (SnC/SnT), alongside the original SC/ST could be used. SnC/SnT was developed similarly to SC/ST: RrgA is a subunit from a pilus adhesin protein from *Streptococcus pneumoniae* (35) and contains a D4 Ig-like domain. This domain houses the isopeptide bond between lysine-742 and asparagine-854. SnoopTag’s reactive lysine contrasts SpyTag’s aspartic acid and means no cross-reactivity is seen between the two systems, allowing them to be used simultaneously (36). A major advantage with these systems, when comparing them to alternative orthogonal systems, is the robustness of the covalent isopeptide bonds that are formed. Both systems can be used in several buffers, in the presence of common detergents, and convey resistance to boiling and provide the protein complex with mechanical strength (29,36).

## Coiled-Coil Driven Assembly

*De novo* protein design is the bottom-up approach to designing and building proteins (37). Coiled-coils (CC), first pioneered *de novo* by Bob Hodges (38), can be defined as multiple amphipathic helices that combine to bury their hydrophobic residues, forming a supercoiled helical structure. Naturally occurring coiled-coils have found roles in restriction enzymes (39), chromosome segregation (40), cargo transport (41), and structural proteins (42,43). The basic structure of CC is a heptad repeat composed of hydrophobic and polar residues in the pattern: HPPHPPP (37), and their sequence to structure relationship is well defined. This has led to their use in oligomeric assemblies on several occasions, as well as the recombination of peptides to reform functional proteins. Two highly cited papers have used antiparallel leucine heterodimers, a common CC motif, to successfully reconstitute green fluorescent protein (GFP) (44,45) , returning its fluorescent property. Achieving the desired oligomerisation state is key to building multi-protein complexes. Attachment of intact GFP to the termini of CC subunits has been shown to cause issues in their ability to combine and form their expected oligomerisation state (46). Simply by increasing the length of the linking region between the CC subunit and GFP was sufficient to rectify this in two of four cases however, a point echoed in a paper investigating the use of CC to produce synthetic transcription factors (47). As is to be expected, when adding a large protein to the end of a CC subunit, linker length will play a key role in allowing the correct folding of both the CC subunit and the “cargo” protein. The concept of using coiled-coils to bring together photosynthetic proteins is once that has been previously explored (48). It was shown that two reaction centres from different strains of *Rbs* could be brought together *in vitro* to form hetero-oligomers, an indication that different proteins could be brought together using this method to form a “hybrid solar energy conversion system”.

# Hypothesis and Objectives

We aim to use an orthogonal linking system to produce a heterooligomeric protein assembly consisting of three photosynthetic proteins originating from different species. Specifically, LHCI and LHCII from (*A*.) *thaliana* combined with the RC from *Rhodobacter sphaeroides*. Crucially we aim to produce this complex in a genetically encodable fashion with a pre-determined stoichiometry of 1:1:1, and with the ability to self-assemble in solution. This will be achieved using a combination of the SC/ST and SnC/SnT protein linking systems. These systems are well documented and understood while also showing a certain robustness, necessary when using large proteins dissolved in a detergent solution.

# Research Proposal

## Expression and purification of the adapted LHCI, LHCII, and RC.

We will produce two forms of LHCI: one wild type (WT) and one consisting of an Lhca4 adapted at the N-terminus with a His-tag and SpyTag. LHCI is a heterodimer so we will also include the expression of Lhca1 that is not adapted with SpyTag or a His-tag. These will be modified genes from (*A*.) *thaliana* producing an apoprotein. *E. coli* will be used to produce the apoproteins after transfection with the pET-28a expression plasmid. Inclusion bodies containing the apoproteins will be purified from cell lysate, dissolved, and mixed with purified pigments chlorophyl A, B, and carotenoids, obtained from spinach leaves (as per Liu *et al.* [2020]). Nickel chromatography will be used to purify the refolded protein away from unreacted Lhca1 before gel filtration chromatography (GFC) is used to further purify according to size. The A470:A674 ratio upon excitation will be used to evaluate the concentration of protein in the fraction and the lowest ratios will be kept.

