THE UNIVERSITY
of EDINBURGH

Self-Assembling Protein Structures for Novel Applications in Synthetic Biology

Ella Lucille Thornton

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

The University of Edinburgh

2022

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Ella Lucille Thornton

May 2022

Abstract

Self-assembling protein tools are highly desirable in the field of synthetic biology. They enable the creation of new macromolecular structures for novel applications, by providing users with the ability to ‘build’ with proteins post-translationally: both in vivo and in vitro.

In this thesis, I investigate and describe applications for a range of self-assembling protein tools. These include the BslA protein, SpyTag/SpyCatcher and SnoopTag/SnoopCatcher peptide/protein pairs, mVirD2 protein, and split inteins. For each of these tools I outline their capabilities, and then demonstrate a new application, either individually or in tandem.

BslA is an amphiphilic protein that has the remarkable ability to self-assemble into monolayers on hydrophobic glass surfaces. Using a BslA-SpyTag or BslA-SnoopTag fusion protein, I show that BslA monolayers can be formed and subsequently functionalised via these tags with proteins attached to SpyCatcher or SnoopCatcher respectively. I show proof of principle with the attachment of fluorescent proteins to the glass surface via these specific tags. Fluorescent labelling of the BslA monolayer with these proteins provides a better understanding of the monolayer properties, specifically with the movement of proteins within the monolayer.

Further expansion of the toolbox for functionalising BslA surfaces is achieved using the mVirD2 protein, which forms a covalent bond with DNA at a specific recognition site. This enables DNA to bind to the glass surface through a series of specific and covalent interactions: first to mVirD2-SpyCatcher fusion protein, and then to BslA-SpyTag.

These tools are utilised to test different applications with biotechnological relevance. The applications proposed and tested in this thesis include: screening for new protein/peptide interactions, nanopillars for anti-glint purposes, and real-time cell-free analyte monitoring.

Within the cell-free experiments, I focus on the capabilities of a BslA surface to capture proteins expressed in situ, and test cell-free protein expression from DNA covalently-bound to the surface. Following this, I attach protein biosensors to the BslA

surface via SpyTag/SpyCatcher peptide/protein pairs and also use them to monitor fluctuations in phosphate levels during a cell-free reaction as it proceeds.

I also help with characterisation of an expanded library of split inteins for protein trans-splicing, by testing the efficiency of splicing and comparing the orthogonality of the library in vitro. Subsequently, I demonstrate the use of split inteins in the development of a protocol for the solid-phase assembly of large, extremely repetitive proteins using split intein trans-splicing. I provide two protocols, that enable the ligation of six protein units, using either five orthogonal split inteins simultaneously, or using just two split inteins sequentially. This enables the creation of repetitive proteins to a length that would be difficult to achieve in vivo. In summary, this thesis presents new applications for self-assembling protein systems and further characterises the tools that underpin them.

Lay Summary

Proteins are essential for life. They underpin many biological functions, from processing light signals received by the eyes, to the creation of material for hair and nail formation. Nature has evolved billions of proteins, each with their own distinct roles. They can be found in human bodies, animals, plants, as well as viruses. Some of these proteins have incredible properties that help them perform their roles. In synthetic biology we discover these proteins and investigate new ways that their unusual properties can be used. Like building blocks, proteins can combine to create bigger structures and perform complex tasks.

In this thesis I look at several particularly interesting ‘building blocks’ of proteins, and find new ways they can be used as tools in bigger protein constructions.

The proteins I investigate are useful in different ways, yet all are capable of some form of self-assembly. One, the BslA protein, creates layers on the surface of bacteria, acting as an umbrella to protect the bacteria beneath. We can take this protein and repurpose it to form self-assembling layers on the surface of glass. This allows us to make functional protein surfaces in a way that is much faster and easier than traditional methods.

I also investigate ‘protein glues’ which are capable of binding proteins, either to other proteins or to DNA. Of particular interest is a protein glue called ‘split inteins’ which can be used to glue any two proteins together, irreversibly. This quality is really useful for ‘building’ with proteins. In this thesis I present and characterise a library of ten split inteins which can be used simultaneously, and which, in theory, have the capability to glue 20 different proteins together without cross reacting with each other.

I show how valuable these protein ligation tools can be, by using them to join multiple SasG proteins together. SasG proteins are found on the surface of one type of bacteria, and are used to ‘reach out’ and connect with other bacteria. For this role, the proteins are required to be very strong, yet elongated. By using split inteins to join multiple SasG proteins together, I’m able to create even longer proteins, which has the potential for exciting biomaterials applications, such as spinning the proteins into strong fibres.

Many other useful self-assembling protein tools are also examined and tested in this thesis. It provides a collection of experimental work to demonstrate the great potential of self-assembling protein tools for novel applications.

Acknowledgements

First and foremost, I couldn't have completed my studies without the help of my academic supervisors. I am grateful that they have always believed in my capabilities and encouraged me to see that myself. Baojun, thank you for the guidance and encouragement you provided in my first few years of my PhD. My industry supervisor, Karl, has also provided crucial support and guidance throughout my studies, thank you.

My primary supervisor, Lynne, has been a wonderful supervisor and mentor in science. Thank you for being so generous with your time, and for sharing your joy for science with me. Working in your lab has transformed my perception of research. Thank you for all the many ways you have supported and encouraged me during my studies.

Working alongside so many incredibly smart and kind people has been a privilege. Louise, Filipe, Xinyi, thank you for always being patient and teaching me so much. Pop, Maggie, Yang, Trevor, Rossana, Zoe, thank you for being wonderful lab mates. Curran, thank you for your endless enthusiasm for science that has undoubtably changed me as a scientist.

Edinburgh is a beautiful city, but the friends I have made here is what makes it feel like a home. Lewis and Simone, thank you for making our flat a home, life living apart from you is far less hilarious. Meg, Emily, Sonny, Arielle, George and Cat, thank you for all of the great memories we have shared together in this city. And a special thank you to Yousef, for all of your support while I was writing this thesis.

I feel very lucky to be so uplifted by my friends who are not physically close but who have still supported me so much. Kate and Rhian, your friendship has meant so much to me and helped me through the hardest parts of the last few years. Niki, thank you for your friendship and the endless hours of conversation about nothing in particular. Vicky and Louisa, thank you so much for your friendship.

Finally, I could not have survived my PhD without photo updates of our family cat, Hunter. I would like to thank my beloved family for providing me with these photographs. Most importantly, I would like to thank my parents, Peter and Anita, along with my sisters, Ariel and Tuesday, for their continued love and support.

Contents

List of Figures

List of Tables

Abbreviations

AA	Amino acid
ATP	Adenosine-5’-triphosphate
CDS	Coding sequence
CFP	Cyan fluorescent protein
CFPS	Cell free protein synthesis
CV	Column volume
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FP	Fluorescent protein
FRAP	Fluorescence recovery after photobleaching
FRET	Förster resonance energy transfer
FT	Flow through
GFP	Green fluorescent protein
GST	Glutathione S-transferase
His	Histidine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MW	Molecular weight
NAD	Nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
OD	Optical density
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
Pi	Inorganic phosphate
POI	Protein of interest
PURE	Purified recombinant elements
RBS	Ribosome binding site
RFP	Red fluorescent protein

RNA	Ribonucleic acid
RNAP	RNA polymerase
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SnC	SnoopCatcher
SnT	SnoopTag
SpC	SpyCatcher
SpT	SpyTag
TCEP	Tris(2-carboxyethyl)phosphine
TEM	Transmission electron microscopy
TMR	Tetramethylrhodamine
tRNA	Transfer RNA
TX-TL	Transcription-translation
UP	Ultra-pure
WT	Wild-type
YFP	Yellow fluorescent protein

Chapter 1 General Introduction

1.1 Synthetic biology and self-assembling materials

Protein-protein interactions underpin structures and systems crucial to life on earth. Through the evolution of life, billions of proteins and interactions have been formed to achieve niche applications and functions in nature. As we begin to scratch the surface into understanding these incredible molecules and their roles in biology, synthetic biology acts as an engineer of these proteins toward new applications (Papapostolou & Howorka, 2009; Qian et al., 2020).

Synthetic biology, taken literally, is a discipline that aims to synthesise biology, this spans the re-purposing of natural devices to the de novo design of new ones, which stands on the shoulders of fundamental biology (Benner & Sismour, 2005). Seminal work in engineering biology at the DNA level to encode logic-gate like behaviour within living organisms has been a tool at the forefront of synthetic biology's development over the past decade (Purnick & Weiss, 2009; Tamsir et al., 2011; Wang et al., 2011). More recently, huge strides toward understanding protein folding and behaviour have been made (Jumper et al., 2021), which has enabled better design and prediction of protein behaviour, thus enabling engineering of biology at the protein level.

Engineering biology at the protein scale in vivo requires specificity, due to the complexity and crowded nature of the cellular milieu. Two approaches to the development of affinity driven self-assembling protein tools for engineering are generally followed. The first approach can be viewed as repurposing of proteins found in nature with unique and useful properties to non-native applications. The other, to start from the beginning, de novo, is enabled through our understanding of protein science over the past century (Huang et al., 2016). Both provide valuable tools for protein science used in synthetic biology today.

De novo design of proteins and protein-protein interactions has been revolutionised by the release of AlphaFold, capable of predicting 3D protein structure from sequence to a degree of accuracy not attainable previously (Jumper et al., 2021; Thornton et al., 2021). De novo protein design was already a field exploding in synthetic biology, and now the growth increases at an even faster rate, with incredibly exciting work demonstrating the capabilities of de novo protein design unveiled at a breakneck speed, largely by David Baker's group at The University of Washington (Chen et al.,

2019, 2020; Edgell et al., 2020; Hsia et al., 2021; Langan et al., 2019; Quijano-Rubio et al., 2021)

Alternatively, self-assembling protein tools can be found in natural protein structures, and repurposed for synthetic biology purposes. Affinity driven self-assembly of proteins has been found and exploited for applications extensively, most notably in the development of a toolbox of coiled-coil protein pairs (Fletcher et al., 2012; Thompson et al., 2012). Greater control over the directed assembly of proteins is possible with a wide array of triggers for protein self-assembly, for example, by modulation of protein charge, temperature, or post-translational modifications (Arpino & Polizzi, 2020)(Piraner et al., 2019; Winter et al., 2020).

Directed, non-covalent assembly of proteins via any of the aforementioned methods remains incredibly useful for applications where reversibility is crucial, such as protein cage assembly and disassembly, helical filament assembly, biosensor development, or even super-resolution imaging methods (Jones et al., 2021; Oi et al., 2020; Quijano-Rubio et al., 2021; Shen et al., 2019).

Nevertheless, the irreversible directed self-assembly of proteins remains a particular goal. To permanently tether two proteins together is a sought-after tool for building permanent structures both inside and outside of cells (Lange & Polizzi, 2021). From the discovery of natural protein units capable of irreversibly gluing two protein domains, protein glue methods have been engineered to create arrays of orthogonal tools for directed and irreversible protein assembly (Pinto et al., 2020; Veggiani et al., 2016). These tools are useful for the creation of macroscale structures for applications in vaccine development, biomaterial creation, and biosensing applications (Bowen et al., 2018; Brune et al., 2016). The library of available irreversible protein ligation tools remains significantly smaller than those for reversible protein ligation tools. Therefore, there remains considerable scope for further expansion and development of this particular toolbox of protein tools.

The intersection of engineering, chemistry, biology, physics and art is where the discipline of synthetic biology lies, and the great potential of synthetic biology comes from the merging of perspectives from all these disciplines to form new and exciting ideas. It would appear that while there is a phenomenal increase in self-assembling protein tools available to the community, and a greatly increasing understanding of the protein science underpinning these tools, real-world applications for these tools

lag behind speed of development. This is a continual hurdle for synthetic biology (Cameron et al., 2014), but I believe that by utilising this wide array of protein tools in new combinations we can arrive at new intersections of technology capable of solving the most pressing problems synthetic biology faces.

1.2 Aims of my work

In this thesis, I aim to investigate protein engineering tools found in nature and apply these tools to new applications. I aim to achieve this by first, testing existing tools and characterising them, then applying these tools to new applications, and finally, combine a number of tools in tandem to further advance capabilities of their applications.

The initial protein tool of interest is the BslA protein, which is capable of self-assembly into monolayers on glass surfaces (Bromley et al., 2015; Williams et al., 2018). I aim to further exploit the fascinating properties of this protein as a method for surface immobilisation, and the many potential applications that brings. Surface immobilisation of proteins and DNA is an integral part of many biochemistry techniques, yet often is inaccessible to users due to the high costs of components (such as antibodies) or complexity of method. The use of a fully protein-based and self-assembling surface for protein and DNA immobilisation holds great potential to circumvent these issues. In Chapter 2 I aim to demonstrate the feasibility of such a system, by combining the BslA technology with genetically encodable protein-protein and protein-DNA glues, Tag/Catcher peptide/protein pairs, and mVirD-2 protein, respectively. Both attachment tools are originally found within proteins from nature, and incredibly provide covalent attachment between the two molecules of interest (Bernardinelli & Högberg, 2017; Zakeri et al., 2012). In order to demonstrate this effectively, I also aim to characterise the properties of the BslA monolayer further, to give users a clearer understanding of how the system works and what it may be compatible with.

Beyond proof of principle of BslA monolayers as a useful tool, I aim to show their compatibility with cell-free (CF) systems as a potential application in Chapter 3. Surface immobilisation of proteins synthesised in CF has been shown as early as 2001, and serves as a useful practical tool for high throughput protein functionality assays, enabling each protein to be captured and further examined for functionality in the same location it was synthesised (Manzano-Román & Fuentes, 2019). As CF

systems are used by more scientists worldwide, the accessibility of the platform and relevant tools remains an important goal to continue allowing its implementation in diverse settings. I aim to show the BslA monolayers are compatible with CF in situ protein capture, and can provide an inexpensive and simple alternative to other surface capture methods.

Another self-assembling protein tool investigated in this thesis within Chapter 4 is that of split-inteins. Split-inteins are found in nature, and are very desirable tools for protein ligation as they leave a small scar site, self-excising the splicing domains from the final protein construct (Aranko et al., 2014). While there has been great interest in split-inteins in synthetic biology and biochemistry, the characterised split inteins available for use have been assessed by individual labs, often under differing conditions. This makes direct comparison of split inteins difficult, and assessment of compatibility for one-pot reactions hard. In this final results chapter of my thesis, I aim to characterise a library of split inteins under the same conditions, assessing functionality and orthogonality. I then aim to demonstrate the potential of these findings by testing multiple inteins in one reaction to ligate different proteins in one pot. Previously, split inteins had been used for biomaterial applications, however they were only demonstrated using one intein at a time (Bowen et al., 2018). The ability to use more than one at once would enable even more applications.

Together, the chapters of my thesis aim to show a diverse range of protein tools and demonstrate that they can be used in tandem toward new purposes. In each results chapter, I provide an in-depth introduction to the technology utilised and novel application I plan to demonstrate, and conclude the results with discussion and conclusions for each chapter. In Chapter 6, I provide a concluding discussion to the entire work presented in this thesis and comment on future perspectives.

Chapter 2 Self-assembling BslA monolayers

2.1 Introduction

2.1.1 Surface immobilisation of macromolecules

Immobilising proteins and DNA molecules on surfaces is central to many techniques ubiquitous in biochemistry. The attachment of molecules to a surface allows for isolation from a heterogeneous mixture, enabling downstream modifications of the molecule of interest, or detection of the molecule via specific antibodies, to occur (Berrade et al., 2011).

Ordered and functional display of macromolecules on a surface is difficult to achieve, and cannot be done by simply adhering the molecule to the surface. The reasons for this are three-fold: firstly, adhering a molecule to a surface can cause denaturation of the molecule itself. Secondly, the orientation is not controlled, meaning that that binding site for a ligand of interest could be inaccessible on the surface due to steric hindrance. Thirdly, the non-specific nature of the adherence would require molecules be fully purified before coating of the surface, which can be difficult and too time consuming for some applications. Ideally, fully folded and functional proteins of interest can be displayed with controlled orientation on a surface. This is not achievable by random adhesion of a molecule to a surface and requires other technology to achieve (de Marco, 2018; D. Kim & Herr, 2013; Yang et al., 2018).

As such, it is desirable to develop a reliable technique for controlled and specific attachment of biomolecules to surfaces, as it underpins a wide array of more mature technologies.

While site-specific chemical modification of biomolecules can be used to attach them to chemically coated surfaces, this approach requires labelling techniques often not possible in vivo (Meldal & Schoffelen, 2016; Spicer & Davis, 2014). Unnatural amino acids can be introduced to proteins for site specific surface immobilisation, however this process requires expensive reagents and can suffer drawbacks such as low yield or slow reaction efficiency (Bednar et al., 2019; Raliski et al., 2014). The most widely used method for protein-based specific surface attachment currently is the Avidin-biotin system, and the functional and structural analogues Streptavidin-biotin and NeutrAvidin-biotin systems (Marttila et al., 2000; Sano et al., 1998). Their widespread use has led to the development of robust and reliable protocols for the display of biomolecules. The system utilises the strong interaction between Avidin and biotin,

the latter of which can be chemically added to DNA or attached to proteins in vitro or in vivo in a process known as biotinylation. These biotinylated molecules can then be attached to a surface coated with Avidin (Ramachandran et al., 2004).

Avidin is a large (67 kDa) tetrameric protein that binds biotin, a vitamin, with the strongest known non-covalent bond interaction between protein and ligand ($K_d = 10^{-15}$ M) (Green, 1975). The strength of the interaction makes this pair impractical for applications such as affinity protein purification, as very harsh conditions are required to disrupt the interaction and elute the pure protein. These harsh elution conditions can cause protein denaturation. However, for surface attachment where retrieval of the protein of interest is not required, the interaction is ideal (Steen Redeker et al., 2013). The fact that Avidin is tetrameric is also an advantage for attaching molecules to surfaces, because each Avidin tetramer can bind four biotin molecules; this produces a surface with a high density of biotin binding sites (Livnah et al., 1993).

While the Avidin/biotin technology is robust and widely used, there are some disadvantages to the system. Firstly, the protein or DNA molecule of interest must be modified to include biotin. In the case of DNA, biotin is attached via chemical modification of the nucleotide, a process that cannot take place in vivo. Therefore, to attach biotin to DNA, the DNA must be synthesised. In the case of proteins, the most common technique for biotinylation requires the POI (protein of interest) be fused to a genetically encoded peptide tag, AviTag, that is biotinylated post-translationally via biotin ligase, BirA. Biotinylation of protein via the AviTag can be done in vivo or in vitro (Fairhead & Howarth, 2015). Secondly, the creation of the functional surface requires multiple binding, blocking, and washing steps for specific binding and coating of the surfaces. These can be labour and time intensive, with many steps during the procedure where mistakes can be made or errors introduced.

The very complex interaction of proteins with surfaces frequently introduces variability to results of protein microarrays, and even when application is mechanically controlled there is still huge potential for discrepancy between samples (Clancy et al., 2019). Thus, the development of reproducible and simple protocols for directed surface coating of functional and folded proteins remains a significant goal in biochemistry (de Marco, 2018).

2.1.2 The BslA protein

I propose to use the self-assembling protein BslA to display proteins and DNA on glass surfaces. Not only does BslA self-assemble into a protein monolayer, a highly convenient feature for coating surfaces, it is also thought to block non-specific adsorption of proteins to its surface (Williams et al., 2018).

BslA is a small (15 kDa) amphiphilic protein capable of self-assembling into monolayers at hydrophobic/hydrophilic interfaces. It is natively found in biofilms of *Bacillus subtilis*, where it protects the colony like an 'umbrella' (Hobley et al., 2013). The protein is a bacterial hydrophobin and its assembly into a monolayer is promoted by its immunoglobulin-like fold and hydrophobic cap (Figure 2.1) (Bromley et al., 2015).

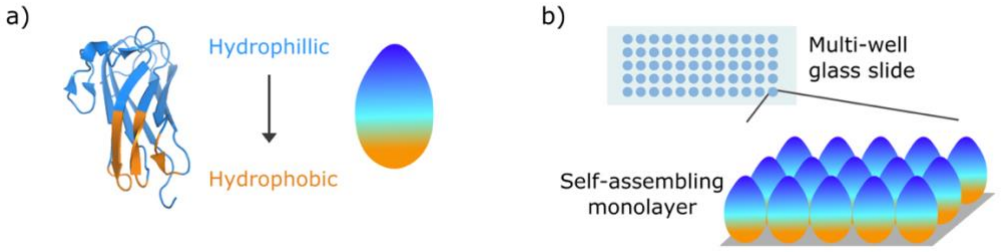


Figure 2.1 BslA protein can form self-assembling monolayers. (**a**) Ribbon representation of BslA protein (PDB ID code: 4BHU), next to simplified cartoon used in figures throughout this thesis. The hydrophilic side of the protein is indicated with blue colouring, and the hydrophobic cap is indicated with orange colouring. (**b**) Schematic of a BslA monolayer coating a glass surface in multi-well layout for higher throughput.

Because BslA is genetically encodable, the protein monolayer can be functionalised with protein ligation tools, added to BslA as fusions. In this work two peptide/protein pairs that spontaneously form covalent bonds upon mixing, SpyTag/SpyCatcher and SnoopTag/SnoopCatcher, are commonly used. Fusion of SpyTag or SnoopTag to BslA, and SpyCatcher or SnoopCatcher to the POI, enables a BslA surface to be functionalised with POIs, as shown previously (Schloss et al., 2016; Williams et al., 2018). To attach DNA to the BslA monolayer, mVirD2, a protein that covalently attaches to a specific 12 nt DNA recognition sequence (Bernardinelli & Högberg, 2017), is first attached to the BslA surface using SpyTag/SpyCatcher or SnoopTag/SnoopCatcher. The DNA of interest can then be covalently attached to

mVirD2. I propose that this approach can be used to covalently attach ssDNA, dsDNA, or proteins of interest to a BslA monolayer.

2.1.3 Attachment of proteins to BslA via covalent peptide/protein pairs

SpyTag/SpyCatcher is a peptide/protein pair capable of spontaneous covalent bond formation (Figure 2.2.), discovered and developed by the Howarth lab in Oxford (Zakeri et al., 2012). The technology was born from the observation that the FnaB protein from *S. pyogenes* contains an isopeptide bond between non-adjacent residues, which is rare in nature. The Howarth lab split the protein at the junction of the isopeptide bond formation, and tested whether, when mixed, the two halves can reconstitute the isopeptide bond spontaneously. This approach succeeded and the authors showed that only a short 13 AA (Amino Acid) peptide sequence was required one side of the reaction site. This short 13 AA peptide is called SpyTag and its cognate protein partner is called SpyCatcher. The SpyTag/SpyCatcher technology has revolutionised synthetic biology since its invention 10 years ago (Keeble & Howarth, 2020). SpyTag/SpyCatcher has been used for many applications, including vaccine development, stabilisation of enzymes, hydrogel creation, macroscale structure formation, imaging techniques, cellular labelling, poly-protein chain formation, protein purification, and many more applications (Berckman & Chen, 2021; Brune et al., 2016; Hinrichsen et al., 2017; Khairil Anuar et al., 2019; F. Sun et al., 2014; X. B. Sun et al., 2019; Veggiani et al., 2016).

The appeal of SpyTag/Catcher as a protein technology is clear: few other protein ligation tools are capable of spontaneous covalent bond formation, and the small peptide on one side of the reaction is especially appealing when a non-perturbative attachment to POI is desired. It also has several other desirable characteristics: a very fast reaction time (80% fully reacted within 30 minutes), functionality in a wide range of buffers, pH (5 – 9), temperatures (4°C – 37°C) and even in the presence of detergents (Zakeri et al., 2012).

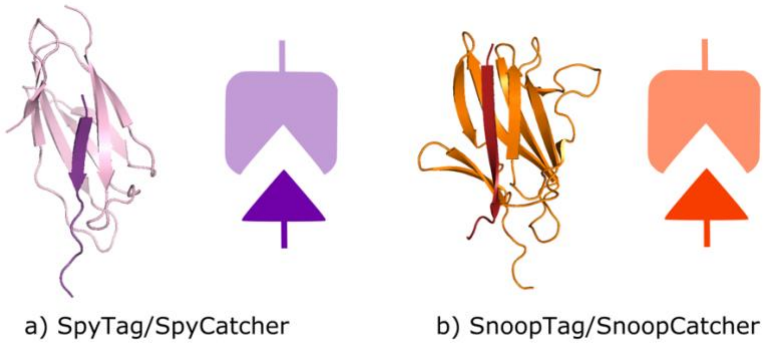


Figure 2.2 Orthogonal protein glues SpyTag/SpyCatcher and SnoopTag/SnoopCatcher. (**a**) Ribbon representation of SpyTag/SpyCatcher from FnaB protein (PDB ID code: 4MLI) next to simplified cartoon used in figures throughout this thesis. The peptide (SpyTag) is coloured dark purple, and the protein pair (SpyCatcher) is coloured light pink in both ribbon and cartoon representations (**b**) Ribbon representation of SnoopTag/SnoopCatcher from RrgA protein (PDB ID code: 2WW8). The peptide (SnoopTag) is coloured dark orange/red, and the protein pair (SnoopCatcher) is coloured light orange in both ribbon and cartoon representations.

Other orthogonal peptide/protein pairs that form covalent bonds when mixed can be used in tandem with SpyTag/SpyCatcher. The most popular orthogonal complement to SpyTag/SpyCatcher is the SnoopTag/SnoopCatcher peptide/protein pair, which was engineered by splitting the bacterial adhesin RrgA protein into a SnoopTag peptide (residues 734 – 745) and partner protein SnoopCatcher (residues 749 – 860) (Figure 2.2). The reaction is similarly robust to SpyTag/SpyCatcher: capable of bond formation in a range of different buffers, across temperatures and pH. When SnoopTag is in 2-fold or higher molar excess of SnoopCatcher, the reaction reaches 100% within 30 minutes (Veggiani et al., 2016).

SpyTag/SpyCatcher and SnoopTag/SnoopCatcher technology is especially useful for the application of surface attachment, as simplicity, speed and strength of binding is of the utmost importance. The availability of orthogonal peptide/protein pairs for adhesion enables greater control over the attachment of different proteins to the surface in either mixed or homogeneous molecular mixtures.

2.1.4 Attachment of DNA to BslA via mVirD2

For attachment of DNA to BslA surfaces, the mVirD2 protein is utilised. Developed as a synthetic biology tool by the Högberg lab in 2017 (Bernardinelli & Högberg, 2017), mVirD2 is a truncated version of VirD, which plays a crucial role in the pathogenic life

cycle of *Agrobacterium tumefaciens* (Pansegrau et al., 1993). VirD is utilised by the plant pathogen to transport the viral DNA into the host nucleus, enabling replication within the cell. For this function, VirD must grasp the DNA strongly, to ensure its grip is not loosened within the cellular matrix. VirD forms a covalent attachment to DNA via a tyrosine residue in a transesterification reaction, catalysed by Mg^{2+} (Van Kregten et al., 2009). Analysis of the mechanism for attachment has allowed for the non-essential domains of the protein to be removed, leaving a small, 204 residue protein, mVirD2, to be used as a DNA tagging protein that recognises a specific 12 nucleotide sequence. mVirD2 attaches to DNA in a two part process, beginning first with non-covalent association with the target sequence, followed by covalent linkage and cleavage of DNA sequence upstream of the cleavage site. This cleavage of DNA leaves just a four nt 'scar', alongside the downstream nt of interest (Figure 2.3).

In comparison with other DNA-protein conjugation techniques, mVirD2-DNA conjugation provides several advantages. Importantly, the bond is covalent, ensuring an irreversible pairing. The DNA to be specifically ligated can be single or double stranded, and only needs to be modified through the addition of a specific 12 nt recognition sequence at the desired site of ligation. In comparison to techniques where chemical modifications of DNA are required, DNA can be synthesised in vivo, enabling much easier and cheaper production. Finally, the only scar left on the DNA sequence after conjugation to mVirD is very short (four nt).

Figure 2.3 Design of protein and DNA oligonucleotide used to validate mVirD2 functionality. The fusion protein mVirD2-SpyCatcher can be conjugated to any DNA of interest containing the 12 nt recognition sequence. For ease of visualisation, Cy3 labelled DNA was used to show proof of concept in this thesis. When the mVirD2-SpyCatcher protein is in proximity with DNA containing the recognition sequence (in the presence of Mg^{2+}), association spontaneously occurs, which then results in cleavage of the 5' end of DNA and covalent attachment of DNA to protein.

In this work, the mVirD2 protein is genetically fused to SpyCatcher (Figure 2.3), thus serving as an intermediate protein in the process of DNA attachment to the surface.

Alongside the generalisable applications of attaching both proteins and DNA to surfaces, more specific applications for BslA monolayers are explored herein. The first is to use the BslA surface as the foundation of a screen for protein-peptide interactions. The second is to use the surfaces as a platform for building structures with 3D topographies, useful in this case for eliminating glint from glass via arrays of nanopillars arranged perpendicular to the surface.

2.1.5 Application of BslA surfaces: Protein/peptide interaction screening

The presentation of correctly folded and functional proteins on a surface is especially important when the goal is for downstream screens or for in situ characterisation of proteins. The user must feel confident that there is a pure layer of the protein of interest, homogenously presented, and all in the native, folded state. I hypothesise that proteins presented on the BslA monolayer will have these desirable characteristics, and present a proof-of-concept example to test peptides binding to a POI presented on the BslA surface.

Together with Professor Manfred Auer's laboratory, I worked on a collaborative project to adapt functionalised BslA surfaces for protein display and interacting peptide screening. The Auer lab had worked extensively on developing methods to elucidate new peptide/protein binding pairs, and examined these interactions in vitro.

A protein of particular interest to them is Survivin (Figure 2.4), a protein overexpressed in many cancers (Velculescu et al., 1999; Wheatley & Altieri, 2019). The protein is an inhibitor of apoptosis (IAP), and more specifically, inhibits caspase activation thereby preventing apoptosis or programmed cell death. Inhibiting Survivin thus has significant potential for the development of anti-cancer therapeutics (Verdecia et al., 2000).

Figure 2.4 Survivin protein. Ribbon representation of Survivin (PDB ID code: 1E31) next to simplified cartoon used in figures throughout this thesis.

The lab had already identified known peptide binders to Survivin (Jeyaprakash et al., 2011). I set out to display functional Survivin on the BslA surface with the goal of providing a new format in which to screen for inhibitors of Survivin-protein interactions.

2.1.6 Application of BslA surfaces: Anti-glint nanopillars

A further application of BslA surfaces explored in this thesis is of attachment of protein-based nanopillars for anti-glint coatings. The BslA coating provides an opportunity to self-assemble ordered structures on the surface of glass, with the potential to finely control the spacing of attached molecules by the titration of BslA-WT molecules into the monolayer for more dispersed molecules. This makes the BslA monolayer an appealing foundation for 'building' surfaces in a controllable way on the surface of glass.

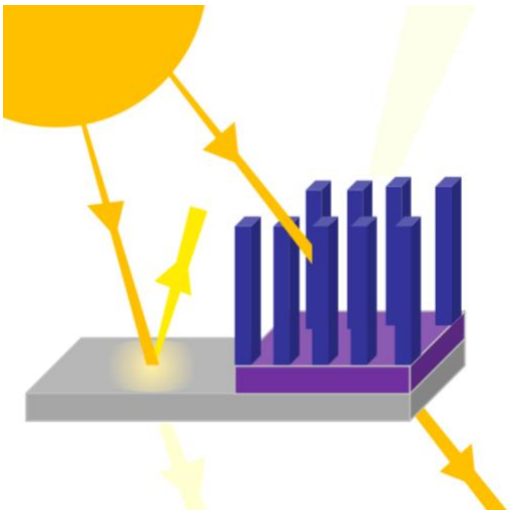


Figure 2.5 Nanopillars arranged on the surface of glass can reduce light reflected. This simplified schematic shows light from a source (yellow) hitting a glass surface (grey). In the areas with no coating, a significant amount of light is reflected from the surface. However, in the areas of glass with nanopillar structures (blue) perpendicular to the surface, the structures

form a gradient refractive index between the air and surface, enabling more light to pass through the glass and less to be reflected compared to a surface with no coating.

Moth-eyes exhibit incredible anti-reflective properties, owing to the sub-wavelength nanopillars that coat the surface of the eye (Stavenga et al., 2006). These arrays of nanopillars create a gradient refractive index between the air and the surface of the eye, reducing the amount of light reflected and instead increasing transmission of light through the surface and into the eye. Taking inspiration from these structures, moth-eye-like nanopillar arrays have been successfully fabricated on glass surfaces, and were found to confer the same desirable properties as in nature: reflectance is reduced and transmission is increased (Figure 2.5) (Ji et al., 2012; J. Sun et al., 2018; Zhang et al., 2017). Interestingly, compared to other anti-reflective coatings, moth eye nanostructures can offer broadband anti-reflectance properties, at high incidence angles of light, both of which are extremely advantageous factors for applications. This technology has been used in industrial applications, including coating solar panels, where the retention of light is of crucial importance, but also for any application where reflection of light from a glass surface is disadvantageous, such as on glasses or on car windscreens (Raut et al., 2011).

Previously, moth-eye anti-glint surface coatings have been created using a top-down approach, where a substance is etched to create the nanopillars by removing the negative space between them (Ji et al., 2013; Raut et al., 2011; Tommila et al., 2010). While this has been successful in creating the desired surface properties, there are a number of drawbacks to top-down approaches. The process is limited by the resolution of the equipment or technique used, meaning that the synthesis of nanopillars with very high aspect ratio, or closely packed pillars can be difficult to achieve, depending on the sophistication of the equipment used. The techniques used are often very expensive, with slow fabrication times and physical limitations, such as the shape of the surface (e.g. curved or complex in topography) (Raut et al., 2011). Finally, it has been found that arrays of randomly arranged nanopillars with diverse heights and widths is actually the most effective arrangement of the pillars on the surface for anti-reflective properties (Siddique et al., 2015). However, this would be extremely difficult to achieve with traditional nanolithography techniques.

Here, I aim to circumvent these problems by providing the building blocks required to synthesise nanopillar arrays with a bottom-up approach. As the building blocks are already at the nanoscale, higher resolution is much simpler to achieve. There are no

restrictions on the shape or type of surface to be coated, the only limitation is in the synthesis of the nanopillars themselves.

To realise this idea, molecules with high length to width aspect ratio are required, ideally with lengths below the wavelength of light desired to be targeted for reduced reflection (Ji et al., 2013) (in this case, optical light: 400 – 700 nm). The molecules should be rigid in structure, so that when assembled on the surface they are able to stand upright and mimic the distinctive 'forests' of nanopillars as seen on the moth eye. I focus on two candidates for nanopillar applications: first an elongated protein, SasG, which has a rigid, elongated structure, and secondly, mini-M13 filamentous bacteriophage, which are also rod-shaped and 350 nm in length. A key requirement of these molecules is that they are capable of being genetically encoded to contain the SpyCatcher protein, and thus can be bound to the BslA surface via SpyTag.

2.1.7 Anti-glint nanopillars: SasG nanorods

One of the candidate molecules for constructing biological nanopillars is the protein SasG (*Staphylococcus aureus* surface protein G), an elongated protein involved in cell adhesion (Figure 2.6) (Gruszka et al., 2012, 2015).

The SasG protein is a multi-domain protein, yet is distinct in its characteristics, being both elongated and very strong - two traits that are usually mutually exclusive in multi-domain proteins. This phenomenon is due to the strong interdomain interfaces, which facilitate long-range cooperativity. The majority of the protein consists of the domains G5 and E, which alternate in order throughout the protein. The longest experimentally characterised SasG protein is termed SasG7, inclusive of the first G5 domain until the 7th G5 domain. SasG7 has been characterised as maintaining rod-like structure in solution, with a length of 67 nm (Gruszka et al., 2015).

Here, I genetically fuse SpyCatcher to the N-terminus of SasG7 (Figure 2.6), to test surface attachment and subsequent conference of anti-reflective properties to the glass. Later in this thesis (Chapter 4), I develop a method to further increase the length of SasG proteins, using an in vitro seamless protein ligation technique.

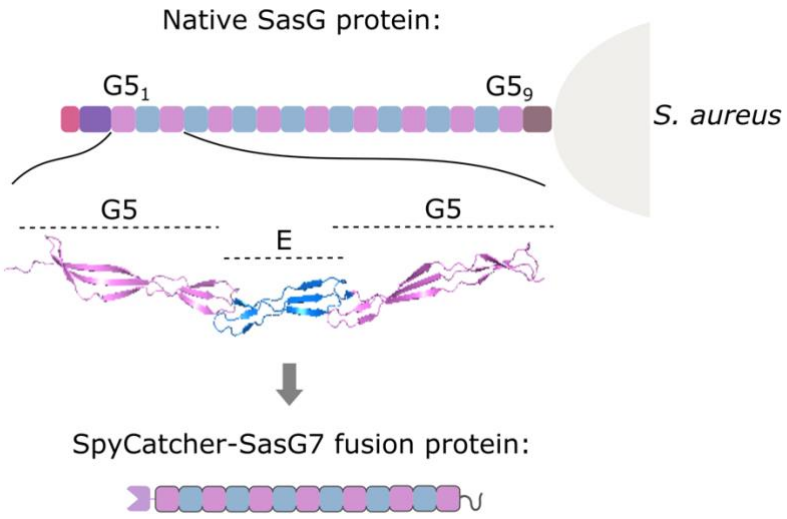


Figure 2.6 Rod-like protein SasG is found on the cell wall of *S. aureus*. Cartoon representation to show placement of native SasG protein on cell wall of S. aureus. The protein is comprised of two repeating domains named G5 (shown in light pink) and E (shown in light blue). The crystal structure of these first three domains together is shown and coloured to match the schematic (PDB ID code: 3TIQ). Other domains of the SasG protein shown in the cartoon: dark pink = signal sequence, dark purple = adhesion domain, brown = C-terminal wall region and LPKTG signal for cell wall attachment. The sequence of native SasG protein was taken from G5(1) to G5(7), and termed SasG7. This was modified with the SpyCatcher protein to allow for surface attachment.

2.1.8 Anti-glint nanopillars: filamentous bacteriophage

To expand the library of nanopillars even further, I looked to filamentous bacteriophage as additional candidates. M13 phage are around 900 nm in length, and 6 nm in diameter. These dimensions, with high aspect ratio, are optimal for the application of nanopillars, yet are slightly too long for arrangement as anti-reflective nanopillars effective for optical light (Ji et al., 2013).

Here, using technology developed by Dr Stanley Brown and colleagues (Brown et al., 2015), initially intended for DNA origami applications, I was able to produce SpyCatcher functionalised mini-M13 phage at a length of 350 nm. The authors truncated the M13 chromosome to only contain the essential replicating components and signals for the phage proteins to pack around: shortening the single-stranded DNA ‘backbone’ of the bacteriophage. They place the coding sequence for the essential phage proteins into a separate plasmid, one which has no ssDNA origin of replication, and so will not be packaged into the phage body (Figure 2.7). This results

in the production of mini-M13 phage: approximately 350 nm in length, with scope to create bacteriophage with lengths anywhere in between 350 and 900 nm through the addition of DNA to the phage plasmid with ss ori. The P3 proteins on the end of the bacteriophage are often genetically modified for phage display: here I modify them for attachment to the surface by the addition of SpyCatcher.

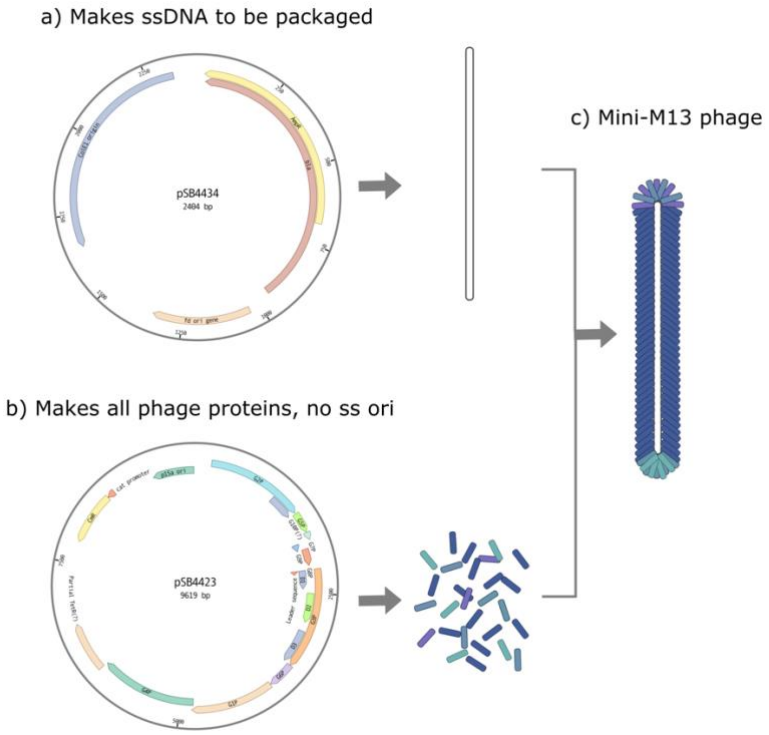


Figure 2.7 Schematic of mini-M13 filamentous bacteriophage production. The M13 genome is split into two plasmids: (**a**) contains minimal DNA sequence to be packaged as single-stranded 'backbone' of phage. (**b**) contains coding sequence for production of all phage proteins required for packaging of DNA into final phage structure. It does not contain a single stranded origin of replication, so this DNA cannot be packaged into phage. Together, these two plasmids create all essential components for the production of the Mini-M13 phage (Brown et al., 2015).

2.2 Results

2.2.1 Protein capture to BslA monolayer

Assembly of fluorescent proteins to BslA surfaces via SpyTag has been shown previously by Professor Lynne Regan's lab group (Williams et al., 2018). In this publication, assembly of the proteins into a monolayer was achieved using a Langmuir trough, which is a device for the formation of surfaces at the air-water interface. After BslA formed a monolayer on the surface of water in the trough, a hydrophobic slide was carefully lowered to meet the water, and deposit the monolayer onto the glass surface via a technique known as Langmuir Schaeffer deposition.

In this work, I wanted to spontaneously assemble the proteins to a monolayer on the surface of glass without the need for a Langmuir trough. Therefore, in the following experiments, pure BslA protein was added directly to very clean and hydrophobic glass for direct formation of monolayers. To test functionality of the monolayers for protein binding, fusion proteins BslA-SpT (SpyTag) and BslA-SnT (SnoopTag) were used for monolayer formation and functionalised with fluorescent fusion proteins containing corresponding Catcher domains (SpyCatcher and SnoopCatcher). Wild-Type BslA monolayers were also frequently used as a control to test for non-specific protein interactions.

Before assembly of proteins to the BslA monolayer, the fluorescent protein fusions were tested in solution for covalent attachment to BslA protein fusions (Figure 2.8). Three different BslA proteins (WT, SnT fused, SpT fused) were reacted with two different protein partners (mCherry-SnC and GFP-SpC). The proteins were mixed in solution and incubated for one hour at RT. Successful covalent attachment of two proteins together is seen as the appearance of a new band on the SDS-PAGE gel. As seen in Figure 2.8, protein band shifts were observed only when the specific Tag/Catcher fused proteins were reacted, indicating that the specific covalent reaction is able to occur in solution with BslA fused to the Tag, and a FP fused to the Catcher. As shown in Figure 2.8, and by others (Veggiani et al., 2016), the SpyTag/SpyCatcher and SnoopTag/SnoopCatcher reaction reaches approximately 80% completion when incubated for one hour at room temperature, and therefore these conditions are used consistently for solution reaction experiments and surface binding experiments herein, unless otherwise specified.

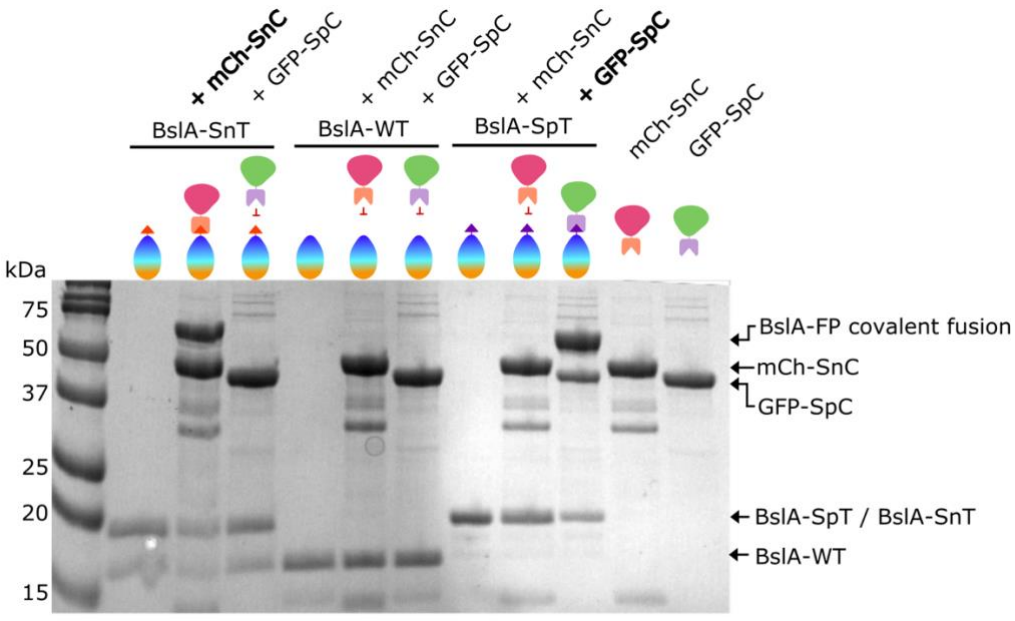


Figure 2.8 SDS-PAGE to show BslA proteins binding specifically to FPs by covalent bond. Three BslA fusion proteins are tested for reactivity with mCherry-SnoopCatcher (44.8 kDa) and GFP-SpyCatcher (41.8 kDa): BslA-SnoopTag (18.3 kDa), BslA-WT (16.5), BslA-SpyTag (18.4 kDa). All proteins used in this thesis have full amino acid and DNA sequences reported in the appendix. An additional protein band can be observed at a higher molecular weight (approximately 60 kDa) when two proteins are joining by a covalent bond. Proteins were combined in buffer and allowed to react at room temperature for 1 hour. The reaction was stopped by the addition of Laemmli buffer, samples were then boiled and loaded onto SDS-PAGE gel.

Following validation of covalent protein ligation in solution, I tested if FP-Catcher could covalently attach to BslA-Tag post monolayer formation. After BslA monolayer formation on a glass surface, I added different concentrations of FP-Catcher proteins to the surface, incubating for one hour before washing away unbound FP-Catcher (Figure 2.9). The results show that BslA surfaces can specifically bind proteins of interest via Tag/Catcher systems. When FP-Catcher proteins are added to BslA-WT surfaces, fluorescence is not retained, indicating that there is very low, or no non-specific binding of the fusion proteins to BslA-WT. When incubated with a corresponding BslA-Tag surface, however, a significant fluorescent signal is seen, indicating specific attachment to the surface via tag/Catcher interaction.