LHCII will follow a similar production method, using *E. coli* as an expression system for the apoprotein, before refolding with the necessary pigments. LHCII will also have a WT version coded by the Lhcb1 gene from (*A*.) *thaliana*, with a His-tag added to the C-terminus*.* The adapted version of LHCII will consist of a shortened Lhcb1 gene, missing 12 amino acids at the N-termini, and a His-tag and SnoopTag immediately before the start codon for Lhcb1. The purification process and evaluation of protein content will be performed in the same was as for LHCI but with the appropriate extinction coefficient for LHCII.

To adapt the RC, a His-tag and SnC will be added to the expression plasmid immediately before the start codon for the PufL subunit and SC will be added immediately after PufM subunit, attaching it to the C-terminus end rather than the more obstructed N-terminus that would require a linker sequence (48). Again, a WT version of the apoprotein will be expressed with a His-tag at the C-terminus of PufM. Both will be expressed separately in a Rbs strain deficient in light-harvesting proteins. Similarly to LHC nickel affinity followed by gel filtration chromatography will be used to purify the products from cell lysate.

## Production of Complexes and Purification

From here we will perform a one-pot reaction of ST#LHCI, SnT#LHCII, and SnC#RC#SC (the # indicate the position of the isopeptide bond) to make a heterotrimeric complex (figure 3). We will mix a three-fold molar excess of LHCI with a two-fold molar excess of LHCII compared to that of RC. GFC will be used to separate unreacted protein from the complex. Absorbance spectroscopy will be used to analyse the stoichiometry of the fractions.

Diagram

Description automatically generated

Figure 3: graphic representation of the base unit vthat we aim to produce using combinations of the SC/ST and SnC/SnT. Created with BioRender.com

## Evaluation of Products

Sucrose density gradients will be performed using combinations of WT RC, WT LHC, and adapted versions of both. We expect to see two lines when running a WT RC with any single LHC and three lines when running a WT RC with both LHC and finally, a single line that has migrated further when using all three adapted proteins, indicating the formation of a heterotrimer. BlueNative PAGE can be used to indicate the molecular weight of differing products. The expected results can then be compared to analyse oligomer formation. Finally, SDS-PAGE and Western blotting can be used to investigate if the formation of the complex is because of isopeptide bond formation.

## Transmission Electron Microscopy

Transmission electron microscopy can be used to assess the morphology of the expected complex in relation to unreacted WT products. WT RC will be mixed with ST#LHCI and SnT#LHCII, producing a mixture of roughly uniformly sized objects, indicating no complexes formed. Mixing adapted RC and adapted LHC as separate assays will indicate the expected object size and shape of incomplete complex formation of RC with only one LHC. This will allow us to compare it to mixing of all three components to assess if a complete heterotrimeric complex has formed.

## Spectroscopic Analysis

To probe if the complex has formed and if it functions as we want it to, excitatory light at 650nm will be used to assess if emission from the LHCs is quenched by the adjoined RC when compared to a mixture of WT RC with adapted LHC. Assuming a drop in emission is seen, we would need to investigate if this was caused by absorption of excitation light by the RC. By varying the wavelength of the excitation light, we can ensure the drop in emission is due to absorption of emission light by the RC from the LHC rather than direct absorption of excitation light. Finally, we would measure photooxidation of the reaction centre’s first electron carrier, P870. If an increase in the photooxidation of P870 is seen alongside a decrease in LHC light emission, we can assume the energy has been transferred from LHC to RC.

# Timeline



# Discussion

We plan to utilise a combination of the SpyCatcher/SpyTag and SnoopCatcher/SnoopTag methods to produce a heterotrimer consisting of LHCI, LHCII, and a purple bacterial reaction centre. One of the key strengths to the method outlined above is that the scientific basis for the work is routed in prior published work, utilising a method that has been well developed and documented. The biggest challenge will be whether the proteins are still able to function when forced into an unnatural configuration. Provided they are within 10nm of each other, Förster resonance energy transfer should be able to occur between them. By attaching additional protein components to two termini of the RC and complexing with two other adapted proteins it may affect their structure too greatly and prevent light absorption and energy transfer. Unforeseen issues may also prevent the complex from self-assembling in general, at which point we would alter the conformation of the Tag and Catcher attachments to experiment with a variety of final products. For example, varying the attachment site of SnC or SC on the subunits of the RC, or addition of linker sequences is a simple way to alter the experiments and investigate unexpected results.

If successful we will have been able to link together two different light harvesting proteins with a reaction centre from a different species, building on the work performed by Liu et al. (2020). In doing so we will have shown the potential for this technology and its use in producing polyprotein assemblies with functional importance. We will have shown the ability to complex photosynthetic proteins from different species while retaining functionality. The option of experimenting with phycobiliproteins (PBP) is exciting. By using PBPs we will capture energy from parts of the spectrum that chlorophyl containing LHCs cannot, namely the ~525-625nm range. Crucially the energy level of the light absorbed by the different pigments in PBPs must be considered to direct energy in a flow to the RC. Positioning the high energy absorbing phycoerythrin or phycoerythrocyanin at a further distance to the RC than the red-light absorbing phycocyanin (49). Further work with other LHCs such as Lhcx, a protein that confers stability through photoprotection (50), could begin to add additional layers complexity that have not been seen in previous work.

Our work will have a great deal of impact in the synthetic biology field and the development of large, structurally significant multi-protein complexes, as well as the biotechnological industry when looking to the production equipment that utilise biological components to produce photovoltaic and photocurrents. This would mark a turning point in our dependency on archaic energy sources and opens a greener future in technology.

# References:

1. Coyle, E. D., & Simmons, R. A. (2014). *Understanding the global energy crisis*. Purdue University Press.
2. Tollefson, J. (2021, November 5). Carbon emissions rapidly rebounded following COVID pandemic dip. *Nature* <https://www.nature.com/articles/d41586-021-03036-x#:~:text=It%20predicts%20that%20carbon%20emissions,recovery%20than%20many%20scientists%20expected>.

(2022, February 18) Fossil fuels and climate change: the facts. *ClientEarth Communications* [*https://www.clientearth.org/latest/latest-updates/stories/fossil-fuels-and-climate-change-the-facts/#:~:text=The%20Intergovernmental%20Panel%20on%20Climate,from%20fossil%20fuels%20and%20industry*](https://www.clientearth.org/latest/latest-updates/stories/fossil-fuels-and-climate-change-the-facts/#:~:text=The%20Intergovernmental%20Panel%20on%20Climate,from%20fossil%20fuels%20and%20industry)*.*

1. Höök, M. & Tang, X. (2013). Depletion of fossil fuels and anthropogenic climate change—A review. *Energy Policy*. Volume 52, January 2013, Pages 797-809.
2. Welsby, D., Price, J., Pye, S., & Ekins, P. (2021). Unextractable fossil fuels in a 1.5 °C world. *Nature*, *597*(7875), 230–234.
3. Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., Summerhayes, C. P., Barnosky, A. D., Cornell, S. E., Crucifix, M., Donges, J. F., Fetzer, I., Lade, S. J., Scheffer, M., Winkelmann, R., & Schellnhuber, H. J. (2018). Trajectories of the Earth System in the Anthropocene. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(33).
4. Nicolopoulou-Stamati, P., Maipas, S., Kotampasi, C., Stamatis, P., & Hens, L. (2016). Chemical Pesticides and Human Health: The Urgent Need for a New Concept in Agriculture. *Frontiers in public health*, *4*, 148.
5. Stoytcheva, M. (Ed.). (2011). *Pesticides in the modern world: effects of pesticides exposure*. BoD–Books on Demand.
6. Donley N. (2019). The USA lags behind other agricultural nations in banning harmful pesticides. *Environmental health : a global access science source*, *18*(1), 44.
7. Rodriguez-Mozaz, S., Marco, MP., Lopez de Alda, M. *et al.* (2004). Biosensors for environmental monitoring of endocrine disruptors: a review article. *Anal Bioanal Chem* 378**,**588–598.
8. Swainsbury, D. J., Friebe, V. M., Frese, R. N., & Jones, M. R. (2014). Evaluation of a biohybrid photoelectrochemical cell employing the purple bacterial reaction centre as a biosensor for herbicides. *Biosensors & bioelectronics*, *58*(100), 172–178.
9. Pokhrel, L.R. & Dubey, B. (2013) Global Scenarios of Metal Mining, Environmental Repercussions, Public Policies, and Sustainability: A Review, *Critical Reviews in Environmental Science and Technology*, 43:21, 2352-2388

Ding, Y. et al. (2019). Recovery of precious metals from electronic waste and spent catalysts: A review. *Resources, Conservation and Recycling*, Volume 141, Pp. 284-298.