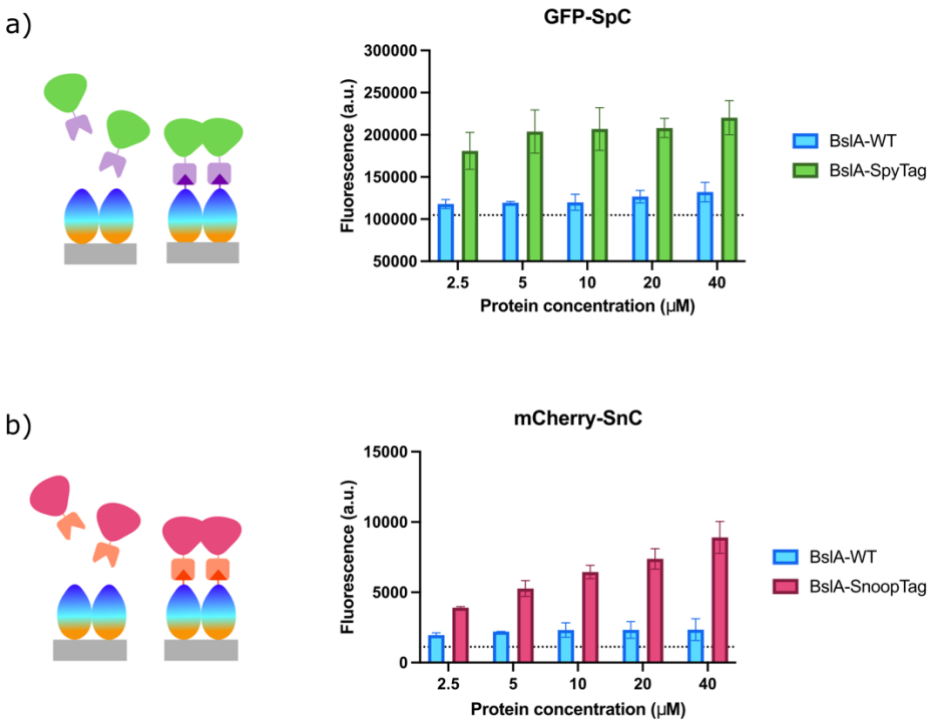


Figure 2.9 Optimising protein binding concentration to functionalised BslA monolayer. Various concentrations of purified GFP-SpC (**a**) or mCherry-SnC (**b**) protein are added to BslA-WT and BslA-SpT (a) or BslA-SnT (b) monolayers, incubated for one hour at room temperature and washed away. The remaining fluorescent signal is imaged with a plate reader. Fluorescence is read using appropriate laser/filter settings for each proteins (For GFP = excitation 485 nm, emission 520 nm, for mCherry = excitation 584 nm, emission 620 nm), with Gain set at 2000. Error bars, s.d. (n=2). Blank values are shown with the dashed line intersecting the y-axis. Here, the blank value is from wells coated with BslA-WT and incubated with buffer.

When the two fluorescent proteins were mixed and added to the surfaces as a heterogenous solution, the BslA surface was still capable of specific capture of the proteins of interest (Figure 2.10). Fluorescence corresponding to the matching FP-Catcher protein is seen only when the surface displays the matching tag: for example, when mCherry-SnoopCatcher is added to BslA-WT or BslA-SpT there is no retention of red fluorescence seen after washing, but when added in the same concentration to BslA-SnT, a significantly higher red fluorescence signal could be seen. This result shows further that the capture of proteins to the surface is specific. Interestingly, I found that the BslA-SnT and mCherry-SnC pairs provided the best results in terms of specific signal to non-specific background fluorescence, in comparison with BslA-SpT and GFP-SpC pairs (Figure 2.10). It is noted that when GFP-SpC is added to BslA-

WT, there is more non-specific binding with the surface than when incubated with a BslA-SnT surface.

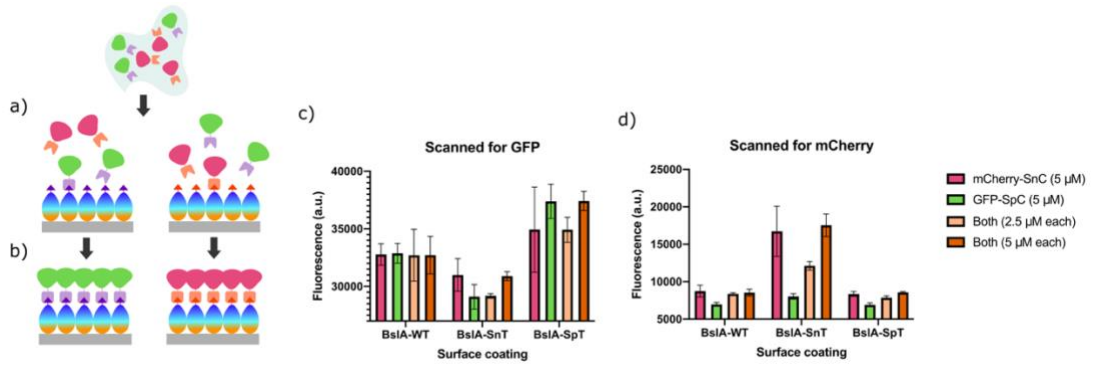


Figure 2.10 Fluorescent proteins bind to BslA surfaces via specific tags. BslA monolayers were formed on hydrophobic glass surfaces within one hour. (**a**) Various concentrations and mixtures of protein were added to monolayers and allowed to react for one hour. After aspiration of unbound protein and washing with 50 mL water (**b**), slides were imaged in the plate reader to assess specificity of fluorescent protein binding. (**c**) Slides scanned for GFP signal (excitation 485 nm, emission 520 nm, gain 1500). Error bars, s.d. (n=4). (**d**) Slides scanned for mCherry signal (excitation 584 nm, emission 620 nm, gain 2500). Error bars, s.d. (n=4).

The results I present here give evidence for proteins specifically attaching to the BslA surface via tags, however they do not necessarily prove that the attachment is covalent. The SpyTag/SpyCatcher and SnoopTag/SnoopCatcher reactions are two step: first there is a non-covalent association step, followed by covalent attachment of protein and peptide (Zakeri et al., 2012). Verifying the attachment to be covalent is an important element of the technology to demonstrate, to ensure proteins are irreversibly attached and not able to be washed away.

Therefore, I completed a Western Blot of the surface attachment of mCherry-SpC to a BslA-SpT monolayer (Figure 2.11). Western Blot was essential over staining techniques such as Coomassie as the protein quantities in each well are very low. Samples of mCherry-SnC protein added to different BslA surfaces were prepared and run on SDS-PAGE, alongside purified mCherry-SnC as control. Probing the membrane with an anti-mCherry antibody shows a visible shift in the size of the mCherry band, only in the wells where mCherry-SnC has been added to BslA-SnT. This indicates the mCherry-SnC protein has bound covalently to the BslA-SpT protein in the monolayer. Some non-specific association of mCherry to the surface can be

seen, as there are faint bands for mCherry-SnC in the lanes showing WT and SpT surfaces. These bands correspond with the size of the purified mCherry-SnC protein, and therefore are unreacted protein.

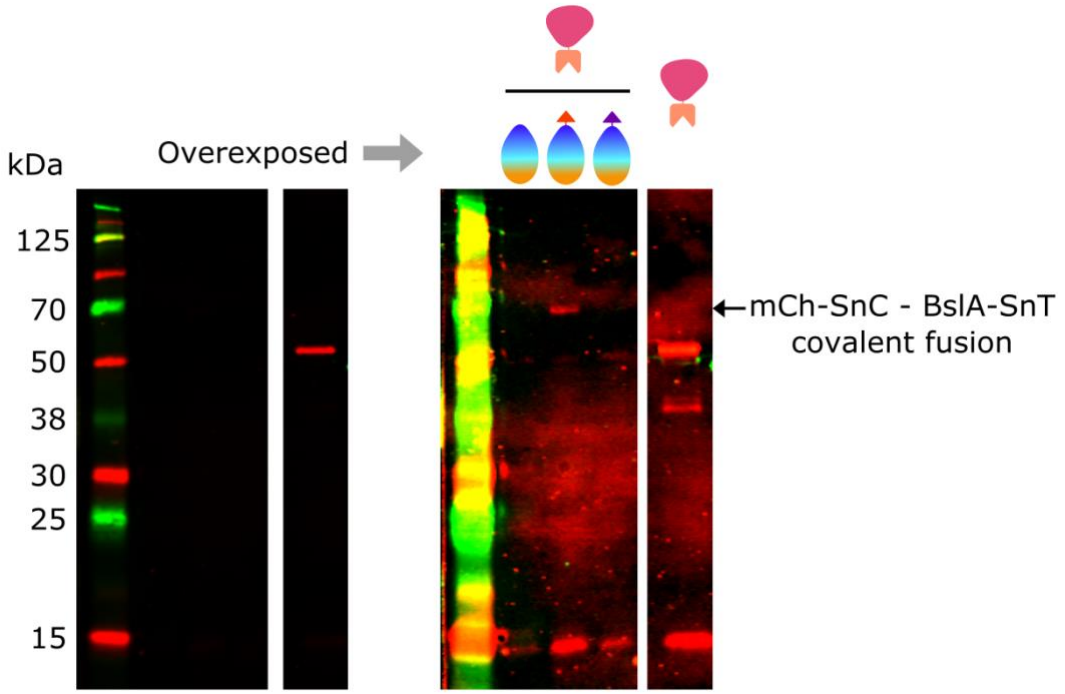


Figure 2.11 Western blot of BslA surfaces reacted with mCherry-SnC. The same Western Blot is shown in both panels, the image has been overexposed on the right hand side to show proteins present at very low concentrations. Samples were prepared by forming BslA-WT, SnT, SpT monolayers on hydrophobic glass surfaces, then reacting each surface with mCherry-SnC for one hour. Excess protein was aspirated, and surfaces washed to remove unbound proteins. These surfaces were then vigorously washed with 10 µL Laemmli buffer, which should disrupt the monolayer and collect all proteins bound to the surface via SnT/C. These samples were then run on a protein gel alongside purified protein stock of mCherry-SnC as reference. The membrane was probed with an anti-mCherry antibody. A band visible with the anti-mCherry antibody at approximately 60 kDa can be seen in lane 3, where the monolayer was formed by BslA-SnT. This shift in MW from the size of unbound mCherry-SnC (44.8 kDa) indicates a covalent binding of the protein to BslA-SnT.

A second band in the Western blot can be seen when probing with the mCherry antibody, around 15 kDa. This could be due to an internal RBS-like sequence in the protein coding sequence, followed by a Methionine, leading to the translation of a truncated protein still detectable by the anti-mCherry antibody, as recently described (Fages-Lartaud et al., 2021).

Together, the results presented in this sub-chapter indicate BslA-Tag proteins can covalently bind Catcher fused proteins, both in solution and in a monolayer.

2.2.2 Movement of BslA proteins in the monolayer

Working with the BslA monolayer for protein attachment, and brainstorming new applications for the technology led to the rise of several questions about the BslA monolayers morphology, in order to better inform which applications it could be suited to. One such question was whether the BslA proteins move in the monolayer once formed. While this question has not been specifically answered in the literature, one paper (Hobley et al., 2013) completes a series of pendant droplet tests with a droplet of solution coated in a BslA protein monolayer. As they contract the size of the droplet by removal of liquid, they observe the wrinkling of the BslA coating. If the BslA proteins were able to move in and out of the monolayer, then when the volume of the droplet is decreased, one would expect the surface coating to contract proportionally to fit the new volume. However, here, they observe wrinkling, which therefore implies the BslA proteins cannot readily move in and out of the monolayer.

To investigate this question further, I used Fluorescence Recovery After Photobleaching (FRAP). FRAP involves photobleaching a fluorescent monolayer using a high-powered laser to bleach the fluorescent molecules in a given area. By tracking the fluorescent signal in that area, you can determine whether the molecules are able to freely diffuse: if they can, you will observe recovery of fluorescent signal as unbleached fluorescent molecules diffuse into the bleached area, whereas if there is no movement of molecules, you will not observe recovery of fluorescent signal and the surface will remain ‘dark’ (Figure 2.12).

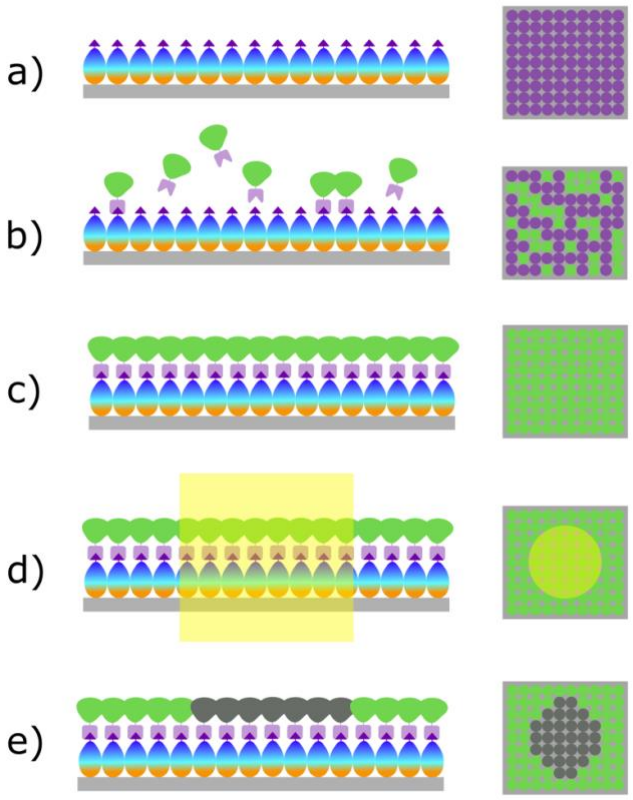


Figure 2.12 Overview of set up for FRAP experiments of BslA surfaces. The left-hand panel of this figure shows the side view of protein assemblies to the glass substrate. The right-hand panel shows a top-down view of these proteins within the microscope's field of view (molecular detail not to scale). (**a**) BslA-SpyTag proteins self-assemble on a hydrophobic glass coverslip. (**b**) GFP-SpyCatcher protein is incubated with glass slide for one hour, allowed to bind to BslA-SpyTag surface. (**c**) Excess GFP-SpyCatcher is washed away, leaving GFP covalently attached to monolayer via SpyTag/SpyCatcher. Slide is placed on microscope and data collection begun. (**d**) A second, high power laser illuminates spotlight in centre of field of vision for a matter of seconds, photobleaching GFP molecules in spotlight. (**e**) Laser is switched off, revealing photobleached GFP in area of illumination. Fluorescence in this area is monitored over several minutes to examine fluorescence recovery.

To use this method for assessment of BslA monolayers, I formed BslA-SpT monolayers on the surface of glass, and added GFP-SpC, to specifically label the proteins in the monolayer with fluorescence. As shown in Figure 2.13, I saw no recovery of fluorescence after photobleaching the GFP proteins in a given area. Fluorescence was monitored for 25 seconds post photobleaching in the data presented in Figure 2.13, however longer time-scale experiments were also conducted to ensure the movement of the proteins was not occurring on a longer time

scale. No recovery of fluorescence was seen after five minutes. Before photobleaching (0 – 2.5 seconds in Figure 2.13), there are visible differences in the fluorescence of the surfaces, consistent with expectations of surface attachment. Importantly, fluorescence is significantly higher when surfaces are composed of BslA-SpyTag proteins, indicating that GFP-SpyCatcher is specifically binding to the surface only when SpyTag is present. This is important because without specific fluorescent labelling of the monolayer molecules, you cannot infer movement of the molecules from the results. Higher fluorescence is seen on the blank slide sample than BslA-WT surface. This is expected, as proteins are likely to stick to a hydrophobic surface in a non-specific manner. Crucially, the BslA-SpyTag surface has much higher fluorescence than either negative control. Together, these results provide further confidence that the BslA surfaces are specifically binding proteins via SpyTag/Catcher, and indicate that the BslA proteins do not move in the monolayer once formed.

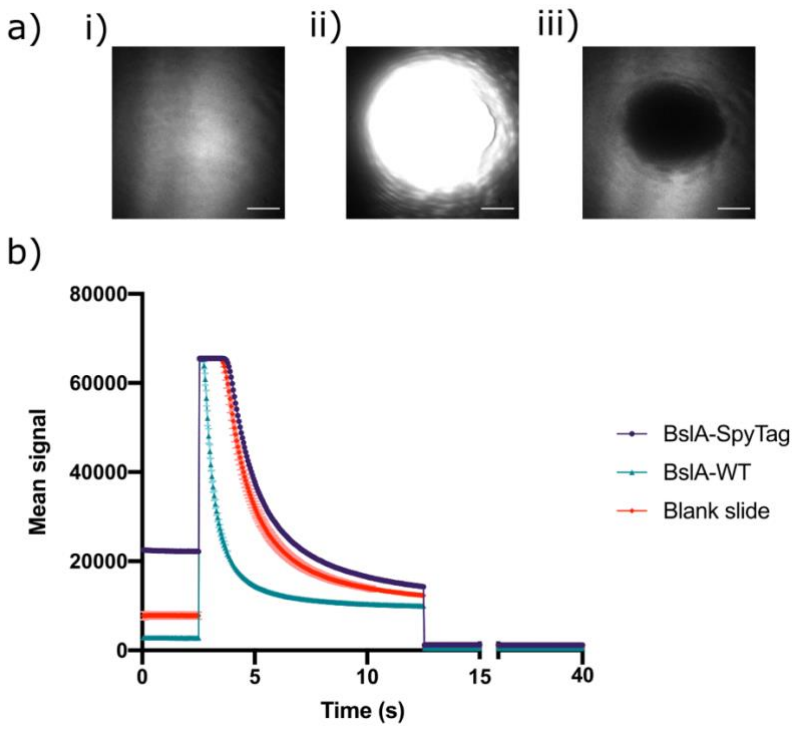


Figure 2.13 FRAP data from fluorescent BslA monolayer. (**a**) Representative images of the monolayer at different stages of the experiment presented in (**b**). Initially, the whole surface appears fluorescent, or 'bright' (**i)** as the whole surface is evenly illuminated by a low power laser. The second, high power laser illuminates an area in the centre of the slide (**ii**). Once switched off, the photobleaching of these surface bound proteins is visible (**iii**) as the area

appears ‘dark’. Data is continually collected to look for recovery of this ‘darkness’. In all images, scale bar represents 10 um in length. (**b**) GFP signal over time of BslA surfaces. Data is collected as an average signal within a defined square area in the centre of the field of view. BslA-WT and blank slide serve as negative controls to the functionalised BslA-SpyTag surface. The first two seconds show fluorescent signal pre-bleaching, and corresponds to a) i). Seconds 2 - 12 show signal when the second, high power laser is activated, corresponding to a) ii). After 12 seconds, the laser is switched off and data collected to detect recovery of fluorescence. This corresponds to image a)iii). Error bars, s.d. (n=4).

2.2.3 DNA capture to BslA monolayer

Following the success of attachment of proteins to BslA surfaces via Tag/Catcher chemistry, I sought a mechanism to bind DNA covalently to the BslA surface. This was possible with the use of the mVirD2 protein.

mVirD2 is a truncated version of mVirD protein, containing only the essential domains for covalent attachment to DNA prompted by association with a short recognition sequence. Here, I use it as a mediator between the BslA surface and DNA of interest (Figure 2.14). I added the SpyCatcher domain to mVirD2, allowing it to be displayed on BslA surfaces where a SpyTag is present. I overexpressed the fusion protein in *E. coli* BL21 and purified it via the His-tag under denaturing conditions.

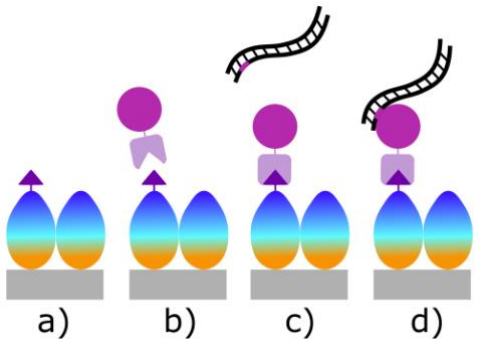


Figure 2.14 Schematic to show DNA binding to BslA surface via mVirD protein. The three-step process is visualised in sequence as following: (**a**) BslA-SpyTag fusion protein forms monolayer on glass, (**b**) mVirD-SpyCatcher fusion protein binds to monolayer covalently, (**c**) DNA binds to mVirD protein via specific recognition sequence, resulting in (**d**) DNA covalently linked to protein monolayer on the surface of glass. Each incubation step (a) through (c) takes one hour.

I first wanted to test if the mVirD2-SpC fusion protein would behave as previously characterised in my hands. While the original paper focuses on attachment of ssDNA

to mVirD2, I wanted to test covalent attachment to dsDNA too. Given mVirD2's native role in the cell as a relaxase (Van Kregten et al., 2009), I inferred it should be capable of attachment to dsDNA via one strand.

To test this I completed a binding experiment with a Cy3 labelled oligo encoding the mVirD2 recognition sequence (Figure 2.15). Following the defined protocols from the original paper (Bernardinelli & Högberg, 2017), I incubated the mVirD2-SpC protein with various DNA for one hour at 37ºC with 2 mM $MgCl_2$. The samples were then mixed with Laemmli buffer before being separated by SDS-PAGE on a denaturing gel (Figure 2.15). This process should ensure that any band shifts seen on the gel are due to covalent bond formation between DNA and protein, an important distinction from the DNA only associating with mVirD2. I scanned the gel using three different techniques to visualise both DNA and protein. In Figure 2.15(a), a Cy3 labelled band shift is seen when the protein is incubated with both ssDNA (oligo alone) and dsDNA (labelled oligo annealed to complementary oligo).

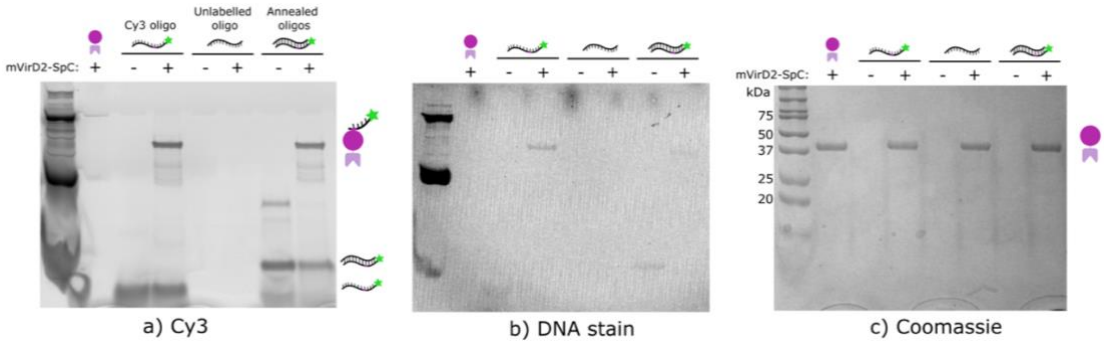


Figure 2.15 SDS-PAGE analysis of mVirD2-SpC fusion protein covalently binding DNA. mVirD2-SpC protein was incubated in combination with different DNA samples: Cy3 oligo (ssDNA with label), unlabelled oligo (complementary oligo with no Cy3 label), or annealed oligos (labelled and unlabelled oligo annealed to produce dsDNA sample with Cy3). The protein and DNA samples were also run separately on the gel as controls. Visualising the same gel for (a) Cy3 signal, (b) DNA stain and (c) Coomassie enabled the visualisation of both DNA and proteins. MW markers are only labelled in (c) as these bands could not be visualised with Cy3 or DNA stain imaging techniques. The gels are aligned vertically so location of bands should be comparable between gels. Expected size of DNA and protein complexes is labelled with cartoons. mVirD2-SpC = 37.1 kDa, each oligo = 35 nucleotides. Gels were imaged and stained in the order reported.

As the protein ladder loaded on the gel does not contain DNA MW standards, it's difficult to be sure which of the larger, additional bands correspond to dsDNA in the annealed oligo section of 2.15(a). Roughly estimating the size of the annealed DNA, here I assume that the lower Cy3 labelled band is the two DNA oligos annealed. Importantly, the appearance of a Cy3 labelled band at the correct size is seen only when mVirD2-SpC is added to the mixture. Assessment of protein content via Coomassie staining shows there is no shift in the size of protein bands when DNA has attached, possibly due to the small size of the DNA being bound to the protein, probably only increasing the size of the protein complex by 2 – 4 kDa. This aligns with results seen in previous work (Bernardinelli & Högberg, 2017), where authors have commented that this shift in protein size is difficult to observe via standard protein gel imaging methods.

These results indicated that ssDNA and dsDNA can covalently bind to mVirD2-SpC protein in solution. Next, I wanted to test this binding capability with the BslA surface. I incubated the BslA-SpT surface with mVirD2-SpC for one hour at room temperature as previously specified, then removed excess protein and incubated the surface with the cy3 labelled oligo (Figure 2.16) for one hour at 37 ºC with 2 mM $MgCl_2$. Following this, excess DNA was removed, and the surface washed. Removing the microplate adaptor allowed for better visualisation in a slide scanning device. Figure 2.16 shows that Cy3 signal is retained on the surface only when the monolayer is formed by BslA-SpT. This seems dependant on concentration of DNA added: below 0.63 µM DNA added the difference in cy3 signal between BslA-WT and BslA-SpT surfaces is barely visible. From these experiments, I concluded that BslA surfaces can bind DNA via the mVirD2 protein, yet the concentration of DNA added for binding should be equal to or higher than 0.63 µM. This value approximately aligns with the binding rate of oligo to protein reported in the supplementary of (Bernardinelli & Högberg, 2017), where they approximately measure Kd by gel shift assay to be 468 nM +/- 68.

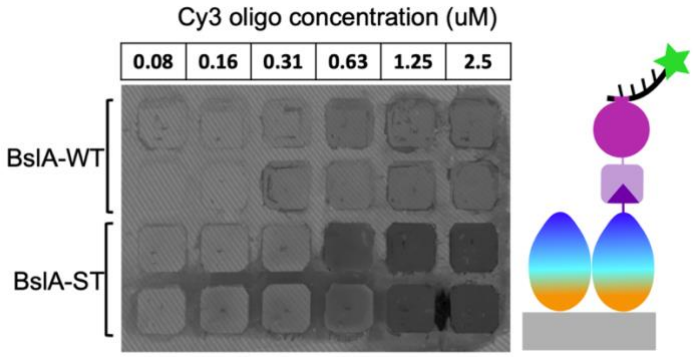


Figure 2.16 BslA surfaces specifically bind labelled DNA via mVirD2 protein. In each well, BslA (WT or SpyTag fused) forms a monolayer, followed by covalent attachment of mVirD2 via SpyCatcher. Cy3-labelled DNA is then covalently attached to the surface via mVirD2, and the slide is imaged for Cy3 signal. Schematic of these attached protein/DNA components is shown on right hand side. Six DNA concentrations were tested with each protein monolayer in duplicate. Signal is expected to increase as higher concentrations of cy3 oligo are added to BslA-SpT-mVirD-SpC surfaces. Background signal is also seen to increase in wells with monolayers only formed with BslA-WT, this could be due to some non-specific sticking of mVirD2-SpC to BslA-WT rather than non-specific binding of cy3-oligo to BslA-WT.

2.2.4 BslA monolayers for protein interaction screening

Display of functional proteins on the BslA monolayer could allow for downstream screening of interactors. In collaboration with Manfred Auer's lab, I explored this idea using the medically relevant protein Survivin as a test-case, known to specifically bind the peptide AKER (Jeyaprakash et al., 2011). This idea is visualised in Figure 2.17. While the initial proof of concept (Figure 2.17(b)) was successfully shown, further work on identifying novel interactors with Survivin (Figure 2.17(c)) has not occurred, mainly due to the closure of the Auer lab. Nevertheless, I present this work in my thesis as it helps to demonstrate the potential of BslA monolayers as self-assembling protein tools for novel applications, even though evidence presented is only initial proof-of-concept.

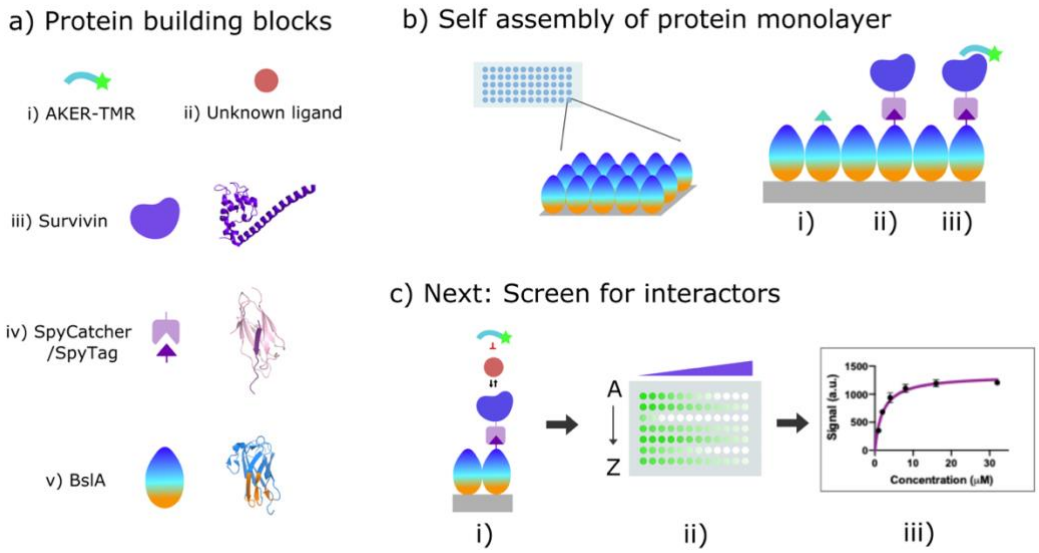


Figure 2.17 Overview of BslA surfaces as a platform for screening Survivin interactors. **(a)** Cartoon representations (plus ribbon representations of protein structures, where known) of the protein 'building blocks' used in this work. **(b)** Assembly of protein components to build a protein scaffold in micro-wells on a glass coverslip: **(i)** BslA (with, and without SpyTag) is added to the hydrophobic glass surface. The protein self-assembles into a monolayer, displaying the SpyTag for binding with its protein pair. **(ii)** SpyCatcher-Survivin protein is added and the spontaneous covalent bond formation allows for irreversible binding of functional Survivin protein on the BslA surface. **(iii)** AKER-TMR is a labelled peptide which can bind with to the Survivin protein. It can bind when Survivin is displayed on the surface in this fashion. **(c)** Workflow for next hypothetical experiments to screen for novel interactors: **(i)** Following stepwise creation of the surface, I would compete against AKER-TMR with a library of unknown ligands to find novel interactors. **(ii)** Binding test of different ligands across a range of concentrations, inferring binding from loss of TMR signal. **(iii)** Analysis of this read-out to understand which ligands look most promising for binding to Survivin.

The design of the Survivin fusion protein was directed by existing Survivin fusion proteins and analysis of the protein structure. Previously, Survivin fusion proteins had purification tags, and even fluorescent proteins at the N terminal of the protein, which was still capable of native functionality. Looking at the crystal structure for Survivin (Fig 2.17(aiii)), there is a long alpha-helix at the C-terminal, which is therefore an appealing site for attachment of a functional domain e.g. SpyCatcher, as the tag should remain exposed from the rest of the protein and available for binding to the SpyTag. Following this reasoning, no flexible linker was used, instead SpyCatcher is

directly fused to Survivin. Therefore the fusion protein expressed and purified contained a coding sequence of 6xHistidine, 3C cleavage site, Survivin, SpyCatcher.

I purified the Survivin-SpyCatcher fusion protein, along with other proteins required for this study, achieving high yields of all recombinant proteins (10 – 50 mg per litre of cell culture). To confirm functionality of the SpyCatcher domain, I reacted the Survivin-SpyCatcher protein with BslA-SpyTag in solution for one hour at room temperature, over a range of molar ratios (Figure 2.18). The appearance of a new protein band at the expected molecular weight of Survivin-SpC combined with BslA-SpT implies the proteins can indeed be linked by a covalent bond using SpyTag-SpyCatcher, and the reaction is efficient.

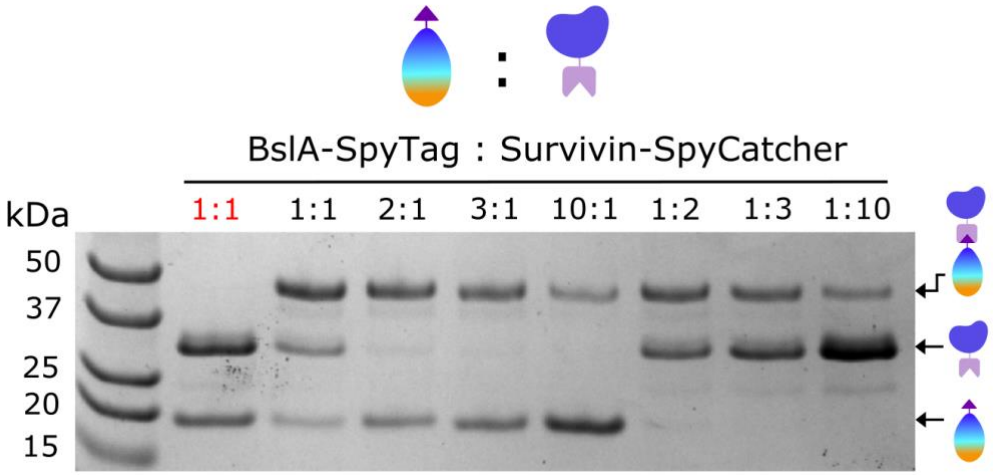


Figure 2.18 SDS-PAGE to show BslA-SpT covalently binding Survivin-SpC. BslA-SpyTag and Survivin-SpyCatcher were mixed at varying molar ratios, and the formation of a covalent linkage between the two proteins was detected as the formation of a new protein band, corresponding to 49.4 kDa, on a denaturing SDS gel. Lane 1 = MW marker. Lane 2 = Unreacted proteins, where proteins were mixed directly into Laemmli buffer which inhibits the covalent bond formation. Lanes 3 – 9 = varying ratios of proteins reacted together for one hour.

To validate the binding of Survivin-SpC to BslA-SpT after monolayer formation, I used a two-step competition fluorescence-based binding test. I hypothesised that reacting the BslA-SpT surface with Survivin-SpC before binding of GFP-SpC would result in low fluorescence, as no SpyTag would be available for GFP-SpC to bind to. Conversely, if Survivin-SpC was not first reacted with the BslA-SpT surface, then all SpyTag should be available for binding to GFP-SpC, resulting in high fluorescent

signal. As a negative control, I included a surface composed only of Wild-Type BslA, with no BslA-SpyTag included. This surface should not bind GFP-SpyCatcher via formation of a ST-SC bond. Any fluorescence from this surface would represent a non-specific background 'sticking' to the BslA surface.

As shown in Figure 2.19, pre-incubation with Survivin-SpC prevents GFP-SpC from binding to the BslA-SpT surface. From this result I deduce that Survivin-SpC binds to the BslA-SpT surface, specifically, via covalent attachment to the SpyTag.

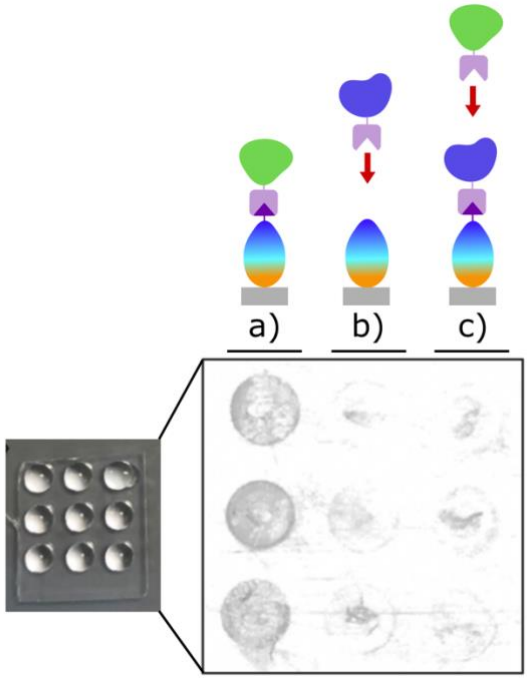


Figure 2.19 BslA-SpT monolayer can specifically bind Survivin-SpC. Microwell slides (left photograph) were incubated with BslA to form a monolayer, then functionalised with different proteins. Slides were scanned for GFP signal, this image is shown on the right. **(a)** GFP-SpC added to BslA-SpT monolayer, and results in a high fluorescent signal. **(b)** GFP-SpC added to BslA-WT, and results in low fluorescent signal. **(c)** Survivin-SpC added to BslA-SpT surface, followed by GFP-SpC, and results in low fluorescent signal. I hypothesise that all binding sites are taken by Survivin-SpC, therefore GFP cannot bind and the fluorescent signal is low.

Having established that Survivin can be presented on the BslA monolayer, I next tested if that surface bound Survivin is correctly folded, by testing if it can bind to a known ligand: the AKER peptide. The Auer lab has previously shown that this fragment of a natural protein ligand of Survivin binds specifically to Survivin, with a K_D of approximately 5 µM.

I prepared the BslA-SpyTag surface as described, and incubated it with Survivin-SpC for one hour. Following this, I incubated the slide with different concentrations of AKER-TMR. As seen in Figure 2.20, I only see TMR signal when the slide has been incubated with BslA-SpT, Survivin-SpC and the peptide sequentially. The lack of signal in the negative control implies that there are minimal, or no, non-specific interactions between either the Survivin-SpyCatcher protein, or the AKER-TMR molecule, and the BslA surface.

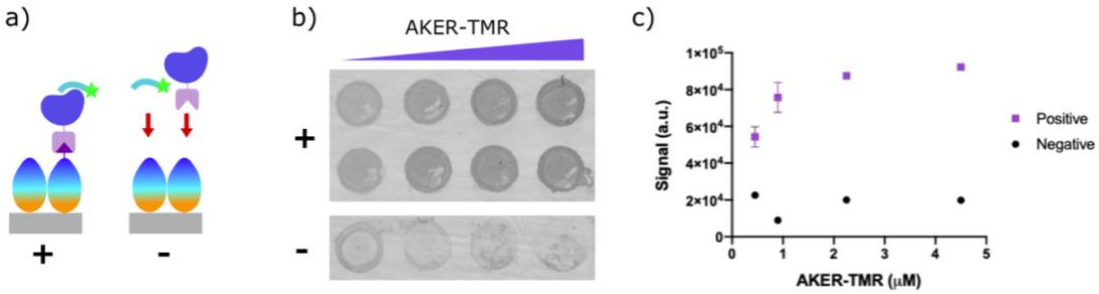


Figure 2.20 AKER-TMR peptide can bind to Survivin displayed on BslA surface. (**a**) Schematic to describe protein components on surface: the positive (+) sample is a BslA monolayer displaying Survivin via the SpyTag/SpyCatcher interaction, which can then be bound by the labelled AKER peptide. As a negative (-) control, I incubated Survivin-SpyCatcher followed by AKER-TMR, but with a BslA-WT surface. The negative control tested for non-specific interactions between Survivin-SpyCatcher and AKER-TMR to BslA-WT. (**b**) Positive samples were prepared in duplicate. AKER concentration increases in the wells from left to right. The slide was then scanned for TMR signal and (**c**) signal plotted against concentration of peptide added. Error bars, s.d. (n=2).

In conclusion, the results presented here show preliminary evidence for BslA monolayers as tools to present medically-relevant proteins for screening of interacting molecules.

2.2.5 BslA monolayers for anti-glint applications

A further proposed application of BslA surfaces was to ‘build’ nanopillar arrays on the surface of glass (Figure 2.21). The successful creation of such nanoscale arrays should theoretically provide anti-reflective properties to the glass, based on the characterised properties of the moth eye: consisting of arrays of sub-wavelength nanostructures that created a gradient refractive index on the surface, increasing transmission of light through the glass and reducing reflection.

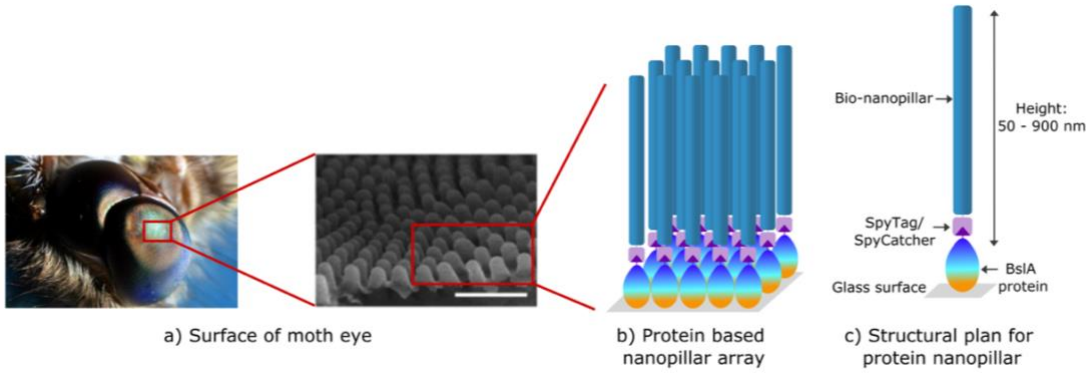


Figure 2.21 Overview of nanopillars for anti-reflective properties and key components used in this study. (**a**) The sub-wavelength structures in moth eyes exhibit fascinating antireflective properties over the broadband wavelength region and at large incident angle by generating an air-mixed heterogeneous optical interface. Left-hand photo © Rick Cowen. Right hand photo adapted from (Ko et al., 2011), scale bar = 500 nm. (**b**) I aim to mimic these unique structures found in nature using a protein based nanopillar array. (**c**) Bio-nanopillars can be produced between the lengths of 50 – 900 nm and covalently attached to the BslA protein via the SpyTag/SpyCatcher interaction. The BslA protein can readily coat hydrophobic glass surfaces, thus providing a foundation for the structures.

To create nanopillar arrays on the surface of BslA monolayers, I required nanopillars capable of SpyCatcher fusions for attachment to the BslA-SpyTag protein. I sought nanopillar like protein-based materials that were rod-like and rigid, under 800 nm in length and with a high aspect ratio. Two different nanopillars were developed to meet those criteria, ranging in length between 50 – 900 nm: SasG protein (50 – 200 nm), and filamentous bacteriophage (350 – 900 nm).

The SasG7 protein was selected as the longest native characterised SasG protein (Gruszka et al., 2015). It is characterised as being rod-like in solution, with a length of 67 nm. Showing proof of concept with SasG7 seemed the best initial test for the system, with the capability to test shorter SasG proteins (SasG3, 4, 5, 6 have also been characterised as rod-like in solution), or longer SasG proteins via ligation using split inteins (detailed further in Chapter 4).

A SasG7-SpC fusion protein was expressed and purified from *E. coli* BL21 with strep tag at the C-terminus for purification purposes and SpyCatcher at the N-terminus for attachment to the BslA-SpyTag monolayer (Figure 2.22). This fusion protein was

capable of covalent attachment to BslA-SpyTag in solution, so I continued to test attachment of SasG7-SpC to BslA-SpyTag in the monolayer.

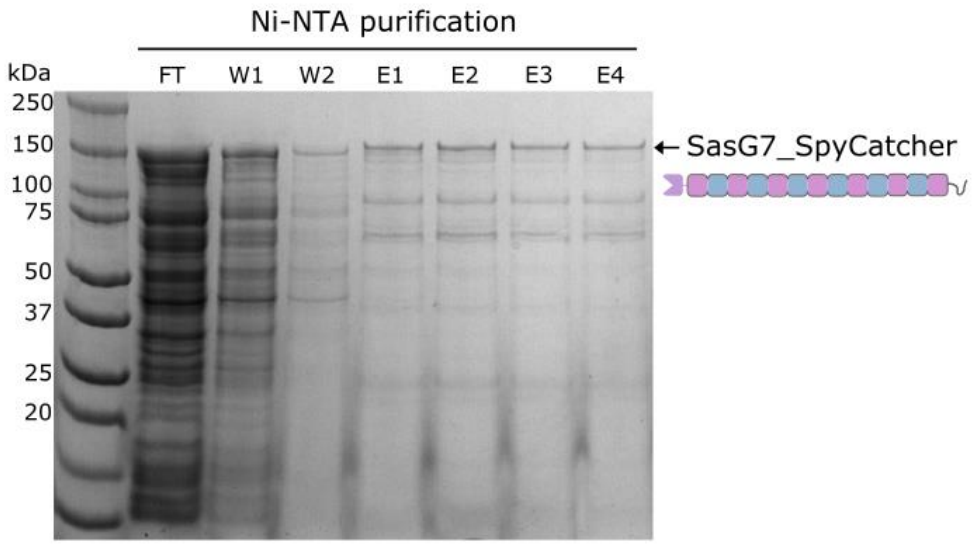


Figure 2.22 SDS-PAGE analysis of SasG7-SpC fusion protein purification. The fusion protein of interest can be seen to be eluted at the expected MW (given the understanding that SasG proteins run at a higher MW than expected). Theoretical MW of protein = 117 kDa.

The second bio-nanopillar developed for the purpose of anti-reflective nanopillars was filamentous bacteriophage. Using a unique system of phage production developed by Dr Stanley Brown (Brown et al., 2015), I was able to produce and purify bacteriophage with a length of 350 nm, as validated by TEM (Figure 2.23).

I modified the phage for functionalization of BslA surfaces by fusion of SpyCatcher to the P3 protein, a phage capsid protein with 3 to 5 copies present per phage. Successful functionalisation of the phage with BslA was tested via Western Blot, by incubation of P3-SpC phage with BslA-SpT for one hour and separating samples on a denaturing SDS-PAGE gel. The membrane was probed with an anti-P3 antibody, looking for a shift in the size of the protein band. As seen in Figure 2.23(c), two bands can be seen when phage are functionalised with SpyCatcher, indicating that attachment of SpyCatcher to P3 is not 100% efficient. This is not uncommon in the modification of phage for phage-display, and some attribute it to phage protein cleavage systems. It is unclear from this data if a fraction of the phage population have all P3 proteins modified with SpyCatcher, or if all phage have a fraction of their P3 proteins fused to SpyCatcher. After incubation of the modified phage solution with

BslA-SpT, a third, higher band can be seen on the Western Blot (Figure 2.23(c)), indicating successful covalent bond formation with BslA-SpT. Unreacted P3-SpC fusions to BslA-SpT could be due to steric hindrance of the arrangement on the phage. Again, it is difficult to say how these functionalised P3 proteins are distributed within the phage population: ideally, at least one P3 protein is successfully functionalised with SpyCatcher and able to bind to SpyTag for attachment to the surface.

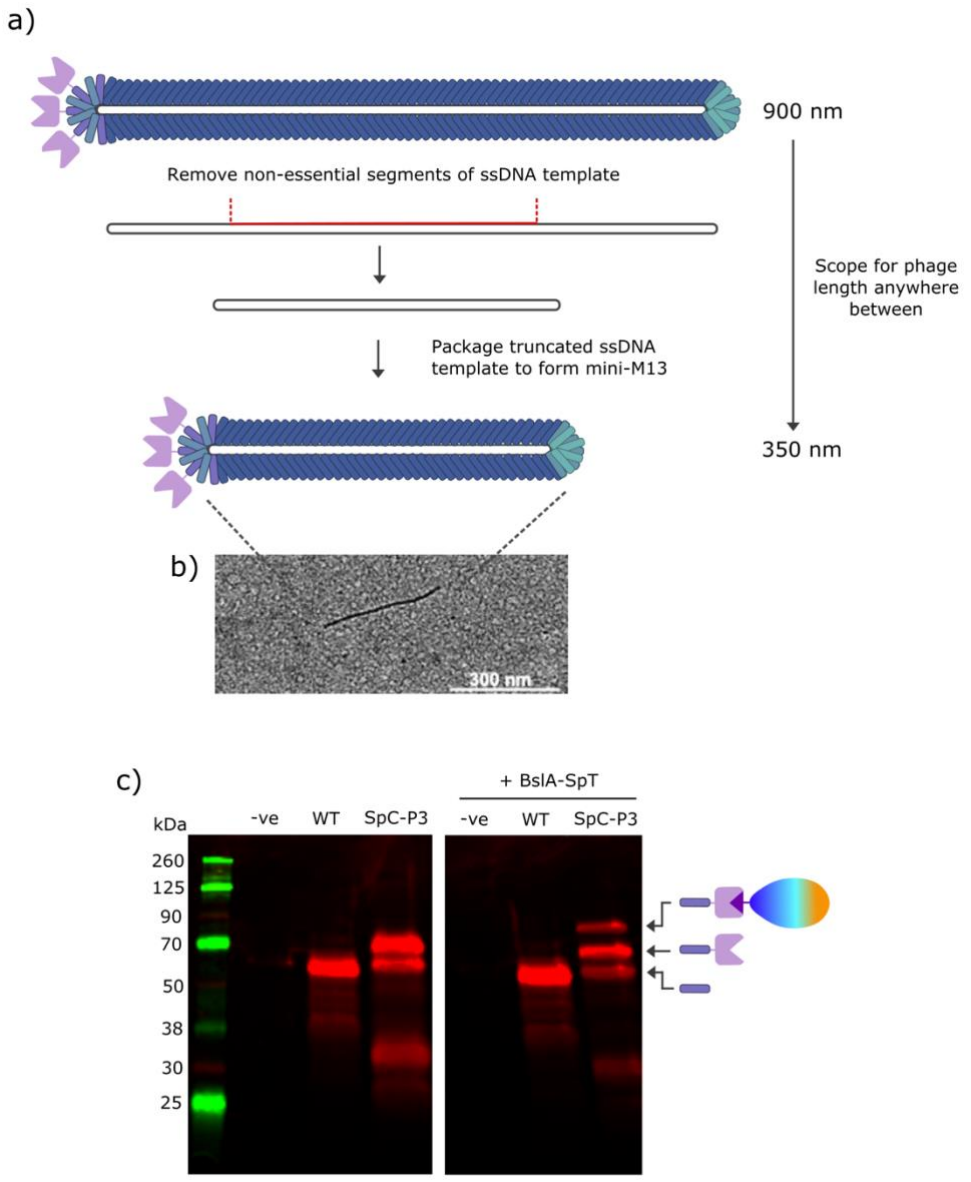


Figure 2.23 Design and production of modified mini-M13 bacteriophage. (**a)** To produce filamentous bacteriophage of customizable lengths, I followed previous work by Dr. Stanley

Brown. Removing non-essential ssDNA from the phage genome and repackaging in a separate plasmid allows for the packaging and creation of a truncated M13 phage. The size is reduced from 900 nm in length to 350 nm, yet via incorporation of DNA back into the genome, a length anywhere between these two could feasibly also be created. (**b**) I validated that the described production technique was effective, even when P3 protein was modified with SpyCatcher by negative stain TEM, where filamentous phage 350 nm long were observed. (**c**) Western Blot using Anti-M13 P3 antibody to analyze functionality of SpyCatcher fused to phage. Here, the negative control is *E. coli* cells containing only one of the essential two plasmids for phage production. WT refers to mini-M13 phage with unmodified P3. SpC-P3 refers to phage samples with SpC genetically fused to P3. Expected MW of proteins (kDa): P3 = 42.5 (known to run ~60 kDa), SpC-P3 = 55, BslA-SpT = 18.3, SpC-P3 + BslA-SpT = 73.3. Comparing WT to SpC-P3, two bands are shown with the anti-M13 P3 antibody; the P3 protein alone, and at a larger MW, the P3 protein with SpyCatcher attached. When BslA-SpyTag is added to all three samples, it's possible to observe the appearance of a third band, indicating that the P3-SpyCatcher protein has successfully bound to BslA-SpyTag by covalent bond. This additional band does not occur for the negative control or WT mini-M13 phage.

To use bacteriophage for biomaterials applications, I aimed to purify the phage away from cellular proteins as completely as possible. I tested various combinations of PEG/NaCl concentrations in tandem with isoelectric precipitation techniques (Figure 2.24), however I ultimately found that many contaminating cellular proteins remained. Likely this is because most of the techniques are developed for phage display, where absolute purity of phage is not of high importance. To overcome this, I applied protein purification techniques to the purification of phage (Zakharova et al., 2005), treating the bacteriophage as a MDa sized protein complex. Using size exclusion chromatography as a final clean up step, I was able to separate the phage from contaminating proteins, and obtain a high purity phage sample (Figure 2.24). The double peak observed in the chromatography read-out is likely an artefact from saturation of the detector: I checked the fractions of this peak for capability to bind with BslA-SpyTag (Figure 2.24(ci)) and they all were capable of binding to a similar degree, indicating that these two peaks are not due to one population of the phage containing the P3-SpyCatcher fusion and the other P3-WT.

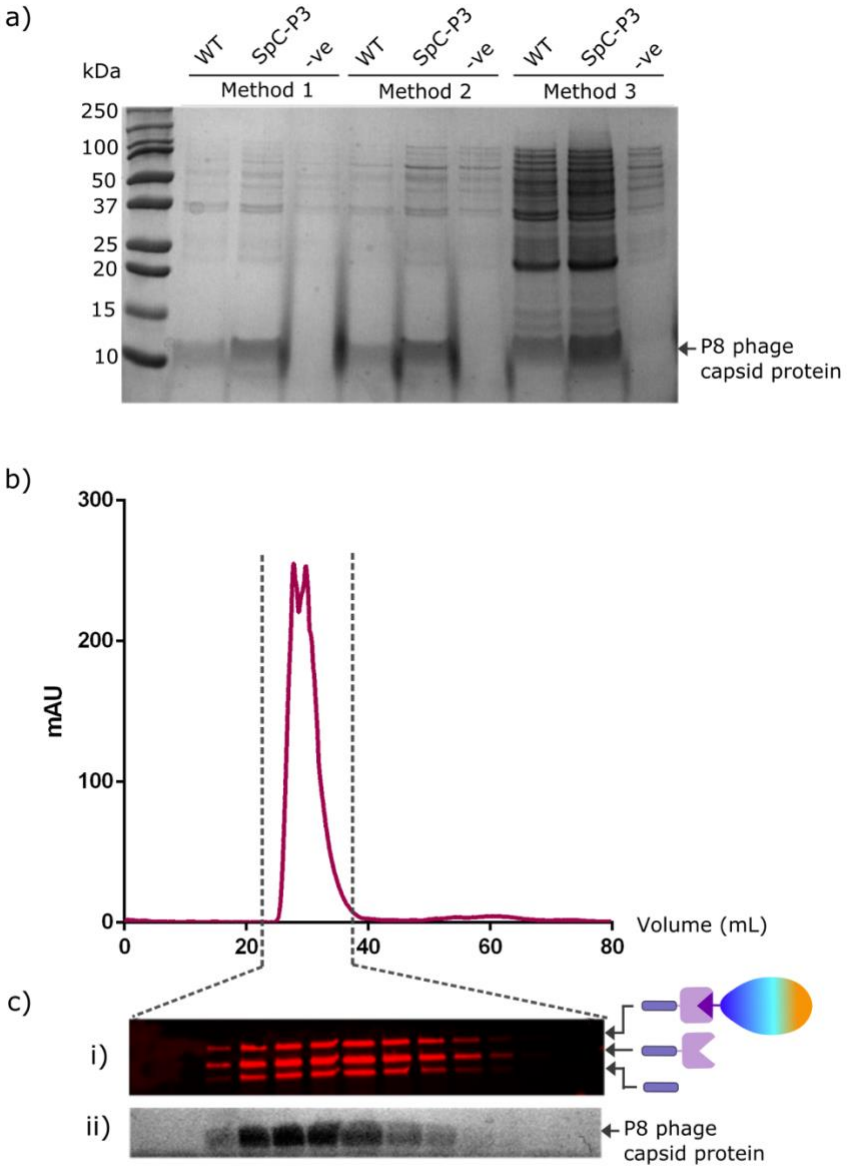


Figure 2.24 Optimisation of bacteriophage purification. (**a**) SDS-PAGE gel to assess purity following common phage purification methods. Three samples of cells were prepared, to be compared simultaneously: WT mini-M13 phage, SpC-P3 mini-M13 phage, and a negative control of *E. coli* cells expressing only one of the two required plasmids for production of mini-M13. Method 1 = High MW PEG precipitation followed by two rounds of isoelectric precipitation. Method 2 = Low MW PEG precipitation followed by two rounds of isoelectric precipitation. Method 3 = Three rounds of isoelectric precipitation. The corresponding band to the P8 phage capsid protein, the most abundant capsid protein with 1000s of copies per phage, is annotated on the gel, indicating the presence of phage in the sample. (**b**) Purification of phage by Size Exclusion Chromatography. Sephacryl S-400 was used to separate phage

particles (peak between 20 and 40 mL) from contaminating cellular proteins (peak at 60 mL). (**c**) Fractions of the peak corresponding to the lines shown on the graph were analysed by both Western-Blot (**i**) and Stain-free (**ii**). An aliquot from each fraction collected from the column was reacted with BslA-SpyTag for one hour, and samples were then run on SDS-PAGE and probed using an anti-P3 antibody, where the covalent attachment of P3-SpC to BslA-SpT is observed by the appearance of a third, higher MW band as previously shown in Western-Blots (Figure 2.23). The same fractions are shown in (**ii**), where the gel was scanned using stain-free before transfer to the membrane for Western-Blotting. Visible is the P8 phage capsid protein, providing a marker as to where the phage are present in the fractions.

Purification and validation of binding in solution of the two nanopillars provided confidence to move to the next stage of work: application of the nanopillars to a surface and measurement of anti-reflective capabilities. Figure 2.25 shows reflection and transmission measurements of visible light from these experiments. The samples were created following previously defined protocols: protein monolayer formation via incubation of pure BslA protein with a hydrophobic glass surface, attachment of protein of interest via the SpyCatcher, washing of excess from the surface, and imaging. A successful anti-reflective surface would decrease the amount of light reflected from the surface, yet increase the amount of light transmitted through the glass.

In Figure 2.25 I see no significant difference in the reflection and transmission of light through the surfaces when the bio-nanopillar assembly is present. The only significant difference between the samples is seen between the blank slide control and all other samples. The completely blank slide has significantly higher reflectance properties than the others, however this is a predictable trend as ‘dirtying’ a glass surface slightly is known to make it less reflective or ‘shiny’ as small molecules coat it. Addition of reflectance and transmission values for each sample equals a value higher than 100%, which is technically impossible. As this difference is consistent between all samples: summing to equal 101-102%, I attribute this to inaccuracies resulting from equipment used for measurement. Crucially, there appears to be no difference in reflection or transmission when the nanopillars have been assembled on the glass surface compared to negative controls without nanopillars assembled.

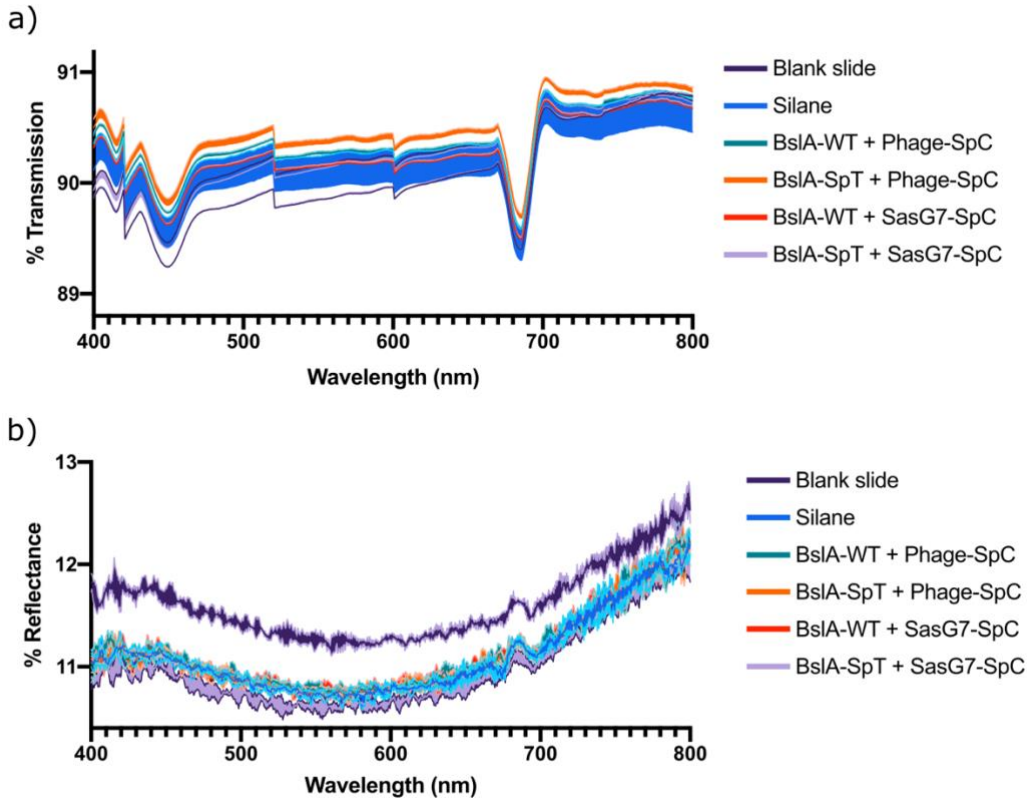


Figure 2.25 Measured reflection and transmission of visible light through glass slides coated with protein assemblies. Where Blank slide = completely untreated glass, Silane = glass slide treated with silane to confer hydrophobic properties, and various BslA monolayers with nanopillar combinations are shown with other samples. Nanopillars are added to BslA-WT surfaces as a negative control. (**a**) Percent transmission of light through the glass is measured between 400-800 nm wavelengths. Error bars, s.d. (n=3). (**b**) Percent reflectance of light from the surface of glass is measured between 400 and 800 nm wavelengths. Error bars, s.d. (n=3).

2.3 Discussion

Immobilisation of proteins and other biological molecules to surfaces is widely utilised in biochemical methods. The simplification of such a technique has clear benefits, and one way to simplify the method is by implementing self-assembling systems. In this chapter, I demonstrate that BslA can self-assemble into monolayers and be subsequently functionalised with the protein or DNA of interest. I test varied applications for this technology, including a protein-peptide screening platform and anti-glint nanopillar array.

Creation of BslA-Tag surfaces and subsequent functionalisation with both proteins and DNA via covalent bonds has been successful. BslA-Tag surfaces can capture the protein of interest from a pure or mixed solution, and BslA-WT surfaces do not promote non-specific adherence of proteins to the surface.

In contrast with different surface immobilisation techniques, BslA provides several advantages.

One key requirement of surface immobilisation of proteins is that they are specifically attached, and therefore correctly orientated on the surface and functional (Meldal & Schoffelen, 2016). While this is made possible with the use of antibodies, they are very expensive and can therefore make protocols inaccessible. An entirely genetically encodable system, such as of BslA, enables all proteins to be made recombinantly in *E. coli* cells and therefore improves the accessibility of the protocol. Further, the assembly of the monolayer is simple and fast, meaning no prior experience is required.

I verify that the attachment of DNA and proteins to the surface is covalent, which is a highly desirable trait for surface immobilisation to improve robustness of attachment (Meldal & Schoffelen, 2016).

While the work presented in this chapter hypothesises that the surface coating should be homogenous due to the nature of the protein monolayer itself, this is an area of the work that could be explored more. Specifically, the next experiments I think would be beneficial are the establishment of saturation of binding to the surface, and percent coverage. It would be interesting to see how diluting tagged BslA proteins with WT-BslA proteins in the monolayer could enable even spacing of the proteins

functionalising the surface, another desirable characteristic of surface immobilisation of proteins (Bednar et al., 2019).

These qualities explored in this chapter confirm that BslA is an excellent candidate protein for self-assembling monolayers and specific capture of proteins from solution.

Interestingly, I find that attachment of mCherry-SnC to BslA-SnT surfaces is consistently more reliable and less likely to cause non-specific interactions with the BslA-WT monolayer, in comparison with GFP-SpC attaching to BslA-SpT. With the evidence at hand, I cannot deduce whether this is due to the inherent properties of the mCherry protein (e.g. charge, protein sequence, fluorescent signal) compared to GFP protein, or in fact is due to differences in binding between the SnoopTag/SnoopCatcher and SpyTag/SpyCatcher pairs. The reasons for this difference in behaviour could be elucidated by comparing binding of GFP-SpC to BslA-SpT with GFP-SnC to BslA-SnT surfaces.

I further characterise BslA monolayers via FRAP experiments, providing evidence that the BslA proteins do not move in the monolayer once formed. This information is relevant for future applications of BslA technology. For example, one might want to attach molecules to the surface in specific positions on the monolayer, and having confidence they will not move from those defined positions is crucial. A specific example of this could be the display of DNA and proteins in separate chambers of a microfluidic device, or to pattern different proteins on the surface (Humenik et al., 2021).

Using the mVirD2-SpC protein, I show that DNA can be covalently attached to the BslA surface in a specific manner. While dsDNA and ssDNA has been shown to covalently bind to mVirD2-SpC in solution, only ssDNA has thus far been shown to bind to the BslA-SpT surface via mVirD2-SpC, but by extension of the results shown in Figure 2.15 it seems possible that dsDNA will also be able to. This is also corroborated by recent research utilising mVirD in vivo (Ali et al., 2020).The potential applications for DNA covalently bound to the surface are exciting, not only as a tool for biosensing capacities, but also for presentation of DNA for transcription and subsequent translation of protein of interest. This has been demonstrated by others previously (Manzano-Román & Fuentes, 2019; Swank et al., 2019), but using the NeutrAvidin/biotin system, which involves a more complex assembly protocol (rounds of washing and specific functionalisation of DNA with biotin) in comparison with the

one step self-assembling BslA monolayer. Utilising the BslA surface for this application, in tandem with protein capture to the surface for cell-free applications, is explored further in Chapter 3.

The first application for BslA monolayers I test is of protein display and subsequent screening for interacting peptides. In this proof-of-concept study I show preliminary data for the Survivin protein displayed on the monolayer and a known binder, the AKER peptide binding specifically. The data presented only shows evidence for the method with one protein and a known binder, and has not been extended to screen for unknown binders. This work could be continued to take the idea beyond proof of concept, and apply this system for screening of novel interactors or inhibitors of protein binding. A similar approach to screen for Survivin interactors has been demonstrated using a bead-based screening approach recently (Ambrose et al., 2021). This is a useful resource to see if the same results can be replicated using the surface immobilised technique described here, and to compare sensitivities.

The second application I test is of protein based nanopillar arrays for anti-glint purposes. Production and purification of two bio-nanopillars, SasG7 and mini-M13 phage, capable of covalent attachment to BslA-SpT in solution, provided encouraging data. The range in lengths of the potential bio-nanopillars is also a useful asset in the development of this application, as previously (Siddique et al., 2015), it has been reported that random arrays of height/width nanopillars actually more closely mimics the natural moth eye structure, and thus improves the surfaces anti-reflective ability. Such randomness would be more difficult to achieve with a nanolithography approach (Raut et al., 2011).

Alongside previous characterisation of BslA monolayers shown in this chapter, all building blocks seem to function well independently. However, when assembled on the glass surface, the desired effect is not achieved, as reflectance and transmission does not change. One problem with the experimental plan for this application is the inability to test for the presence of nanopillars on the surface. I rely on the functionality of the protein building blocks to function as they did in solution when on the surface, and assemble in the expected manner. This is one potential explanation for the unexpected results: it could be that the nanopillars simply have not assembled onto the surface. Another explanation is that the nanopillars may be present as expected, yet not oriented on the surface as desired. In order for the surface to confer anti-

reflective properties, it is thought the nanopillars present should be standing perpendicular to the surface, to create a 'forest' of pillars. Random arrangements of nanopillars lying flat on the surface have not been characterised as conferring anti-reflective properties. Therefore, it is possible the nanopillars could not be arranged in the optimal conditions to promote this. This could be investigated further in the project by atomic force microscopy (AFM) or a different physical surface characterisation technique to determine the orientation of nanopillars. Both AFM and SEM have been successfully used previously to characterise synthesised moth eye surfaces (Z. W. Han et al., 2016; Raut et al., 2011)

To overcome this problem, there are potential routes to explore, for example one recent paper describes a technique for orderly packing of M13 phage into arrays of nanopillars on gold surfaces by polarisation of P8 coat protein and using novel methods for assembly (J. H. Lee et al., 2019). Unfortunately, this work was not in the scope of this project, however it provides a promising trouble-shooting route that could be explored in the future.

Interestingly, despite the wave of innovation in materials sciences the moth eye nanostructures has inspired in the last 40 years, I found it difficult to locate research into the underlying molecular and biochemical basis of these protrusions of the corneal surface of moth eyes. Most research into this topic is rather interested in what the pillars look like, and how to mimic them from an engineering perspective, but not how the pillars are actually formed. However, one recent paper does thoroughly describe the biochemical basis (Kryuchkov et al., 2020). In this paper, they identify the proteins and lipids responsible for creating the Turing patterns that govern nanopillar formation and show that these building blocks can be produced and combined to form materials on glass surfaces with similar properties. This paper provides a fascinating insight into the biochemical basis of the nanopillar formation, and hopefully this will encourage more research to replicate the structures using biochemical tools.

The work I present in this chapter demonstrates the potential for BslA surfaces to display both protein and DNA for novel applications, and further characterises the properties of the monolayer itself.

Chapter 3 BslA surfaces for cell-free applications

3.1 Introduction

3.1.1 Cell-free protein synthesis

One key technology at the forefront of the emergent field of synthetic biology is cell-free biotechnology (Tinafar et al., 2019; Villarreal & Tan, 2017). Cell-free protein synthesis offers practical advantages to in vivo protein synthesis. Without cell walls or membranes, the 'open' nature of the reaction allows for continual monitoring of the environment and easier engineering of the reaction (Laohakunakorn, 2020; Lu, 2017). Additionally, it can enable the production of proteins or chemicals that would ordinarily be toxic to living cells (Khambhati et al., 2019; Villarreal & Tan, 2017). Concerns about biosafety and biocontainment have been raised in the use of whole cells for real-world synthetic biology applications, however many of these risks are minimised in the used of cell-free systems, as self-replication is not possible, thus providing a route for synthetic biology to be more readily employed outside of the lab (Brooks & Alper, 2021; J. W. Lee et al., 2018).

While the principle of performing cellular tasks - such as DNA transcription, RNA translation and metabolic pathway synthesis – outside of the cell is not a new idea, there has been a remarkable surge in new scientific work pushing the limits of cell-free technology, exploring the basic science underpinning it, and already applying this knowledge to real-world applications (He, 2008; Hodgman & Jewett, 2012; Meyer et al., 2021).

One of these recent outstanding applications is demonstrated by Dr Zoe Swank, Dr Nadanai Laohakunakorn, and colleagues (Swank et al., 2019). Miniaturisation of the cell-free reaction to pico-litre amounts allows for thousands of reactions to be run in parallel within one microfluidic device. NeutrAvidin is used as a coating on a device surface, which is then functionalised in individual chambers with mutagenized DNA sequences modified with biotin, resulting in thousands of wells displaying DNA sequences for a library of mutations. The cell-free solution flows into these individual wells and begins the process of in vitro transcription and translation, producing a library of mutated proteins in-situ. In this example, the protein of interest being mutated was being examined for binding capabilities to a different surface bound protein. One of the most exciting elements of this work is the capabilities for high-throughput: while cost remains a limiting factor for cell-free work, scaling down the

reactions to the smallest size possible enables many more reactions to be run in the same volume, hence decreasing cost. In order to realise this technology, the attachment of the proteins and DNA to the surface was essential. To have the DNA immobilised meant it could be kept compartmentalised in the device, and templates prevented from mixing between wells. To have the proteins immobilised not only meant they could be kept in discrete areas, but also that they could be purified away from the other cell-free reaction components. This allows for downstream assessment of protein functionality, for example by washing of the surface or the addition of a secondary binder or substrate. I detail the technology presented in this paper as just one example of synthetic biology applications combining cell-free protein synthesis with surface immobilisation of DNA and proteins.

3.1.2 Surface immobilisation of DNA and proteins in CFPS

Surface immobilisation of proteins expressed in cell-free reactions has been demonstrated as early as 2001, where the method is described as Protein In Situ Array, or PISA (He & Taussig, 2001), and immobilisation of expressed protein of interest is achieved by affinity attachment to a nickel coated surface via the His-tag.

Following this, in 2004, cell-free expression from immobilised DNA and subsequent capture to the surface is reported with the NAPPA method (Nucleic Acid Programmable Protein Array) (Ramachandran et al., 2004). In NAPPA, biotinylated DNA is immobilised on a surface coated with both avidin and polyclonal anti-GST antibodies. The latter is used to capture GST-tagged proteins to the surface expressed in the cell-free reaction. Over a decade later, many more advanced protocols based on the PISA technique have been developed and implemented, improving signal, yield and resolution (Manzano-Román & Fuentes, 2019). Interestingly, despite the optimisation of all other components of the method, e.g. cell-free constituents and high-throughput, evolution of the surface attachment protocol has been minimal.

One paper in 2016 describes the adaptation of NAPPA systems for use with expression of HaloTagged-proteins for increased density of immobilised protein binding to HaloTag ligand chemically adhered to the surface, demonstrating it gives higher yields of active protein compared to attachment via anti-GST antibodies used originally in NAPPA (Yazaki et al., 2016). In a recent review of NAPPA technology, they evaluate methods for surface attachment of proteins, focusing on the tags used

for directed attachment. They describe the desired characteristics of surfaces for protein capture from a cell-free reaction, including: ease of assembly, robustness, cost and time efficiency, and capability of retaining correct protein folding (Kilb et al., 2014).

Given the previous characterisation of BslA surfaces I had completed (Chapter 2.1), it seemed a natural fit to apply the BslA surface coating technology to the field of cell-free protein synthesis. Utilising the mVirD2 protein to covalently attach DNA to the BslA monolayer, and the SpyTag/Catcher technology to covalently join proteins to the BslA monolayer, I proposed the adaptation of BslA monolayers for display of DNA for expression of POI that can subsequently be captured to the surface. I expect there would be several advantages using BslA monolayers in this application above current technology: primarily, the simplicity of the self-assembling nature of the platform, but also the advantage that template DNA does not need to be chemically synthesised, plus greater flexibility with the type of DNA bound to the surface: single or double stranded, linear or circular. Attachment of circular DNA to a surface is especially useful in the context of cell-free systems because efficiency of transcription and thus protein production is much higher from circular DNA than linear DNA. This is due to the fast degradation of linear DNA within a cellular environment – even ex-vivo.

As I began work to implement this idea, I noticed how difficult it was to untangle the complexity of the cell-free system to troubleshoot expression problems. While there have been huge leaps in the capabilities of cell-free systems within the past decade, there are still many challenges that the field faces, especially relating to the sometimes unreliable nature of cell-free reactions (Dopp et al., 2019). The reactions exist in a delicate equilibrium of ingredients required for protein synthesis. Considering the high cost of reagents required for cell-free protein synthesis, they are often set up in extremely small batch reactions, each of which requires <10 µL volumes pipetted together, often completed by hand. Taking these factors into consideration it is perhaps not too surprising that variability between cell-free reactions can be extremely high, not only between experiments, but also between the researcher completing the work, the lab the work is completed in, and the exact reagents used (Cole et al., 2019).

3.1.3 Monitoring metabolism of cell-free reactions

Therefore, in the process of realisation of this application, I also embarked upon a cell-free optimisation odyssey, which is detailed further in the first section of this results chapter. This process led to production of functional cell-free reactions, but also provided a valuable first hand insight into the difficulties of cell-free work. This became relevant in the search for an application for the cell-free BslA platform.

Most commonly, the measured 'success' of cell-free reaction optimisation is an increase in total product formation at the end point. And while that may indeed be the end goal of the optimisation process – achieving higher yields of the product – it does not reveal exactly 'how' or 'why' the applied optimisation has or hasn't worked. Greater real-time analysis of the metabolic components could help to open the 'black box' of the reaction, and see beyond final product as the measure of success. To understand why the optimisation techniques are or aren't working provides tools to further improve those techniques themselves and provides a greater chance of success in troubleshooting, alongside increasing the total product formed.

Endpoint metabolomic assessment of cell-free reactions has previously been used to successfully optimise product yields in cell-free (Miguez et al., 2019). Additionally, real-time analysis of cell-free reaction metabolomics has been achieved by real-time-NMR (Panthu et al., 2017) to optimise mammalian cell-free systems. The information learned from these analytical techniques is fed back into decisions on how to optimise the CF reaction from time point zero in the next round of optimisation, thus closing the loop of the classic synthetic biology Design-Build-Test-Learn cycle. Although the analytical techniques deployed here are extremely accurate, they involve the use of specialist equipment which is generally inaccessible and expensive to use. Thus, the development of a more accessible technique to monitor cell-free analytes in situ would prove useful in the DBTL cycle for optimising cell-free reactions.

The idea I propose is to attach protein-based biosensors to the BslA surface capable of detecting crucial analytes involved in cell-free metabolism. Once the protein biosensors are bound via SpyTag/Catcher, the surface can be washed, removing any unbound proteins, and the cell-free reaction can begin, with the biosensor functionalised monolayer coating the base of the well, reporting the metabolic state of the reaction in real-time.

3.1.4 Biosensors for cell-free metabolite detection

To realise this idea, I first completed a literature review to collect suitable biosensor candidates. In the selection process, I sought well characterised protein-based biosensors that were capable of sensing a small molecule that fluctuates in cell-free metabolism. I required the dynamic range to increase by at least two-fold, to ensure the output would be detectable within the system, and for the operating range of the sensor to be at a concentration relevant to CF reactions.

Three potential biosensors were identified from the literature: for the sensing of inorganic phosphate (P_i), NAD+, and pH. Interestingly, each sensor uses a different design principle for protein-based biosensors. The P_i sensor is created by addition of two fluorescent proteins (CFP and YFP) on either side of a phosphate binding domain found in nature (PiBP) (Gu et al., 2006). When P_i binds, the sensor undergoes a conformational change, moving the two FP further apart, thus reducing FRET signal of the sensor. The original paper reports a library of five FLIPPi sensors, each with different P_i binding affinities. Three were selected for testing in this study, with P_i operating ranges relevant to cell-free reactions (FLIPPi-30m, FLIPPi-200u, and FLIPPi-5u). The selected NAD+ sensor is created by the insertion of an NAD+ binding domain into cpVenus, causing loss of Venus fluorescence when NAD+ binds (Cambronne et al., 2016). The sensor was shown to be very sensitive to NAD+, yet not to other to precursor molecules of NAD, such as NADPH or ATP. Finally, pHRed has been characterised as a pH sensor with extremely high dynamic range (~10-fold change), and is tested for use in this study (Tantama et al., 2011). pHRed is a monomeric protein, capable of sensing pH changes by an intrinsic changes in peak excitation wavelengths of the protein. With lower pH, excitation is optimal at 440 nm, whereas with higher pH, excitation is optimal at 585 nm.

The final section of this chapter explores the production and characterisation of these biosensors (FLIPPi, NAD+, pHRed) with the end goal of real-time monitoring of analytes in cell-free reactions via protein-based biosensors bound to BslA monolayers via SpyTag/SpyCatcher.

3.2 Results

3.2.1 Optimisation of homemade cell-free systems

Variation in cell-free experiments is a well-known problem in the community (Cole et al., 2019; Dopp et al., 2019). In my first experiments using cell-free systems I encountered many problems with reproducibility and functionality: I found the expression level of my protein of interest was far lower than expected, sometimes not detectable, and the functionality of the system was wildly variable between experiments. Therefore, I began optimising each of the three core components in the homemade cell-free system I used, looking for the source of the variability and to increase total protein production in the system. Figure 3.1 provides an overview of the components required and how the functionality of the system and components are measured.

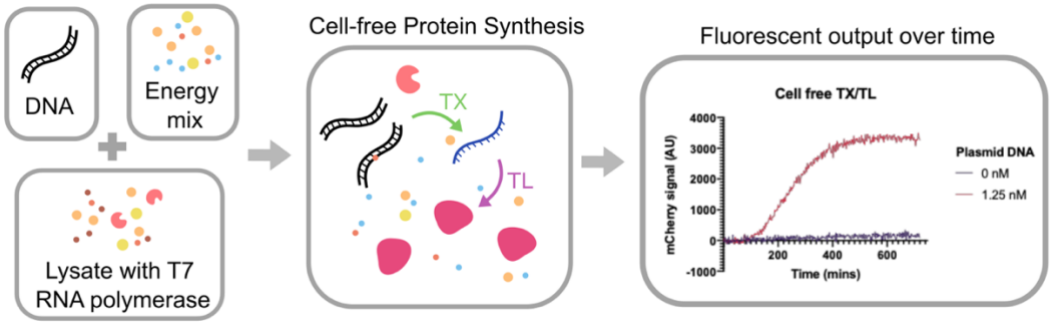


Figure 3.1 Components required in homemade cell-free system. DNA, energy mix, and lysate with T7 RNA polymerase are the three key components required for a functional cell-free reaction in the application explored herein. Together, they contain the required constituents to synthesise a protein of interest in a Cell-Free Protein Synthesis (CFPS) reaction. In the subsequent experiments shown in this subchapter, I use expression of mCherry protein as a marker of 'functional' cell-free systems. Observing the increase of red fluorescence in the reaction over time indicates that the system has been successful in producing the POI from DNA added to the reaction. In this sense, I equate red fluorescence at the reaction endpoint (12 hours) to functionality of the cell-free system.

I began with the optimisation of the energy mix, as anecdotally this component has been very problematic for other users in our lab. This fact is not surprising, as the recipe to make the energy mix is comprised of 27 components (Figure 3.2), many of which are both extremely expensive and only required in very small quantities. Both aforementioned qualities require the chemicals to be weighed out in extremely small quantities, introducing a heightened risk of human error. Discussion with other cell-

free users helped me to highlight the most likely problematic ingredients of the recipe; tRNA and amino acids. I also suspected that as an inexperienced user of cell-free, there may be error or variance introduced by myself when preparing the solutions, therefore to control for 'operator', I compared functionality of the energy mix when made by myself or an experienced user, in this case Nadanai Laohakunakorn (N.L.). Remaking stocks of the most troublesome components with extreme care, and comparing these to stocks made by N.L. allowed me to examine the variation in accuracy of stock solution preparation, the difference between an experienced user preparing stocks to an inexperienced user, and ultimately the difference between an energy solution made entirely by myself, compared to one made entirely by a more experienced operator, N.L..

a)

Component	Final concentration (mM)
HEPES	200
Water	
ATP	6
GTP	6
CTP	3.6
UTP	3.6
tRNA	0.8
CoA	1.04
NAD	1.32
cAMP	3
Folinic acid	0.27
Spermidine	4
3-PGA	120
Amino acids	6
Tyrosine	3
PEG-8000	8%
Mg-glutamate	42
K-glutamate	400
DTT	1

b)

Number	Composition
1	N.L. made by ET
2	N.L. + E.T. AA (old)
3	N.L. + E.T. AA (new)
4	N.L. + E.T. tRNA (old)
5	N.L. + E.T. tRNA (new)
6	E.T. made by E.T.
7	N.L. made by N.L.

c)

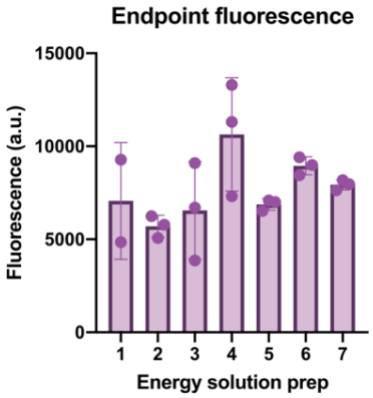


Figure 3.2 Optimising energy solution for cell-free functionality. (**a**) Table of components required in energy solution. Full details are in methods. Highlighted in red are the components advised as being potentially troublesome. (**b**) Guide to seven different energy solutions prepared to examine variability. Here, N.L. = Nadanai Laohakunakorn (experienced user), and E.T. = Ella Thornton (myself, inexperienced user). In rows 1, 6, and 7, the variable being tested

is the experience level of the user who made the stock solutions, against the experience level of the user who put all stocks together into one energy solution. E.g. in row 1, N.L. made by E.T. refers to an energy solution comprised of N.L. stock solutions, but pipetted together into one energy solution by E.T. (**c**) Endpoint fluorescence for mCherry production. CFPS reactions supplemented with 2.5 nM DNA and incubated for 12 hours at 37ºC. Fluorescence measured at 584 nm excitation and 610 nm emission, Gain = 2500. Data is blank subtracted. Error bars, s.d. (n=3). For sample 1, n=2.

As seen in Figure 3.2, changing the stock solutions and operators did not significantly affect the functionality of the system. Comparing the energy solution completely made by N.L. to one made myself, the endpoint signal is not significantly different. This comparison should indicate if there are any significant differences in functionality of the energy mix overall, as all variables have been changed here: all stock solutions and the user composing the final solution. Similarly, no significant difference is seen in the output of all other solutions. From these results I conclude that the protocol followed for energy mix preparation is robust enough to not introduce significant variation to the system functionality.

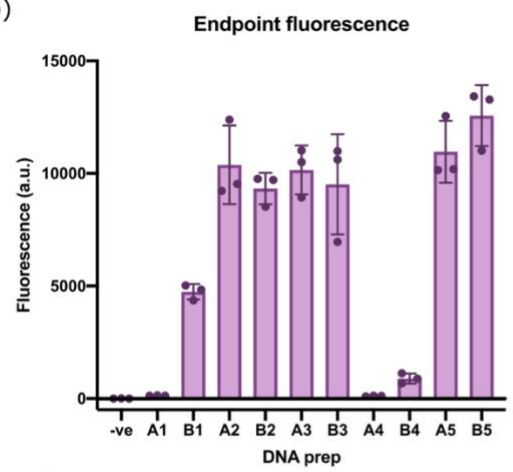
While this result was positive; as my energy solution seemed comparable to one made by a more experienced user, it was ultimately dissatisfying as I had not identified the source of the variation in my cell-free experiments. Therefore I continued my process of optimisation to the next component of the reactions: DNA.

I noticed that there was a large variation in protocols used to prepare plasmid DNA for cell-free reactions in the literature. Some only used a miniprep kit to extract the DNA, whereas others used maxiprep kits or included a secondary purification step. This prompted me to compare my current protocol to others that were frequently cited in the literature (Figure 3.3a). I selected a maxiprep kit and secondary clean-up kit listed in the methods of prominent cell-free papers in the field (Gregorio et al., 2020; Wandera et al., 2020). I compared these kits with my currently used protocol, and other kits we had available in the lab: Monarch and Promega secondary clean-up kits. While following the protocol for the Zymo clean-up kit, I noticed the protocol recommended using water at pH 7.5 to elute. I included this as a variable in my experiments, preparing DNA eluted either at pH 7.5 specifically, or at the naturally more acidic state of water. I was unsure if this would have an effect on yield or purity of the resulting DNA. It had no effect on either.

a)

Code	DNA prep technique
First round (extract from cells)	
A	Miniprep (Qiagen)
B	Maxiprep (Invitrogen)
Second round (cleanup)	
1	No further cleanup
2	Zymo
3	Monarch
4	Promega
5	Zymo (pH 7.5)

b)



c)

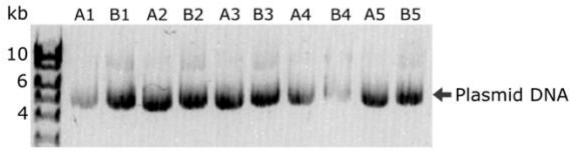


Figure 3.3 Optimisation of DNA clean-up protocol for cell-free reactions. (**a**) Table to show coding system for different DNA prep techniques tested. All DNA was extracted from one large culture of cells, then divided into smaller fractions at each stage of the preparation. DNA was either extracted from the cells using a miniprep (code A) or maxiprep kit (code B), and then either used directly in the cell-free rection, or further purified using one of three kits (code 1, 2, 3, 4 or 5). (**b**) The DNA preps were diluted to the same concentration (2.5 nM) and added to a cell-free reaction in triplicate reactions. This was monitored for 12 hours with incubation at 37ºC, and the endpoint fluorescence is plotted here. Negative control (no DNA added) is shown as the first bar (named -ve). Fluorescence measured at 584 nm excitation and 610 nm emission, Gain = 1500. Data is blank subtracted. Error bars, s.d. (n=3). (**c**) Samples taken of each DNA prep and analysed by gel electrophoresis. Plasmid DNA = 5.9 kb. Each lane coordinates with the DNA sample directly above it in the bar chart. Samples A4 and B4 contain less final sample owing to contamination of ethanol in the final eluate, which caused the sample to float out of the well during loading. This experiment was also repeated on another day, purifying DNA from another culture of cells. The same trends were observed.

The data presented in Figure 3.3 demonstrates the importance of DNA preparation technique used for cell-free purposes. The large variation in functionality is surprising as all reactions have the same concentration of DNA added, and because using standard techniques to check purity of these DNA samples they appeared to have similar levels of purity, e.g. by gel electrophoresis (Fig 3.3c), or by spectrophotometry. Yet there are evidently contaminants present in some of these DNA preparations

which can inhibit the cell-free reaction from proceeding, resulting in lower levels of POI at the reaction endpoint. Here, I deduce 'purity' of DNA prep from final mCherry signal produced in the cell-free reaction.

Most importantly, the protocol I had been previously using: miniprep of DNA from cells using a Qiagen kit, produces DNA of an insufficient purity for cell-free reactions. This could be due to toxins or residual salts from the cells which have not been purified away from the DNA in the process. Using a maxiprep kit significantly increases the purity of the DNA, yet this purity level is variable between different batches. Zymo and Monarch secondary clean-up steps all significantly increase purity of DNA from the initial extraction step to a similar degree.

In contrast, the Promega DNA clean-up kit seems to introduce contaminants into the final DNA prep, as functionality of CFPS from DNA decreases when this kit is used as an extra step. This can be seen most dramatically when comparing samples B1 (Maxiprep, no further clean-up) and B4 (Maxiprep, Zymo clean-up), where fluorescent signal 'decreases' by approximately 5 fold when the additional purification step is included. As shown in Figure 3.3(c), the DNA samples produced with Promega clean-up kits were thought to contain residual ethanol, as the samples were lower in density and floated out of the well, leading to lower intensity bands shown on the gel. Ethanol could therefore have been the inhibiting factor present in the DNA prep. Although there are commonly used tips and tricks to reduce ethanol contamination in final elutions, these were not used in these experiments. All protocols were followed exactly as specified by the manufacturer, for consistency, between kits.

From these experiments, I chose a new standard protocol to follow for DNA preparation for cell-free experiments. I first purify using a Maxiprep kit, and then complete a secondary clean-up step using the Zymo kit (water for elution pH unchanged). This is the best technique for reliable and reproducible purity of DNA for functional cell-free experiments in my experience.

Following this, I systematically examined the third and last component of the cell-free reaction: the lysate. I began by assessing the protocol I used and comparing it to protocols described most commonly in the literature (Cole et al., 2020). This identified three key steps to be varied: the growth of the cells (which *E. coli* cell type to use: BL21 or RosettaGami-2, and whether they are grown to fixed OD or fixed time), the

inclusion of a run off step (37ºC, 220 RPM for 90 minutes), and the inclusion of a dialysis step (Figure 3.4).

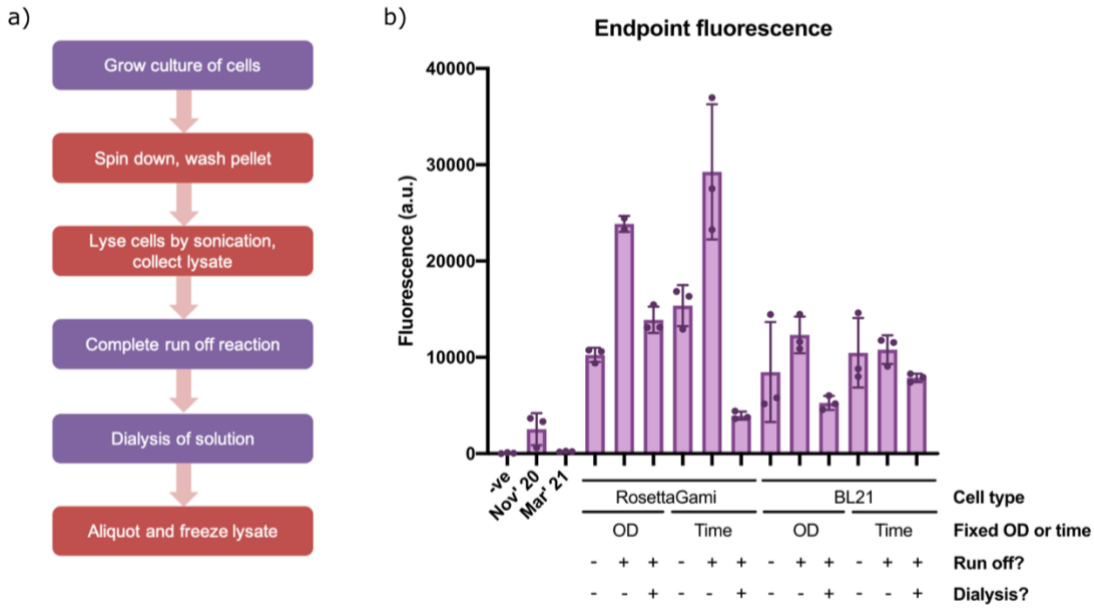


Figure 3.4 Optimisation of lysate production for cell-free reactions. (**a**) Key steps in the method for lysate preparation. In purple are the steps I explore variations of in these experiments. In red are the steps kept constant throughout experiments. (**b**) Endpoint fluorescence for mCherry production with different preps of lysates. CFPS reactions supplemented with 2.5 nM DNA and incubated for 12 hours at 37ºC. Negative control (no DNA added) is shown as the first bar in purple (named -ve). Lysates produced using old protocol are annotated as Nov' 20 and Mar' 21, referring to the month they were made. Conditions used for the preparation of each lysate are shown in the table below the x axis. Fluorescence measured at 584 nm excitation and 610 nm emission, Gain = 2000. Data is blank subtracted. Error bars, s.d. (n=3).

Similarly to experiments for optimisation of DNA purity, it was surprising to see how variable the functionality of lysates for cell-free was under the conditions tested. As seen in Figure 3.4(b), there are some key trends observed from this data. Firstly, there is a consistent trend that *E. coli* RosettaGami-2 cells make a more functional lysate than *E. coli* BL21 cells.

Secondly, growing cells to fixed OD or for a fixed time did not seem to affect functionality. This is beneficial to the user, as growing cells for a fixed time of 3 hours compared to fixed OD = 0.5 yields many more cells and subsequently much larger

quantities of lysate. As the protocol is time-consuming, the fewer times it is required the better.

Thirdly, I observed an interesting trend in my data regarding the additional steps of run-off and dialysis. These two steps tend not to be recommended when making lysate with T7 RNA polymerase, they are instead used most often when preparing a cell-free lysate for use with native *E. coli* promoters. However, in previous papers optimising cell-free lysates they tend to group these steps together, rather than examining the effect of each independently (Silverman et al., 2019). When separated, you can see that in fact dialysis after run-off does decrease the functionality of the lysate, however, including the run-off step alone actually improves the lysate functionality. Best conditions for post-lysis processing is a matter of great contradiction within the cell-free literature, and I will discuss the context of my result in the literature further in the discussion of this chapter.

One suggested cause of the variation in lysate prep functionality was the total protein concentration: it was speculated that the more functional lysates could be a result of a higher total concentration. To test this, I measured the total protein concentration of lysates via a Bradford assay (Figure 3.5). Comparing the protein concentration of the lysates to their functionality (Figure 3.4), there are no trends observed between the two data sets. For example, the most functional lysate tested in Figure 3.5 (A5 and B5: RosettaGami cells grown for fixed time, followed by run off but no dialysis) has a total protein concentration very similar to the worst lysate in terms of functionality (Mar' 21). Similarly, looking at the highest concentrated lysate (B10: BL21 cells grown for fixed time with no additional steps), this corresponds to a cell-free functionality almost three fold lower than my best tested lysate. Overall, I found no significant relationship between lysate functionality and total protein concentration.

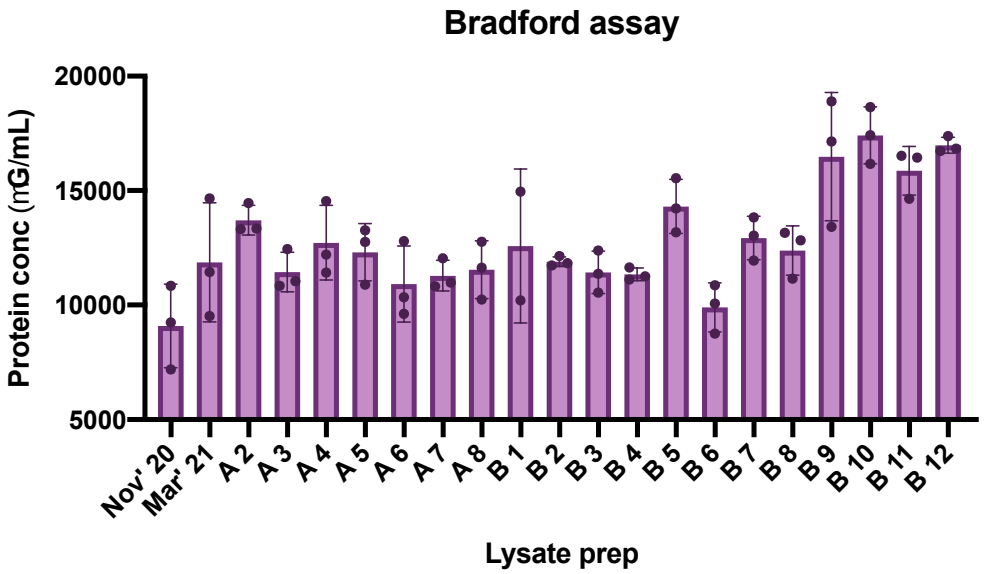


Figure 3.5 Total protein concentration of prepared lysates for cell-free. Protein concentration of lysates were measured via a Bradford assay. Three aliquots of each lysate were taken and combined with Bradford reagents in a 384 well plate, before reading absorbance values in a plate reader. Absorbance values were converted to protein concentration using a standard curve of known protein concentrations. Not all lysate preps contained sufficient remaining sample to be tested. Those with sufficient volume were tested and shown here. The letter (A or B) refers to batches of lysates prepared on different days. The number (1 to 12) refers to the lysate prep out of the total 12 conditions tested, as read from left to right on the x axis of Figure 3.4, e.g. 1 = RosettaGami cells, grown to fixed OD with no further steps, 7 = BL21 cells, gown to fixed OD with no further steps. Error bars, s.d. (n=3).