Suresh, L. et al. (2019). Biodegradable Protein-Based Photoelectrochemical Cells with Biopolymer Composite Electrodes That Enable Recovery of Valuable Metals. *ACS Sustainable Chemistry & Engineering*, *7* (9), 8834-8841

Lee, B., Ko, Y., Kwon, G., Lee, S., Ku, K., Kim, J., & Kang, K. (2018). Exploiting biological systems: toward eco-friendly and high-efficiency rechargeable batteries. *Joule*, 2(1), 61-75.

Liu, J., Friebe, V.M., Frese, R.N. and Jones, M.R. (2020) Polychromatic solar energy conversion in pigment-protein chimeras that unite the two kingdoms of (bacterio)chlorophyll-based photosynthesis. *Nature Communications* 11, 1542.

Shah, N. H., & Muir, T. W. (2014). Inteins: Nature's Gift to Protein Chemists. *Chemical science*, *5*(1), 446–461.

Shah, N.H., Muir, T.W. (2011). Split Inteins: Nature's Protein Ligases. *Isr J Chem*, 51(8-9):854-861.

Perler, F.B. (2002). InBase: the Intein Database. *Nucleic Acids Res.,* 30:383.

Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., Stevens, T. H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*, 250, 651.

Wang, Z. A., & Cole, P. A. (2020). Methods and Applications of Expressed Protein Ligation. *Methods in molecular biology (Clifton, N.J.)*, *2133*, 1–13.

Mootz, H.D. (2009). Split inteins as versatile tools for protein semisynthesis. *Chembiochem. Nov 2; 10(16):2579-89.*

Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., & Benkovic, S. J. (1999). Production of cyclic peptides and proteins in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13638–13643.

Lu, W., Sun, Z., Tang, Y., Chen, J., Tang, F., Zhang, J., & Liu, J. N. (2011). Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage. *J8ournal of chromatography. A*, *1218*(18), 2553–2560.

Olschewski, D., Seidel, R., Miesbauer, M., Rambold, A. S., Oesterhelt, D., Winklhofer, K. F., Tatzelt, J., Engelhard, M., & Becker, C. F. (2007). Semisynthetic murine prion protein equipped with a GPI anchor mimic incorporates into cellular membranes. *Chemistry & biology*, *14*(9), 994–1006.

1. Zettler, J., Schütz, V., & Mootz, H. D. (2009). The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS letters*, *583*(5), 909–914.
2. Dassa, B., Amitai, G., Caspi, J., Schueler-Furman, O., & Pietrokovski, S. (2007). Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry*, *46*(1), 322–330.

Pinto, F., Thornton, E. L., & Wang, B. (2020). An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nature communications*, *11*(1), 1529.

Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T., & Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), E690–E697.

Hatlem, D., Trunk, T., Linke, D., & Leo, J. C. (2019). Catching a SPY: Using the SpyCatcher-SpyTag and Related Systems for Labeling and Localizing Bacterial Proteins. *International journal of molecular sciences*, *20*(9), 2129.

Reddington, S. C., & Howarth, M. (2015). Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. *Current opinion in chemical biology*, *29*, 94–99.

Akiba, H., Takayanagi, K., Kusano-Arai, O., Iwanari, H., Hamakubo, T., & Tsumoto, K. (2020). Generation of biparatopic antibody through two-step targeting of fragment antibodies on antigen using SpyTag and SpyCatcher. *Biotechnology reports (Amsterdam, Netherlands)*, *25*, e00418.