Together, these results provided me with confidence to determine an optimised protocol for the creation of all relevant components of the cell-free reaction. The decided protocol from all options tested is summarised in Figure 3.6(a) and 3.6(b). Comparison of the functionality of the cell-free reaction following these protocols with the original protocols I followed is shown in Figure 3.6(c) and 3.6(d).

a)

Cell strain	Grown to?	Run off?	Dialysis?
Rosetta-Gami	OD = 0.5	No	No
		Yes	Yes
	Time = 3 hours	No	No
		Yes	Yes
BL21	OD = 0.5	No	No
		Yes	Yes
	Time = 3 hours	No	No
		Yes	Yes

b)

Initial prep	Secondary prep
Miniprep	None
Maxiprep	
Miniprep	Zymo
Maxiprep	
Miniprep	Monarch
Maxiprep	
Miniprep	Promega
Maxiprep	
Miniprep	Zymo pH 7.5
Maxiprep	

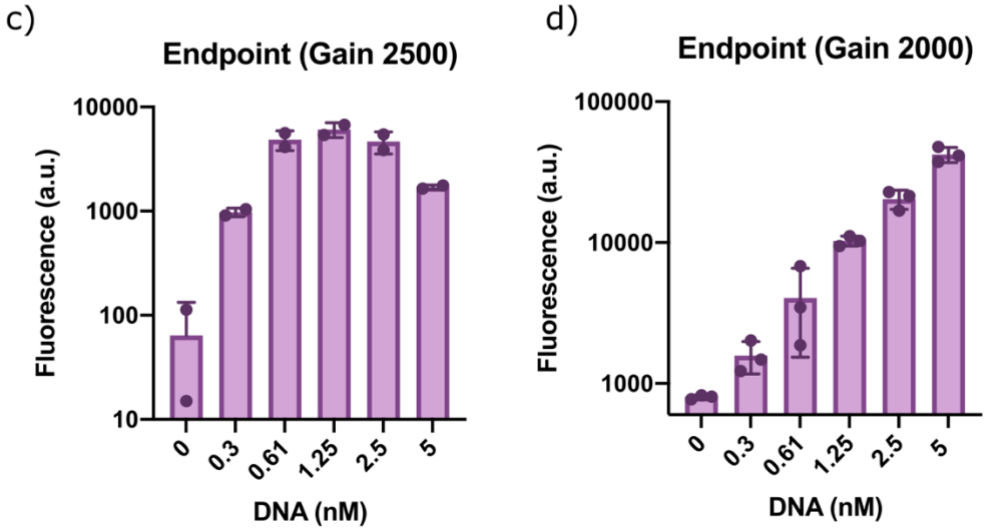


Figure 3.6 Conditions decided for protocols associated with homemade cell-free preparation. (**a**) Table to show conditions tested in cell-free lysate preparation. The chosen conditions are highlighted in red: Rosetta-Gami-2 cells are used for growth, grown for three hours, with run off step included but dialysis omitted. (**b**) Table to show conditions tested in optimisation of DNA clean-up protocol. Chosen conditions are highlighted in red: DNA is extracted from cells via maxiprep, and further purified using Zymo clean-up kit. (**c**) mCherry expression in cell-free from titrations of DNA encoding the protein before optimisation of components. The response to DNA concentration is non-linear: the amount of protein produced decreases at the highest and lowest DNA concentrations. Fluorescence measured at 584 nm excitation and 610 nm emission, Gain = 2500. Error bars, s.d. (n=2). (**d**) Expression of mCherry from DNA titrations encoding the protein after optimisation. The response to DNA concentration is now more linear: at higher concentrations more protein is produced, and at lower concentrations less protein is produced. Fluorescence measured at 584 nm excitation

and 610 nm emission, Gain = 2000 Error bars, s.d. (n=3). For both (c) and (d), CFPS reactions were incubated for 12 hours at 37ºC.

The data shown in Figure 3.6 was taken using different Gain settings on the plate-reader, as fluorescence levels in Figure 3.6(d) were several times higher than in 3.6(c). Therefore numerical values should not be directly compared. Gain can be compared to exposure, with higher Gain values increasing the 'value' of fluorescence recorded. Therefore, if Figure 3.6(c) were also measured at Gain 2000, the recorded values would be much lower, further increasing the difference in measured fluorescence between 3.6(c) and 3.6(d).

Interestingly, in the first dataset (Figure 3.6(c)) a non-linear response to DNA concentration can be seen, where higher concentrations of DNA (5 nM) actually seem to inhibit the cell-free reaction from proceeding. In the context of my DNA prep optimisation experiments (Figure 3.3), this could be understood as the toxins present in the miniprep inhibiting the reaction at higher concentrations. After optimisation, there is a linear response of protein produced with DNA concentration added. This is the expected trend within the DNA concentrations tested, and aligns with data in the literature (Silverman et al., 2019).

Overall, the results presented show a systematic approach for optimisation of preparation of cell-free components. A significant increase in protein production (i.e. functionality) is seen under all optimised conditions in comparison with data collected pre-optimisation.

3.2.2 Specific capture of proteins from CFPS to BslA surfaces

Following optimisation of homemade cell-free protein synthesis protocols, I began implementing the cell-free system together with BslA coated surfaces. The idea was to use BslA surfaces to both display the DNA to be transcribed and capture the protein translated (Figure 3.7). To reach this goal, I split the idea into two key experiments: capture of proteins expressed in CFPS to a surface, and expression from DNA bound to BslA surface.

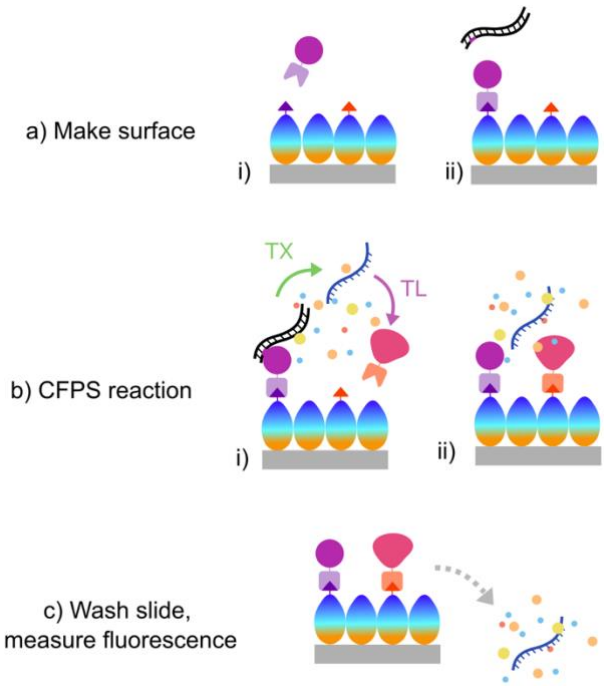


Figure 3.7 Potential application of CFPS with BslA presentation of DNA and capture of translated protein. (**a**)**i**) BslA monolayer is made, with both BslA-SnT and BslA-SpT. mVirD-SpC is covalently attached, and (**a)ii**) DNA encoding POI is subsequently bound. (**b**)**i**) CFPS occurs, creating mCherry-SnC, which is then (**b**)**ii**) specifically captured to the surface. (**c**) Wells are washed, with red fluorescence retained only in wells with both SnT and SpT BslA proteins comprising the surfaces.

I started with the experiment to show protein capture to the BslA surface expressed in cell-free (Figure 3.8). Figure 3.8(a) shows a schematic of the key idea of the experiment: a cell-free reaction synthesises mCherry-SnC atop a BslA-SnT surface. Therefore, synthesised mCherry-SnC is captured to the BslA monolayer via SnT. When this happens, you expect to see retention of red fluorescence on that surface after washing away non-specific binders and residual cell-free components. Whereas the same cell-free protein synthesis reaction taking place atop a BslA-WT surface will not lead to retention of red fluorescence due to the lack of immobilised SnT. (Figure 3.8(aiv)).

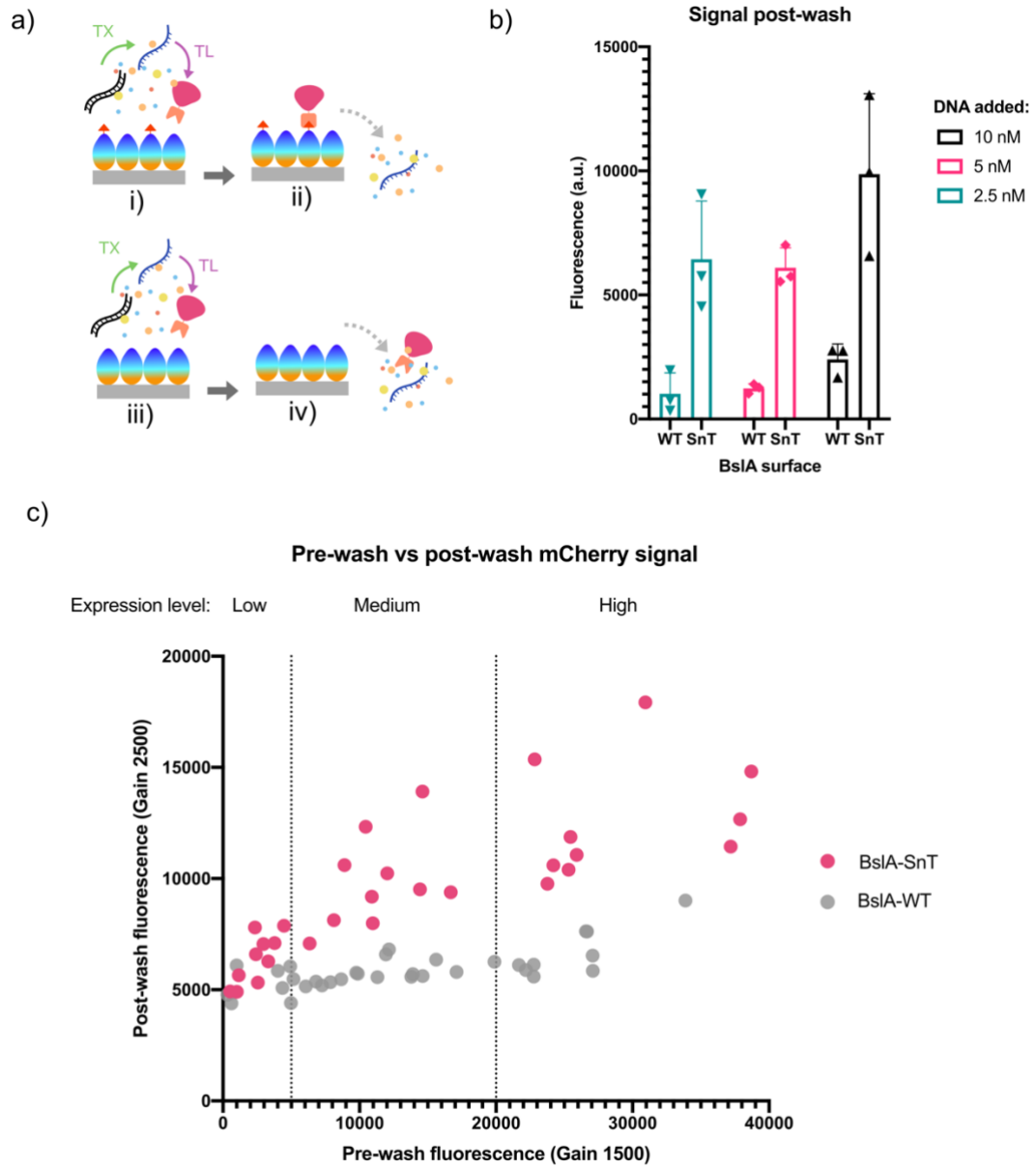


Figure 3.8 BslA surfaces can specifically capture protein produced in a cell-free reaction. (**a**) Schematic to show key idea of experiment: (**i**) BslA-SnT surface is made on surface of glass slide, and mCherry-SnC is produced by CFPS. After washing (**ii**), mCherry-SnC is retained on the surface via covalent attachment to BslA-SnT monolayer. Alternatively, BslA-WT is used to coat a glass slide (**iii**), which after CFPS of mCherry-SnC, does not retain the mCherry protein as there is no SnT to be bound to. Therefore red fluorescence is removed from the well upon washing (**iv**). (**b**) Signal from wells after CFPS reaction (12 hours at 37ºC) and following thorough washing with water. Fluorescence measured at 584 nm excitation and 610 nm emission, Gain = 2500 Error bars, s.d. (n=3). (**c**) Signal in each well plotted before (y axis) and after washing (x axis). Each point represents one datapoint. Samples where

mCherry-SnC has been made in situ above a BslA-SnT surface is shown in pink, and samples where mCherry-SnC has been made above a BslA-WT surface are shown in grey. Dotted lines intersecting x axis indicate arbitrary values dividing low, medium and high expression of mCherry-SnC in CFPS. These lines help to indicate that higher expression of the protein of interest in situ leads to more protein binding to the BslA-SnT surface, as there is higher red fluorescence retained in these wells. Conversely, if not enough protein is synthesised in the wells (samples with low expression), there is not substantial retention of red fluorescence in these wells.

Figure 3.8(b) shows fluorescence of the BslA surfaces after incubation with CFPS for 12 hours, followed with washing with water to remove non-specific interactions. High fluorescence is seen from wells coated with BslA-SnT, but low fluorescence is seen with BslA-WT surfaces. As DNA concentration increases, and more protein is made, the background levels of mCherry-SnC non-specifically interacting with BslA-WT slightly increase, but the difference between BslA-SnT and BslA-WT surfaces is still substantial. Following this experiment, I wanted to check that the signal post-wash was not related to signal in each well pre-wash (i.e. dependent on the concentration of protein produced in situ). Therefore I plotted the signal pre-wash against signal post-wash for each well, to determine if there was a trend in the data to suggest this (ie. a positive correlation and no difference between BslA-WT and BslA-SnT surfaces). As can be seen in Figure 3.8(c), there is a separation between the two sets of data: for BslA-WT and BslA-SnT. For BslA-WT surfaces, fluorescence post-wash does not increase as fluorescence pre-wash does. This indicates that even when there are high concentrations of target protein present, it can be removed from the BslA-WT surface with washing. Looking at the BslA-SnT datapoints, you can see high fluorescence post-wash increase as higher quantities of protein are made in each well. This could imply that the saturation point of proteins binding to the surface has not been reached yet, and that adding more protein could increase signal post-wash.

Next, I tested binding of DNA to the surface for expression of proteins in CF (Figure 3.9). As before, the main test of the system's functionality was a comparison of two surfaces: one containing all active components required for the reaction (Figure 3.9(a)), and the other, a negative control with only a BslA-WT surface which should repel non-specific interactions (Figure 3.9(b)).

In this experiment, the mCherry-SnC protein is only being expressed in solution, there is no subsequent capture to the surface or washing of the slide. What I hope to see is

expression of mCherry (increase in red fluorescence) in the wells only where DNA has been specifically and covalently attached to the surface. An increase in fluorescence in BslA-WT wells would indicate expression is occurring via DNA that has not been washed from the surface successfully. The data shown in Figure 3.9(c) indicates that DNA is non-specifically retained to both BslA-WT and BslA-SnT surfaces as you see mCherry signal in both. This non-specific attachment is only seen as higher concentrations of DNA are added to the surface.

It could be possible to uncover the optimal washing conditions for the surfaces to enable non-specifically bound DNA to be removed from the surfaces, yet this has not been possible in the scope of these experiments. Therefore, it has not been possible to show expression of protein from DNA bound to BslA surfaces.

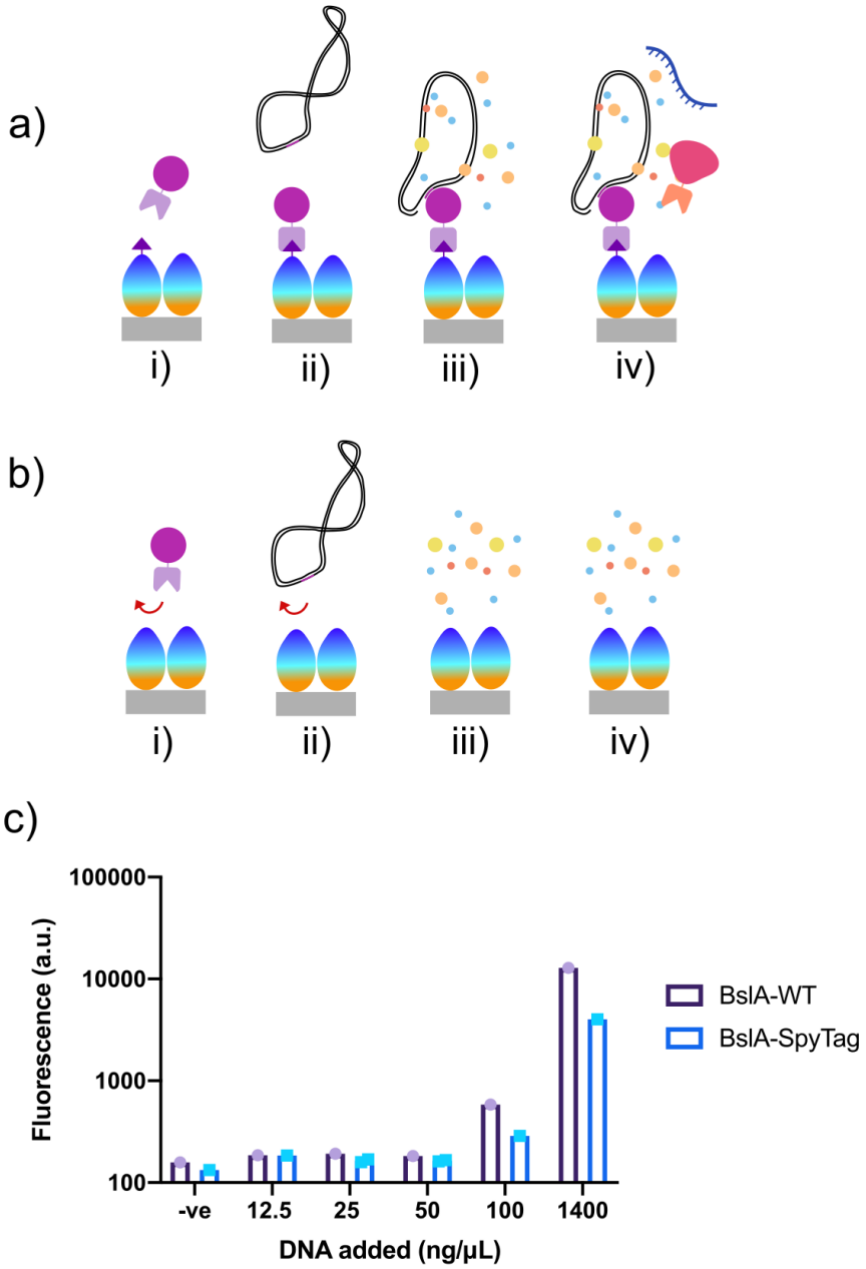


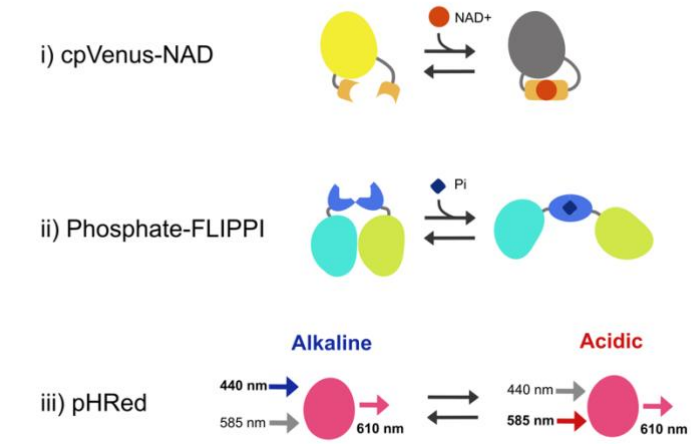
Figure 3.9 Test of mCherry-SnC expression from DNA bound to BslA surfaces. (**a**) Schematic of positive samples, whereby DNA is specifically bound to BslA surfaces via mVirD protein. (**i**) Formation of BslA monolayer and attachment of mVirD protein via SpyTag/SpyCatcher. Non-specifically bound mVirD is washed away, and (**ii**) plasmid is added encoding mCherry-SnC with mVirD recognition sequence for binding, plus 2 mM $MgCl_2$. (**iii**) Plasmid is nicked and covalently attached to mVirD protein on surface, the surface is washed to remove non-specifically associated DNA, and cell-free reagents are added for reaction to begin. (**iv**) CFPS begins, with mCherry-SnC produced in well, resulting in an increase in red fluorescence. (**b**) When these same steps are followed (addition of mVirD-SpC (i), addition of

DNA (**ii**), addition of CF reagents and incubation (**iii**)), but on a BslA-WT surface, we expect there to be no increase in red fluorescence, as there is no expression of mCherry-SnC (**iv**). (**c**) Fluorescence endpoint data after 12 hours of incubation at 37ºC with CF reagents. BslA-SpT surfaces, shown in blue, were treated as described in (a), and BslA-WT surfaces, shown in purple treated as described in (b). The amount of DNA incubated with the surfaces is described on the x-axis. Fluorescence measured at 584 nm excitation and 610 nm emission, Gain 1500. Single data points shown on bar chart, n = 1 or 2 per condition, as shown.

3.2.3 Sensing of CF metabolites via biosensors

I continued to the next application for this work: protein-based biosensors for monitoring cell-free analytes in real-time. Figure 3.10 details the sensors used in these studies, taken from the literature. The protein-based sensors selected were capable of detecting inorganic phosphate (Pi), NAD+, or pH (Cambronne et al., 2016; Gu et al., 2006; Tantama et al., 2011). The paper describing the Pi sensor in fact characterises a whole library of sensors, each with different affinity to Pi, and thus different effective operating ranges (Gu et al., 2006). I select three of the described Pi sensors for testing here, thus in total I test five protein-based biosensors to detect three different analytes.

a) Three candidate sensors:



b) Sensor properties:

Sensing:	Sensor design	kD	CFPS range	Response
NAD+	Decrease in fluorescence from cpVenus	65 μM	< 300 mM	5 – 500 μM, 2-fold change
Phosphate	Decrease in FRET (sensing domain in-between)	260 nM – 30 mM	< 10 mM	Up to 2-fold, depending on sensor
pH	Change in fluorescent ratio	N/A	pH 5.5. - 9	pH 5.5 – pH 9, 10-fold change

Figure 3.10 Diagram to show function of three biosensors of interest. (**a**) Three candidate protein-based biosensors were selected from the literature for use. (**i**) Sensor for NAD+. This sensor was created by insertion of NAD sensing domain into circular permuted fluorescent protein Venus. Upon NAD binding, fluorescence is lost from the protein, therefore the higher the NAD concentration, the lower the fluorescence of the sensor. (**ii**) Sensor for inorganic phosphate. Two proteins, eCFP and eYFP, are linked by a phosphate binding domain. With no phosphate bound, the two fluorescent proteins are in close proximity and FRET can occur from eCFP to eYFP. Once phosphate binds, the conformational change of the protein structure distances the two proteins, and the FRET signal decreases. Therefore, the higher the phosphate concentration, the lower the FRET signal. (**iii**) Sensor for pH. This is a monomeric fluorescent protein that emits at 610 nm, but has two excitation peaks, at 440 nm and 565 nm. The ratio between excitation at these two peaks corresponds to pH of the solution. (**b**) Table to show properties of sensors in further detail.

To enable attachment to the surface, I genetically encoded SpyCatcher or SnoopCatcher to each sensor. Following this, I purified fusion proteins via affinity chromatography. Expression and purification of cpVenus_NAD_SpC and FLIPPi_SpC sensors was successful, but pHRed_SnC expression failed. Ultimately, the maturation time was far too long. However the lack of red fluorescence in the cell culture was initially thought to be an expression problem, so further conditions for expression were researched and tested (Figure 3.11). I compared the expression of the original pHRed protein with my fusion protein pHRed_SnC under different common protein expression protocols. Expression in TOP10 from a T7 promoter is very uncommon, yet it was listed as the protocol of choice for expression of pHRed in the original paper (Tantama et al., 2011). Therefore it was also tested.

Expression of both pHRed and pHRed_SnC was successful in *E. coli* BL21 and Rosetta-Gami-2 cells under all tested conditions, seen by the presence of a protein band at the expected molecular weight (Figure 3.11). No protein expression was seen under expression in TOP10. However, as can be seen from the photographs of cell pellets, there was still no fluorescence detected, even when relatively high concentrations of target protein were expressed (Figure 3.11). Through correspondence with the authors of the paper, I later learned that maturation of this FP could take several days. This is impractical for the desired application, as ideally I would like to be able to synthesise the sensors for use in situ. Therefore work with the pHRed sensor was discontinued, despite its high potential impact.

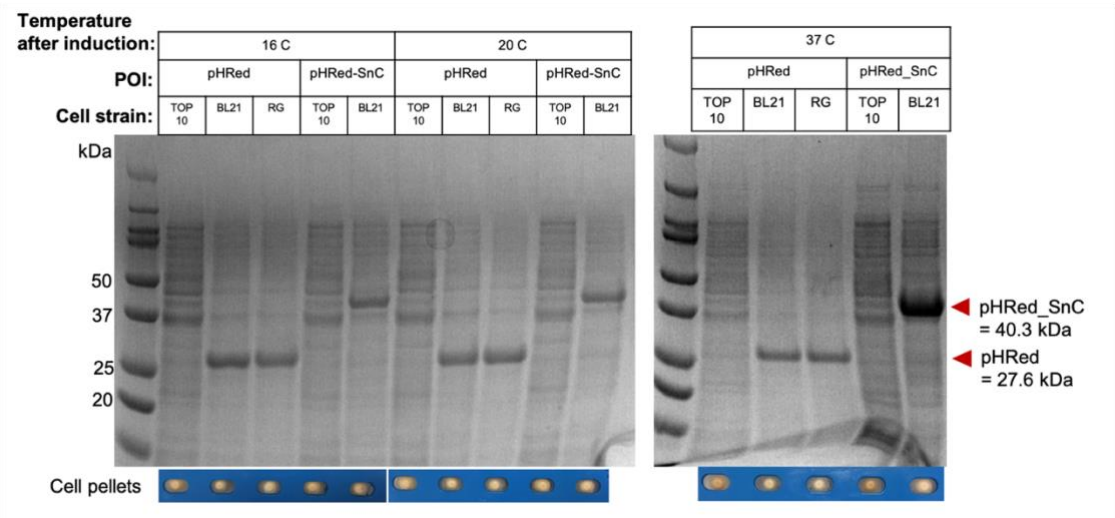


Figure 3.11 SDS-PAGE to show expression test of pHRed and pHRed-SnC. A comparison of three *E. coli* cell strains were used for protein expression: TOP10 (according to original

paper methods), BL21 and RosettaGami-2. Cultures were grown at 37ºC, and induced at OD = 0.6 - 0.8, before addition of 1 mM IPTG and reduction of temperature to 16, 20, or 37ºC. Photographs of cell pellets from each cell culture shown on the SDS-PAGE gel are displayed under the corresponding lane (blue colouring in photograph is from tube rack holding tubes). There is no visible red colour to the pellets, which would usually be expected with successful expression of a red fluorescent protein in the cells.

The purified fusion protein sensors for NAD+ and Pi were tested in solution for response to the specified analyte. Figure 3.12 shows characterisation of FLIPPIi_SpC sensors in solution. The sensors are able to respond to the analyte in the range characterised in the original paper (Gu et al., 2006), even with the added SpyCatcher domain.

FLIPPi-30m_SpC was consistently able to respond to inorganic phosphate in the desired range, between different purifications of the sensor. However, for FLIPPi-200u_SpC and FLIPPi-5u_SpC fusion proteins, the response was very variable between protein preparations. After initial purification of the sensors using a buffer containing phosphate by mistake, I found that sensor functionality for FLIPPi-200u_SpC and FLIPPi-5u_SpC could not be restored even after several rounds of dialysis of the final eluate to remove phosphate. When re-purified using only Tris-buffers, all sensors were capable of sensing phosphate in desired range (Figure 3.12).

This finding would imply that binding of phosphate to these sensors is irreversible, as once saturated by the phosphate in the purification buffer, the sensor was unable to release the molecules from the binding domain and restore sensing function. However this would be surprising, not only as in the original paper the sensors are capable of binding phosphate reversibly, but also because the sensors are expressed inside cells, where phosphate is in abundance, and therefore the buffers used for purification should not make a difference.

It is also noted that when responses of all sensors are plotted on the same graph (Figure 3.12(b)), the dramatic difference in dynamic range is visible: FLIPPi-30m changes in fluorescent ratio from 0.8 to 1.3 (change of 0.5), whereas FLIPPi-200u and FLIPPi-5u change only by a ratio of 0.1 and 0.2 respectively. Therefore, the FLIPPi-200u and FLIPPi-5u sensors were not continued further in these experiments due to unreliable nature. Serendipitously, the most functional sensor, FLIPPi-30m, is

in fact the one with the sensing range most useful to us (1 – 10 mM). Therefore, this biosensor was continued in the experimental workflow further.

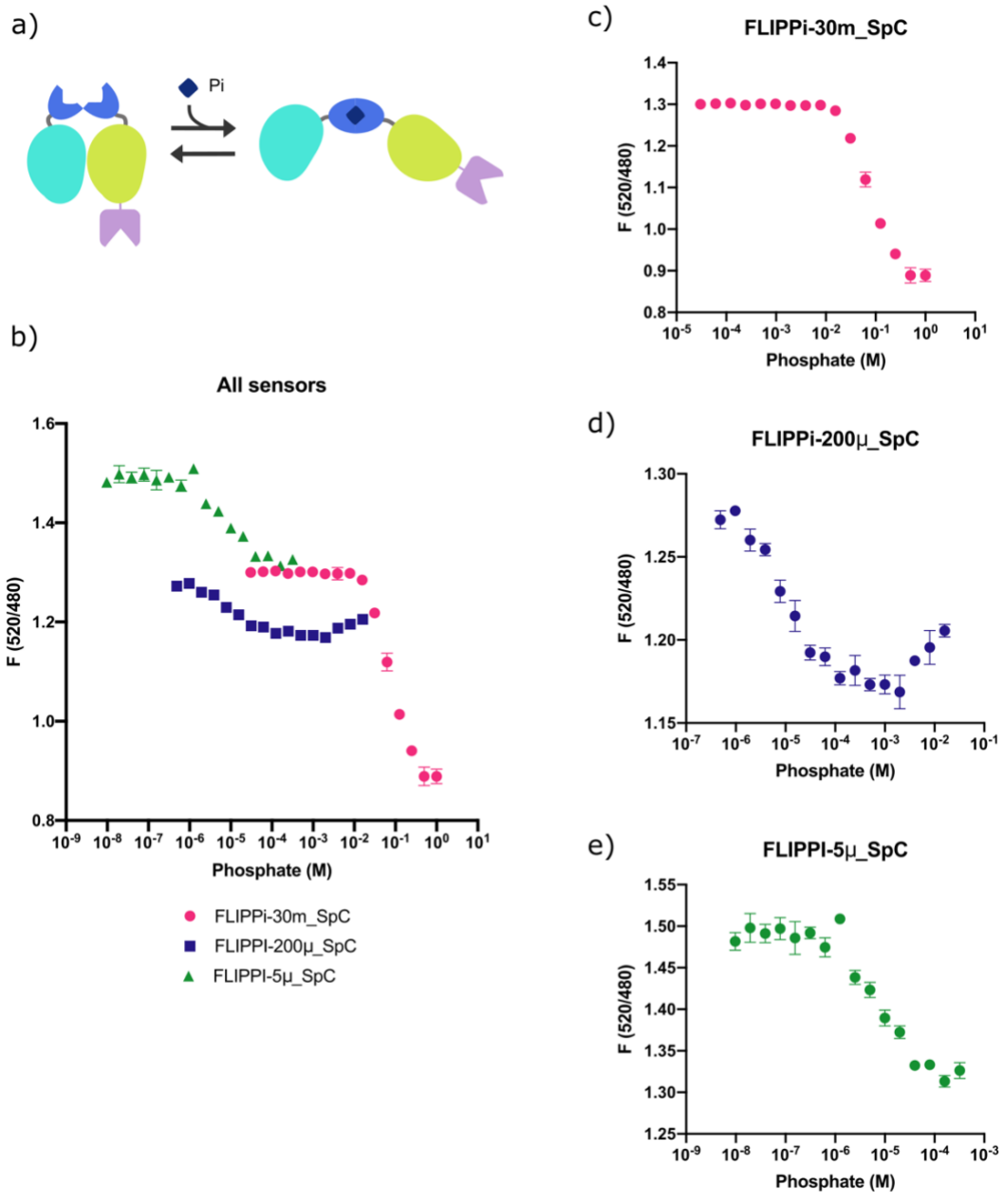


Figure 3.12 FLIPPi-SpC fusion proteins can sense Pi in solution. (**a**) Cartoon schematic of FLIPPi_SpC proteins. SpyCatcher domain was added at C-terminus of protein biosensor. (**b**) Sensor response curves for all three sensors combined on one plot. Purified protein sensors (final concentration 1 µM) were mixed with phosphate solutions. Data was blanked against H_2O, and fluorescence signal at (ex/em) (430/520) was divided by fluorescence signal at (430/480), blank = 1000. Error bars, s.d. (n=3). All sensors show response to phosphate within defined range, as FRET signal decreases as phosphate concentration increases. Data

shown in (b) is plotted on separate graphs for each sensor: FLIPPi-30m_SpC on (**c**), FLIPPi-200u_SpC on (**d**), and FLIPPi-5u_SpC on (**e**).

Characterisation of NAD+ biosensor with SpyCatcher domain demonstrated the fusion protein is capable of sensing NAD+ to a similar degree as shown in original paper (Cambronne et al., 2016) (Figure 3.13), concluding the SpyCatcher domain does not interfere with sensing of NAD+.

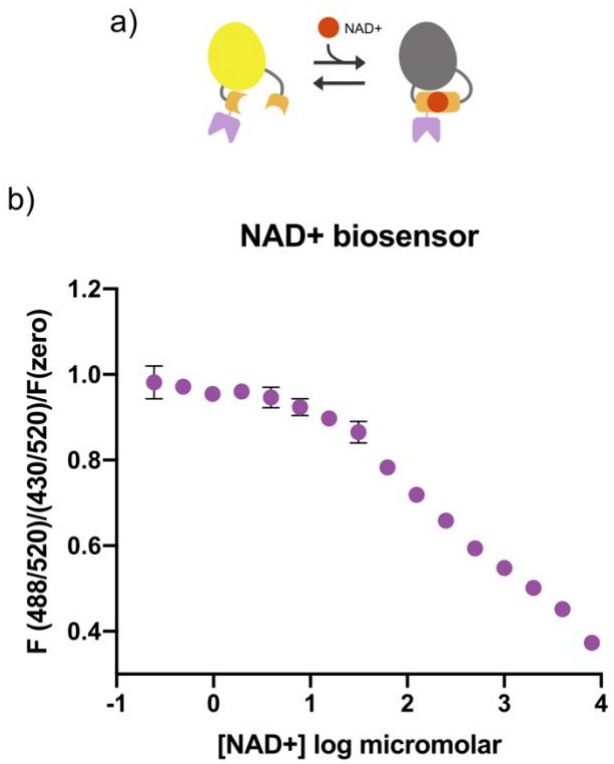


Figure 3.13 NAD-SpC fusion protein can sense NAD+ in solution. (**a**) Cartoon schematic of NAD_SpC fusion protein. The SpyCatcher domain was added at the N-terminus of the protein. Increasing concentrations of NAD+ binding to the sensor reduces fluorescence. (**b**) Purified NAD_SpC protein (final concentration 1 µM) is mixed with serial dilutions of NAD+ and imaged for fluorescence. Data here is analysed as in original paper (Cambronne et al., 2016): Fluorescence measured at (ex/em) 488/520 is divided by fluorescence at 430/520, this value is then divided by the signal when no NAD+ is present. Gain = 2000. Data was blanked against H_2O. Error bars, s.d. (n=3).

FLIPPi-30m_SpC and NAD_SpC were characterised as the most promising sensors to be trialled in the next stages of experiments. As the ultimate goal of this work is to sense analyte levels from surface-bound proteins, purified sensors were tested for

binding to BslA surfaces via SpyTag/SpyCatcher (Figure 3.14). A range of protein concentrations were tested for application to the surface, as previously I found the optimum binding concentration to differ from protein to protein (Chapter 2.1.3). FLIPPi30m_SpC was capable of specific attachment to BslA surfaces via the SpyTag, as seen in Figure 3.14 by the significant high fluorescent signal in wells with BslA-SpT monolayers opposed to low fluorescent signal on BslA-WT monolayers. 15 μM of protein added to BslA surfaces is found to be the optimum concentration tested here. Higher concentrations of FLIPPi-30m_SpC reacted with the BslA monolayers should be tested next.

The NAD_SpC protein was not capable of specific attachment to the BslA monolayer (Figure 3.14(d)). Data collected shows unexpected trends: fluorescent signal is close to blank values in most wells, with the only the lowest tested concentration (2.5 μM) of NAD_SpC protein for attachment showing high fluorescent signal on BslA-SpT surfaces. This requires further characterisation, as results do not follow a consistent trend and error bars are relatively large on the data sets.

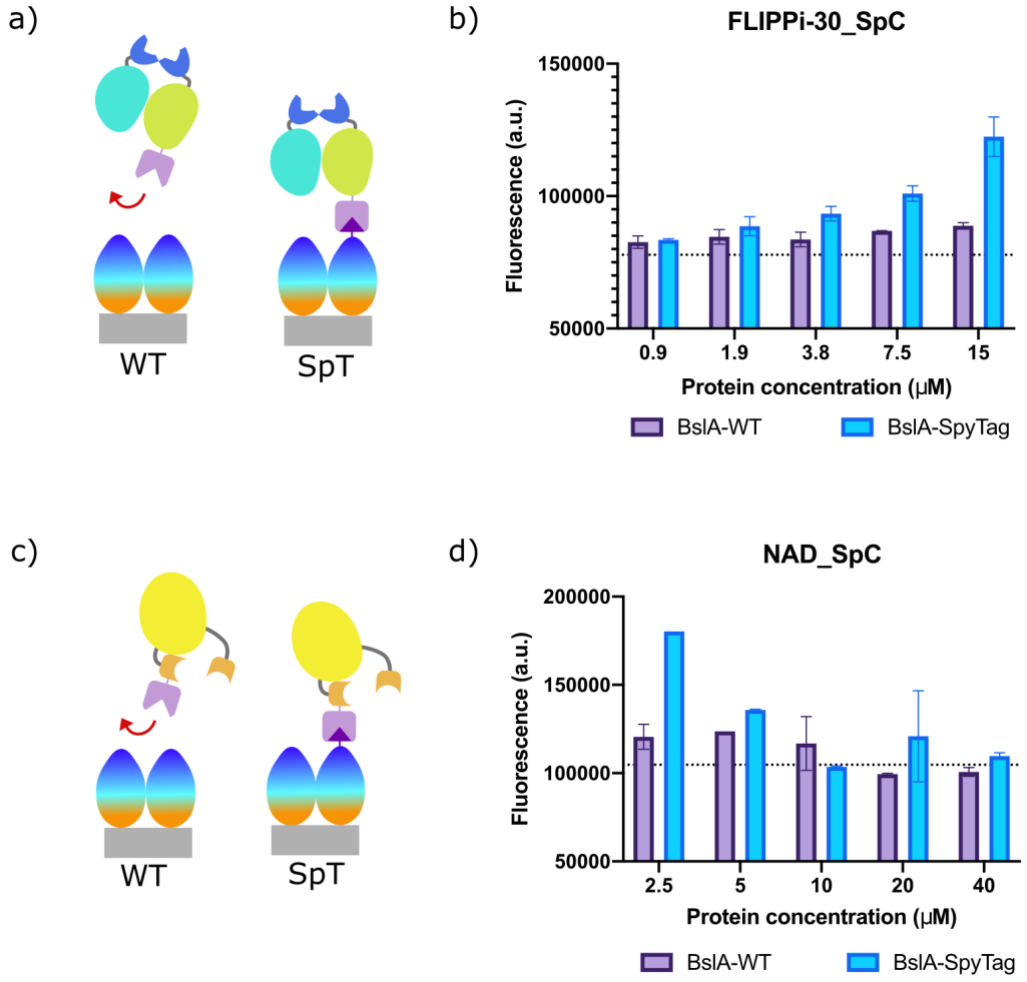


Figure 3.14 Binding of purified FLIPPi-30m-SpC and NAD-SpC to BslA surfaces. (**a**) Cartoon to illustrate FLIPPi-30m_SpC binding specifically to surfaces composed of BslA-SpT but not to BslA-WT. (**b**) Fluorescent signal from wells coated with BslA-WT or BslA-SpT monolayers, incubated for one hour with varying concentrations of FLIPPi-30m_SpC. Wells were thoroughly washed, and 10 µL H_2O added to every well before imaging in plate reader. (**c**) Cartoon to illustrate NAD_SpC binding specifically to surfaces composed of BslA-SpT but not to BslA-WT. (**d**) Fluorescent signal from wells coated with BslA-WT or BslA-SpT monolayers, incubated for one hour with varying concentrations of NAD_SpC. Wells were treated as in (b). Average blank values are indicated by the dashed line cutting the y axis, (blank = BslA monolayer in well incubated with buffer). Fluorescence of both biosensors was measured using the 'control' wavelengths of light that do not vary with analyte concentration. Error bars, s.d. (n=2).

Given the unexpected preliminary result with NAD_SpC, I proceeded with using the FLIPPi-30m_SpC biosensor for experiments testing real-time sensing of Pi in cell-free

protein synthesis reactions (Figure 3.15). Purified FLIPPi-30m_SpC was combined with DNA encoding mCherry_SnC in a cell-free reaction and fluorescence output corresponding to mCherry and FLIPPi-30m Pi sensing was measured over 20 hours. As seen in Figure 3.15(a), mCherry fluorescence increases over time when encoding DNA is present, indicating synthesis of protein. Simultaneously, Pi levels steeply increase, as seen by a decrease in FRET signal between the two proteins in the sensor (Figure 3.15(b)). Combined on the same graph (c), it's possible to compare the inferred accumulation of phosphate simultaneously with the appearance of mCherry fluorescence. The steepest change in both FRET signal and mCherry signal take place in the first two hours, which is likely to be the most active phase of the cell-free reaction. When the FLIPPi-30_SpC sensor is added to buffer alone, no change in red fluorescence or FRET signal is seen across the 20 hours, indicating the dynamic changes in fluorescence we see in other samples are due to activity of the cell-free reaction. Interestingly, when the sensor is added to a cell-free reaction with no plasmid DNA present, thus no mCherry synthesised, the FRET signal does not decrease as rapidly as when mCherry is overexpressed, and generally remains at a higher level (equating to a lower phosphate concentration) throughout. This would align with expectations from previous literature (Caschera & Noireaux, 2014), where inorganic phosphate production is largely driven by protein synthesis in the cell-free reaction.

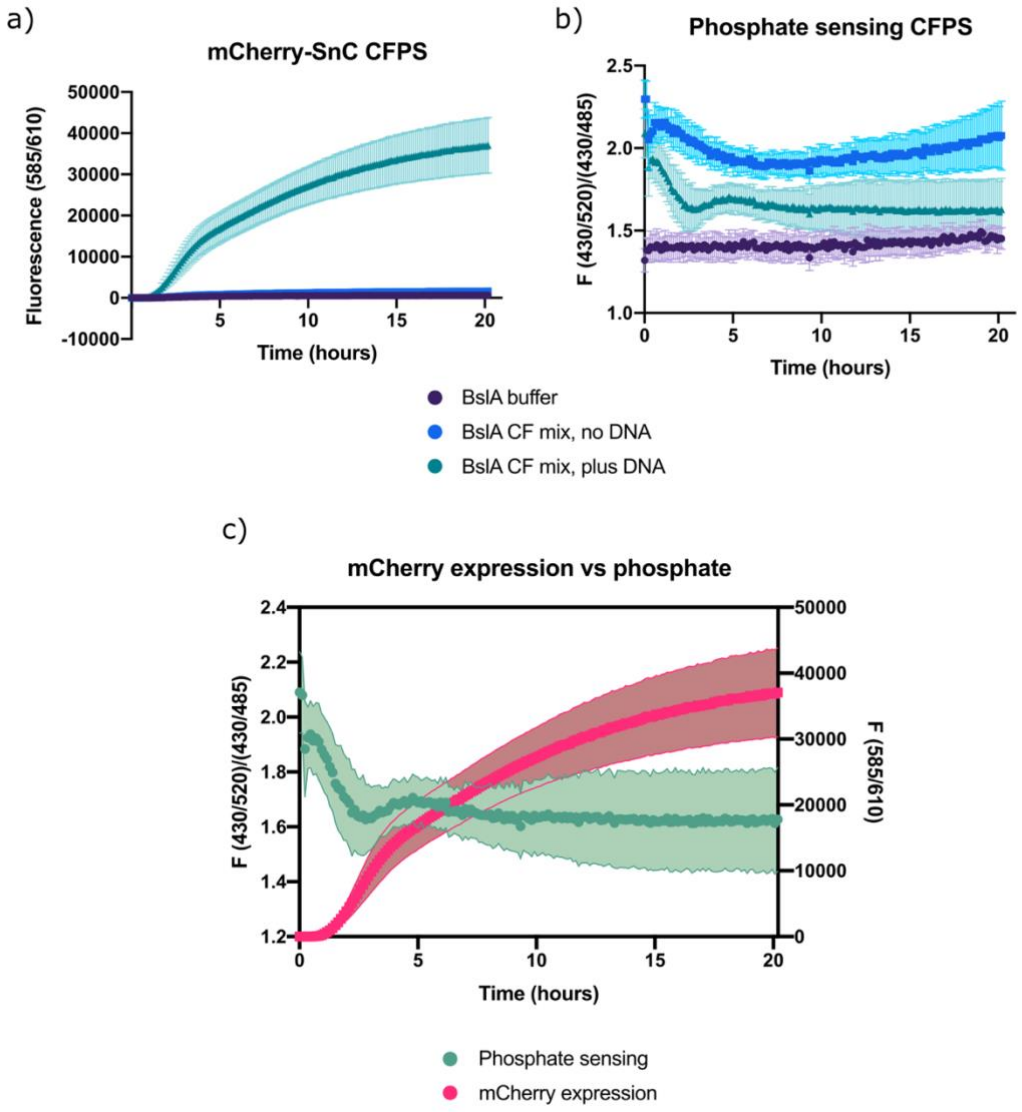


Figure 3.15 Simultaneous Pi biosensing and mCherry expression tracked in the same reaction. (**a**) Red fluorescence, corresponding to expression of mCherry-SnC in CFPS over 20 hours in plate reader. (**b**) FRET signal from phosphate sensor present in the same CFPS reactions over time. The ratio decreases as more phosphate binds to the sensor. (**c**) An overlay of mCherry expression (a) and phosphate sensing (b) over time. Phosphate sensing is shown on left hand side y axis, mCherry expression is shown on right hand side y axis. For all graphs, error bars, s.d. (n=3).

The results presented in this chapter show promising preliminary data toward the realisation of protein-based biosensors for real-time cell-free monitoring of inorganic phosphate levels.

3.3 Discussion

Cell-free technology is a field of fast development in synthetic biology. Since its advent in the 1940's (Doerner et al., 2014), it has evolved into a key technology used for a myriad of applications in both research and industry. During the evolution of the technology, much effort has been focused on how to increase reproducibility within the field, and how to optimise the reaction to prolong the reaction time and increase efficiency to enable higher yields of molecule of interest (Dopp et al., 2019; T. W. Kim et al., 2007; Takahashi et al., 2021; Vezeau & Salis, 2021)

In this chapter, I present results from a systematic study optimising three core components of homemade cell-free systems. I pair the optimised cell-free protocol with prior characterisation of BslA monolayers (Chapter 2.1), demonstrating the capabilities of BslA monolayers to specifically capture proteins synthesised in CFPS. Further to this, I present preliminary results for an adaptable method to monitor cell-free analytes in real-time.

Through my optimisation of the protocols for homemade cell-free, I present data demonstrating the delicacy of a functional cell-free reaction, which emphasises the importance of user optimisation for implementation of homemade cell-free protein synthesis protocols.

I find no significant difference in the optimised preparation of energy solution, when controlling for user and human error. However, I show that preparatory techniques for isolation of DNA from cells vary significantly in the production of a functional template for CFPS. Most surprisingly, these differences in DNA functionality, or purity, could not be detected by standard methods of purity assessment (gel electrophoresis, spectrophotometry). This suggests that to screen for DNA functionality in CFPS, one must test the DNA in the reaction itself. I conclude that preparation of plasmid DNA from *E. coli* cells by maxiprep followed by a secondary clean-up step using a Zymo kit provides the most reliable, functional preparation of DNA for cell-free reactions. This increases CF expression of POI 100 fold, when compared to expression from miniprepped DNA, the preparatory technique I originally used.

In the literature, protocols and optimisation of the cell-free reaction focuses almost entirely on protocols for lysate production (Cole et al., 2020). The few references to DNA preparation methods I could find are contradictory: two papers report that secondary clean-up steps for plasmid DNA have no effect on functionality of the

reaction (Levine et al., 2019; Z. Z. Sun et al., 2013), and one reports that secondary clean-up of plasmid DNA should always be included (Wandera et al., 2020). In other papers, although not directly referenced or explained, the methods used for plasmid purification vary significantly: use of miniprep, midiprep or maxiprep, and inclusion or exclusion of secondary clean-up steps. I conclude that there is large variability in functionality and purity of DNA resulting from kits made by different manufacturers. This would explain why for some research groups, DNA extraction via miniprep provided sufficiently pure DNA, whereas my results show that miniprep of DNA using a Qiagen kit never produce DNA pure enough for CFPS. This also makes sense in the context of my results for secondary DNA clean-up steps: some kits (Zymo, Monarch) improved purity and functionality of DNA, whereas one (Promega) severely decreased DNA functionality in CF.

The absence of a comparative test for plasmid DNA purification techniques for cell-free reactions seems a major gap in the research. If a user comes to the field with no prior experience (as I did), there is no consensus on recommended conditions for DNA preparation, and thus they may waste precious weeks and months understanding why reactions fail (as I unfortunately did). To fill this gap, a more comprehensive test of DNA purification methods for cell-free protein synthesis could be tested. I would propose the purification of plasmid DNA using all popular commercial purification kits, and compare expression of protein from these preparations across the most popular cell lysate systems. As CFPS continues to be implemented in more and more labs worldwide, accessibility of the technology for new users is important.

Preparatory techniques for CFPS lysate are similarly variable, and I show in this chapter that optimising each of the key steps in the lysate production protocol can significantly affect the final output of POI. The chosen cell strain for lysate production can influence the functionality of the reaction, yet growing cells for a fixed time (3 hours) or to fixed OD (0.5) does not significantly influence functionality of the system. Interestingly, separation of the two most common post-lysis processing steps identifies the beneficial effects of one (run-off reaction) over the other (dialysis) when used in this protocol. Most commonly, these post-lysis steps are tested together in the literature, rather than independently. When decoupled, it’s possible to see that one has an advantageous effect on protein synthesis above the other.

There is much argument about the benefits of run-off and dialysis steps in the literature (Cole et al., 2020). The majority of papers report that dialysis and run-off improves protein yield from native promoters, yet decreases or does not affect yield from T7 promoters (T. W. Kim et al., 2006; Kwon & Jewett, 2015; Silverman et al., 2019). Although in the minority, some papers do support my findings, that run-off can increase protein synthesis under T7 expression in CFPS (Liu et al., 2005), and dialysis can decrease activity (J. Kim et al., 2019). However, the factor that complicates comparison of these results even further is that many of the studies were completed using lysates prepared from different cell strains, which appears to have a large influence over the effects of these post-lysis steps. The papers in agreement of my findings, that run-off improves activity and dialysis decreases it, used *E. coli* cell strains A19 and MG1655 to create lysates. The lack of consensus in the literature about whether post-lysis steps are beneficial, and exactly for which strains of *E. coli* could be due to the lack of complete understanding of the mechanisms of their action. The speculated role of the run-off reaction is to degrade sheared host mRNA and DNA, as well as releasing ribosomes from bound mRNA. However, the reasons for the wildly differing effects this step can have on functionality of CFPS remain largely unexplained and therefore prompt the question of whether the role is truly understood, and how it can relate to the genotype of the strain used so significantly. Similarly, it is not explained in the literature how dialysis can have such a dramatic effect on functionality between different cell strains. It could be speculated that the decrease in functionality of lysates when expressing protein via T7 RNA polymerase infers that during dialysis an important small molecule involved in T7 RNAP driven transcription is lost. However, this would still leave the question of why the inclusion of dialysis could improve functionality for native expression. Evidently, a more detailed understanding and analysis of the effect of both dialysis and run-off steps on CFPS lysate production is required.

After optimisation, my homemade cell-free system is able to consistently produce above 60 µg/mL protein. This is around 10-fold lower than the best described example in the literature: 900 µg/mL (Levine et al., 2019), and from a commercial kit: 500 µg/mL using the S30 T7 High-Yield Protein Expression System (Promega), adapted originally from (Zubay, 1973). This indicates that while I have significantly improved cell-free expression protocols in my hands, there is still room for further optimisation of the components to achieve higher protein yields. One aspect of the reaction I did not examine in this study is the composition of the additives included in the cell-free

reaction: such as buffer conditions or components in the energy mix. This could potentially be explored further, along with more thorough optimisation of all components tested here, to improve final yields of POI.

For a more thorough understanding of cell-free lysate functionality, these experiments should be repeated with expression of a different protein of interest. As the DNA sequence and subsequent protein of interest was kept constant through these experiments, it is possible the chosen conditions may not be optimal for expression of other proteins of interest. I would recommend a similar optimisation process be undertaken at the start of any project using cell-free technology for the first time in the lab.

After optimisation of the homemade cell-free components to produce my protein of interest, I combine this knowledge with prior optimisation of BslA monolayers to capture proteins from heterogenous solutions (Chapter 2.1.1) for the capture of proteins expressed in a cell-free reaction to the BslA monolayer. I demonstrate this idea through expression of mCherry_SnC protein in cell-free and subsequent capture to a BslA-SnT surface. I believe the use of BslA monolayers as a tool for in situ capture of proteins from CFPS will enable CFPS to be more accessible to others, due to the advantages outlined in detail in Chapter 2.3, such as ease of use, low cost of materials, and robustness of attachment to the surface. These are criteria that have been listed as desirable changes for CFPS surface immobilisation previously (Kilb et al., 2014). The ability to immobilise the protein synthesised on the surface enables downstream assessment of protein functionality, for example to test for binding proteins, or to potentially have one protein coating the surface as another is made in a second round of CFPS. These are both exciting avenues of work I would like to see demonstrated next to follow on from this work.

My previous work in binding DNA to a BslA surface (Chapter 2.1.4) provided the rationale that DNA encoding POI could be tethered to the surface via mVirD, and subsequently transcribed from to produce POI. Unfortunately, I was unable to demonstrate this idea experimentally. Expression from surface bound DNA was achieved, but it was clear the DNA was stuck non-specifically at such high concentrations. Exploration of optimal DNA concentrations added to the surface, reaction times, and more stringent washing techniques could reveal a protocol where specific attachment of plasmid DNA to a BslA monolayer via mVirD2 can be achieved

for protein expression. To help solve this issue it would also be important to gain a better understanding of the number of molecules present in a BslA monolayer: how do the proteins pack together in a given area, and if all of these BslA proteins are tagged (SpT or SnT), how many of them can be functionalised by protein or DNA? The answer to the latter I would expect to be dependent on binding rates and steric hindrance of the molecules on the surface. A better understanding of the monolayer properties would help to understand how efficiently it is working and how it can be further improved for potential applications.

Finally, I investigate novel techniques for real-time monitoring of cell-free reactions. FLIPPi-30m, a FRET-based biosensor for inorganic phosphate, is capable of reporting inorganic phosphate levels in the cell-free reaction in real-time (Gu et al., 2006). As the sensor uses wavelengths of light orthogonal to those for detection of mCherry, inorganic phosphate accumulation and appearance of mCherry fluorescence can be plotted together, revealing interesting trends. mCherry is known to have a relatively slow maturation time in cell-free protein synthesis (Balleza et al., 2018; MacDonald et al., 2012; Moore et al., 2018), which means the appearance of fluorescence seen in this graph does not necessarily correlate to protein production. It is likely that protein synthesis has been terminated long before maximum fluorescence of the reaction is reached. To use these tools for commentary on real-time interplay between protein production and phosphate accumulation (or other analyte of interest), one would require a protein with an extremely fast maturation time, or to characterise the maturation rate of the protein in the cell-free environment, and then apply this rate to the fluorescence values measured over time. However, without those additional steps, I still expect this tool to be useful for tracking of analyte accumulation in real-time in cell-free reactions, independent of protein production rate. For example, it would be interesting to compare the rate of accumulation, and final values of Pi in different cell-free systems: namely a homemade lysate-based system compared with PURE and MyTXTL systems (containing an inorganic phosphate regeneration mechanism). In the future, it could also be exciting to see how these analytes fluctuate in cell-free systems made from organisms other than *E. coli*, for example those derived from *V. natriegens* or *Streptomyces*-based systems, both of which are capable of high yields of target protein, thus are good targets for optimisation for industry applications (Des Soye et al., 2018; J. Li et al., 2017).

After completion of this thesis, new information was obtained regarding the pH sensitivity of the phosphate sensor. It seems that pH can drastically affect the FRET output of the sensor, and as pH can change in CFPS during the course of the reaction, this is a major problem for using this sensor in a cell-free setting. For future development of this idea, the pH of the cell-free reaction would need to be closely monitored or controlled, to be certain it is only the phosphate level that the sensor is reporting. This is a major limitation of many protein based biosensors due to the intrinsic sensitivity of many proteins to pH. There are paths to overcome this, however, for example by using a protein based pH sensor simultaneously to monitor pH levels.

If controlling for pH sensitivity can be achieved, pairing of this technology with the rapidly advancing field of modelling of cell-free systems could help to optimise bottlenecks in the metabolism of the reaction (Batista et al., 2021), toward increasing final product yield in cell-free systems.

Ultimately, I would be excited to see the implementation of this system with multiple protein based bio-sensors bound to the surface in different defined locations (a possibility given our understanding that BslA proteins do not move in the monolayer, Chapter 2.1.2), reporting analyte accumulation in real-time for one cell-free reaction as it proceeds. As the sensors selected are reversible, and capable of attachment to the surface, this enables real-time read-out of analyte concentration in a flow state, for example in a microfluidic device. It could potentially enable the reuse of such devices, depending on the robustness of BslA monolayers over longer time scales.

Chapter 4 An expanded split intein library for in vitro nanorod synthesis

Sections 4.2.1 and 4.2.3 are adapted from (Pinto et al., 2020)*. The experiments in Section 4.2.1 were completed together with Dr Filipe Pinto; specifically, I helped with cloning and initial characterisation of in vitro trans-splicing, and optimisation of the mCherry platform and protocols. Filipe obtained final data shown in this section. All work in Section 4.2.2 and 4.2.3 was completed by myself.*

4.1 Introduction

4.1.1 Protein ligation tools

Catalysts for the irreversible ligation of proteins in vivo and in vitro are a desirable tool in synthetic biology. Directing the association and ligation of proteins in a complex environment enables specified localisation and the creation of novel macromolecular complexes. While large libraries of non-covalent protein-protein interactors (e.g. coiled-coil pairs (Chen et al., 2019; Lebar et al., 2020)) exist, orthogonal tools for irreversible protein binding are much more limited.

Currently, the most prominent protein ligation tool used currently in synthetic biology is SpyTag/SpyCatcher, capable of linking two proteins by covalent attachment (Zakeri et al., 2012) (introduced in greater detail in Chapter 2.1). The development of this technology in the 9 years since its discovery has led to the creation of several new versions and related peptide/protein pairs, namely the SpyTag/SpyCatcher versions 002 (Keeble et al., 2017) and 003 (Keeble et al., 2019), capable of faster ligation and higher efficiency, and the SnoopTag/SnoopCatcher peptide/protein pair, also capable of covalent ligation, yet orthogonal to SpyTag/SpyCatcher (Veggiani et al., 2016). SnoopTag/SnoopCatcher were developed from the protein RrgA, which contains another native isopeptide bond. This isopeptide bond has been exploited as another protein glue technology, split to give proteins Jo and In, which, when mixed, form a spontaneous covalent bond between them (Bonnet et al., 2017). Jo and In are relatively slow to react (only reaching 100% completion after 3 hours), and are composed of two quite bulky protein domains (Jo is 78 residues and In is 134 residues). Cross reactivity of Jo and In with Spy technology has not been published. Therefore, while development of these tools has been widespread, the toolbox of verified orthogonal tools remains limited, standing at only two. A further drawback of these tools is the large protein domain retained between the two ligated proteins. This

is impractical in applications where native-like final protein assemblies are desired, and it can also be perturbative to function of proteins.

4.1.2 Split inteins for protein ligation

Split inteins present an alternative to the covalent protein ligation toolbox. Inteins are protein segments capable of association and excision from their flanking polypeptide sequences (exteins), catalysing the irreversible ligation of the two exteins in the process of self-excision (Figure 4.1) (Shah & Muir, 2014). Inteins are found across all domains of life and in viruses, and can exist as three types: full length inteins, which contain an endonuclease domain not required for intein splicing, mini-inteins, which lack the endonuclease domain, but exist as one protein sequence, or split inteins, where the intein is split into two separately encoded fragments, capable of association and splicing in trans (Aranko et al., 2014; Novikova et al., 2016; Saleh & Perler, 2006). Split inteins are found in nature, but can also be created by the splitting of full-length inteins or mini-inteins (Shah & Muir, 2014).

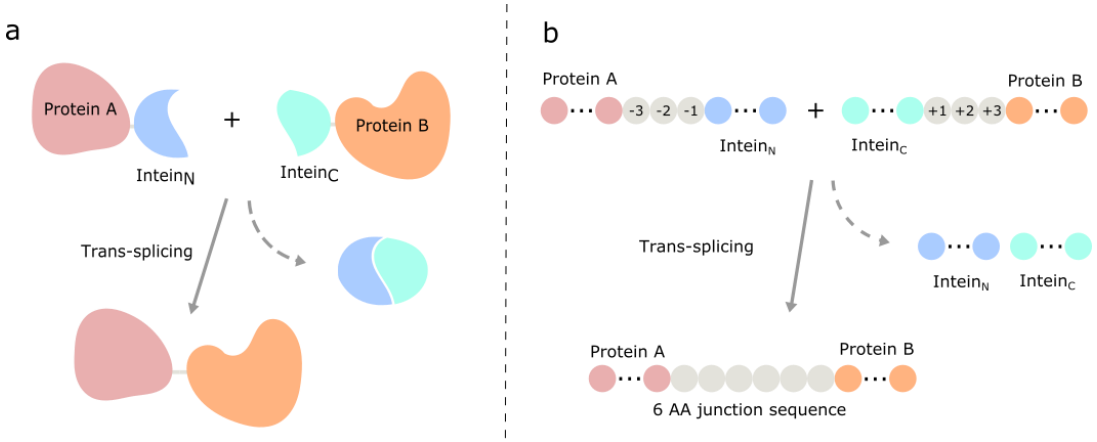


Figure 4.1 Protein ligation via split intein trans-splicing. (**a**) Two proteins (protein A and protein B), can be spontaneously ligated through the association and subsequent excision of split intein from the C and N terminus of protein A and B respectively. (**b**) Amino acid view of splicing reaction. For efficient splicing, the native junction sequence of each split intein is required between the split intein and protein of interest (-3 and +3). After trans-splicing and excision of split intein from protein complex, a six AA 'scar site' is left at the site of ligation between protein A and B.

The functionality and efficiency of inteins is dictated in part by the extein residues either side of the intein domains. The most crucial is the first residue of the C-extein, which is required to contain a thiol or hydroxyl group for catalysis of the reaction. The

influence of the first three residues of the C-extein (+1, +2, +3) and last three of the N-extein (-3, -2, -1) over the efficiency of the splicing reaction varies between inteins; for some, native extein residues are essential for splicing to occur, whereas other inteins are more promiscuous, capable of splicing with only one or two original extein residues (Saleh & Perler, 2006; Stevens et al., 2017). The engineering of split inteins to allow for high splicing efficiency with non-native extein sequences has been a field of research in itself (Appleby-Tagoe et al., 2011; Oeemig et al., 2012).

While the amino acids left between the two ligated proteins is much smaller with inteins than with other protein ligation tools (i.e. 6 AA compared to ~100), it is important to note the reaction is not completely scarless. The most common catalytic residue required, cysteine, can sometimes be problematic to introduce into protein sequences. Therefore, inteins should be checked for compatibility with the desired proteins to be ligated.

Split inteins are useful for a wide range of applications. They have been used in the development of bi-specific IgG antibodies, underwater adhesives, antimicrobials, protein purification systems, gene therapy, engineering of ion channels, and most recently, in the development of a platform for protein/peptide interaction screening (L. Han et al., 2017; E. Kim et al., 2018; Y. Li, 2015; López-Igual et al., 2019; Sarkar et al., 2021; Truong et al., 2015; Wood & Camarero, 2014).

While there has been considerable attention on split inteins as a prominent field of biotechnology, most inteins have been characterised independently by different lab groups, and therefore under different conditions. In this chapter, I present data from my paper with Dr Filipe Pinto (Pinto et al., 2020), where we comprehensively characterise an expanded library of split inteins all under the same conditions, both in vivo and in vitro. As aforementioned, required extein junction sequences can be problematic for some protein ligations, where the inclusion of non-native AA in the final sequence may inhibit functionality. Therefore, we hope this large library will increase the chances of users finding a compatible intein for their desired application. Alongside this, the characterised library allows for more thorough understanding of how the inteins function, detailing rates of splicing and testing orthogonality between inteins.

4.1.3 Synthesis of elongated proteins

One particularly interesting application for this technology is for the post-translational synthesis of extremely elongated proteins (Bowen et al., 2018, 2021; E. Kim et al., 2018). The distinct properties of rod-like, elongated proteins: strength, high length to diameter aspect ratio, make them ideal candidates for several biomaterials applications. As mentioned in further detail in Chapter 2.1, rod-like behaviour of proteins has the potential to mimic nanopillar like arrays found in nature, employed as anti-glint nanostructures (Raut et al., 2011). They are also useful for crosslinking, for example as agents for hydrogel formation (F. Sun et al., 2014). Beyond this, there has been significant interest in the use of rod-like proteins for mimicking spider-silk fibres and the production of silk-like biomaterials (Humenik et al., 2011; Xia et al., 2010).

In vivo recombinant expression of extremely long proteins would be difficult, partly as eventually they would be too big for the cell, but also as rod-like proteins often contain repetitive domains, which makes genetic modification extremely difficult due to genetic instability and toxicity (Bowen et al., 2018). Proteins with other interesting biomaterials properties, outside of rod-like morphology, also tend to consist of repeating domains, such as elastin, keratin, collagen and resilin (Abascal & Regan, 2018). Therefore, the development of new methods for post-translational assembly of proteins with biomaterials applications is highly desirable.

Here, the test protein is SasG, as described in Chapter 2, an extremely strong and elongated protein native to *S. aureus* (Gruszka et al., 2015, 2016). SasG is set out to be further elongated for biomaterials applications, by assembly of SasG protein units by split intein ligation.

In this chapter I present an in vitro application for the characterised library of split inteins: assembly of elongated proteins via intein trans-splicing. I present two protocols to assemble SasG proteins, one by recursive solid-phase assembly using only two orthogonal split inteins, and the other by a one-pot reaction of five orthogonal split inteins. Both protocols result in the production of SasG18, assembled from five protein building blocks of SasG3 proteins.

4.2 Results

4.2.1 Split intein library in vitro characterisation

Following the identification and characterisation of 24 functional inteins in vivo by Dr Filipe Pinto (Pinto et al., 2020), we worked to analyse the trans-splicing ability of these inteins in vitro. We created a simple split fluorescent protein platform, which enabled parallel analysis of intein trans-splicing activity over time (Figure 4.2). mCherry was split between residues 159 and 160, producing two protein fragments unable to produce fluorescence separately, yet capable of reconstituting fluorescence if brought into close proximity of each other and ligated.

To assess each split intein's ability to trans-splice two proteins together, we inserted it between the mCherry halves. Once these two proteins were mixed ($mCherry_N$-SI_N + SI_C-$mCherry_C$), if fluorescence was observed it implied that successful trans-splicing had been achieved by the intein. The fluorescence of the reaction could be monitored over time, allowing for speed of splicing to be determined as well as end-point splicing efficiency. Split intein fused proteins do not require purification for ligation to occur, and can react efficiently in cell-lysates. This is an important advantage of the method we developed, as it allowed for fast screening of multiple inteins, from the protein expression in *E. coli* to the analysis of activity in vitro, in contrast with methods requiring purification from cells which is much more time consuming.

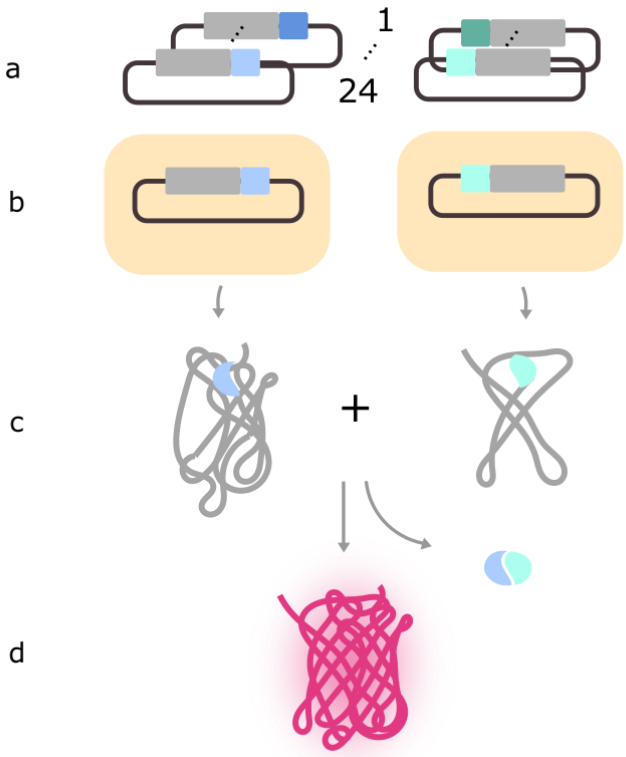


Figure 4.2 Principle of in vitro platform for assessment of split intein functionality. (**a**) mCherry is split into two halves, with split inteins fused at the split site. 48 plasmids are created: to encode for 24 split inteins on N and C terminus of mCherry. (**b**) Each plasmid is transformed into an *E. coli* protein expression cell strain, and cells are grown to express POI. (**c**) Cells are lysed and cleared lysate collected containing POI. (**d**) Lysates containing N- and C- fragments of mCherry are combined in splicing buffer. Fluorescence of mCherry is tracked over time, signalling the successful trans-splicing activity of split inteins and successful reassembly of mCherry.

Addition of any AA sequence to a protein can dramatically alter expression levels, as we discovered here, where although the split intein fragments added to N and C terminus halves of mCherry were relatively small, optimal expression and extraction conditions differed drastically between proteins. Intein fusion to proteins has previously been shown to negatively influence solubility (Aranko et al., 2014; Otomo et al., 1999). Initially, all proteins were tested for expression via the pBAD promoter in *E. coli* TOP10 cells. Coding sequence of proteins with low expression or low solubility were subsequently cloned into pET11 plasmids, and tested for expression in *E. coli* Origami cells via the T7 promoter. Still, some proteins remained largely insoluble, therefore extraction under denaturing conditions was preferable to obtain

all proteins. This resulted in expression of all fragments in either TOP10 or Origami cells, and extraction of all proteins under mild denaturing conditions (Figure 4.3).

Previously, split inteins have been characterised in vitro under different buffer conditions between different papers. We wanted to characterise all inteins under the same conditions, with the goal of providing a library of orthogonal inteins that can be used efficiently in tandem within the same reaction. Analysis of optimal splicing conditions for all inteins in vitro allowed us to find the best-fit buffer, which all subsequent experiments were completed in: 100 mM Tris-HCl pH 9.0, 100 mM NaCl, 2 mM DTT.

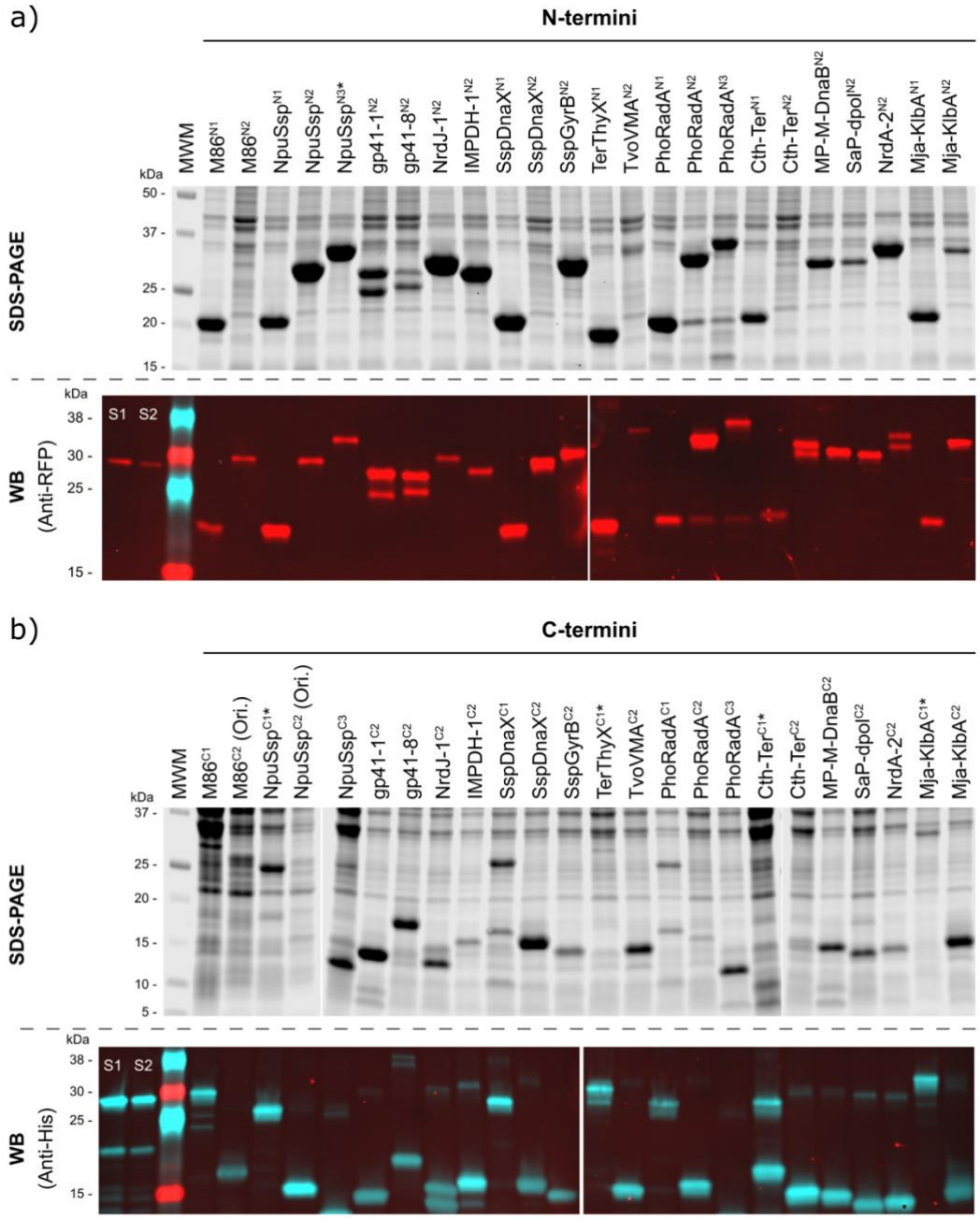


Figure 4.3 Validation of extraction of all split mCherry fusion proteins under mild denaturing conditions. All fusion proteins can be seen at desired MW, by SDS-PAGE (top panel), or Western blot (lower panel). Inteins can be split at three different positions: S1 (N1 and C1), S2 (N2 and C2), or S3 (N3 and C3). (**a**) N-termini fragments of split mCherry fused to split intein N-terminus. These fusion halves are detected by the Anti-RFP antibody, visualised as red in this Western blot. S1 and S2 = standards of purified mCherry-His, 2.5 pmol and 1.25 pmol, respectively. (**b**) C-termini fragments of split mCherry fused to split intein N-terminus, with C-terminus His-tag. These fusion halves are detected by the Anti-His antibody, as visualised here in blue. Purified mCherry-His is visible with both antibodies. MWM =

molecular weight marker. * = inteins with flexible linker at the canonical split site. Ori = *E. coli* Origami.

Cell lysates containing pairs of split mCherry-intein fusions were combined in the reaction buffer at room temperature. Red fluorescence of each reaction was measured at 5 min 24 second intervals over 21 hours, this data is plotted in Figure 4.4. While there is an initial delay in fluorescence output, likely due to time taken for mCherry maturation after reconstitution, many inteins are able to reconstitute fluorescence very quickly, with five (gp41-1, gp41-8, NrdJ-1, IMPDH-1 and SspGyrB) achieving 75% of the maximum spliced product within one hour of mixing.

The final end point fluorescence of these reactions is validated as protein trans-splicing via Western blot (Figure 4.5(a)). The overlapping signal of both antibodies (Anti-RFP recognising N-termini, Anti-His recognising C-termini) is shown where successful assembly of mCherry has been achieved. Fluorescence observed in the mCherry platform can be seen to proportionally correlate to quantity of spliced mCherry protein via Western Blot, thus confirming the accuracy of our technique and the assembly of mCherry. The 12 best performing split inteins were further assessed for splicing activity at additional time points: after 1 hour, 4 hours, 20 hours and 67 hours (Figure 4.5(b)). While most inteins reach maximum splicing after 20 hours, for PhoRadA at split site 1, splicing is still incomplete after 20 hours. Slow, yet highly efficient inteins such as PhoRadA could have niche applications in bioengineering, such as when timing of cellular interactions are in interplay with each other.

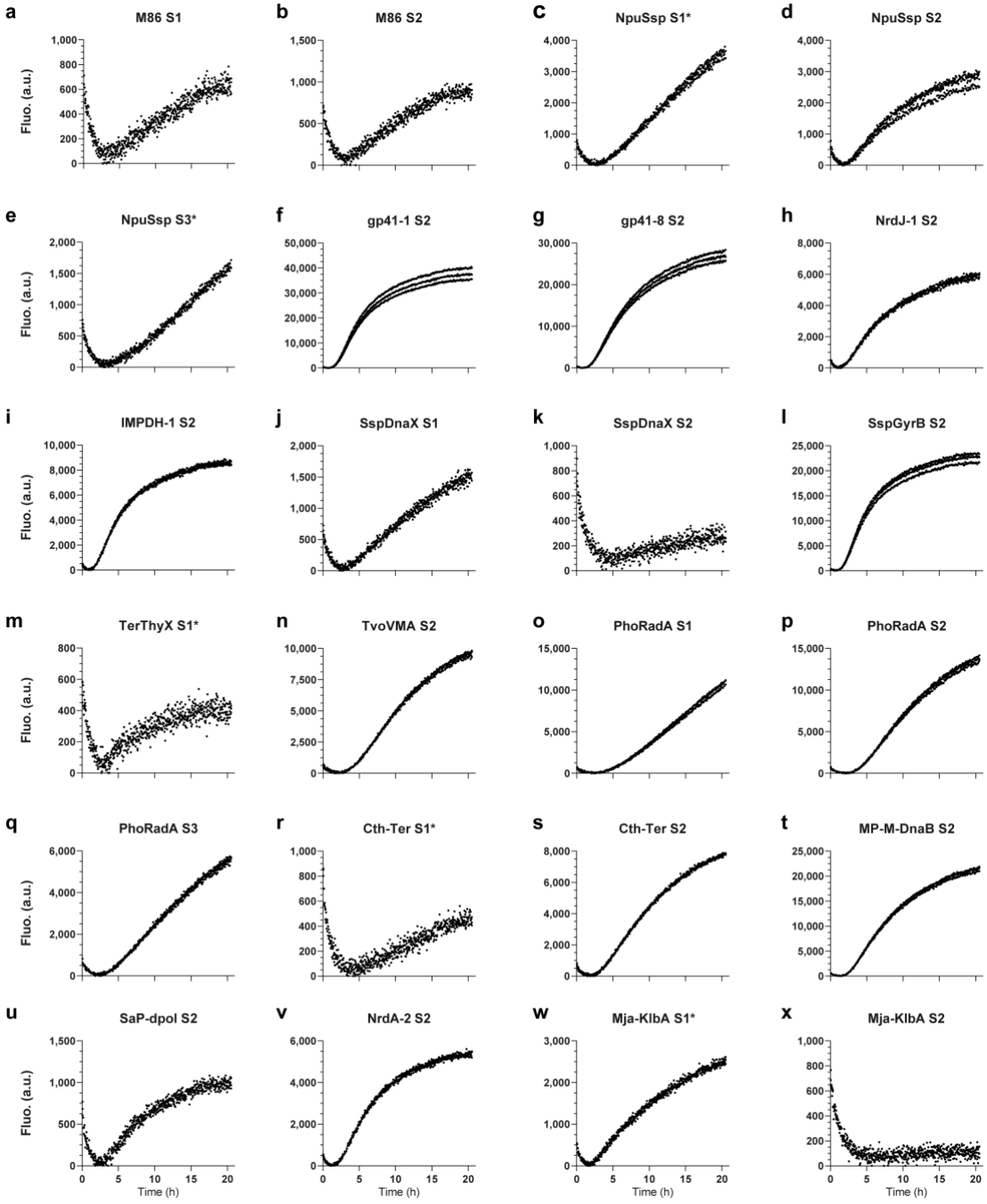


Figure 4.4 Analysis of split intein splicing activity by monitoring of mCherry fluorescence reconstitution over time. Cleared cell lysates containing corresponding N and C fragments of intein-mCherry fusions are combined in splicing buffer (100 mM Tris-HCl pH 9.0, 100 mM NaCl, 2 mM DTT) at equal concentrations estimated by SDS-PAGE. Reactions are allowed to react for 20.5 hours at room temperature, and fluorescence is measured every 5 minutes 24 seconds by the plate reader. Data from 3 independent replicates is shown here. Data was normalised against the lowest value for each sample. * Inteins with a flexible linker at the canonical split site.

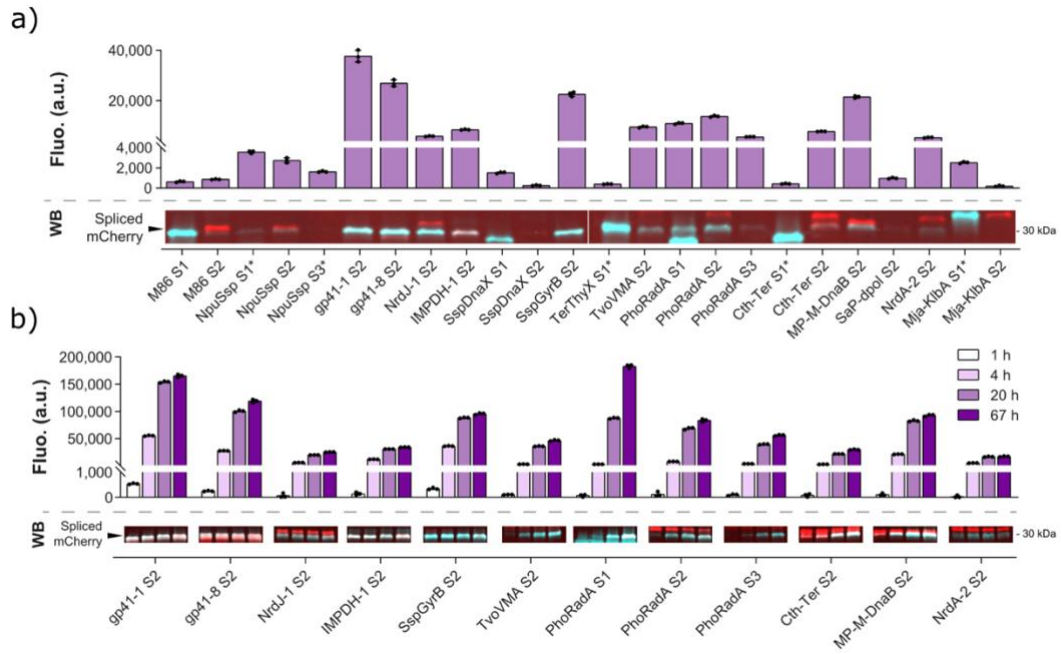


Figure 4.5 Splicing activity of split inteins validated by Western blot. (**a**) Endpoint fluorescence values (top panel) of mCherry assembly in vitro from cell lysates over 20 hours, with Western-blot validation (lower panel) of mCherry assembly for each sample shown. Red signal is from Anti-mCherry antibody and blue signal is from Anti-His antibody, overlap of these two bands is shown with white colour, indicating the successful assembly of mCherry halves at the correct MW (28 kDa). (**b**) Selection of 12 most functional split inteins for characterisation over longer time course. Fluorescent plate reader measurements (top panel), and Western blot analysis (lower panel) shown. Reacted cell lysates were incubated for 67 hours in total, with fluorescence measurements and samples removed for WB analysis at intermediate time points. Error bars, s.d. (n=3).

Toward the goal of one-pot multi-intein reactions, we tested orthogonality of the 12 best performing inteins using the split mCherry platform (Figure 4.6). Each $mCherry_N$-SI_N was mixed with all 12 intein $mCherry_C$-SI_C and vice versa. Fluorescence was quantified after 20 hours, where absence of fluorescence indicates no cross reaction between corresponding inteins, and the presence of fluorescence between two non-paired intein halves indicates cross-reactivity. As seen in Figure 4.6, 10 out of the 12 tested split inteins were shown to be highly orthogonal. Combinations of inteins with high fluorescent signal were further tested by Western blot validation of splicing (Figure 4.6(b)). Only in the case of PhoRadA (split at multiple sites) was fluorescence due to intein trans-splicing. In the other cases, N- and C-terminus fragments can be seen separately on the Western blot, with no band present at the expected MW for

spliced mCherry. We conclude that in these cases, inteins are non-specifically associating, enabling mCherry halves to unite and cause fluorescence in an impermanent manner.

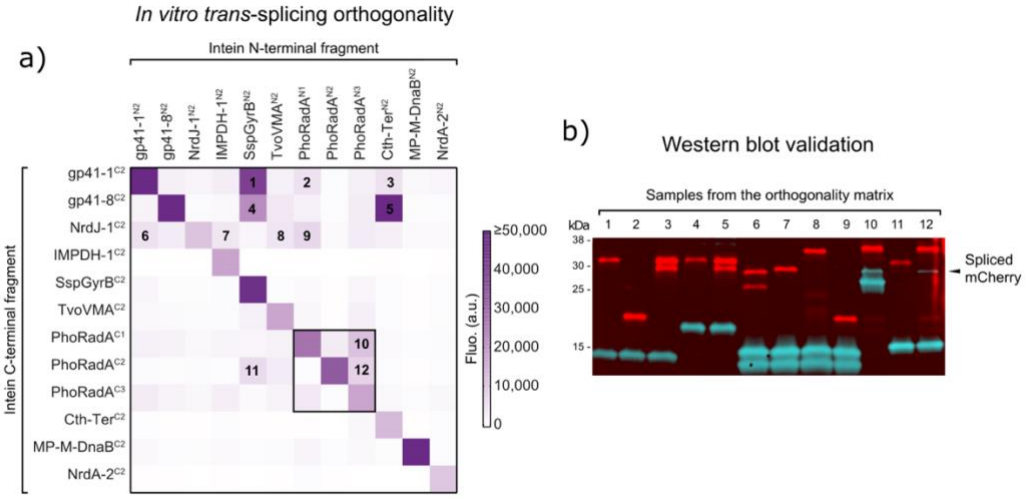


Figure 4.6 Orthogonality matrix of 12 most functional split inteins. (**a**) Cell lysates containing N and C terminal halves were mixed in all possible combinations to test for cross-reactivity. Colour gradient represents fluorescence value, indicating efficiency of intein splicing. Black lines outline split inteins derived from the same intein, split in different locations. Values shown are mean values of three independent replicates. (**b**) High fluorescent signal seen between non-matching protein fragments in the orthogonality matrix were further assessed by Western blot. Numbers at top of gel correspond to labelled squares within matrix. Spliced mCherry is seen by overlap of red and blue signal at 28 kDa in a white colour.

In conclusion, the data presented describes a novel split mCherry platform for in vitro characterisation of split inteins. We use the platform to test 24 split intein pairs, with 12 capable of efficient splicing within 21 hours. 10 of these 12 are found to be highly orthogonal, thus providing a well-defined toolbox of orthogonal protein ligation tools. Explored next in this chapter is the utilisation of this toolbox toward the creation of elongated proteins for biomaterial applications.

4.2.2 Split intein mediated splicing of SasG7 proteins

I set out to use split inteins to ligate long, repetitive proteins together in vitro, with the goal to reach protein lengths not feasible by recombinant protein expression.

To investigate the relationship between repetitive protein length and yield within cells when overexpressed, I induced expression of five SasG proteins within *E. coli* with increasing length. The SasG protein is comprised of repeating domains, and shortened versions of the native protein after each repeating G5 domain have been shown to retain native-like folding and rod-like structure (Gruszka et al., 2015). Figure 4.7 shows overexpression of these proteins in *E. coli* BL21 cells. Comparing the protein content of the uninduced with induced cells, it is clear there is overexpression of protein with each construct. Protein bands correspond to the expected size of the SasG proteins, considering the expected running size of SasG proteins is slightly higher than the corresponding MW, a phenomenon previously mentioned by others working with SasG proteins.

From Figure 4.7, it can be seen that protein yield decreases as protein length increases. This provided motivation for my studies, as it is evident that producing even longer SasG proteins recombinantly in vivo would result in extremely low yields of final protein product. Therefore, in this chapter I employ split inteins to enable post-translational assembly of longer proteins in vitro.

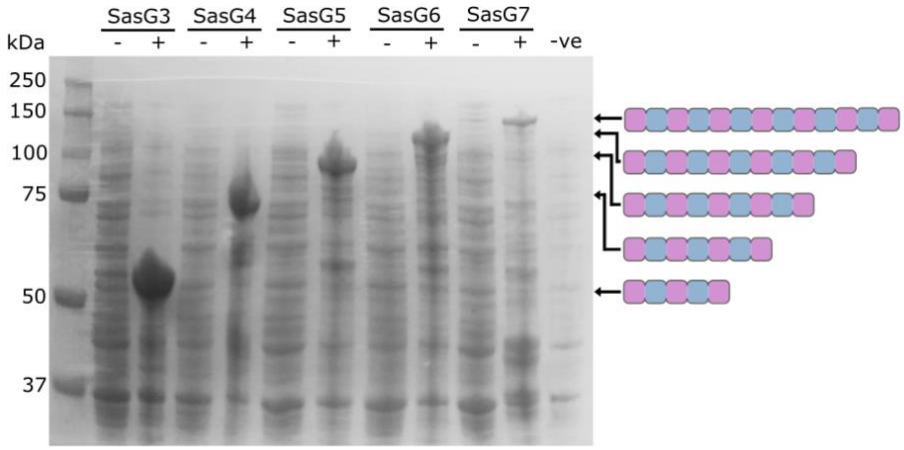


Figure 4.7 SDS-PAGE analysis of *E. coli* BL21 cells expressing SasG proteins of increasing length. Cells harbouring plasmids encoding for SasG proteins are grown with (+) and without (-) IPTG induction, and samples are shown here to assess the appearance of protein bands at the correct MW, indicating overexpression of protein of interest. The last lane

shows a negative control (-ve), cells grown with an empty pET28 plasmid under the same conditions. Cartoons of the final proteins produced are shown on the right hand side, indicating the height protein band they correspond to. Theoretical protein sizes (kDa): SasG3 = 41, SasG4 = 55, SasG5 = 69, SasG6 = 83, SasG7 = 97.

Initially, I began with design and splicing of three SasG7 proteins together using two split inteins (Figure 4.8). I designed these constructs to enable purification of final product away from intermediates: His-tags are present on each fusion protein, but are removed from the final construct in the splicing process. Strep-tag is present on only one domain, and this is retained after splicing. I hypothesised that the spliced mixture could be run through Ni-NTA resin to 'catch' any unreacted fusion proteins, and then purified in a final step via the strep-tag.

Visible by SDS-PAGE in Figure 4.8, trans-splicing of these purified SasG7 fusion proteins can be achieved to create a SasG21 protein. Final yields are low, and the band indicating three way trans-splicing is barely visible. In part, this is due to low quantity of starting materials. Many problems arose in the purification of SasG proteins, such as inefficient binding to resin, loss of protein in washing and concentrating steps. Consequently, POI was lost at every stage of purification and final yields were very low.

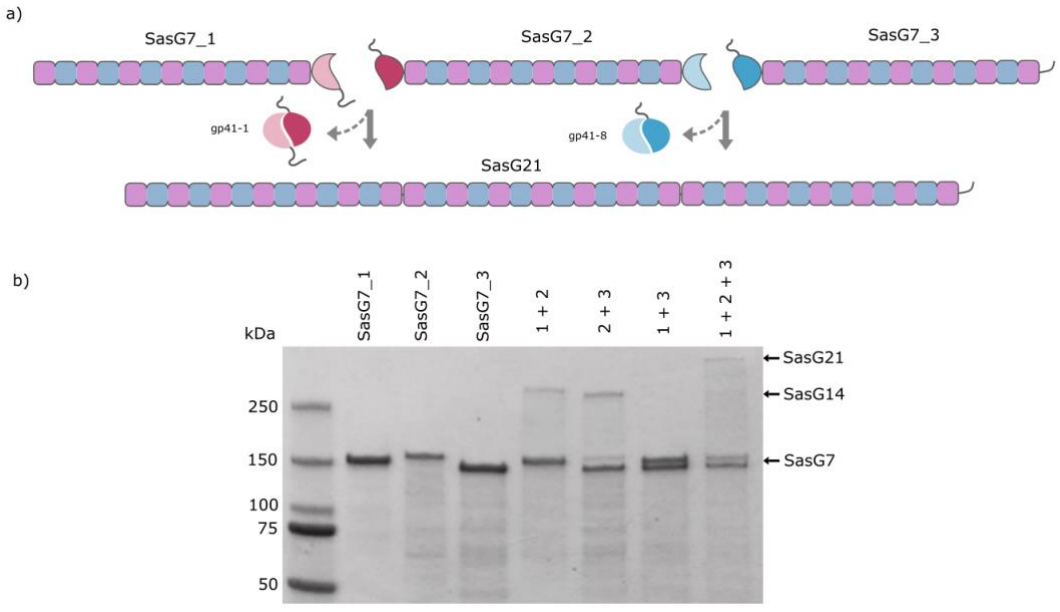


Figure 4.8 SDS-PAGE gel to show trans-splicing of purified SasG7 split intein fusion proteins. (**a**) Cartoon representations of SasG7 fusion proteins. SasG7_1 is fused to gp41-

1N at the C-terminus. SasG7_2 is fused to gp41-1C at the N terminus and gp41-8N at the C terminus, allowing for trans-splicing with both SasG7_1 and SasG7_3. SasG7_3 contains the C-terminus half of gp41-8. All SasG7 units contain a His-tag for purification purposes attached to the split intein. This does not affect the splicing efficiency, yet functions as a convenient way to remove the tag from the final protein product. SasG7_3 also includes a strep-tag at the C-terminus of the protein, which should allow for purification of the final spliced product away from intermediate assemblies. (**b**) SDS-PAGE gel to show purified SasG7 fusion proteins ligating via intein trans-splicing. Purified SasG7 proteins are shown separately and reacted together in all possible combinations. The numbering system of protein assemblies corresponds to the number of the SasG7 fusion proteins being reacted together, e.g. (1 + 2) = (SasG3_1 + SasG3_2). Theoretical protein sizes (kDa): SasG7 = ~100, SasG14 = ~195, SasG21 = 283.