Li, L., Fierer, J. O., Rapoport, T. A., & Howarth, M. (2014). Structural analysis and optimization of the covalent association between SpyCatcher and a peptide Tag. *Journal of molecular biology*, *426*(2), 309–317.

Liu, J., Friebe, V. M., Frese, R. N., & Jones, M. R. (2020). Polychromatic solar energy conversion in pigment-protein chimeras that unite the two kingdoms of (bacterio)chlorophyll-based photosynthesis. *Nature communications*, *11*(1), 1542.

Nelson, A. L., Ries, J., Bagnoli, F., Dahlberg, S., Fälker, S., Rounioja, S., Tschöp, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., Holden, D. W., Rappuoli, R., Normark, S., Barocchi, M. A., & Henriques-Normark, B. (2007). RrgA is a pilus-associated adhesin in Streptococcus pneumoniae. *Molecular microbiology*, *66*(2), 329–340.

Veggiani, G., Nakamura, T., Brenner, M. D., Gayet, R. V., Yan, J., Robinson, C. V., & Howarth, M. (2016). Programmable polyproteams built using twin peptide superglues. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), 1202–1207.

Woolfson D. N. (2021). A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *Journal of molecular biology*, *433*(20), 167160.

Hodges, R. S., Saund, A. K., Chong, P. C., St-Pierre, S. A., & Reid, R. E. (1981). Synthetic model for two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization of an 86-residue analog of tropomyosin. *The Journal of biological chemistry*, *256*(3), 1214–1224.

Kim, J. S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., & Kim, S. H. (2005). Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3248–3253.

Wei, R. R., Sorger, P. K., & Harrison, S. C. (2005). Molecular organization of the Ndc80 complex, an essential kinetochore component. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5363–5367.

Schmidt, H., Zalyte, R., Urnavicius, L., & Carter, A. P. (2015). Structure of human cytoplasmic dynein-2 primed for its power stroke. *Nature*, *518*(7539), 435–438.

Bragulla, H. H., & Homberger, D. G. (2009). Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. *Journal of anatomy*, *214*(4), 516–559.

Brodsky, B., & Persikov, A. V. (2005). Molecular structure of the collagen triple helix. *Advances in protein chemistry*, *70*, 301–339.

Magliery, T. J., Wilson, C. G., Pan, W., Mishler, D., Ghosh, I., Hamilton, A. D., & Regan, L. (2005). Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *Journal of the American Chemical Society*, *127*(1), 146–157.

Ghosh, I., Hamilton, A.D., & Regan, L. (2000) Antiparallel Leucine Zipper-Directed Protein Reassembly:  Application to the Green Fluorescent Protein. *J. Am. Chem. Soc.,* 122, 23, 5658–5659.

Cristie-David, A. S., Sciore, A., Badieyan, S., Escheweiler, J. D., Koldewey, P., Bardwell, J. C., ... & Marsh, E. N. G. (2017). Evaluation of de novo-designed coiled coils as off-the-shelf components for protein assembly. *Molecular Systems Design & Engineering*, *2*(2), 140-148.

Hussey, B. J., & McMillen, D. R. (2018). Programmable T7-based synthetic transcription factors. *Nucleic acids research*, *46*(18), 9842–9854.

Swainsbury, D. J., Harniman, R. L., Di Bartolo, N. D., Liu, J., Harper, W. F., Corrie, A. S., & Jones, M. R. (2016). Directed assembly of defined oligomeric photosynthetic reaction centres through adaptation with programmable extra-membrane coiled-coil interfaces. *Biochimica et biophysica acta*, *1857*(12), 1829–1839.

Watanabe, M., & Ikeuchi, M. (2013). Phycobilisome: architecture of a light-harvesting supercomplex. *Photosynthesis research*, *116*(2-3), 265–276.

Buck, J. M., Sherman, J., Bártulos, C. R., Serif, M., Halder, M., Henkel, J., Falciatore, A., Lavaud, J., Gorbunov, M. Y., Kroth, P. G., Falkowski, P. G., & Lepetit, B. (2019). Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom Phaeodactylum tricornutum. *Nature communications*, *10*(1), 4167.