Therefore, a new approach was tested, where cleared cell lysates were mixed for the trans-splicing reaction to occur, and final spliced products were purified from this mixture (Figure 4.9). This began with two rounds of affinity chromatography: first using Ni-NTA resin to 'capture' any unreacted single units of SasG7, and then using Strep-Tactin resin to purify the assemblies via a Strep-tag on the C-terminus of the final protein assembly. This was successful, and elutions were combined, concentrated and buffer exchanged. However, there were still intermediate proteins eluted with the target protein. As these proteins assemblies are doubling and tripling in size, it was expected they could be separated by size exclusion chromatography (SEC). The concentrated SasG21 sample was run on a Sephacryl S-400 column, which has a separation range of 20 kDa – 8 MDa. The SasG proteins are well within this size limit (SasG7 = ~100 kDa, SasG14 = ~195 kDa, SasG21 = 283 kDa) and so separation of all proteins should be possible.

The fractions collected from SEC, in order of elution from the column, are shown on the SDS-PAGE gel in Figure 4.9(c). The first lanes in the SDS-PAGE gel display proteins that ran through the column first, and therefore should be the largest. The first several fractions are from the void of the column, meaning these proteins are larger than the upper size limit of this column (8 MDa). However, the first proteins eluted from the column are in fact SasG7, followed by SasG21, SasG14, and SasG7 again. This is surprising, and could indicate association of the SasG proteins to each other.

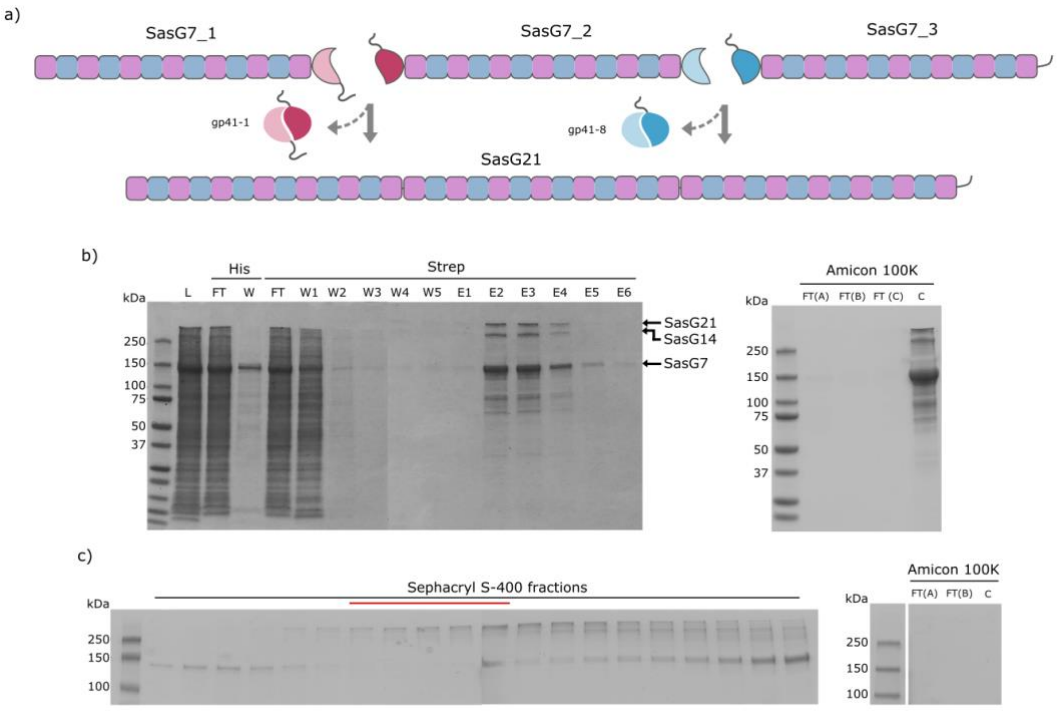


Figure 4.9 Purification workflow of SasG21 spliced product from cell lysate mixture of three SasG7 precursor proteins. (a) Cartoon representation of three SasG7 proteins splicing together to create SasG21. Here, they are mixed while present in a heterogeneous protein mixture of the cleared cell lysate. (**b**) Purification of trans-spliced product via affinity chromatography. The three cleared cell lysates are mixed and allowed to react overnight at room temperature (shown in lane labelled with L for lysate). This allows for three way trans-splicing to occur. The mixture is then passed over Ni-NTA resin, with the goal to capture any unreacted SasG7 proteins as the final desired product, SasG21 contains no His-tag. The FT (Flow-Through) from this resin is then bound to Strep-Tactin resin, aiming to purify the final SasG21 assembly via the strep-tag. (W = Wash, E = Elution). Bands corresponding to the SasG proteins are annotated by the arrows to the right hand side of the protein gel. Following purification via strep-tag, elutions 2 – 6 are combined and simultaneously buffer exchanged and concentrated using an Amicon 100K concentrator column. The flow through from the three rounds of buffer exchange performed in this column (FT(A), FT(B), FT(C)), and resulting concentrated sample (C) are shown in the protein gel to the right hand side of this panel. (**c**) Concentrated SasG21 protein sample is then further purified by size exclusion chromatography (SEC) with the column Sephacryl S-400. Fractions from the main peak are run on an SDS-PAGE gel for analysis of purity and success of separation. Fractions marked with the red line are combined for further concentration via a new Amicon 100K centrifugal unit. However, as seen in the right hand gel, no protein bands can be seen on the SDS-PAGE, indicating loss of POI during this stage of the protocol.

The fractions where the SasG21 protein assembly looks purest were combined (shown in red on the gel image), then dialysed and further concentrated for downstream applications. Unfortunately, the protein was completely lost at this stage of the purification process. This is a phenomenon I have repeatedly seen with the SasG protein: once well purified, they are able to flow through the Amicon filtration columns with pore sizes much smaller than their total size in kDa. Even when the concentration step was repeated with an Amicon 3K filtration unit, which should only allow molecules smaller than 3 kDa to pass through the pore, SasG was seen in the flow through, and was not retained in the column.

I hypothesise that the difficulties I faced in purification of the SasG proteins are due to the high aspect ratio of the protein dimensions. It would appear the protein can behave simultaneously as an extremely large and extremely small molecule, perhaps depending on whether the length or diameter of the protein is considered. The difficulties faced with SasG concentration imply the protein is behaving as a molecule < 3 kDa, yet the difficulties faced with SEC imply the protein is behaving as a molecule > 8 MDa.

Following this work with SasG7, I set out to use SasG3 proteins as an example for multi-intein assembly of protein units. I wanted to demonstrate the capabilities of the split intein library with a smaller, better expressed SasG proteins to enable higher yield of final products. I used the knowledge gained in this sub chapter to direct the design of experiments with SasG3 proteins.

4.2.3 Split intein mediated splicing of SasG3 proteins

I aimed to create extended versions of the repetitive protein SasG by post-translationally assembling shorter and easier to express units of SasG. I truncated the SasG protein after the third G5 domain, named herein as SasG3, and fused split inteins to either side in different combinations, depending on the desired protocol to be used (Figure 4.10). The inteins selected for the assembly techniques were chosen based on splicing efficiency and orthogonality as characterised by the mCherry platform (Figure 4.6). As shown in Figure 4.10(d), all SasG3 fusion proteins were expressed in *E. coli* BL21 and were soluble.

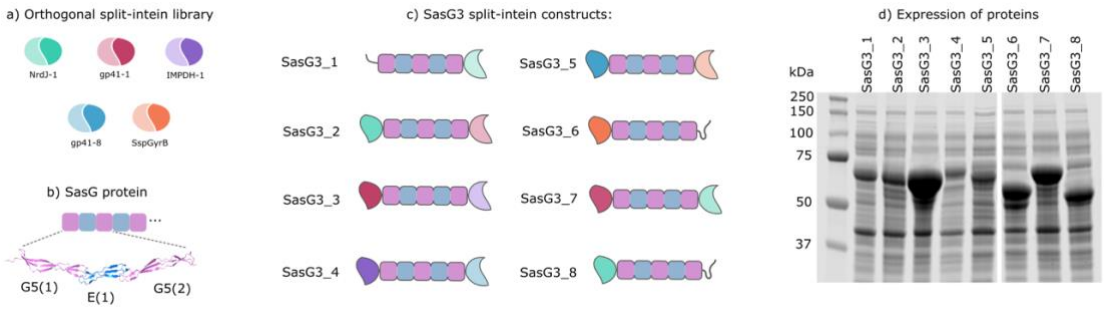


Figure 4.10 Design and expression of SasG3 fusion proteins. (**a**) Five of the most efficient and orthogonal split inteins previously characterised in our studies were selected: Nrdj-1, gp41-1, IMPDH-1, gp41-8, and SspGyrB for use in SasG assembly. (**b**) The SasG protein was truncated at the third G5 domain in the repeating structure, thus termed SasG3 as there are three repetitive units comprising it. (**c**) Split inteins were genetically encoded to N- and C-terminus of SasG3 proteins in a total of 8 fusion constructs. SasG_1 inclusive to SasG_6 were designed for use in a one-pot assembly reaction, using five orthogonal inteins in one reaction. A Strep-tag on the first unit, SasG_1, and a His-tag on the last unit, SasG_6, allowed for purification of the fully assembled protein unit. SasG fusion proteins SasG_1, SasG_2, SasG_7 and SasG_8 were used for recursive solid-phase assembly, alternately using two efficient and orthogonal inteins for repeating assembly steps on Ni-NTA resin. (**d**) SDS-PAGE analysis of SasG3 fusion protein expression. Cleared lysate from *E. coli* BL21 cells overexpressing protein of interest is shown. The bands corresponding to proteins of interest can be seen at the expected molecular weight for each protein (between 50 and 75 kDa markers).

The library of SasG-intein fusion proteins were initially screened for cross-reactivity to each other and with their corresponding partner (Figure 4.11). SasG3_2 and SasG3_7 proteins are able to react ‘infinitely’ as they contain both N- and C-terminus intein segments for two inteins, enabling end-to-end trans-splicing without limit. All units reacted with required partners as expected but not via other inteins, even when the proteins were left to react overnight. Most pairs also reached maximum splicing activity after just one hour. Following this result, I test two novel assembly methods for multi-way SasG3 splicing.

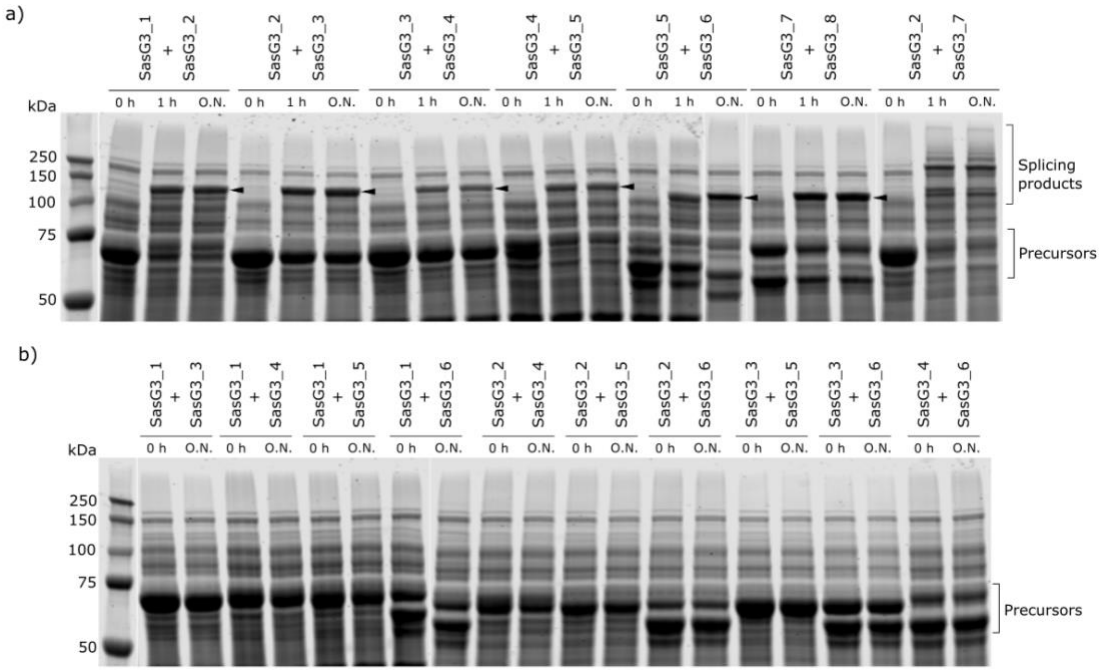


Figure 4.11 Testing cross-reactivity of SasG3 split intein fusion proteins in soluble cell extracts. (**a**) SasG3 fusion proteins are reacted in pairs, with samples run on SDS-PAGE at zero hours (mixed directly into Laemmli buffer to inhibit reaction), and after reacting at room temperature for both one hour and overnight. Arrowheads point to the spliced products resulting from two SasG3 proteins ligating via successful intein trans-splicing. Mixing of SasG3_2 and SasG3_7 together results in endless head to tail ligation by design, as they contain both the N- and C-termini of two split inteins. Protein precursors can be seen at zero hours, before trans-spliced together. (**b**) Testing cross reactivity of SasG3 fusion proteins. Orthogonal SasG3 proteins tested for non-specific splicing by reacting at room temperature overnight. Precursors are indicated on the SDS-PAGE, but no spliced products can be seen.

Initially, I tested a one-pot assembly technique. 6 units of SasG were assembled using 5 orthogonal inteins in one reaction (Figure 4.12). Protein affinity tags present on both the N- and C-terminus of the final protein allowed for purification of only the fully ligated product . This approach is valuable in its simplicity, as all 6 proteins, contained in cleared lysates, can be added together with a reducing agent (1 mM TCEP rather than DTT, to avoid nickel reduction) with Ni-NTA resin and left overnight. The elutions from this column are then pooled and bound to Strep-Tactin resin for 1 hour, to remove any incomplete side-products of the reaction.

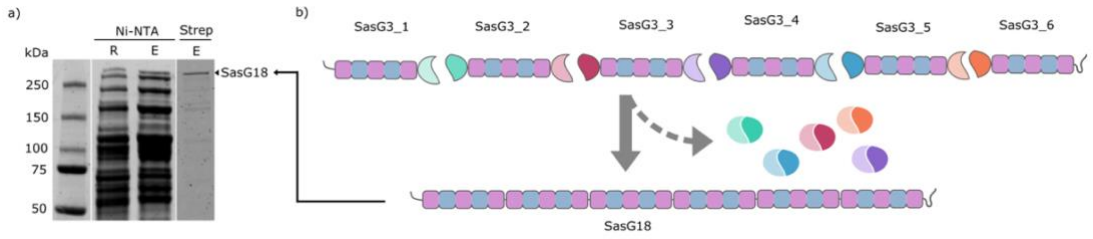


Figure 4.12 One pot assembly of SasG3 proteins using five orthogonal split inteins. (**a**) SDS-PAGE gel to show one pot assembly of SasG3 proteins and purification. SasG3 fusion proteins are mixed in one reaction and left at room temperature overnight to allow for trans-splicing of the split inteins. A sample of Ni-NTA resin (R) is saved for analysis by SDS-PAGE. Protein assemblies are eluted (E) and then further purified via strep-tag. Elution from the strep resin results in only one protein band, seen in Strep (E). This corresponds to the final protein, named here as SasG18. (**b**) Schematic of the fusion proteins to be assembled in this reaction, with arrows depicting the trans-splicing reaction that occurs overnight, excising the split intein domains from the fusion proteins and resulting in one extended SasG protein with His tag at one end and strep-tag at the other.

The limitation of this approach lies in the weakest link of the chain, the final yield is only as good as the least efficient intein. This may not be problematic if all inteins behaved at close to 99% efficiency, yet I was not able to select 5 orthogonal inteins that could reach those standards. This problem may not be necessarily relevant for some applications, as total efficiency is not always important, and precursor proteins may be in high enough abundance to obtain substantial concentrations.

However, to overcome this limitation, I proposed a further approach: recursive solid-phase assembly (Figure 4.13). This approach uses only two inteins, which enables the total splicing efficiency to be raised considerably. In total, four protein units are required to assemble proteins comprising 6 protein units: in this case, to make SasG18 as with previous method demonstrated in Figure 4.12. The ‘starting’ SasG3 fusion protein (SasG3_8, Figure 4.13a) contains a His-tag, and this is used to anchor the assembly to the Ni-NTA resin. Lysate containing SasG3 protein fused with a corresponding split intein half is added to the resin and allowed to react with the tethered SasG3 unit, enabling the tethered protein unit to ‘grow’ by trans splicing, thus becoming a SasG6 protein. This process is repeated step-wise with SasG3 fusions containing orthogonal split inteins, increasing the size of the tethered SasG protein one SasG3 unit at a time. With high enough efficiency, assembly of these units is technically infinite. I began by testing this approach with five cycles of assembly,

creating SasG18 (Figure 4.13), and then pushed this further by increasing the number of cycles completed to nine. This allowed for the synthesis of SasG30 (Figure 4.14).

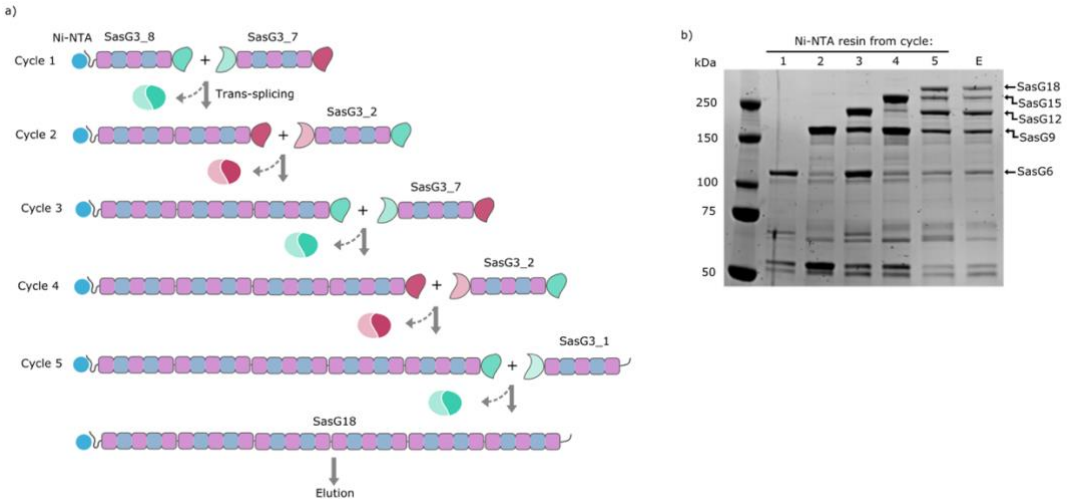


Figure 4.13 Recursive solid-phase assembly of SasG3 proteins using only two split inteins: five cycles creating a SasG18 protein. (**a**) Schematic of fusion proteins assembled in a step-wise manner on the Ni-NTA resin. First, the starting fusion protein SasG_8 is bound to the Ni-NTA resin via the His-tag. This enables the entire assembly to be tethered to a solid-phase material and for purification of the final construct from the heterologous mixture of proteins. In cycle 1, lysate containing SasG3_7 is added to the resin and incubated for one hour at room temperature, allowing the trans-splicing of protein domains on the resin via Nrdj-1 split intein. The resin is then washed, and any unbound proteins are removed, along with excised intein segments from the assembly. In cycle 2, the same principle is applied, but here using lysate containing SasG3_2, which allows for extension of the SasG protein via the split intein gp41-1. These steps are repeated for a total of five cycles, which results in a protein termed here as SasG18. This is eluted from the Ni-NTA resin. (**b**) SDS-PAGE gel analysis of samples taken from resin after each incubation cycle, and following washing of non-specific proteins. Proteins shown in these samples represent those tethered to the Ni-NTA resin.

Assembly of extended SasG proteins via both five (Figure 4.13) and nine (Figure 4.14) cycles were successful. Elongation of the protein was achieved at each assembly cycle in a controlled manner. However, the main difficulty with this approach is the inability to separate the longest protein assembly from the intermediate assemblies. If splicing is not 100% efficient, there will always be partial assemblies on the column, which will then be available for reaction with the ‘end’ SasG3 unit, SasG3_1. Therefore, this approach is beneficial as it uses less protein units and it has potential for much longer assemblies, but the best application for this method would be where

a heterogeneous mixture of assemblies is not problematic. As seen in Figure 4.13(b) and 4.14(b), the different assemblies are themselves pure from the rest of the cell lysate proteins, but they are not separated independently.

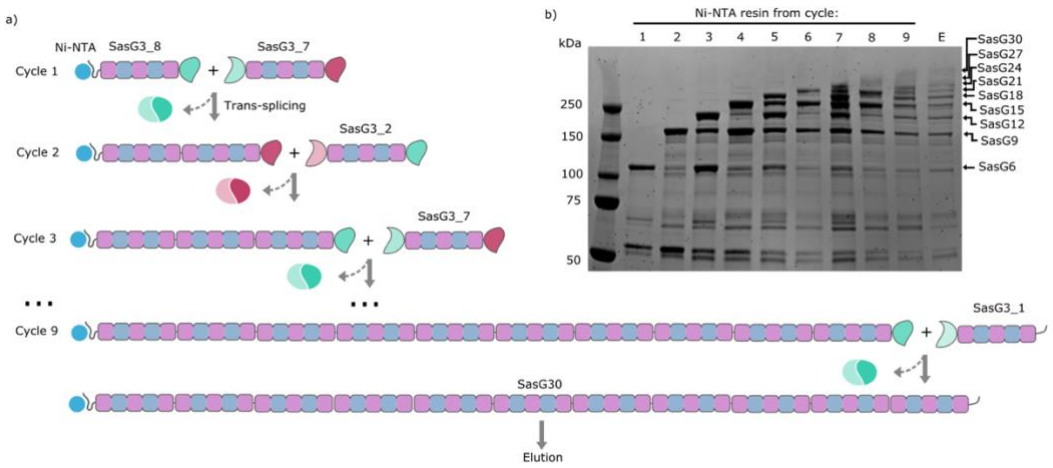


Figure 4.14 Recursive solid-phase assembly to create SasG30 via nine cycles of protein addition. (**a**) Schematic to show sequential protein addition of SasG3 units to create SasG30. Following the same principles set out in Figure 4.13, two SasG_3 fusion proteins are added sequentially to one SasG3 protein bound to Ni-NTA resin. This time, the cycles of SasG unit addition are continued for nine cycles, creating a final protein of SasG30. This is then eluted from the Ni-NTA resin (E). (**b**) SDS-PAGE gel of samples taken from Ni-NTA resin after washing in each cycle. Proteins shown in these samples represent those tethered to the Ni-NTA resin. Therefore, bands in each cycle should correspond to the stepwise growth of the SasG protein assembling on the resin.

Together, the design and execution of the two methods presented here provide approaches for simple and fast assembly of protein units directly from lysates under mild conditions. The vast potential for the method is demonstrated in the assembly of 10 protein units to create an elongated repetitive protein otherwise difficult to express using heterologous expression systems.

4.3 Discussion

The work presented here shows characterisation of an expanded split intein library in vitro. We develop a split FP platform for the fast and simple validation and characterisation of split inteins in vitro. In comparison with previous methods employed for split intein characterisation, primarily SDS-PAGE analysis and Western blot, this platform enables higher throughput and more direct analysis of intein activity. Using this platform, we were able to screen functionality of hundreds of samples per day, a volume of samples not feasible by traditional methods.

The presented platform enables time course analysis of 24 split inteins in parallel, with reconstitution of red fluorescence indicating speed of splicing. We validate accuracy of the platform by Western blot analysis of samples for assembled mCherry, and find that the platform is accurate in its reporting of splicing activity between pairs of split inteins. We select 12 of the best performing split inteins for further analysis and examination of orthogonality. Non-specific association of the split intein halves can cause reconstitution of mCherry signal in some cases, although by Western blot we see that in most cases, trans splicing of the inteins has not occurred. The exception is with PhoRadA, an intein capable of trans splicing when split in three different positions, yet we find these three different versions are not orthogonal to each other. This indicates a possible limitation of the system, as occasionally further validation of orthogonal splicing activity between non-paired split inteins may be required.

We find 10 out of the 12 split inteins to be highly orthogonal in vitro. In comparison with the most popular protein ligation tool currently used in synthetic biology, SpyTag/SpyCatcher, which only has validated orthogonality with one other tool: SnoopTag/SnoopCatcher (Veggiani et al., 2016), this is a dramatically increased library size of orthogonal tools. As all split inteins are characterised under the same conditions, this provides the synthetic biology community with a toolbox of protein ligation tools ready for implementation in tandem.

One advantage that SpyTag/SpyCatcher systems hold over the use of split inteins is the ability for ligation to occur in both oxidising and reducing conditions. Split inteins are redox-dependent, due to the catalytic Cysteine residue, which is required to be in reducing conditions for trans splicing to occur (Aranko et al., 2014). While one cysteine-less, and therefore, redox insensitive split intein has been discovered (Bhagawati et al., 2019), the vast majority of split inteins require addition of a reducing

agent for efficient in vitro trans splicing. Although this factor should not be limiting for most biotechnology applications, it could be problematic when the POI to be ligated is sensitive to reducing agents, or if the splicing reaction is required to take place in an oxidising environment in vivo, such as the bacterial periplasmic space. Therefore, the expansion of the orthogonal split intein library to include redox-insensitive split inteins would further increase the possible applications for split inteins.

When selecting split inteins for specific applications, one must also consider the junction sequence that will be inserted into the final protein product. The sequence inserted with split inteins is far smaller than with other protein ligation tools (6 AA, compared to ~100 AA with SpyTag/SpyCatcher (Zakeri et al., 2012)), however, addition of any residues can sometimes perturb protein function or morphology. We hope that the presentation of this extended library of inteins, each with different preferred extein sequences, and thus junction sequences, will provide a compatible option for ligation of any two proteins according to the specific needs of each application. A recently published paper (Ho et al., 2021), provides an impressive approach to determine the best insertion sites for split inteins within any POI. This approach could be used in parallel with the methods and information described in this chapter, to not only find the best possible split intein for the desired application, but also the best possible location for it to be inserted within the protein of interest.

To demonstrate the potential applications enabled by this library, I employ split inteins to ligate repetitive protein SasG in vitro, constructing proteins unattainable by recombinant protein expression.

Initially, I test splicing of the longest native SasG protein, SasG7, using the gp41-1 and gp41-8 split inteins. This is successful, yet proves difficult to purify, likely due to the unusual biophysical properties of the SasG protein (Gruszka et al., 2015). I include these results in this chapter to demonstrate the difficulties associated with working with extremely long proteins. This work was taking place in tandem with the split intein library characterisation. Therefore, when a rethink of the method was required, the tools characterised in the split intein library were vital in the evolution of the method.

I apply the five best performing, orthogonal, split inteins to the application of SasG protein assembly. Two methods for assembly are presented here, capable of ligating six SasG protein units via either two or five split inteins. Five split inteins are used in tandem to ligate six SasG protein units in one reaction of lysates overnight, followed

by purification of final protein assembly away from all intermediates. Using just two split inteins, I present a method for solid-phase recursive assembly of the SasG3 protein units. This method has potential to ligate as many protein units as required, simply by increasing the number of assembly cycles, demonstrated here in the assembly of 10 SasG protein units by 9 cycles of assembly. Importantly, the methods described do not require prior purification of the protein units, enabling simple and fast assembly of the desired protein following expression of protein units in cells. Additionally, as this method is only tested here with the SasG protein, use of the method to assemble other POI could enable higher splicing efficiency, and greater ease for separation of final productions in purification. As seen in Chapter 4.2.2., the purification of SasG proteins has consistently been problematic and therefore could be hindering the demonstration of the method here.

Assembly of long proteins by split inteins for biomaterials purposes has been recently demonstrated, for spider silk (Bowen et al., 2018) and titin proteins (Bowen et al., 2021). These examples demonstrate further the capabilities of split inteins and their importance and relevance in biomaterial synthesis. However in these examples, just one split intein is used, and the protein units are allowed to ligate end-to-end infinitely. While this approach is appropriate in these cases, as fibre length is required to be as long as possible, the methods I demonstrate in this chapter present a more controllable way to extend protein length with the use of multiple orthogonal split inteins in tandem. A recent publication demonstrated assembly of SasG protein units using coiled-coils for ligation (Jasaitis et al., 2020). This approach differs from the protocols presented in this chapter as the coiled-coil interactions are reversible, and as the protein units are reacted in an uncontrolled manner for the production of long protein nanofibers. Nevertheless, it emphasises the interest in SasG proteins for biomaterials purposes, and elongated proteins in the literature.

Selection of the most appropriate method for protein ligation in vitro must be considered on a case by case basis, depending on the desired application of the final assembled protein. The advantages of the methods I present here are primarily the simplicity and almost scarless nature of the protein assembly. Using the recursive solid-phase assembly method, there is no limit to the number of protein units assembled, however owing to incomplete protein splicing, the final elution of proteins consists of intermediate assemblies alongside the final assembly. For complete purification of the final assembly, the one-pot assembly method can be used, but here

the final yields are much lower due to the splicing ability of the least efficient intein. Use of SnoopTag/Catcher and SpyTag/Catcher sequentially for solid-phase assembly of protein units has been demonstrated (Veggiani et al., 2016), and with higher ligation efficiencies, fewer intermediate assemblies are present in the final elution. However, the use of these tools comes with the insertion of relatively large protein domains at each ligation site. Where the final protein sequence, and subsequent protein morphology, is important (e.g. for biomaterials purposes), this method would not be appropriate. Therefore, each of the two methods described here, in comparison with others from the literature, brings different advantages and disadvantages: the relevance of which are dependent on final application of the protein assembly.

In conclusion, in this chapter I present an expanded library of characterised split inteins, and demonstrate their potential as tools for in vitro protein ligation, relevant for protein engineering and biomaterial production applications.

Chapter 5 Materials and Methods

5.1 Materials

5.1.1 Growth media

Cells were routinely grown using media listed in Table 5.1. All media was sterilised by autoclaving before use.

Media	Purpose	Ingredients
Lysogeny Broth	Bacterial growth in liquid culture.	10 g/L peptone, 5 g/L NaCl, 5 g/L yeast extract
LB Agar	Bacterial growth on solid media.	10 g/L peptone, 5 g/L NaCl, 5 g/L yeast extract, 16 g/L agar
Terrific Broth	Bacterial growth in liquid culture.	24 g/L yeast extract, 20 g/L tryptone, 4 mL/L glycerol
2xYPTG	Cell growth for cell-free lysate production.	16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L KH_2PO_4 3 g/L K_2HPO_4. 18 g/L glucose was filter sterile and added after autoclaving of other buffer components.

Table 5.1 Composition of media used for cell growth.

5.1.2 Antibiotics

Antibiotics used for bacterial selection are listed in Table 5.2. Stock solutions were prepared and passed through a 0.22 µm filter before storage at -20°C.

Antibiotic	Solvent	Stock	Working
Kanamycin	ddH_2O	50 mg/mL	50 µg/mL
Ampicillin	ddH_2O	100 mg/mL	100 µg/mL
Chloramphenicol	Ethanol	15 mg/mL	15 µg/mL
Tetracycline	ddH_2O	10 mg/mL	10 µg/mL

Table 5.2 Antibiotics used for bacterial selection.

5.1.3 Bacterial strains

E. coli strains were selected for use depending on desired product; DNA, protein or bacteriophage. Details of all strains used are used listed in Table 5.3.

Bacterial strain	Purpose	Genotype	Source
TOP10	General DNA cloning	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lacX*74 *nup*G *recA*1 *araD*139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rps*L(Str^R) *end*A1 λ$^-$	Lab stock
BL21 Gold(DE3)	Protein expression	F$^-$ *omp*T *hsdS*$_B$($r_B^-m_B^-$) *dcm* (Tet^R) *gal* λ(DE3) *endA* Hte	Lab stock
Rosetta-gami 2	Protein expression	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F′[lac$^+$ lacI^q pro] gor522::Tn10 trxB pRARE2 (Cam^R, Str^R, Tet^R)	Professor Chris French (University of Edinburgh)

Origami B(DE3)pLysS	Protein expression	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm lacY1 ahpC* (DE3) *gor522*:: Tn*10 trxB* pLysS (Cam^R, Kan^R, Tet^R)	Merck (70839)
S2188	Bacteriophage production	F^- $lacI^Q$ ΔlamB endA hsdR17 supE44 thi1 relA1 gyrA96 ΔfimB-H::kan	Dr Stanley Brown (University of Copenhagen)

Table 5.3 Details of bacterial cell strains used in this study.

5.1.4 Buffers

Buffers were used for a range of biochemical techniques, and the recipe for all are listed in Table 5.4. When sterile buffers were required, they were either autoclaved or passed through a 0.22 µm filter.

Name	Ingredients	Purpose	Source
Splicing Buffer A	50 mM Tris-HCl (pH 7) 300 mM NaCl 1 mM EDTA 10% glycerol 2 mM DTT	For trans-splicing of purified split inteins (Chapter 4.2.2)	(Carvajal-Vallejos et al., 2012)
Splicing Buffer B	100 mM Tris-HCl 5 mM DTT 1 mM EDTA	For resuspension of cell pellets and direct trans-splicing of inteins (Chapter 4.2.2)	(Cheriyan et al., 2013)
SasG SEC buffer	200 mM NaCl 1 mM EDTA 20 mM Tris-HCl (pH 7.5)	For separation of SasG proteins in SEC (Chapter 4.2.2)	(Gruszka et al., 2015)
Lysis Buffer for His purification	50 mM NaH_2PO_4 300 mM NaCl 10 mM imidazole (optional)	Purification of proteins via His-tag	QIAexpressionist manual
Wash buffer for His purification	50 mM NaH_2PO_4 300 mM NaCl 20 mM imidazole	Purification of proteins via His-tag	QIAexpressionist manual
Elution buffer for His purification	50 mM NaH_2PO_4 300 mM NaCl 250 mM imidazole	Purification of proteins via His-tag	QIAexpressionist manual
Buffer W for Strep purification	100 mM Tris-HCl (pH 8) 150 mM NaCl 1mM EDTA	Purification of proteins via Strep-tag.	Strep-Tactin purification manual
Buffer E for Strep purification	100 mM Tris-HCl 150 mM NaCl 1 mM EDTA 2.5 mM desthiobiotin	Purification of proteins via Strep-tag	Strep Tactin purification manual
Equilibration/Wash Buffer for GST purification	50 Mm Tris-HCl 150 mM NaCl pH 8.0	Purification of proteins via GST-tag.	GST Agarose manual

Elution Buffer for GST purification	50 mM Tris 150 Mm NaCl 10 mM reduced glutathione pH 8.0	Purification of proteins via GST-tag	GST Agarose manual
Regeneration buffer #1 for GST resin	0.1 M Tris 0.5 M NaCl 0.1% SDS pH 8.5	Purification of proteins via GST-tag	GST Agarose manual
Regeneration buffer #2 for GST resin	0.1 M Sodium acetate 0.5 M NaCl 0.1% SDS pH 4.5	Purification of proteins via GST-tag	GST Agarose manual
BslA storage buffer	25 mM phosphate buffer pH 7.0	Storage of purified BslA	(Morris et al., 2017)
Buffer A for His purification under denaturing conditions	100 mM NaH_2PO_4 10 mM Tris·Cl 6 M GuHCl Adjust pH to 8.0 using NaOH immediately, do not autoclave.	Purification of proteins via His-tag under denaturing conditions	QIAexpressionist manual
Buffer B for His purification under denaturing conditions	100 mM NaH_2PO_4 10 mM Tris·Cl 8 M urea Adjust pH to 8.0 using NaOH immediately, do not autoclave.	Purification of proteins via His-tag under denaturing conditions	QIAexpressionist manual
Buffer C for His purification under denaturing conditions	100 mM NaH_2PO_4 10 mM Tris·Cl 8 M urea Adjust pH to 6.3 using HCl immediately, do not autoclave.	Purification of proteins via His-tag under denaturing conditions	QIAexpressionist manual
Buffer D for His purification under denaturing conditions	100 mM NaH_2PO_4 10 mM Tris·Cl 8 M urea Adjust pH to 5.9 using HCl immediately, do not autoclave.	Purification of proteins via His-tag under denaturing conditions	QIAexpressionist manual
Buffer E for His purification under denaturing conditions	100 mM NaH_2PO_4 10 mM Tris·Cl 8 M urea Adjust pH to 5.9 using HCl immediately, do not autoclave.	Purification of proteins via His-tag under denaturing conditions	QIAexpressionist manual
Buffer A for CFPS	10 mM tris acetate (pH 8.2) 14 mM magnesium glutamate	Cell-free reactions	(Kwon & Jewett, 2015)

	60 mM potassium glutamate		
Protease cleavage buffer	50 mM Tris-HCl (pH 7.0) 150 mM NaCl 1 mM EDTA 1 mM DTT	For cleavage of GST tag via PreScission Protease	PreScission Protease manual
FLIPPi sensor characterisation buffer	20 mM Tris-HCl (pH 7.0)	For characterisation of FLIPPi-SpC sensors	(Gu et al., 2006)
NAD+ sensor characterisation buffer	100 mM Tris-HCl (pH 7.4) 150 mM NaCl 1 mM EDTA 0.5 mM DTT	For characterisation of NAD-SpC sensors	(Cambronne et al., 2016) The paper specifies 100 µM PMSF and 50% glycerol be used, but this was omitted for my experiments.

Table 5.4 Details of buffers used in this study.

5.2 *E. coli* general techniques

5.2.1 Standard overnight growth

E. coli cells were picked from a single colony on an LB agar plate into 5 mL LB in a universal tube with the appropriate antibiotic and grown overnight at 37°C with shaking (180 RPM).

5.2.2 Competent cell preparation

A single colony was inoculated into 5 mL LB in a universal tube and grown overnight at 37°C with shaking (180 rpm). The next day, 500 µL from the overnight culture was used to inoculate 200 mL LB in a 500 mL flask, and grown at 37°C with shaking until OD_{600} was between 0.3 – 0.4. At this point, cells were transferred into pre-frozen tubes and incubated on ice for 20 minutes before collection by centrifugation at 2,500 xg for 10 minutes at 4°C. Supernatant was discarded and cells gently resuspended with 40 mL pre-chilled 50 mM $CaCl_2$. Cells were centrifuged and resuspended as before, then incubated on ice for 30 minutes before centrifugation at 0°C for 6 minutes at 2,500 xg. Supernatant was discarded and the cell pellet gently resuspended in 8 mL pre-chilled 50 mM $CaCl_2$ 15% glycerol and incubated on ice for 2 hours. Following this, competent cells were separated in 100 µL aliquots into pre-frozen microtubes and immediately stored at -80°C until use.

5.2.3 Heat shock transformation

DNA was routinely transformed into *E. coli* cells by heat shock transformation. Competent cells were thawed for 10 minutes on ice. 1 µL DNA was mixed with 50 µL

cells by gentle aspiration and incubated on ice for 30 minutes. Cells were heat shocked using a dry heat block at 42°C for 1 minute before incubation on ice for 2 minutes. 500 µL LB was used to resuspend the cells, before they were allowed to recover at 37°C for 1 hour with shaking. 100 µL of recovered cells were spread evenly on an LB agar plate with the appropriate antibiotic and incubated at 37°C overnight.

5.3 DNA design and synthesis

5.3.1 DNA synthesis and sequencing

DNA fragments were synthesised as GeneArt Strings (Thermo Fisher) when required. Plasmid sequences were verified by DNA sequencing performed either by DNA Sequencing & Services (www.dnaseq.co.uk) or Source Bioscience (www.sourcebioscience.com).

5.3.2 Primer design

All primers used in this study are listed in Table 6.1. Primers were designed using tools available on online software Benchling against template DNA sequences. Annealing temperatures were calculated using the integrated annealing temperature calculator on Benchling, set to the Modified Breslauer (1986) algorithm compatible with Phusion polymerase.

5.3.2.1 Primers for sequencing

Typically, primers provided by the sequencing company were used, commonly T7F and T7R, the sequences of which can be found in Table 6.1. To sequence DNA outside of this region, primers were designed following these guidelines: 60°C annealing temperature to template DNA, GC clamp at 3’ end and total GC content close to 50%.

5.3.2.2 Primers for Gibson assembly

The following general principles were followed as closely as possible when designing primers to create compatible DNA fragments for Gibson assembly: 65°C annealing temperature to template, overlap annealing temperature to complementary primer = 50°C, GC clamp at 3’ end, and total GC content close to 50%.

5.4 DNA production, manipulation and purification from *E.coli*

5.4.1 PCR

DNA fragments were routinely amplified using Polymerase Chain Reaction (PCR). All PCR was completed using Applied Biosystems Verti™ Thermal Cycler or ProFlex PCR System.

5.4.1.1 Phusion PCR

General amplification of DNA for cloning purposes was completed using Phusion polymerase (NEB) according to the manufacturer's instructions. Purified template DNA was added to each reaction to a final concentration of 1 – 10 ng. Each reaction was completed in 50 µL and run in a thermocycler using the annealing temperatures corresponding to the primer pair used and extension time determined by length of transcript (30 seconds per kb).

5.4.1.2 Colony PCR

To verify successful insertion of DNA fragments into plasmids by cloning, colonies of *E. coli* cells were screened for successful incorporation. GoTaq® polymerase (Promega) was used according to manufacturer's instructions. The colony to be screened was picked using an inoculation loop and briefly touched into a tube containing the complete PCR mixture. The cells were lysed by a 5 minute incubation step at 98°C before continuation with recommended cycling conditions for PCR with GoTaq®, considering primer annealing temperature and transcript length.

5.4.2 Agarose gel electrophoresis

DNA was separated by size via agarose gel electrophoresis. If the DNA to be visualised was smaller than 1 kb in size, 1% agarose (w/v) gel was used, and larger than 1 kb DNA fragments were visualised on a 0.8% agarose gel. TBE (Tris-Borate-EDTA) buffer was combined with agarose powder at the desired ratio and microwaved until all powder had dissolved in solution. The flask was cooled until comfortable to touch by hand and SBYR® Safe DNA Gel Stain mixed thoroughly throughout solution to recommended final concentration. The solution was poured into a gel cast with a gel comb in place and left to cool for at least 40 minutes, until set. DNA samples were prepared by mixing with loading dye in an appropriate volume. Hyperladder™ was loaded to one well of each gel as a molecular weight marker. Once loaded with samples, gels were run in 1x TBE buffer at 120 mV for 30 minutes, or until sufficient

separation of DNA had occurred. DNA bands on gels were visualised using the Bio-Rad Gel Doc XR+ system with filter 1 and SYBR Safe mode.

5.4.3 Gibson assembly

All cloning was completed using standard Gibson Assembly® protocols. DNA fragments were obtained by PCR (5.4.1.1), purified (5.4.4), and mixed with Gibson Master Mix (NEB) according to described protocols. After incubation at 50°C for 1 hour, the solution was mixed with Dpn1 (New England Biolabs) at 37°C to digest any original template plasmid, either for 1 hour or overnight. 1 µL of this mixture was used to transform TOP10 cells (5.2.3) and plated on an LB agar plate with the appropriate antibiotic. Colonies were screened by colony PCR (5.4.1.2) for incorporation of desired DNA fragment and promising colonies were cultured overnight (5.2.1), DNA extracted the next day (5.4.4), and pure DNA solutions sent to be sequenced (5.3.1).

5.4.4 DNA purification and quantification

Plasmids replicated in *E. coli* cells were purified following standard protocols described by Qiagen using the QIAprep® Spin Miniprep Kit. Linear DNA was purified from PCR mixture or agarose gels using the Promega Wizard® SV Gel and PCR Clean-up System following described protocols. All purified DNA was eluted from the column using ddH_2O. When purifying a medium-to-low copy number plasmid, or when looking to obtain a highly concentrated stock of DNA, elution from the column was completed using 25 µL ddH_2O. Purified DNA solutions were quantified using a DeNovix DS-11+ spectrophotometer, before storage of DNA solutions at -20°C.

5.5 Protein expression and purification

Unless otherwise stated in 5.5.12, proteins were produced and purified following the relevant standard protocols listed here.

5.5.1 Standard growth and expression protocol

E. coli cells harbouring plasmids encoding for POI were routinely cultured and expression induced to produce protein for use. Overnight cultures were diluted 100-fold into LB containing the appropriate antibiotic and grown at 37 °C with shaking (180 rpm) until OD_{600} = 0.6 – 0.8. Protein expression was induced with 1 mM IPTG and growth continued for a further 20 hours at 20 °C with shaking (180 rpm). Cells were collected by centrifugation at 6,000 xg for 10 minutes and pellets stored at -20 °C until needed.

5.5.2 Cell lysis and clarification

5.5.2.1 Under native conditions

Cells were resuspended in lysis buffer with cOmplete Protease Inhibitor Cocktail (4693132001, Sigma-Aldrich) at a standard ratio of 1:50 buffer to cell culture volume. Resuspended cells were sonicated (Soniprep 150, MSE) on ice for 30 seconds, followed by a 30 second rest period. This sonication-rest cycle was repeated until sufficient cell lysis was achieved. Clarified cell lysates were prepared by centrifugation at 10,000 xg for 30 minutes at 4°C, then stored on ice or at 4°C.

5.5.2.2 Under denaturing conditions

Cells were resuspended in 8M urea and gently vortexed until solution became transparent and lysis was complete. Extract was clarified by centrifugation at 10,000 xg for 30 minutes at room temperature.

5.5.3 Affinity chromatography

POI were initially separated from native *E. coli* proteins by affinity chromatography. Proteins were purified using either Ni-NTA, Strep-Tactin® or glutathione agarose, depending on the tag present: hexahistidine, strep, or GST respectively. For mVirD2-SC, purification occurred under denaturing conditions as the protein was insoluble. Buffer recipes relating to all purification methods can be found in Table 5.4.

5.5.3.1 Purification by hexahistidine tag

Clarified cell extract was incubated with pre-equilibrated Ni-NTA agarose (with lysis buffer) (30230, Qiagen) with gentle agitation overnight at 4 °C to allow for binding. Flow through was collected and the resin washed with lysis buffer three times with 5 CV (column volume) before elution six times with 1 CV elution buffer. Fractions of all steps were collected for analysis by SDS-PAGE. Pure protein elutions were pooled and dialysed against desired storage buffer according to methods in 5.5.9.

5.5.3.2 Purification by hexahistidine tag under denaturing conditions

Clarified cell extract prepared under denaturing conditions was incubated with pre-equilibrated Ni-NTA resin (using buffer C) at room temperature for one hour to allow for binding. Flow through was collected and the resin was washed twice with 8 CV buffer C. Protein was eluted with 1 CV buffer D four times, followed by elution with 1 CV buffer E four times. Fractions of all steps were collected for analysis by SDS-PAGE. Pure protein elutions were pooled and dialysed against desired storage buffer according to methods in 5.5.9.

5.5.3.3 Purification by strep tag

Clarified cell extract was incubated with pre-equilibrated Strep-Tactin® resin with gentle agitation overnight at 4 °C to allow for binding. Flow through was collected and the resin washed five times with 1 CV Buffer W. Protein was eluted six times with 0.5 CV buffer E. Fractions of all steps were collected for analysis by SDS-PAGE. Pure protein elutions were pooled and dialysed against desired storage buffer according to methods in 5.5.9.

5.5.3.4 Purification by GST tag

Clarified cell extract was incubated with pre-equilibrated GST agarose (ThermoFisher) (with lysis buffer) with gentle agitation overnight at 4 °C to allow for binding. Flow through was collected and the resin was washed three times with 10 CV lysis buffer. The column was capped and protease cleavage buffer added with PreScission Protease (GE Healthcare) and left overnight with gentle agitation at 4 °C to allow for cleavage of GST tag. Flow-through containing purified protein with no GST tag was collected. The resin was washed three times with 10 CV equiblibration/wash buffer, and bound proteins eluted three times with 2 CV elution buffer. Fractions of all steps were collected for analysis by SDS-PAGE. Pure protein elutions were pooled and dialysed against desired storage buffer according to methods in 5.5.9.

5.5.4 Small scale protein assembly on affinity column

5.5.4.1 One-pot assembly

One-pot assembly of SasG3 proteins was achieved by first mixing of cell lysates containing all six required protein units at equal concentrations to a final volume of 6 mL with 5 mM TCEP. This mixture was incubated at room temperature overnight with gentle agitation. The next day, 10 mM imidazole (56750, Sigma-Aldrich) was added to the solution and the mixture was incubated with 1 mL Ni-NTA resin (Qiagen). This solution was left overnight at 4°C, and the next day the assembled protein was purified according to manufacturer's protocols. Elutions of assembled protein were pooled, and concentrated by ammonium precipitation (5.5.8.2), the pellet of which was then resuspended in 150 µL Strep-Tactin® buffer W (2-1002-001, IBA GmbH). This sample was incubated in a 1 mL spin column (10351454, Thermo Fisher) with 50 µL Strep-Tactin® Sepharose (2-1201-010, IBA GmbH) for 1 hour at 4°C. Flow through was collected by gentle centrifugation of the column. The column was washed four times with 50 µL buffer W, and eluted six times with 12.5 µL Strep-Tactin buffer E (2-1002-001, IBA GmbH). All samples were retained for subsequent analysis by SDS-PAGE.

5.5.4.2 Recursive solid-phase assembly

Recursive solid-phase assembly of SasG3 proteins was achieved by sequential addition of lysates containing four fusion protein units. First, 200 µL of SasG3_8 lysate and 400 µL of SasG3_7 lysate were incubated with 250 µL Ni-NTA resin with TCEP (5 mM) and imidazole (10 mM) in a 1 mL spin column. This mixture was incubated overnight at room temperature with gentle agitation to allow for protein binding to the resin. Flow through from the column was collected by gentle centrifugation (900 xg for 10 seconds). The next day, the column was washed four times with 500 µL of wash buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl and 10 mM imidazole). 400 µL of the next protein lysate was then added to the column, together with 200 µL splicing buffer and TCEP to a final concentration of 5 mM. The column was incubated for 1 hour at room temperature with gentle agitation for each binding step of SasG-intein protein. The cycles of column washing, and next protein addition were repeated until the desired protein length was reached. 20 µL samples of resin were taken at the penultimate washing step in the protocol for analysis by SDS-PAGE. After the last washing step, the final protein assembly was eluted six times with 250 µL elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, and 200 mM imidazole) and analysed by SDS-PAGE.

5.5.5 High throughput protein extraction

To assess splicing capabilities of all split inteins in vitro, chimeric proteins were tested for expression and solubility in 5 mL of TB medium. Protein expression in *E. coli* TOP10 cells was induced by addition of 0.83 mM arabinose to the culture, and incubation overnight with shaking at 37°C (160 rpm). Protein expression from Origami cells was induced by addition of 1 mM IPTG after 3 hours of growth at 37°C (160 rpm), followed by growth at 20°C overnight (200 rpm). To assess protein expression, samples of the culture were taken directly and added to Laemmli buffer for SDS-PAGE analysis. To assess protein solubility, cleared lysates were prepared by centrifugation of cell culture (5 min at 4696 xg) and resuspended in TE buffer (100 mM Tris-HCl pH 8.0, 1 mM EDTA). 0.25 g of 425–600 µm glass beads (G8772, Sigma-Aldrich) were added to each tube, and cells were disrupted by vortexing at maximum speed for four cycles of 5 minutes, with 1 minute rest period on ice between each cycle. Clarified cell lysates were separated by centrifugation for 20 minutes at 17,000 xg at 4°C. Inclusion bodies were then solubilised by resuspension in 250 µL extraction buffer (100 mM Tris-HCl, 2M urea, pH 12.5), and incubated at room temperature with gentle agitation for 30 minutes. Samples were then diluted with TE buffer to a final

volume of 1 mL, correcting the pH to 8 with concentrated HCl. Samples were incubated at room temperature for 10 minutes and then centrifuged at 4°C for 20 minutes at 17,000 xg. Cleared lysate and solubilised inclusion body samples were then compared by SDS-PAGE to assess protein solubility. For in vitro splicing tests, all proteins were prepared under mild denaturing conditions. This was achieved by resuspension of cell pellets in 250 µL extraction buffer, together with 0.25 g of glass beads. Cells were disrupted by vortexing at maximum speed for four cycles of 5 min, with 1 min rest period on ice between each cycle. Samples were then incubated at room temperature with gentle agitation and each diluted to a final volume of 1 mL using dilution buffer (100 mM Tris-HCl, 1 mM EDTA, pH 9.0, cOmplete Protease Inhibitor Cocktail). The pH was adjusted to pH 9.0, and tubes were incubated for a further 10 minutes at room temperature. Cleared cell lysates were obtained by centrifugation (20 minutes at 17,000 xg at 4°C) and protein quantity was assessed by SDS-PAGE analysis for relative quantification. Western blots were also performed to validate presence of target proteins.

5.5.6 Size Exclusion Chromatography (SEC)

Further purification of proteins by SEC was used when affinity chromatography could not provide sufficient purity. All columns were provided by Edinburgh Protein Production Facility (EPPF) and run using an in-house ÄKTA Pure 25 system following standard protocol from column manufacturer. Protocols were created and data initially visualised using UNICORN 7.1 software.

5.5.7 Quantification of protein

Concentration of protein solution was determined by measuring absorbance at 280 nm using a DeNovix DS-11+ spectrophotometer and using the absorbance coefficient of each protein calculated using tools available on Benchling. When further verification of protein concentration was required, it was determined using either a Pierce BCA Protein Assay kit (10678484, Thermo Fisher) or Pierce Coomassie (Bradford) Protein Assay Kit (23200, ThermoFisher) according to manufacturer's instructions.

5.5.8 Concentration of protein

5.5.8.1 Centrifugal filters

Buffer exchange and simultaneous concentration of protein solutions was frequently completed using Amicon® Ultra Centrifugal Filters (Merck, various sizes). Columns

were selected based on the sample volume and POI size. As a general rule, pore size was chosen to be at least 1.5x smaller than the POI size in kDa. Columns were centrifugated (4,000 xg at 4°C) for at least 15 minutes, depending on protein concentration within the sample.

5.5.8.2 Ammonium sulfate precipitation

Ammonium sulfate was added gradually to the protein sample to reach 80% saturation (0.56 g mL^{-1} of protein solution) and after complete solubilization it was incubated at 4 °C overnight. The tube was then centrifuged at 4 °C for 10 min at 17,000 × *g* and the pellet was resuspended in buffer of choice.

5.5.9 Dialysis

Buffer exchange of protein solutions was completed by dialysis using SnakeSkin Dialysis Tubing (ThermoFisher, various sizes) according to manufacturer's instructions. Typically, samples were dialysed into 1 L buffer overnight at 4°C with stirring, followed by dialysis into 1 L fresh buffer twice more, for a period of 1 hour each at 4°C with stirring. For continuous buffer exchange, two peristaltic pumps were used. Samples were placed in 1 L buffer close to their composition and 4 L of the desired buffer was slowly filtered in while filtering out the same volume, with constant mixing at 4°C. This reduced the speed of buffer exchange inside the sample and prevented aggregation.

5.5.10 SDS-PAGE

Protein expression, purification and reactivity was frequently analysed by SDS-PAGE. Samples in 1x Laemmli buffer were boiled at 100°C for 10 minutes using a Thermocycler and spun down for 1 minute using a Mini-Fuge. Depending on expected protein concentration, 2 – 20 µL each sample was analysed on SDS-PAGE gels alongside Precision Plus Protein Dual Xtra Prestained Protein Standards as a molecular marker. Both pre-cast (Bio-Rad Any kD, 7.5%,12% or 4-20% TGX Stain-free protein gels) and homemade gels were used. Homemade gels were prepared at the desired percentage according to standard methods outlined by Amersham Biosciences Hoefer system protocols. SDS-PAGE gels were run at 200 mV for 30 minutes, or until protein samples had fully separated. Protein bands were visualised by staining InstantBlue® Coomassie Protein Stain according to manufacturer's instructions, and subsequent imaging with Bio-Rad Gel Doc XR+ system with white light filter and Coomassie mode.

5.5.11 Western blot

Protein samples were prepared and loaded onto an SDS-PAGE gel as described in 5.5.10, with Chameleon Duo Prestained Protein Ladder included in each gel as a molecular weight marker. SDS-PAGE gels were run at 200 mV for 30 minutes, or until protein samples had fully separated. Western blot procedure was followed as according to “Near-Infrared Western Blot Detection Document” (Doc. #988-13627). Antibodies recognising mCherry (Anti RFP-tag, pAb, Rabbit; A00682, GenScript) or the hexahistidine tag (THE His Tag Antibody, mAb, Mouse; A00186, GenScript) were used as primary antibodies at dilutions of 1:3000 and 1:5000, respectively. Antibodies recognising M13 P3 coat protein (Anti-M13, pAB, Rabbit; 68007-T32, SinoBiological) were used as primary antibodies at dilutions of 1:3000. IRDye 680RD Goat anti-Rabbit (925-68071, Li-cor) and IRDye 800CW Goat anti-Mouse (925-32210, Li-cor) were used as secondary antibodies at dilutions of 1:20,000. Blotted membranes were visualised using an Odyssey CLx Infrared Imaging System (Li-cor) and analysed by Image Studio Lite software (Li-cor).

5.5.12 Protein-specific considerations

5.5.12.1 BslA Wild-Type and fusion proteins

During purification of BslA proteins, solution containing concentrated and pure protein could often look opaque and white in colour. When this happened, the solutions were left to stand at room temperature to allow for re-solubilisation of BslA protein, which could take between 5 – 30 minutes, depending on concentration of solution. Freeze thaw cycles were always kept to a minimum with pure protein solutions, but even more stringently with BslA: pure protein was aliquoted into small volumes before freezing at -20°C. For experiments involving functionalisation of surfaces, BslA solutions were only thawed once.

5.5.12.2 SasG Wild-Type and fusion proteins

After purification of SasG proteins, buffer exchange or concentration was not possible by use of centrifugal units. Even very small pore size filter units (3 kDa) used to concentrate SasG proteins up to 100 kDa in size always resulted in protein flowing through filters and not retained. Therefore, buffer exchange was always performed on column, by dilution, or by ammonium precipitation (5.5.8.2).

5.5.12.3 mVirD2-SpyCatcher protein

Expression of mVirD2-SpyCatcher protein always resulted in the majority of protein present in the insoluble fraction of cell lysate, both in BL21 and RosettaGami2 *E. coli* cells with different induction techniques. Therefore, the protein was purified following methods outlined in 5.5.3.2 and buffer exchanged using continuous dialysis as outlined in 5.5.9. Due to the very high expression levels, aggregation often occurred when the protein solution was too concentrated. When this occurred in the purification process, the solution was spun at 4,000 rpm for 10 minutes at 4°C and supernatant transferred into a fresh tube to continue the purification process.

5.6 Cell free protein synthesis

5.6.1 Lysate production

2xYTPG was inoculated with 1/200 dilution of overnight cultures of cell strain of interest. Cultures were grown for 2 hours at 37°C with shaking (220 rpm), then induced with 0.4 mM IPTG and grown for a further 2 hours in the same conditions, before growth arrest by placing on ice. Cells were harvested by centrifugation at 10,000 xg for 10 minutes at 4°C, supernatant was discarded and cell pellets were resuspended with 80 mL Buffer A per 400 mL cells harvested. Cells were collected by centrifugation at 2,500 xg for 10 minutes at 4°C. The washing and cell harvesting process was repeated twice more, and cell pellets stored at -80°C for future downstream processing. Cell pellets were resuspended with 1 mL buffer A per 1g wet cell mass and homogenised by vortexing. 1.5 mL aliquots were taken from the total mixture and sonicated until total energy output of 556 J was achieved, while incubated in ice for 10 sec on, 10 sec off time. Lysate was clarified by centrifugation at 12,000 xg at 4°C for 10 minutes. Supernatant was removed and re-spun under the same conditions. Clarified supernatant was placed in a clean 1.5 mL microcentrifuge tube, and incubated at 37°C for 1.5 hours with shaking (220 rpm) in a run-off reaction. Following this, samples were centrifuged at 12,000 xg at 4°C for 10 minutes. Supernatant was removed and aliquoted into 25 ul samples in PCR tubes, before storage at -80°C.

5.6.2 Energy solution production

Energy solution was assembled from stock solutions of all constituents. Amino acid stock solution was made of final concentration 50 mM each of the following amino acids: Alanine, Arginine, Asparagine, Aspartate, Cysteine, Glutamate, Glutamine, Glycine, Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Proline,

Serine, Threonine, Tryptophan, Valine. Tyrosine was prepared separately, in an acidic solution (pH ~5.2) also at a final concentration of 50 mM. Stock batches of energy solution were prepared in volumes of 3 mL, with the following recipe:

Component	Final concentration (mM)
HEPES	200
Water	
ATP	6
GTP	6
CTP	3.6
UTP	3.6
tRNA	0.8
CoA	1.04
NAD	1.32
cAMP	3
Folinic acid	0.27
Spermidine	4
3-PGA	120
Amino acids	6
Tyrosine	3
PEG-8000	8%
Mg-glutamate	42
K-glutamate	400
DTT	1

Table 5.5 Energy solution recipe for cell-free protein synthesis reactions.

5.6.3 Preparation of DNA for cell-free protein synthesis reactions

DNA used in cell-free protein synthesis reactions required a higher level of purity than those for standard molecular biology procedures. DNA was extracted from cells using the PureLink HiPure Plasmid Maxiprep Kit (Invitrogen) following described protocols. DNA was further purified by DNA Clean & Concentrator Kit (Zymo Research) following described protocols. Final DNA concentration was measured by Nanodrop and DNA was stored at -20°C until required.

5.7 Plate reader assays

5.7.1 Split intein in vitro trans splicing assay

Quantification of N and C terminus halves of mCherry proteins by Western Blot allowed for equal amounts of each half of the protein to be combined. They were mixed in 2x TE buffer (pH 9.0) and aliquoted to microplates, the splicing reaction was then started by the addition of splicing buffer containing DTT (100 mM Tris-HCl pH 9.0, 1 mM EDTA, 200 mM NaCl, and 4 mM DTT). The plate was sealed with a breathable film and incubated inside the plate reader at room temperature. Fluorescence measurements were taken at 5 minute 24 second intervals across 20 hours, or for more extended experiments, up to 67 hours. Data was normalised by

subtraction of lowest value from each well. Orthogonality of the 12 most efficient split inteins was assessed via mixing of normalised amounts of all intein halves together in 2x TE buffer (pH 9.0), a total of 144 reactions. These were pipetted in triplicate into wells, and the splicing reaction started by the addition of DTT in splicing buffer (100 mM Tris-HCl pH 9.0, 1 mM EDTA, 200 mM NaCl, and 4 mM DTT). Fluorescence data was taken at 0 hours and 20 hours. Data was normalised by subtraction of the average value for time point zero between the triplicate values. Samples with unexpected high fluorescence (indicating cross-reactivity) were further assessed by Western blot.

5.7.2 Cell-free protein synthesis assays

CFPS reactions were prepared with a final volume of 10 µL in a 384 well microplate. Energy solution, lysate, DNA and buffer A were all combined in a 1:1:1:1 ratio in each well. When required, master mixes were prepared of DNA, energy solution and buffer before dispensing into appropriate wells. Lysate was always added to each well individually, and pipetted onto the side of the well wall. This provided a useful visual note of progress through the plate, but also allowed for the initiation of the cell-free reaction at the same time in every well. Once all combined, the plate was spun at 500 xg for 2 minutes at 4°C. This mixed all components simultaneously and removed any air bubbles from pipetting. After centrifugation, 10 µL BioRad chillout wax was added to each well to prevent evaporation of samples. Plates were sealed with non-breathable film and placed in plate reader.

5.7.3 Biosensor characterisation

Purified protein biosensors were combined with buffer and analyte of interest (Pi or NAD+) in 10 µL reactions within a 384-well microplate. Protein was diluted to give a final concentration in each reaction of 1 µM. Buffers for sensor characterisation were repeated from original methods of papers (detailed in Table 5.4). Samples were prepared in triplicate, by thorough mixing of protein biosensor with analyte in the appropriate buffer within the 384 well plate, and then sealed with a non-breathable seal. Fluorescent values were then read in the plate reader at the appropriate wavelengths (Table 5.6) to deduce fluorescence changes. Note: the wavelengths listed in this table are not necessarily the optimum wavelengths for each sensor, but rather the closest available options to me.

Protein name	Excitation wavelength	Emission wavelength
GFP	485 nm	520 nm

mCherry	584 nm	620 nm
NAD_Venus	Control: 430 nm	520 nm
	Variable: 485 nm	
FLIPPI	430 nm	Control: 480 nm
		Variable: 520 nm

Table 5.6 Wavelengths of light used to detect different fluorescent proteins.

5.8 Bacteriophage production, purification and ssDNA extraction

5.8.1 Strains and plasmid construction

S2188 cell strains were used for expression of bacteriophage. Two co-expressed plasmids were required for production of bacteriophage: 4423 encoded for all required phage proteins, while 4434 contained a single-stranded origin of replication and a minimal version of the phage genome to be packaged as the 'DNA backbone' to the phage. Therefore, to modify exterior proteins of phage, 4423 DNA was modified, and to change ssDNA sequence produced, 4434 DNA was modified, both by techniques described in 5.4.3.

5.8.2 Production of bacteriophage

S2188 cells harbouring the plasmids required for production of phage of interest were cultured in LB with the appropriate antibiotic overnight at 37°C with shaking (180 rpm). Cells were diluted 100 fold in LB the next day and left to grow overnight at 37°C with shaking (180 rpm).

5.8.3 Purification of bacteriophage

Bacteriophage were purified with different levels of stringency depending on the application. Bacteriophage for biomaterials purposes were purified by double PEG precipitation, followed by size exclusion chromatography and isoelectric precipitation. Bacteriophage for ssDNA extraction were purified by single PEG precipitation followed by isoelectric precipitation and subsequent DNA extraction. All purified bacteriophage were stored in buffer (20 mM Tris-HCl, 2 mM EDTA, pH 8.5) at 4°C for up to two weeks.

5.8.3.1 PEG precipitation

Supernatant containing bacteriophage was separated from *E. coli* cells by centrifugation at 5,635 xg at 4°C for 10 minutes. Supernatant was mixed with PEG buffer (12% PEG 4000, 750 mM NaCl) at a 2:1 supernatant to buffer ratio and incubated at 4°C for 2 hours. Precipitated phage was collected by centrifugation at 5,635 xg at 4°C for 10 minutes. Supernatant was discarded and phage pellets

resuspended in buffer (20 mM Tris-HCl, 2 mM EDTA, pH 8.5) in a volume 10 times less than the original culture volume. The resuspended pellet was incubated at 4°C for 2 hours. A second round of PEG precipitation was used to achieve higher purity if required. This also resulted in a more concentrated stock as the phage was concentrated 10 fold again.

5.8.3.2 Size exclusion chromatography

Following two rounds of PEG precipitation, the purified phage in buffer (final volume 1.5 mL) was filtered through a 0.22 µM filter and loaded onto a XK16/40 Sephacryl S-400 HR connected to the AKTA system. The bacteriophage were collected from the column in 2x TE buffer in 1 mL fractions. Samples were taken from fractions across main peak and run on SDS-PAGE, to analyse purity and relative concentration (PVIII protein is visible at 8 kDa). Fractions containing pure phage were combined and stored at 4°C.

5.8.3.3 Isoelectric precipitation

Phage were concentrated by isoelectric precipitation when required using protocol adapted from previous publications (Brown et al., 2015). 4 volumes of buffer (45 mM trisodium citrate, 55 mM citric acid) were mixed with one volume of phage sample and incubated at room temperature for 2 hours. The phage was recovered by centrifugation at 10,000 xg for 20 minutes at 4°C and supernatant removed. Phage pellet was then resuspended in desired volume of storage buffer (20 mM Tris-HCl, 2 mM EDTA, pH 8.5).

5.8.4 ssDNA purification and precipitation

5.8.4.1 Phenol extraction

An equal volume of phenol (pH 8.0) was added to purified phage in buffer and vortexed to mix. Following centrifugation at 17,000 xg for 1 minute, the aqueous top phase of liquid was separated into a new tube. 3 volumes of chloroform were added per volume of separated aqueous layer and mixed by vortexing, before centrifugation at 17,000 xg for 1 minute. The top phase and material at interface was removed by pipette and disposed of. The same volume of chloroform was added again, and subsequent steps followed. After removal of top phase again, the remaining solution was left in the fume hood with the lid open for one hour to allow for evaporation of residual chloroform.

5.8.4.2 Ethanol precipitation

For concentration of ssDNA from bacteriophage, ethanol precipitation was used. The original DNA solution was mixed well with 1/10 volume of buffer (3M sodium acetate, pH 5.2), followed by 2 volumes ice-cold EtOH. The solution was mixed and incubated at -80°C for 1 hour to allow for DNA precipitation. DNA was recovered by centrifugation at 17,000 xg at 0°C for 15 minutes. Supernatant was removed carefully by pipetting, and any remaining liquid was evaporated by incubation of tubes (with lids open and inverted on rack) at 37°C for 30 minutes. The remaining DNA pellet was resuspended in desired volume of liquid, which was most often either ddH_2O or TE storage buffer (20 mM Tris-HCl, 2 mM EDTA, pH 8.5).

5.8.5 TEM

Phage samples were analysed by Transmission Electron Microscopy (TEM) to check dimensions and for any signs of aggregation. Purified phage prepared by techniques outlined in 5.8.3 were diluted 100-fold in 2x TE storage buffer (20 mM Tris-HCl, 2 mM EDTA, pH 8.5). Samples were prepared for TEM by standard negative staining techniques and imaged with Electron Microscope JEOL TEM-1400. Resulting images were analysed and formatted using Fiji software.

5.9 Surface functionalisation

5.9.1 Preparation of glass slides

Ultra-clean, hydrophobic glass microscope slides were prepared following methods adapted from literature (Cras et al., 1999). 150 mL 12 M HCl was slowly added into 150 mL MeOH in a glass beaker and gently swirled to mix. Slides were added to the solution and incubated at room temperature for one hour. Slides were removed and added to a fresh beaker with 300 mL H_2O, gently swirled to mix and this wash step was repeated four more times. Slides were tapped dry (avoiding touching the surface) and dried completely by incubation at 90°C for two hours. The following steps were performed in the fume hood. Dichloromethylsilane was dissolved in Trichloroethylene to a final concentration of 0.05% and incubated with the dried glass slides for one hour. Glass was transferred to a methanol solution and incubated for 1 minute, slides were removed and then incubated again three more times to wash slides. Slides were rinsed twice in ddH_2O and dried before being packaged with lens paper in a sealed container. Slides were still sufficiently clean and hydrophobic for the experiments described after storage for up to 3 months.

5.9.2 Application of proteins, phage, DNA to BslA

Initially, CultureWell Reusable Gaskets (Grace Bio-Labs) were used for creation of wells on glass slides. The protocol was then further optimised by use of ProPlate Multi-Well Chambers (Grace Bio-Labs) for higher throughput. Up to four could be combined in one adaptor plate (ProPlate Multi-Array Slide System (Grace Bio-Labs)) for imaging in the plate reader.

5.9.2.1 BslA monolayer formation

Addition of the adaptors to the hydrophobic coverslips or microscope slides allowed for separation of samples in 10 µL wells. 10 µL BslA protein (22 µM) was incubated in each well at room temperature for 1 hour with high humidity to prevent evaporation. After monolayer formation, excess BslA was removed by pipette.

5.9.2.2 Protein or phage binding

10 µL protein or phage of interest was incubated with the well for 1 hour to allow the SpyTag/SpyCatcher reaction to occur to completion. Excess protein or phage was removed from each well by pipetting, adaptors were removed and the whole slide was washed with 50 mL UP water.

5.9.2.3 Peptide binding

If required, 10 µL AKER-TMR was then incubated with the surface for 1 hour at room temperature to allow for binding. To remove excess AKER-TMR we used gentle washing, where each well was washed slowly with 10 µL UP water, then gaskets were removed and any remaining solution removed via gentle blotting with lens paper.

5.9.2.4 DNA binding

DNA binding to BslA surfaces was facilitated by mVirD2-Spycatcher. 10 µL DNA was added to the well containing the BslA-mVirD2 monolayer and incubated at 37°C for one hour in the presence of 2 mM Mg^{2+}. Excess solution was removed from each well by pipetting, the adaptors were removed, and the whole slide was washed with 50 mL UP water. The slide was blotted dry with lens paper and subsequently imaged.

5.9.3 Signal detection and quantification

Slides were photographed for fluorescence, Cy3 or TMR signal using a FujiFilm scanner FLA-5000. Image files were exported and subsequently analysed via ImageJ to measure pixel intensity in each well. Well dimensions were measured and plotted as a circle, which was then placed in the centre of each well in the image to measure

signal, avoiding potential inconsistent background signal outside of each well. Data was analysed using GraphPad Prism v8.1.2. Slides functionalised using the ProPlate Multi-Array System could be imaged using the FLUOstar Omega plate reader, according to methods described in 5.7.

5.10 FRAP microscopy

Samples for FRAP experiments were prepared in triplicate according to methods in 5.9, on glass coverslips with CultureWell Reusable Gaskets. BslA-WT or BslA-SpT was used at a concentration of 11 µM, and GFP-SpC at a concentration of 5 µM. After incubation of proteins with the surface, the gasket was removed and the coverslip washed thoroughly with ddH_2O. Samples were then imaged using a custom-built TIRF microscope, using a 488 nm laser to excite and illuminate the GFP molecules throughout imaging. Photobleaching of one section of the surface was achieved using a beam size modifier on a 405 nm laser line at a high laser power density, enabling a smaller section in the centre of the slide to be photobleached. Experiments were typically set up to collect data for 2 seconds pre-bleach, bleach for 10 seconds, and then collect data for the 30 seconds following. For some experiments, data was collected post-bleach for 5 minutes. The microscope was set up for each experiment as following. Laser light (Colbalt Diode Laser Systems, Cobalt, Sweden) were aligned and directed parallel to the optical axis at the edge of a 1.49 NA TIRF objective (CFI Apochromat TIRF 60XC Oil, Nikon, Japan), mounted on an inverted Nikon TI2 microscope (Nikon, Japan). The fluorescence was separated from the returning TIR beam by a dichroic mirror (Di01-R405/488/561/635 (Semrock, Rochester, NY, USA)), passed through the appropriate filters (Semrock, NY, USA) and then recorded on an EMCCD camera (Delta Evolve 512, Photometrics, Tuscon, AZ, USA). Each pixel was 103 nm in length. Data was analysed using Fiji to track mean fluorescence over time from the centre of the photobleached area. This numerical data was visualised in GraphPad Prism v8.1.2.

5.11 Optical measurements of glass

Samples were prepared by methods outlined in 5.9 and analysed using a JASCO-V670 spectrophotometer for transmission and reflection of light between the wavelengths of 300 and 800 nm. Each sample was measured in triplicate, in a different location on the slide each time. For analysis of light transmission through samples, the FLH-741 adaptor was used. For analysis of reflection of light from surfaces, integrating sphere adaptor ISV-722 was used, and samples were backed

with black fabric to ensure collection of specular reflection only. Data was exported and analysed using GraphPad Prism v8.1.2.

Chapter 6 Final discussion and conclusions

Self-assembling protein materials are essential tools in the realisation of more sophisticated biomolecular structures and systems in synthetic biology (Grünberg & Serrano, 2010). Many have been adapted from the unique properties of proteins found in nature and repurposed into tools for engineering biology. In this thesis, I have tested several tools and combined them toward novel applications in synthetic biology.

I have presented new ways to combine self-assembling and useful protein structures. The combination of the BslA protein together with SpyTag/SpyCatcher technology or the mVirD2 protein provided the capability to create self-assembling, and functionalisable monolayers with protein or DNA (Chapter 2). I show this technology can be further adapted for applications in cell-free technology (Chapter 3). Additionally, I present the characterisation of a large library of split inteins as protein glues, and demonstrate their use in the assembly of multiple structurally rigid SasG proteins into long protein nanorods (Chapter 4).

In Chapter 2, the self-assembling properties of the natural protein BslA are applied to the construction of protein-based surfaces on glass. The novelty of this technology is further enabled by other unique protein tools: for covalent protein-protein and DNA-protein ligation onto the surface. This work builds on previous data published by Lynne Regan's lab, where assembly of BslA on air-water interfaces is shown (Williams et al., 2018). Here, the protocol is simplified further, with monolayer formation capable within one hour at room temperature, directly onto the hydrophobic glass surface and subsequently functionalised with FP or DNA. I believe that this simplification of the technique will make the approach even more accessible to users within biotechnology for creation of protein and DNA arrays on surfaces.

Following the success of initial proof of concept, I then tested two distinct routes for exploitation of this technology: firstly as a screen for protein/peptide interactions, and secondly as a basis for synthesising 3D structures for the conferment of optical properties to glass.

Using BslA as a tool to screen for protein-peptide interactions, the protein Survivin was functionalised to the BslA monolayer and a complementary peptide was tested for binding (Jeyaprakash et al., 2011). I successfully demonstrated that the peptide can recognise and bind to the complementary protein. This is an exciting first step

toward novel applications for BslA technology, but it also indicates that the proteins displayed on BslA surfaces are correctly folded which is a promising benefit of the BslA system.

To advance this technology further, I would like to complete experiments to more quantitatively understand how many functional molecules are capable of attachment to a BslA monolayer per unit area, and how this number compares with leading methods currently used (e.g. biotin/avidin). Alongside this, I think investigation into how tightly we can control spacing of the immobilised molecules via titration of more BslA-WT proteins would be really interesting, and is generally an underreported capability of surface immobilisation techniques (Meldal & Schoffelen, 2016).

The aim of creating 3D structures for conference of anti-reflective properties was not successfully achieved in my thesis. While I produced and characterised the building blocks of the system to have the correct dimensions and properties, ultimately the assembly had no effect on reflection or transmission through glass. However, I feel optimistic about a synthetic biology approach for nanopillar synthesis, as there have been fascinating advancements in the field towards a thorough biochemical understanding of such physical structures (Kryuchkov et al., 2020).

In Chapter 3, I explore a further application of BslA surfaces: using them to capture protein products from a heterogeneous cell-free solution. This is an application for surface immobilisation of proteins that has been utilised and optimised throughout the past decade (Manzano-Román & Fuentes, 2019). And while many aspects of current techniques has been improved, I think the high cost of materials for implementation of these techniques can serve as a barrier to users. The work presented in this chapter I hope demonstrates the simplicity and accessibility of using BslA proteins toward this goal.

However, the expression of proteins in CF from DNA covalently attached to the BslA monolayer could not be successfully demonstrated. Troubleshooting this problem would unlock a range of further exciting applications for this technology, and is something to be assessed following this work.

Additionally, I examined the use of protein-based biosensors for real-time tracking of CF analytes. It would be interesting to take this further and compare functionality of the sensors when they are surface bound compared to free in solution, and potentially

even pair this with expression from surface bound DNA, enabling in situ production of biosensors for in situ production and functionalisation of a surface for the monitoring of cell-free protein synthesis. This would be an innovative tool that could be used to assess real-time metabolomics in CF at a much cheaper cost than current methods. Current methods for analysis of CF metabolism generally involve specialist equipment (Garenne et al., 2019; Panthu et al., 2017), which therefore means the approach is not generally accessible for users of CF technology.

In Chapter 4, I have presented a characterised library of 10 split inteins, shown as functional and orthogonal in vitro. This is a tool that will be of great interest to the synthetic biology community, to enable the fast and orthogonal ligation of proteins together, with almost seamless assembly. I have applied this self-assembling tool to the construction of long protein nanorods. To further develop this technology, it could be interesting to apply the protein assembly methods to other proteins of interest, to create enzymatic pathways or other fibrous biomaterials. While this has already been shown with inteins previously (Bowen et al., 2018, 2021), the more in depth characterisation of the split inteins presented in this chapter could help to advance this technology even further by use of multiple orthogonal split inteins simultaneously. The long SasG nanorods I create could also be applied to biomaterials purposes, for example in the production and spinning of silk like fibres to create strong textiles, as demonstrated previously with split inteins (Bowen et al., 2018).

For the next challenges in the field of synthetic biology, such as creating artificial cells, or portable biosensors, self-assembling protein materials and structures will undoubtably play a crucial role (Gallup et al., 2021). In the realisation of these exciting applications for synthetic biology, there will be a role for both de novo design of proteins and the repurposing of proteins found in nature (Lange & Polizzi, 2021). In 1977, Francois Jacob referred to evolution as a tinkerer rather than an engineer, gradually modifying molecules for purposes without the foresight of their function, rather than building new molecules from scratch (Jacob, 1977). While the immense power of an engineering approach to protein synthetic biology, de novo protein design, is undisputed, until a total understanding of the complex biological systems is fulfilled, there will always be an important role for tinkering in protein synthetic biology, be that to repurpose natural protein tools, or to combine existing ones in new ways.

7 Appendix

7.1 Split inteins

Split intein	Successful trans-splicing in vivo?	High splicing rate in vitro?	Orthogonality in vitro?
M86	1, 2		
NpuSsp	1, 2, 3		
gp41-1	2	2	2
gp41-8	2	2	2
NrdJ-1	2	2	2
IMPDH-1	2	2	2
SspDnaX	2		
SspGyrB	2	2	
TerThyX	1		
TvoVMA	2	2	2
PhoRadA	1, 2	1, 2, 3	1, 2, 3
Cro-RIR1			
Cro-RPB2			
Cro-Top2			
CIV-RIR1			
CP-Thy1			
Ama-Ter			
BsuP-RIR1			
Cbp-RNR			
Ckl-Ter			
Cth-Ter	1, 2	2	
EP-Pri			
LLP-Pol			
LP-Hel			
MP-B- DnaB	2	2	2
MP-K-gp53			
MP-C-gp206			
MP-M-DnaB			
PP-Phi			
SaP-dpol	2		
SaP-Hel			
NrdA-2	2	2	2
Pfu-RIR1			
Mja-KlbA	1, 2		

Table 7.1 All inteins investigated in original study (Pinto et al, 2020), on which Chapter 4 of this thesis is based. This table shows which inteins performed best at each of the experimental phases: in vivo trans-splicing, in vitro trans-splicing, and orthogonal in vitro trans-splicing. The numbers refer to the split site at which each intein was tested. All were originally tested in 3 different positions, yet it can be seen that split site 2 was the most feasible split site for all inteins.

7.2 DNA oligonucleotides

DNA primers used in Chapter 4.2.1 and 4.2.3 are listed in the supplementary document of the publication (Pinto et al., 2020).

Label	Name	Sequence 5' – 3'	Purpose
P1	M13_Puc_R	AGCGGATAACAATTTCACACAGG	Sequencing
P3	1_SasG7_F	GGAGAAATATACTAGATGGCGCATATTGTAATGGTGGATGCTTACAAACCCACGAAAGGCGGCAGCGGCGGCAGCACCATCACCGAGCTGGAGAAGAAAG	SasG7_intein construct cloning
P4	1_SasG7_R	CCAGACAATAGCCGCTCGGACCATACTCGGTGAGCTCATTGATAG	SasG7_intein construct cloning
P5	Gp41-1_N_F	GTCCGAGCGGCTATTGTCTGGATCTGAAAACCCAGGTTCAGACA	SasG7_intein construct cloning
P6	Gp41-1_N_R	CCGCTACTAGTATTATTAATGATGATGATGATGGCCGCCTTCTTTAACATACAGGCACATACCTTCTTT	SasG7_intein construct cloning
P7	Gp41-1_C_F	GAGGAGAAATATACTAGATGCATCATCATCATCATCACATGCTGAAAAAAATCCTGAAAATCGAGG	SasG7_intein construct cloning
P8	Gp41-1_C_R	ATGGTGCTGCTGCTATTATGGGTCAGAATATCATTGGCATAGAA	SasG7_intein construct cloning
P9	2_SasG7_F	ATAATAGCAGCAGCACCATCACCGAGCTGGAGAAGAAAGTTGAA	SasG7_intein construct cloning
P10	2_SasG7_R	CCGCTACTAGTATTATTATTTTTCAAATTGAGGATGAGACCACGGACCATACTCGGTGAGCTCATTGATAGG	SasG7_intein construct cloning
P12	RBS30_1_SasG7_F	ATTAAAGAGGAGAAATACTAGATGGCGCATATTGTAATGGTGGATGCTTACAAACCCACGAAAGGCGGCAGCGGCGGCAGCACCATCACCGAGCTGGAGAAGAAAG	SasG7_intein construct cloning
P13	RBS30_GP41-1C_F	ATTAAAGAGGAGAAATACTAGATGCATCATCATCATCATCACATGCTGAAAAAAATCCTGAAAATCGAGGAA	SasG7_intein construct cloning
P14	pet28_RBS30_F	GTATTTCTCCTCTTTAATATTATTTCTAGAGGGGAATTGTTATCCGCT	SasG7_intein construct cloning
P15	pet28_R	TAAGGATCCGAATTCGAGCTCCGTCGA	SasG7_intein construct cloning
P16	pet_Gp41-1N_R	CTCGAATTCGGATCCTTAATGATGATGATGATGATGGCCGCCTTCTTTAACATACAGGCACATACCTTCT	SasG7_intein construct cloning
P19	pet_2_SasG7_R	CTCGAATTCGGATCCTTATTTTTCAAATTGAGGATGAGACCACGGACCATACTCGGTGAGCTCATTGATAGG	SasG7_intein construct cloning
P21	T7 prom (F)	TAATACGACTCACTATAGGG	Sequencing
P22	T7 term (R)	GCTAGTTATTGCTCAGCGG	Sequencing
P23	MID-SASG	AAAACCCGCTGACCGGCGAGATCATT	Sequencing
P113	Phagemid_seq_1	GAAGTCAGCCCCATACG	Phagemid sequencing
P114	Phagemid_seq_2	ACTCTCTAATCCTGACCTGTTGG	Phagemid sequencing
P115	Phagemid_seq_3	GTAACATGGAGCAGGTCG	Phagemid sequencing

P116	Phagemid_seq_4	AGGGTGGCGGTACTAAACC	Phagemid sequencing
P117	Phagemid_seq_5	CCTTTTGTCTTTAGCGCTGG	Phagemid sequencing
P118	Phagemid_seq_6	GATTGGTTTCTACATGCTCG	Phagemid sequencing
P119	Phagemid_seq_7	GCAATCAGGCGAATCC	Phagemid sequencing
P120	Phagemid_seq_8	CGATTGAGCGTCAAAATGTAGG	Phagemid sequencing
P121	Phagemid_seq_9	TTCGATCACTGGACCG	Phagemid sequencing
P122	Phagemid_seq_10	TATGTTCCGGATCTGCATC	Phagemid sequencing
P123	Phagemid_seq_11	GGATGAGCATTCATCAGG	Phagemid sequencing
P124	Phagemid_seq_12	TCACTGACTCGCTACGCTC	Phagemid sequencing
P125	Phagemid_seq_13	TAGCGGCGCATTAAGC	Phagemid sequencing
P126	Phagemid_seq_14	AAGCGTGGCGCTTTC	Phagemid sequencing
P127	Phagemid_seq_15	GCAACTTTATCCGCCTCC	Phagemid sequencing
P130	pSB4423_P3_F	GCTGAAACTGTTGAAAGTTGTTTAGC	P3_SpyCatcher cloning
P131	pSB4423_P3_R	GGAGTGAGAATAGAAAGGAACAACTAAAGG	P3_SpyCatcher cloning
P132	pSB4423_P3_SC_F	CCTTTCTATTCTCACTCCGCCATGGTTGATACCTTATCAGG	P3_SpyCatcher cloning
P133	pSB4423_P3_SC_R	CTTTCAACAGTTTCAGCGCTGCCGCCGCTGCCGCCAATATGAGCGTCACCTTTAGTTGCTTTG	P3_SpyCatcher cloning
P136	Phagemid_seq_16	GCATTCCACAGACAACCC	Phagemid sequencing
P137	Phagemid_seq_7_new	GCAAACAGGTGAATCTG	Phagemid sequencing
P141	pET28_stop_F	TAAGGATCCGAATTCGAGCTCC	SasG7_SpyCatcher cloning
P142	SpyCatcher_SasG7_F	AGAAATAATATTAAAGAGGAGAAATACTAGATGGCCATGGTTGATACCTTATCAGG	SasG7_SpyCatcher cloning
P143	SpyCatcher_SasG7_R	TGAACTACCACCACTACCACCAATATGAGCGTCACCTTTAGTTGCTT	SasG7_SpyCatcher cloning
P144	SasG_SpyCatcher_F	TGGTAGTGGTGGTAGTTCAACCATCACCGAGCTGGAGA	SasG7_SpyCatcher cloning
P145	Survivin_backbone_R	ATGGCATCCATGGCAGCCAGCT	Survivin_SpyCatcher cloning
P146	Survivin_Backbone_F	GACGCTCATATTTAAGCAGTCGGTGGCGG	Survivin_SpyCatcher cloning
P147	SpyCatcher_Survivin_R	GCTTAAATATGAGCGTCACCTTTAGTTGC	Survivin_SpyCatcher cloning
P148	SpyCatcher_Survivin_F	CCATGGATGCCATGGTTGATACCTTATCAGG	Survivin_SpyCatcher cloning
P149	SC_F	GCCATGGTTGATACCTTATCAGGT	mVirD_SpyCatcher cloning
P150	SC_VIRD2_R	TTTGCGTTAAATATGAGCGTCACCTTTAGTTGC	mVirD_SpyCatcher cloning
P151	VIRD2_F	CTCATATTTAACGCAAAAAACCCCGC	mVirD_SpyCatcher cloning
P152	VIRD2_R	GGTATCAACCATGGCGCTGCCGCCGCTGCCGCCCCATGGCTGCATCCGC	mVirD_SpyCatcher cloning
P155	VirD2_rec_Cy3	GCTCAAATTACAACGGTATATATCCTGCCAGTCAG(CY3)	Cy3 labelled oligo mVirD2 binding sequence

P165	pPROEX-HTa_virD_rec_2_F	GTACCATATATCCTGTCAATGGCCCGCATAATTTTGTTAAAATTCGC	To add mVirD recognition sequence to pPROEX-HTa plasmids
P166	pPROEX-HTa_virD_rec_2_R	ATTGACAGGATATATGGTACTGTCTGTGAAATACCGCACAGATGC	To add mVirD recognition sequence to pPROEX-HTa plasmids
P169	pPROEX_R	GTGATGGTGATGGTGATGCA	General cloning
P170	pPROEX_SnC_F	TCCGGTGGTAGCGGTATGAAGCCGCTGCGTG	General cloning
P173	Link_SpyCatcher_F	TCCGGTGGTAGCGGTGCCATGGTTGATACCTTATCAGGT	General cloning
P174	SpyCatcher_R	GGTGATGGTAGTACGATTAAATATGAGCGTCACCTTTAGTTGC	General cloning
P176	pPROEX_F	TAATCGTACTACCATCACCATCACCA	General cloning
P177	pPROEX-H_F	TAATCGTACTACCATCACCA	General cloning
P178	22_02_backbone_R	GGTGATGCATGGTCTGTTTCC	General cloning
P179	22_02_backbone_F	TAAGGATCCGGAATTCAAAGGC	General cloning
P187	FLIPPI_R	CTTGTACAGCTCGTCCATGCC	FLIPPi_SpyCatcher cloning
P189	pHRED_R	TCTAGAACCACCAGAGCCACC	pHRed_SnoopCatcher cloning
P190	pRSET_F	TGACTCGAGGGGCCCA	Sensor cloning
P191	pRSET_F_NAD	ACACAGCGCGTAGGTTGACTCGAGGGGCCCA	NAD_SpyCatcher cloning
P192	pRSET_R_His	AACCATGGCATGGTGATGGTGATGGTGATGT	Sensor cloning
P193	SnC_F_pHRed	GGCTCTGGTGGTTCTAGAATGAAGCCGCTGCGTG	pHRed_SnoopCatcher cloning
P194	SnC_R_PRSET	GGCCCCTCGAGTCATTTCGGCGGTATCGGTTC	pHRed_SnoopCatcher cloning
P195	SpC_F_His	CATCACCATGCCATGGTTGATACCTTATCAGGT	Sensor cloning
P196	SpC_Link_FLIPPI_F	GACGAGCTGTACAAGGGTGGCTCTGGTGGTTCTGCCATGGTTGATACCTTATCAGG	FLIPPi_SpyCatcher cloning
P197	SpC_pRSET_R	AGCCGGATCAAGCTTTTAAATATGAGCGTCACCTTTAGTTGC	FLIPPi_SpyCatcher cloning
P198	SpC_R_Linker_NAD	CGGGGCCTTTTCCATGCTGCCGCCGCTGCCGCCAATATGAGCGTCACCTTTAGTTGCT	NAD_SpyCatcher cloning
P199	SpC_Venus_link_F	CCGCCACAACATCGAGGGTGGCTCTGGTGGTTCTGCCATGGTTGATACCTTATCAGG	NAD_SpyCatcher cloning
P200	pRSET_FLIPPI_F	AAGCTTGATCCGGCTGCTAA	FLIPPi_SpyCatcher cloning
P201	PiBP_midway_F	GCCTAGAGCAGGTGATC	Sequencing
P202	NAD_midway_seq_F	CATTACGGAGAATCTCCG	Sequencing

Table 7.2 DNA oligonucleotides used in this study.

Plasmid no.	Protein	Size (kDa)
E4	ST_S7_GPN_H	106.1
E7	H_GP1C_S7_GP8N	108.7
E8	H_GP8C_S7_STP	100.1
E20	STP_S3_NRDN	50.4
E21	NRDC_S3_GP1N	51.9
E22	GP1C_S3_IMPN	53.2
E23	IMPC_S3_SGBN	55.2
E24	SGBC_S3_NRDN	54.3
E25	SGBC_S3_GP8N	52.2
E26	SGBC_S3_H	42.9
E27	GP8C_S3_H	43.1
E28	NRDC_S3_H	41.9
E29	GP1C_S3_NRDN	53.5
E30	NRDC_S3_H	42.6
E36	pIII	42.5
E38	pIII_SpyCatcher	55.0
E41	pmVirD2	24.6
E42	VirD_SC	37.3
E44	Survivin	18.8
E45	Survivin_SC	31.1
E53	mCherry_SnC	44.9
E54	GST-BslA-WT	42.9
E55	GST-BslA-SpyTag	44.8
E57	GST-BslA-SnoopTag	45.0
E59	GFP-SpyCatcher	43.1
E64	SasG7_SpC	117
E79	His_pHRed_SnC	40.4
E80	His_SpC_NAD_cpVenus	77.6
E81	His_ Phosphate_5uM_FRET_SpC	106
E82	His_ Phosphate_200uM_FRET_SpC	106
E83	His_ Phosphate_30mM_FRET_SpC	106

Table 7.3 Molecular weight in kDa of proteins used in this study. Values were estimated using the inbuilt tool from Benchling, calculating MW based on amino acid sequence of the proteins.

7.3 Plasmid sequences

Plasmids used in Chapter 4.2.1 are detailed in the supplementary document of the publication (Pinto et al., 2020).

No	Plasmid name	Purpose
E4	ST_S7_GPN_H	SasG7 three-way splicing
E7	H_GP1C_S7_GP8N	SasG7 three-way splicing
E8	H_GP8C_S7_STP	SasG7 three-way splicing
E20	STP_S3_NRDN	SasG3 multi-splicing
E21	NRDC_S3_GP1N	SasG3 multi-splicing
E22	GP1C_S3_IMPN	SasG3 multi-splicing
E23	IMPC_S3_SGBN	SasG3 multi-splicing
E24	SGBC_S3_NRDN	SasG3 multi-splicing

E25	SGBC_S3_GP8N	SasG3 multi-splicing
E26	SGBC_S3_H	SasG3 multi-splicing
E27	GP8C_S3_H	SasG3 multi-splicing
E28	NRDC_S3_H	SasG3 multi-splicing
E29	GP1C_S3_NRDN	SasG3 multi-splicing
E30	NRDC_S3_H	SasG3 multi-splicing
E35	pSB4434	Mini-M13 production (ssDNA template)
E36	pSB4423	Mini-M13 production (proteins for phage)
E38	4423_SpyCatcher	Mini-M13 with modified P3
E41	pmVirD2	mVirD2 experiments
E42	VirD_SC	mVirD2 experiments
E44	Survivin	Survivin experiments
E45	Survivin_SC	Survivin experiments
E53	VirD_mCherry_SnC	BslA surface binding
E54	BslA-WT	BslA experiments
E55	BslA-SpyTag	BslA experiments
E57	BslA-SnoopTag	BslA experiments
E59	GFP-SpyCatcher	BslA experiments
E64	SasG7_SpC	Single SasG7 protein with SpC
E79	His_pHRed_SnC	Cell-free sensors
E80	His_SpC_NAD_cpVenus	Cell-free sensors
E81	His_ Phosphate_5uM_FRET_SpC	Cell-free sensors
E82	His_ Phosphate_200uM_FRET_SpC	Cell-free sensors
E83	His_ Phosphate_30mM_FRET_SpC	Cell-free sensors

Table 7.4 Plasmids used in this study.

Plasmids listed here are further detailed in the following pages, where DNA sequence and expressed protein sequence from DNA is annotated with a matching colour scheme. 100 bp of DNA sequence is included upstream and downstream of protein coding sequence. The full sequence of every plasmid is available by opening the Benchling link provided for each entry, this folder of sequences is open-access. I commonly use flexible linkers between protein sequences, mainly a combination of Glycines and Serines, these are not individually labelled in the shading key. The shading scheme for annotation of DNA and protein sequences is as follows:

Protein of Interest
Intein
Purification tag
SpyTag/SpyCatcher
Cleavage site
Start codon
Stop codon

Freezer Box	ET Plasmids	**#**	4
Name	ST_S7_GPN_H	**Source**	Self
Resistance	Kanamycin	**Total plasmid size**	8163 bp
Parent Vector	pET28	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-3zhyFxEsrEWRag5AFRPI		

Description
SasG7 protein with SpyTag at N-terminus and Gp41-1N at C terminus for testing of SasG7 trans-splicing via split inteins.

Shading Key
SpyTag–SasG7–Gp41-1N-His

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGCGCATATTGTAATGG
TGGATGCTTACAAACCCACGAAAGGCGGCAGCGGCGGCAGCACCATCACCGAGCTGG
AGAAGAAAGTTGAAGAGATTCCTTTTAAAAAAGAACGTAAGTTCAACCCGGATCTGGCA
CCTGGTACCGAGAAGGTGACACGTGAGGGCCAGAAAGGCGAGAAGACAATCACAACA
CCTACCTTAAAAAATCCTTTAACAGGCGTGATTATCAGTAAAGGTGAACCGAAAGAGGA
GATTACCAAAGACCCGATCAACGAGCTGACAGAGTATGGCCCGGAAACCATCGCACCG
GGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGTTCCTG
GTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAG
CGTTACCAAATACGGTCCTGTTAAAGGCGACAGTATTGTGGAAAAAGAGGAAATCCCGT
TCGAAAAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGT
GAGGGTCAAAAAGGTGAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCG
GCGAGATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAACGA
ACTGACCGAATACGGCCCGGAGACCATTACCCCGGGTCACCGTGACGAGTTCGATCCT
AAGCTGCCTACAGGCGAAAAGGAAGAAGTGCCTGGTAAACCGGGCATCAAGAACCCTG
AAACCGGCGACGTGGTTCGTCCGCCGGTTGACAGTGTTACCAAGTACGGCCCGGTGAA
GGGCGATAGCATCGTTGAGAAGGAAGAAATCCCGTTTGAGAAAGAGCGCAAATTCAAT
CCGGATCTGGCCCCTGGTACCGAGAAGGTTACCCGCGAAGGCCAAAAGGGTGAAAAA
ACAATTACAACACCTACACTGAAAAATCCGCTGACCGGTGTTATTATTAGTAAGGGTGA
GCCGAAAGAGGAAATTACCAAAGATCCGATCAACGAATTAACCGAATACGGTCCGGAAA
CAATCACCCCGGGCCATCGCGATGAATTTGATCCTAAATTACCGACAGGCGAGAAAGA
GGAAGTGCCGGGCAAGCCTGGTATTAAGAATCCGGAGACAGGTGATGTTGTGCGCCC
GCCGGTTGATAGCGTGACAAAGTATGGCCCTGTGAAGGGCGACAGCATCGTGGAAAAA
GAGGAGATCCCGTTCAAAAAGGAGCGCAAATTTAATCCGGACTTAGCACCGGGTACAG
AGAAAGTTACCCGTGAAGGTCAAAAAGGCGAGAAGACCATTACCACCCCTACCTTAAAG
AACCCTCTGACAGGTGAAATCATCAGTAAAGGCGAAAGCAAAGAAGAGATCACCAAAGA
TCCGATCAATGAGTTAACAGAATATGGTCCGGAGACAATCACACCGGGCCACCGCGAT
GAGTTTGACCCGAAGCTGCCGACAGGTGAAAAAGAAGAGGTGCCTGGCAAACCGGGT
ATCAAGAACCCGGAGACCGGTGATGTTGTTCGCCCTCCGGTGGATAGTGTGACAAAAT
ACGGTCCTGTGAAGGGTGATAGTATTGTTGAAAAGGAAGAAATTCCGTTTGAAAAAGAG
CGTAAGTTCAATCCTGATTTAGCCCCTGGCACAGAGAAGGTTACACGCGAAGGTCAGA
AAGGTGAGAAAACAATCACCACCCCGACCCTGAAGAATCCTTTAACCGGCGAAATCATC
AGTAAGGGCGAGAGTAAGGAAGAAATTACAAAAGACCCTATTAATGAATTAACAGAGTA
TGGTCCTGAAACAATCACACCTGGCCACCGTGACGAATTCGATCCGAAACTGCCTACC
GGTGAAAAAGAAGAGGTTCCTGGCAAGCCTGGCATTAAGAACCCGGAAACCGGCGATG
TGGTGCGTCCGCCTGTGGACAGTGTTACAAAATATGGCCCGGTGAAAGGCGATAGCAT
TGTGGAAAAAGAGGAGATTCCTTTTAAGAAGGAGCGTAAATTCAACCCTGACCTGGCCC
CGGGTACAGAAAAGGTGACCCGCGAGGGCCAAAAGGGCGAAAAAACCATCACCACAC
CGACATTAAAAAACCCTTTAACAGGCGAGATCATCAGCAAAGGTGAGAGCAAAGAAGAA

ATCACCAAAGACCCGATTAATGAACTGACAGAGTACGGCCCTGAGACAATCACCCCTG
GTCACCGCGACGAGTTCGACCCTAAGTTACCGACCGGTGAAAAAGAAGAAGTTCCGGG
TAAACCTGGCATCAAGAATCCTGAGACCGGCGACGTTGTTCGCCCTCCGGTGGATAGC
GTGACCAAATATGGTCCGGTTAAAGGTGACAGTATCGTGGAGAAGGAAGAGATTCCTTT
CGAGAAAGAGCGCAAGTTTAATCCGGACCTGGCCCCTGGCACCGAGAAAGTTACACGC
GAGGGCCAGAAGGGTGAAAAGACCATCACAACCCCTACCCTGAAGAACCCGTTAACCG
GTGAAATTATCAGCAAGGGTGAAAGTAAAGAGGAGATCACCAAAGATCCTATCAATGAG
CTCACCGAGTATGGTCCGAGCGGCTATTGTCTGGATCTGAAAACCCAGGTTCAGACAC
CGCAGGGTATGAAAGAAATTTCAAATATTCAGGTGGGTGATCTGGTTCTGAGCAATACC
GGTTATAATGAAGTGCTGAATGTGTTTCCGAAAAGCAAAAAAAAAAAGCTATAAAATCACC
CTGGAAGATGGCAAAGAAATCATTTGTAGCGAAGAACACCTGTTTCCGACCCAGACCG
GTGAAATGAATATTAGCGGTGGTCTGAAAGAAGGTATGTGCCTGTATGTTAAAGAAGGC
GGCCATCATCATCATCATCATTAAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGG
CCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAA
GGAAGCT

Expression Product Sequence

MAHIVMVDAYKPTKGGSGGSTITELEKKVEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTI
TTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHRDEFDPKLPTGEKEEVPGKPG
IKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEK
TITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKG
EKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPG
KPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFKKERKFNPDLAPGTEKVTREGQK
GEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVP
GKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREG
QKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKE
EVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFKKERKFNPDLAPGTEKVTR
EGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGE
KEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKV
TREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPSGYCLDLKTQVQTPQG
MKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHLFPTQTGEMNISG
GLKEGMCLYVKEGGHHHHHH*

Freezer Box	ET Plasmids	#	7
Name	H_GP1C_S7_GP8N	**Source**	Self
Resistance	Kanamycin	**Total plasmid size**	8220 bp
Parent Vector	pET28	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-Tis3uMn0E5uuP2cdGzON		

Description
SasG7 protein with Gp41-1C at N-terminus and Gp41-8N at C terminus for testing of three way trans-splicing of SasG7 proteins via split inteins.

Shading Key
His-Gp41-1C-SasG7-Gp41-8N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGCATCATCATCATCATC
ACATGCTGAAAAAAATCCTGAAAATCGAGGAACTGGATGAACGCGAACTGATTGATATT
GAAGTTAGCGGTAACCACCTGTTCTATGCCAATGATATTCTGACCCATAATAGCAGCAG
CACCATCACCGAGCTGGAGAAGAAAGTTGAAGAGATTCCTTTTAAAAAAGAACGTAAGT
TCAACCCGGATCTGGCACCTGGTACCGAGAAGGTGACACGTGAGGGCCAGAAAGGCG
AGAAGACAATCACAACACCTACCTTAAAAAATCCTTTAACAGGCGTGATTATCAGTAAAG
GTGAACCGAAAGAGGAGATTACCAAAGACCCGATCAACGAGCTGACAGAGTATGGCCC
GGAAACCATCGCACCGGGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAG
AAGGAAGAAGTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTC
GTCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGTATTGTGGAA
AAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCAC
CGAAAAAGTGACCCGTGAGGGTCAAAAAGGTGAGAAGACCATTACAACCCCTACACTG
AAAAACCCGCTGACCGGCGAGATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAA
AGGACCCTATTAACGAACTGACCGAATACGGCCCGGAGACCATTACCCCGGGTCACCG
TGACGAGTTCGATCCTAAGCTGCCTACAGGCGAAAAGGAAGAAGTGCCTGGTAAACCG
GGCATCAAGAACCCTGAAACCGGCGACGTGGTTCGTCCGCCGGTTGACAGTGTTACCA
AGTACGGCCCGGTGAAGGGCGATAGCATCGTTGAGAAGGAAGAAATCCCGTTTGAGAA
AGAGCGCAAATTCAATCCGGATCTGGCCCCTGGTACCGAGAAGGTTACCCGCGAAGGC
CAAAAGGGTGAAAAAACAATTACAACACCTACACTGAAAAATCCGCTGACCGGTGTTAT
TATTAGTAAGGGTGAGCCGAAAGAGGAAATTACCAAAGATCCGATCAACGAATTAACCG
AATACGGTCCGGAAACAATCACCCCGGGCCATCGCGATGAATTTGATCCTAAATTACCG
ACAGGCGAGAAAGAGGAAGTGCCGGGCAAGCCTGGTATTAAGAATCCGGAGACAGGT
GATGTTGTGCGCCCGCCGGTTGATAGCGTGACAAAGTATGGCCCTGTGAAGGGCGACA
GCATCGTGGAAAAAGAGGAGATCCCGTTCAAAAAGGAGCGCAAATTTAATCCGGACTTA
GCACCGGGTACAGAGAAAGTTACCCGTGAAGGTCAAAAAGGCGAGAAGACCATTACCA
CCCCTACCTTAAAGAACCCTCTGACAGGTGAAATCATCAGTAAAGGCGAAAGCAAAGAA
GAGATCACCAAAGATCCGATCAATGAGTTAACAGAATATGGTCCGGAGACAATCACACC
GGGCCACCGCGATGAGTTTGACCCGAAGCTGCCGACAGGTGAAAAAGAAGAGGTGCC
TGGCAAACCGGGTATCAAGAACCCGGAGACCGGTGATGTTGTTCGCCCTCCGGTGGAT
AGTGTGACAAAATACGGTCCTGTGAAGGGTGATAGTATTGTTGAAAAGGAAGAAATTCC
GTTTGAAAAAGAGCGTAAGTTCAATCCTGATTTAGCCCCTGGCACAGAGAAGGTTACAC
GCGAAGGTCAGAAAGGTGAGAAAACAATCACCACCCCGACCCTGAAGAATCCTTTAAC
CGGCGAAATCATCAGTAAGGGCGAGAGTAAGGAAGAAATTACAAAAGACCCTATTAATG
AATTAACAGAGTATGGTCCTGAAACAATCACACCTGGCCACCGTGACGAATTCGATCCG
AAACTGCCTACCGGTGAAAAAGAAGAGGTTCCTGGCAAGCCTGGCATTAAGAACCCGG
AAACCGGCGATGTGGTGCGTCCGCCTGTGGACAGTGTTACAAAATATGGCCCGGTGAA
AGGCGATAGCATTGTGGAAAAAGAGGAGATTCCTTTTAAGAAGGAGCGTAAATTCAACC
CTGACCTGGCCCCGGGTACAGAAAAGGTGACCCGCGAGGGCCAAAAGGGCGAAAAAA
CCATCACCACACCGACATTAAAAAACCCTTTAACAGGCGAGATCATCAGCAAAGGTGAG
AGCAAAGAAGAAATCACCAAAGACCCGATTAATGAACTGACAGAGTACGGCCCTGAGA

CAATCACCCCTGGTCACCGCGACGAGTTCGACCCTAAGTTACCGACCGGTGAAAAAGA
AGAAGTTCCGGGTAAACCTGGCATCAAGAATCCTGAGACCGGCGACGTTGTTCGCCCT
CCGGTGGATAGCGTGACCAAATATGGTCCGGTTAAAGGTGACAGTATCGTGGAGAAGG
AAGAGATTCCTTTCGAGAAAGAGCGCAAGTTTAATCCGGACCTGGCCCCTGGCACCGA
GAAAGTTACACGCGAGGGCCAGAAGGGTGAAAAGACCATCACAACCCCTACCCTGAAG
AACCCGTTAACCGGTGAAATTATCAGCAAGGGTGAAAGTAAAGAGGAGATCACCAAAGA
TCCTATCAATGAGCTCACCGAGTATGGTCCGCTGAACCGTTGTCTGAGCCTGGATACCA
TGGTTGTTACCAATGGTAAAGCCATTGAAATTCGTGATGTGAAAGTTGGTGATTGGCTG
GAAAGCGAATGTGGTCCGGTTCAGGTTACCGAAGTTCTGCCGATTATCAAACAGCCGG
TTTTTGAAATTGTGCTGAAAAGCGGCAAAAAAATCCGTGTTAGCGCCAATCATAAATTCC
CGACCAAAGATGGTCTGAAAACCATTAATAGCGGTCTGAAAGTGGGCGATTTTCTGCGT
AGCCGTGCAAAATAAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTC
GAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCT

Expression Product Sequence

MHHHHHHMLKKILKIEELDERELIDIEVSGNHLFYANDILTHNSSSTITELEKKVEEIPFKKERK
FNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAP
GHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETIT
PGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEK
ERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPET
ITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPF
KKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGP
ETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEI
PFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEY
GPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEK
EEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELT
EYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIV
EKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPIN
ELTEYGPLNRCLSLDTMVVTNGKAIEIRDVKVGDWLESECGPVQVTEVLPIIKQPVFEIVLKS
GKKIRVSANHKFPTKDGLKTINSGLKVGDFLRSRAK*

Freezer Box	ET Plasmids	#	8
Name	H_GP8C_S7_STP	**Source**	Self
Resistance	Kanamycin	**Total plasmid size**	7995 bp
Parent Vector	pET28	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-Yd06hsmeShcltqb9cL1i		

Description
SasG7 protein with Gp41-8C at N-terminus for testing of three-way trans-splicing of SasG7 proteins via split inteins, with strep tag for final purification of fully spliced product.

Shading Key
His-Gp41-8C-SasG7-Strep

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGCATCATCATCATCATC
ACATGTGTGAAATCTTTGAAAACGAGATCGACTGGGATGAAATTGCCAGCATTGAATAT
GTTGGTGTGGAAGAAACCATCGATATTAACGTGACCAATGATCGTCTGTTTTTTGCCAAT
GGTATTCTGACCCATAATAGCGCGGTGACCATCACCGAGCTGGAGAAGAAAGTTGAAG
AGATTCCTTTTAAAAAAGAACGTAAGTTCAACCCGGATCTGGCACCTGGTACCGAGAAG
GTGACACGTGAGGGCCAGAAAGGCGAGAAGACAATCACAACACCTACCTTAAAAAATC
CTTTAACAGGCGTGATTATCAGTAAAGGTGAACCGAAAGAGGAGATTACCAAAGACCCG
ATCAACGAGCTGACAGAGTATGGCCCGGAAACCATCGCACCGGGCCATCGTGATGAGT
TTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGTTCCTGGTAAGCCGGGTATTAAA
AACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAGCGTTACCAAATACGGTC
CTGTTAAAGGCGACAGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAA
TTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGTG
AGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGATCATTAGCAA
GGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAACGAACTGACCGAATACGGC
CCGGAGACCATTACCCCGGGTCACCGTGACGAGTTCGATCCTAAGCTGCCTACAGGCG
AAAAGGAAGAAGTGCCTGGTAAACCGGGCATCAAGAACCCTGAAACCGGCGACGTGGT
TCGTCCGCCGGTTGACAGTGTTACCAAGTACGGCCCGGTGAAGGGCGATAGCATCGTT
GAGAAGGAAGAAATCCCGTTTGAGAAAGAGCGCAAATTCAATCCGGATCTGGCCCCTG
GTACCGAGAAGGTTACCCGCGAAGGCCAAAAGGGTGAAAAAACAATTACAACACCTAC
ACTGAAAAATCCGCTGACCGGTGTTATTATTAGTAAGGGTGAGCCGAAAGAGGAAATTA
CCAAAGATCCGATCAACGAATTAACCGAATACGGTCCGGAAACAATCACCCCGGGCCA
TCGCGATGAATTTGATCCTAAATTACCGACAGGCGAGAAAGAGGAAGTGCCGGGCAAG
CCTGGTATTAAGAATCCGGAGACAGGTGATGTTGTGCGCCCGCCGGTTGATAGCGTGA
CAAAGTATGGCCCTGTGAAGGGCGACAGCATCGTGGAAAAAGAGGAGATCCCGTTCAA
AAAGGAGCGCAAATTTAATCCGGACTTAGCACCGGGTACAGAGAAAGTTACCCGTGAA
GGTCAAAAAGGCGAGAAGACCATTACCACCCCTACCTTAAAGAACCCTCTGACAGGTG
AAATCATCAGTAAAGGCGAAAGCAAAGAAGAGATCACCAAAGATCCGATCAATGAGTTA
ACAGAATATGGTCCGGAGACAATCACACCGGGCCACCGCGATGAGTTTGACCCGAAGC
TGCCGACAGGTGAAAAAGAAGAGGTGCCTGGCAAACCGGGTATCAAGAACCCGGAGA
CCGGTGATGTTGTTCGCCCTCCGGTGGATAGTGTGACAAAATACGGTCCTGTGAAGGG
TGATAGTATTGTTGAAAAGGAAGAAATTCCGTTTGAAAAAGAGCGTAAGTTCAATCCTGA
TTTAGCCCCTGGCACAGAGAAGGTTACACGCGAAGGTCAGAAAGGTGAGAAAACAATC
ACCACCCCGACCCTGAAGAATCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTA
AGGAAGAAATTACAAAAGACCCTATTAATGAATTAACAGAGTATGGTCCTGAAACAATCA
CACCTGGCCACCGTGACGAATTCGATCCGAAACTGCCTACCGGTGAAAAAGAAGAGGT
TCCTGGCAAGCCTGGCATTAAGAACCCGGAAACCGGCGATGTGGTGCGTCCGCCTGT
GGACAGTGTTACAAAATATGGCCCGGTGAAAGGCGATAGCATTGTGGAAAAAGAGGAG
ATTCCTTTTAAGAAGGAGCGTAAATTCAACCCTGACCTGGCCCCGGGTACAGAAAAGGT
GACCCGCGAGGGCCAAAAGGGCGAAAAAACCATCACCACACCGACATTAAAAAACCCT

TTAACAGGCGAGATCATCAGCAAAGGTGAGAGCAAAGAAGAAATCACCAAAGACCCGA
TTAATGAACTGACAGAGTACGGCCCTGAGACAATCACCCCTGGTCACCGCGACGAGTT
CGACCCTAAGTTACCGACCGGTGAAAAAGAAGAAGTTCCGGGTAAACCTGGCATCAAG
AATCCTGAGACCGGCGACGTTGTTCGCCCTCCGGTGGATAGCGTGACCAAATATGGTC
CGGTTAAAGGTGACAGTATCGTGGAGAAGGAAGAGATTCCTTTCGAGAAAGAGCGCAA
GTTTAATCCGGACCTGGCCCCTGGCACCGAGAAAGTTACACGCGAGGGCCAGAAGGG
TGAAAAGACCATCACAACCCCTACCCTGAAGAACCCGTTAACCGGTGAAATTATCAGCA
AGGGTGAAAGTAAAGAGGAGATCACCAAAGATCCTATCAATGAGCTCACCGAGTATGGT
CCGTGGTCTCATCCTCAATTTGAAAAATAAGGATCCGAATTCGAGCTCCGTCGACAAGC
TTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGC
CCGAAAGGAAGCT

Expression Product Sequence

MHHHHHHMCEIFENEIDWDEIASIEYVGVEETIDINVTNDRLFFANGILTHNSAVTITELEKKV
EEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELT
EYGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIV
EKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPIN
ELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGD
SIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKD
PINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVK
GDSIVEKEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEIT
KDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGP
VKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEE
ITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKY
GPVKGDSIVEKEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGES
KEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKG
ESKEEITKDPINELTEYGPWSHPQFEK*

Freezer Box	ET Plasmids	#	20
Name	STP_S3_NRDN	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7009 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-CQzqFSRF4m0A8CZj9sr5		

Description
SasG3 with Nrdj-1N at C-terminus for multi-way assembly of SasG proteins via inteins. Strep tag for purification purposes. Construct used in 2020 intein paper.

Shading Key
Strep-SasG3-Nrdj-1N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGTGGTCTCATCCTCAAT
TTGAAAAAGGCGGCAGCGGCACCATTACAGAGCTGGAGAAGAAGGTTGAGGAAATCCC
GTTCAAGAAGGAGCGCAAGTTTAACCCGGACCTGGCCCCTGGCACAGAAAAAGTGACA
CGTGAGGGCCAGAAGGGCGAGAAGACCATCACAACCCCGACCCTGAAAAATCCTCTGA
CCGGCGTGATTATTAGCAAAGGTGAACCGAAAGAGGAGATCACAAAAGATCCGATCAA
CGAGTTAACCGAGTACGGTCCGGAAACAATCGCACCGGGTCACCGCGATGAGTTTGAC
CCGAAGTTACCGACCGGCGAAAAAGAGGAAGTGCCGGGTAAACCGGGTATCAAAAATC
CTGAGACCGGCGATGTGGTTCGCCCTCCGGTTGATAGCGTGACCAAGTATGGTCCGGT
GAAGGGCGACAGTATCGTGGAAAAAGAAGAGATCCCTTTCGAGAAAGAGCGCAAATTC
AACCCTGACCTGGCACCGGGTACCGAGAAAGTTACACGCGAAGGTCAGAAGGGTGAG
AAGACAATTACAACACCTACCCTGAAGAACCCTTTAACCGGCGAAATCATCAGTAAGGG
CGAGAGTAAAGAGGAAATCACCAAAGACCCGATTAACGAACTGACCGAATACGGTCCT
GAGACCATTACCCCGGGCCATCGTGATGAGTTCGACCCGAAACTGCCGACAGGTGAAA
AAGAAGAGGTGCCGGGCAAGCCGGGTATCAAGAACCCTGAAACCGGTGACGTGGTTC
GTCCGCCGGTTGACAGCGTTACCAAGTACGGCCCGGTGAAAGGCGACAGCATTGTTGA
AAAGGAAGAAATTCCGTTCGAAAAGGAACGTAAGTTCAACCCTGATCTGGCCCCTGGTA
CAGAGAAAGTTACCCGCGAGGGTCAGAAAGGTGAGAAAACCATCACCACCCCTACACT
GAAAAATCCGCTGACAGGCGTGATCATTAGCAAGGGTGAGCCGAAAGAGGAGATTACC
AAAGATCCGATTAACGAGCTCACCGAGTATGGTCCGAATCCGTGTTGTCTGGTTGGTAG
CAGCGAAATCATTACCCGTAATTATGGTAAAACCACCATCAAAGAAGTGGTCGAGATCT
TCGATAACGACAAAAACATTCAGGTGCTGGCCTTTAATACCCATACCGATAATATTGAAT
GGGCACCGATTAAAGCAGCACAGCTGACCCGTCCGAATGCAGAACTGGTTGAACTGGA
AATTGATACCCTGCATGGTGTTAAAACCATTCGTTGTACACCGGATCATCCTGTGTATAC
CAAAAATCGTGGTTATGTTCGTGCAGATGAACTGACCGATGATGATGAACTGGTGGTTG
CAATTTAAGGAGGTACCGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG
CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MWSHPQFEKGGSGTITELEKKVEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETG
DVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTL
KNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPE
TGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTP
TLKNPLTGVIISKGEPKEEITKDPINELTEYGPNPCCLVGSSEIITRNYGKTTIKEVVEIFDNDK
NIQVLAFNTHTDNIEWAPIKAAQLTRPNAELVELEIDTLHGVKTIRCTPDHPVYTKNRGYVRA
DELTDDDELVVAI*

Freezer Box	ET Plasmids	#	21
Name	NRDC_S3_GP1N	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7048 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-JuvHybrsQpHu51ZBDDgd		

Description
SasG3 with Nrdj-1C at N-terminus and Gp41-1N at C-terminus for multi-way assembly of SasG proteins via inteins. Construct used in 2020 intein paper.

Shading Key
Nrdj-1C-SasG3-Gp41-1N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGAAGCCAAAACCTATA
TCGGCAAACTGAAAAGCCGTAAAATTGTGAGCAACGAGGATACCTATGATATTCAGACC
AGCACCCATAACTTTTTCGCCAATGATATTCTGGTGCATAATTCAGAGATCACCATTACA
GAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCAAGTTTAACCCGG
ACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGCCAGAAGGGCGAGAAGACCA
TCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGATTATTAGCAAAGGTGAACCG
AAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACCGAGTACGGTCCGGAAACAA
TCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTACCGACCGGCGAAAAAGAGG
AAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCGGCGATGTGGTTCGCCCTCC
GGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCGACAGTATCGTGGAAAAAGAA
GAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGACCTGGCACCGGGTACCGAGA
AAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTACAACACCTACCCTGAAGAA
CCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAAATCACCAAAGAC
CCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTACCCCGGGCCATCGTGATG
AGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGGGCAAGCCGGGTA
TCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGTTGACAGCGTTACCAAGTA
CGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCCGTTCGAAAAGGAA
CGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCCGCGAGGGTCAGA
AAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCGCTGACAGGCGTGATCATT
AGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGATTAACGAGCTCACCGAGT
ATGGTCCGTCAGGATACTGTCTGGATCTGAAAACCCAGGTTCAGACACCGCAGGGTAT
GAAAGAAATTTCAAATATTCAGGTGGGTGATCTGGTTCTGAGCAATACCGGTTATAATGA
AGTGCTGAATGTGTTTCCGAAAAGCAAAAAAAAAAGCTATAAAATCACCCTGGAAGATG
GCAAAGAAATCATTTGTAGCGAAGAACACCTGTTTCCGACCCAGACCGGTGAAATGAAT
ATTAGCGGTGGTCTGAAAGAAGGTATGTGCCTGTATGTTAAAGAATAAGGAGGTACCGG
ATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCA
ATAACTAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MEAKTYIGKLKSRKIVSNEDTYDIQTSTHNFFANDILVHNSEITITELEKKVEEIPFKKERKFNP
DLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHR
DEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKF
NPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPG
HRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKER
KFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPSGYC
LDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHL
FPTQTGEMNISGGLKEGMCLYVKE*

Freezer Box	ET Plasmids	#	22
Name	GP1C_S3_IMPN	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7075 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-UCl9VseXE5k1PhlRuLZL		

Description
SasG3 with Gp41-1C at N-terminus and IMPDH-N at C-terminus for multi-way assembly of SasG proteins via inteins. Construct used in 2020 intein paper.

Shading Key
Gp41-1C-SasG3-IMPDH-N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGCTGAAAAAAATCCTGA
AAATCGAGGAACTGGATGAACGCGAACTGATTGATATTGAAGTTAGCGGTAACCACCTG
TTCTATGCCAATGATATTCTGACCCATAATTCTAGCAGTACCATTACAGAGCTGGAGAAG
AAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCAAGTTTAACCCGGACCTGGCCCCTG
GCACAGAAAAAGTGACACGTGAGGGCCAGAAGGGCGAGAAGACCATCACAACCCCGA
CCCTGAAAAATCCTCTGACCGGCGTGATTATTAGCAAAGGTGAACCGAAAGAGGAGAT
CACAAAAGATCCGATCAACGAGTTAACCGAGTACGGTCCGGAAACAATCGCACCGGGT
CACCGCGATGAGTTTGACCCGAAGTTACCGACCGGCGAAAAAGAGGAAGTGCCGGGT
AAACCGGGTATCAAAAATCCTGAGACCGGCGATGTGGTTCGCCCTCCGGTTGATAGCG
TGACCAAGTATGGTCCGGTGAAGGGCGACAGTATCGTGGAAAAAGAAGAGATCCCTTT
CGAGAAAGAGCGCAAATTCAACCCTGACCTGGCACCGGGTACCGAGAAAGTTACACGC
GAAGGTCAGAAGGGTGAGAAGACAATTACAACACCTACCCTGAAGAACCCTTTAACCG
GCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAAATCACCAAAGACCCGATTAACGA
ACTGACCGAATACGGTCCTGAGACCATTACCCCGGGCCATCGTGATGAGTTCGACCCG
AAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGGGCAAGCCGGGTATCAAGAACCCT
GAAACCGGTGACGTGGTTCGTCCGCCGGTTGACAGCGTTACCAAGTACGGCCCGGTG
AAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCCGTTCGAAAAGGAACGTAAGTTCAA
CCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCCGCGAGGGTCAGAAAGGTGAGAAA
ACCATCACCACCCCTACACTGAAAAATCCGCTGACAGGCGTGATCATTAGCAAGGGTG
AGCCGAAAGAGGAGATTACCAAAGATCCGATTAACGAGCTCACCGAGTATGGTCCGGG
TGGAGGTTGTTTTGTTCCGGGTACACTGGTGAATACCGAAAATGGTCTGAAAAAAATCG
AAGAAATCAAAGTGGGCGACAAAGTGTTTAGCCATACCGGTAAACTGCAAGAAGTTGTT
GATACCCTGATCTTTGATCGTGATGAAGAGATTATTAGCATCAACGGTATCGACTGCAC
CAAAAACCATGAGTTTTATGTGATCGACAAAGAAAATGCCAATCGCGTGAACGAAGATA
ACATTCACCTGTTTGCACGTTGGGTTCATGCCGAAGAACTGGATATGAAAAAACATCTG
CTGATCGAGCTGGAATAAGGAGGTACCGGATCCGGCTGCTAACAAAGCCCGAAAGGAA
GCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTA
AA

Expression Product Sequence
MLKKILKIEELDERELIDIEVSGNHLFYANDILTHNSSSTITELEKKVEEIPFKKERKFNPDLAP
GTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHRDEFD
PKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDL
APGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDE
FDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNP
DLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPGGGCFVP
GTLVNTENGLKKIEEIKVGDKVFSHTGKLQEVVDTLIFDRDEEIISINGIDCTKNHEFYVIDKEN
ANRVNEDNIHLFARWVHAEELDMKKHLLIELE*

Freezer Box	ET Plasmids	#	23
Name	IMPC_S3_SGBN	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7141 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-Sm0BkOVqAYb6Hn5RgKxc		

Description
SasG3 with IMPDH-C at the N-terminus and SspGyrB-N at the C-terminus for multi-way assembly of SasG proteins via inteins. Solubility tag at N-terminus of protein. Construct used in 2020 intein paper.

Shading Key
Sol-tag-IMPDH-C-SasG3-SspGyrB-N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGACAGATGTAACGATTA
AAAAATTCAAACTGAAAGAGATCACCAGCATCGAAACCAAACACTATAAAGGCAAAGTT
CATGATCTGACCGTGAATCAGGATCATAGCTATAACGTTCGTGGCACCGTTGTTCATAA
TTCGATATGCACCATTACAGAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAG
GAGCGCAAGTTTAACCCGGACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGC
CAGAAGGGCGAGAAGACCATCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGA
TTATTAGCAAAGGTGAACCGAAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACC
GAGTACGGTCCGGAAACAATCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTAC
CGACCGGCGAAAAAGAGGAAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCG
GCGATGTGGTTCGCCCTCCGGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCG
ACAGTATCGTGGAAAAAGAAGAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGAC
CTGGCACCGGGTACCGAGAAAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTA
CAACACCTACCCTGAAGAACCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAA
GAGGAAATCACCAAAGACCCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTAC
CCCGGGCCATCGTGATGAGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGT
GCCGGGCAAGCCGGGTATCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGT
TGACAGCGTTACCAAGTACGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAA
ATTCCGTTCGAAAAGGAACGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGT
TACCCGCGAGGGTCAGAAAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCG
CTGACAGGCGTGATCATTAGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGA
TTAACGAGCTCACCGAGTATGGTCCGGCAGGTGGTTGTTTTTCTGGAGATACATTAGTC
GCTTTAACTGATGGTCGTAGCGTTAGCTTTGAGCAATTGGTTGAAGAAGAAAAACAAGG
AAAACAAAACTTTTGTTATACCATCCGCCATGATGGTTCTATAGGGGTTGAAAAATCAT
CAATGCCCGCAAAACAAAAACTAATGCGAAGGTAATCAAGGTTACGTTGGACAATGGTG
AGTCTATTATTTGCACCCCGGATCATAAATTCATGTTGCGGGATGGGAGCTACAAATGT
GCGATGGATTTAACTCTCGATGATTCGTTAATGCCGTTACACCGAAAAATTTCGACTACG
GAAGATTCTGGTCATGCGTAAGGAGGTACCGGATCCGGCTGCTAACAAAGCCCGAAAG
GAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCT
CTAAA

Expression Product Sequence
MTDVTIKKFKLKEITSIETKHYKGKVHDLTVNQDHSYNVRGTVVHNSICTITELEKKVEEIPFK
KERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPE
TIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIP
FEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYG
PETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEE
IPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEY
GPAGGCFSGDTLVALTDGRSVSFEQLVEEEKQGKQNFCYTIRHDGSIGVEKIINARKTKTNA
KVIKVTLDNGESIICTPDHKFMLRDGSYKCAMDLTLDDSLMPLHRKISTTEDSGHA*

Freezer Box	ET Plasmids	#	24
Name	SGBC_S3_NRDN	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7114 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-FFUBWUz2fcmERDDMyvFj		

Description
SasG3 with SspGyrB-C at the N-terminus and Nrdj-1N at the C-terminus for multi-way assembly of SasG proteins via inteins. Construct used in 2020 intein paper.

Shading Key
SspGyrB-C-SasG3-Nrdj-1N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGAAGCAGTATTAAATT
ACAATCACAGAATTGTAAATATTGAAGCTGTGTCAGAAACAATCGATGTTTATGATATTG
AGGTTCCCCACACCCACAATTTTGCTTTGGCAAGCGGAGTGTTTGTCCATAACAGTGCA
AAAACCATTACAGAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCA
AGTTTAACCCGGACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGCCAGAAGG
GCGAGAAGACCATCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGATTATTAG
CAAAGGTGAACCGAAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACCGAGTAC
GGTCCGGAAACAATCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTACCGACCG
GCGAAAAAGAGGAAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCGGCGATGT
GGTTCGCCCTCCGGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCGACAGTATC
GTGGAAAAAGAAGAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGACCTGGCAC
CGGGTACCGAGAAAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTACAACACC
TACCCTGAAGAACCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAA
ATCACCAAAGACCCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTACCCCGG
GCCATCGTGATGAGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGG
GCAAGCCGGGTATCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGTTGACA
GCGTTACCAAGTACGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCC
GTTCGAAAAGGAACGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCC
GCGAGGGTCAGAAAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCGCTGAC
AGGCGTGATCATTAGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGATTAAC
GAGCTCACCGAGTATGGTCCGTCAAACCTTTGTCTCACGGGCGATGCTAAAATCGACG
TACTTATCGATAATATTCCTATCTCCCAGATATCCCTGGAGGAAGTTGTTAACCTGTTTA
ATGAAGGCAAAGAGATATATGTTTTGTCTTATAATATTGATACCAAGGAGGTTGAATATA
AAGAAATTTCTGACGCAGGTCTCATCAGTGAATCTGCCGAAGTGCTGGAAATCATTGAC
GAAGAAACTGGGCAGAAAATTGTTTGTACCCCTGATCATAAAGTGTATACTCTGAACCG
GGGGTACGTTTCTGCTAAGGATCTCAAAGAAGACGATGAGCTGGTGTTTAGCTAAGGA
GGTACCGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCAC
CGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MEAVLNYNHRIVNIEAVSETIDVYDIEVPHTHNFALASGVFVHNSAKTITELEKKVEEIPFKKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETI
APGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFE
KERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPE
TITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIP
FEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYG
PSNLCLTGDAKIDVLIDNIPISQISLEEVVNLFNEGKEIYVLSYNIDTKEVEYKEISDAGLISESA
EVLEIIDEETGQKIVCTPDHKVYTLNRGYVSAKDLKEDDELVFS*

Freezer Box	ET Plasmids	#	25
Name	SGBC_S3_GP8N	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7063 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-wr21vO107sNfNmk0HIji		

Description
SasG3 with SspGyrB-C at the N-terminus and Gp41-8N at the C-terminus for multi-way assembly of SasG proteins via inteins. Construct used in 2020 intein paper.

Shading Key
SspGyrB-C-SasG3-Gp41-8N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGAAGCAGTATTAAATT
ACAATCACAGAATTGTAAATATTGAAGCTGTGTCAGAAACAATCGATGTTTATGATATTG
AGGTTCCCCACACCCACAATTTTGCTTTGGCAAGCGGAGTGTTTGTCCATAACAGTGCA
AAAACCATTACAGAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCA
AGTTTAACCCGGACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGCCAGAAGG
GCGAGAAGACCATCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGATTATTAG
CAAAGGTGAACCGAAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACCGAGTAC
GGTCCGGAAACAATCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTACCGACCG
GCGAAAAAGAGGAAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCGGCGATGT
GGTTCGCCCTCCGGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCGACAGTATC
GTGGAAAAAGAAGAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGACCTGGCAC
CGGGTACCGAGAAAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTACAACACC
TACCCTGAAGAACCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAA
ATCACCAAAGACCCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTACCCCGG
GCCATCGTGATGAGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGG
GCAAGCCGGGTATCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGTTGACA
GCGTTACCAAGTACGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCC
GTTCGAAAAGGAACGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCC
GCGAGGGTCAGAAAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCGCTGAC
AGGCGTGATCATTAGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGATTAAC
GAGCTCACCGAGTATGGTCCGTTGAACAGGTGTCTGAGCCTGGATACCATGGTTGTTA
CCAATGGTAAAGCCATTGAAATTCGTGATGTGAAAGTTGGTGATTGGCTGGAAAGCGAA
TGTGGTCCGGTTCAGGTTACCGAAGTTCTGCCGATTATCAAACAGCCGGTTTTTGAAAT
TGTGCTGAAAAGCGGCAAAAAAATCCGTGTTAGCGCCAATCATAAATTCCCGACCAAAG
ATGGTCTGAAAACCATTAATAGCGGTCTGAAAGTGGGCGATTTTCTGCGTAGCCGTGCA
AAATAAGGAGGTACCGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCT
GCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MEAVLNYNHRIVNIEAVSETIDVYDIEVPHTHNFALASGVFVHNSAKTITELEKKVEEIPFKKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETI
APGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFE
KERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPE
TITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIP
FEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYG
PLNRCLSLDTMVVTNGKAIEIRDVKVGDWLESECGPVQVTEVLPIIKQPVFEIVLKSGKKIRV
SANHKFPTKDGLKTINSGLKVGDFLRSRAK*

Freezer Box	ET Plasmids	#	26
Name	SGBC_S3_H	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	6805 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-6Ci4CJYVE92E2BEl3D1F		

Description
SasG3 with SspGyrB-C at N-terminus for multi-way assembly of SasG proteins via inteins. His tag at C-terminus for purification. Construct used in 2020 intein paper.

Shading Key
SspGyrB-C-SasG3-His

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGAAGCAGTATTAAATT
ACAATCACAGAATTGTAAATATTGAAGCTGTGTCAGAAACAATCGATGTTTATGATATTG
AGGTTCCCCACACCCACAATTTTGCTTTGGCAAGCGGAGTGTTTGTCCATAACAGTGCA
AAAACCATTACAGAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCA
AGTTTAACCCGGACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGCCAGAAGG
GCGAGAAGACCATCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGATTATTAG
CAAAGGTGAACCGAAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACCGAGTAC
GGTCCGGAAACAATCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTACCGACCG
GCGAAAAAGAGGAAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCGGCGATGT
GGTTCGCCCTCCGGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCGACAGTATC
GTGGAAAAAGAAGAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGACCTGGCAC
CGGGTACCGAGAAAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTACAACACC
TACCCTGAAGAACCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAA
ATCACCAAAGACCCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTACCCCGG
GCCATCGTGATGAGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGG
GCAAGCCGGGTATCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGTTGACA
GCGTTACCAAGTACGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCC
GTTCGAAAAGGAACGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCC
GCGAGGGTCAGAAAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCGCTGAC
AGGCGTGATCATTAGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGATTAAC
GAGCTCACCGAGTATGGTCCGCATCATCACCATCACCACTAAGGAGGTACCGGATCCG
GCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAAC
TAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MEAVLNYNHRIVNIEAVSETIDVYDIEVPHTHNFALASGVFVHNSAKTITELEKKVEEIPFKKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETI
APGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFE
KERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPE
TITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIP
FEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYG
PHHHHHH*

Freezer Box	ET Plasmids	#	27
Name	GP8C_S3_H	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	6808 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-OYBlMQw9Y8BgG2A52OsT		

Description
SasG3 with Gp41-8C at the N-terminus for multi-way assembly of SasG proteins via inteins. His tag at C-terminus for purification. Construct used in 2020 intein paper.

Shading Key
Gp41-8C-SasG3-His

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGTGTGAAATCTTTGAAA
ACGAGATCGACTGGGATGAAATTGCCAGCATTGAATATGTTGGTGTGGAAGAAACCATC
GATATTAACGTGACCAATGATCGTCTGTTTTTTGCCAATGGTATTCTGACCCATAATAGT
GCGGTTACCATTACAGAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAGGAGC
GCAAGTTTAACCCGGACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGCCAGAA
GGGCGAGAAGACCATCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGATTATT
AGCAAAGGTGAACCGAAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACCGAGT
ACGGTCCGGAAACAATCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTACCGAC
CGGCGAAAAAGAGGAAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCGGCGAT
GTGGTTCGCCCTCCGGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCGACAGTA
TCGTGGAAAAAGAAGAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGACCTGGC
ACCGGGTACCGAGAAAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTACAACA
CCTACCCTGAAGAACCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAAGAGG
AAATCACCAAAGACCCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTACCCCG
GGCCATCGTGATGAGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGTGCCG
GGCAAGCCGGGTATCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGTTGAC
AGCGTTACCAAGTACGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAAATTC
CGTTCGAAAAGGAACGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGTTACC
CGCGAGGGTCAGAAAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCGCTGA
CAGGCGTGATCATTAGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGATTAA
CGAGCTCACCGAGTATGGTCCGCATCATCACCATCACCACTAAGGAGGTACCGGATCC
GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA
CTAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MCEIFENEIDWDEIASIEYVGVEETIDINVTNDRLFFANGILTHNSAVTITELEKKVEEIPFKKER
KFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAP
GHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETIT
PGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEK
ERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPHH
HHHH*

Freezer Box	ET Plasmids	#	28
Name	NRDC_S3_H	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	6775 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-wSkBVg4ZRYYcOixQ8kd3		

Description
SasG3 with Nrdj-1C at N-terminus for multi-way assembly of SasG proteins via inteins. His tag at C-terminus for purification. Construct used in 2020 intein paper.

Shading Key
Nrdj-1C-SasG3-His

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGGTTTAAAAATTATAAA
ACGTGAATCAAAGGAACCAGTGTTTGACATTACTGTCAAAGATAATAGCAATTTTTTTGC
CAATAATATCTTGGTGCATAACTGCAATGAGACCATTACAGAGCTGGAGAAGAAGGTTG
AGGAAATCCCGTTCAAGAAGGAGCGCAAGTTTAACCCGGACCTGGCCCCTGGCACAGA
AAAAGTGACACGTGAGGGCCAGAAGGGCGAGAAGACCATCACAACCCCGACCCTGAA
AAATCCTCTGACCGGCGTGATTATTAGCAAAGGTGAACCGAAAGAGGAGATCACAAAAG
ATCCGATCAACGAGTTAACCGAGTACGGTCCGGAAACAATCGCACCGGGTCACCGCGA
TGAGTTTGACCCGAAGTTACCGACCGGCGAAAAAGAGGAAGTGCCGGGTAAACCGGGT
ATCAAAAATCCTGAGACCGGCGATGTGGTTCGCCCTCCGGTTGATAGCGTGACCAAGT
ATGGTCCGGTGAAGGGCGACAGTATCGTGGAAAAAGAAGAGATCCCTTTCGAGAAAGA
GCGCAAATTCAACCCTGACCTGGCACCGGGTACCGAGAAAGTTACACGCGAAGGTCAG
AAGGGTGAGAAGACAATTACAACACCTACCCTGAAGAACCCTTTAACCGGCGAAATCAT
CAGTAAGGGCGAGAGTAAAGAGGAAATCACCAAAGACCCGATTAACGAACTGACCGAA
TACGGTCCTGAGACCATTACCCCGGGCCATCGTGATGAGTTCGACCCGAAACTGCCGA
CAGGTGAAAAAGAAGAGGTGCCGGGCAAGCCGGGTATCAAGAACCCTGAAACCGGTG
ACGTGGTTCGTCCGCCGGTTGACAGCGTTACCAAGTACGGCCCGGTGAAAGGCGACA
GCATTGTTGAAAAGGAAGAAATTCCGTTCGAAAAGGAACGTAAGTTCAACCCTGATCTG
GCCCCTGGTACAGAGAAAGTTACCCGCGAGGGTCAGAAAGGTGAGAAAACCATCACCA
CCCCTACACTGAAAAATCCGCTGACAGGCGTGATCATTAGCAAGGGTGAGCCGAAAGA
GGAGATTACCAAAGATCCGATTAACGAGCTCACCGAGTATGGTCCGCATCATCACCATC
ACCACTAAGGAGGTACCGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGG
CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA

Expression Product Sequence
MGLKIIKRESKEPVFDITVKDNSNFFANNILVHNCNETITELEKKVEEIPFKKERKFNPDLAPG
TEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHRDEFDP
KLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLA
PGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEF
DPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPD
LAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPHHHHHH*

Freezer Box	ET Plasmids	#	29
Name	GP1C_S3_NRDN	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	7087 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-fmghLqcD83c5MUbLrjlw		

Description
SasG3 with Gp41-1C at N-terminus and Nrdj-1N at C-terminus for multi-way assembly of SasG proteins via inteins. Construct used in 2020 intein paper.

Shading Key
Gp41-1C-SasG3-Nrdj-1N

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGCTGAAAAAAATCCTGA
AAATCGAGGAACTGGATGAACGCGAACTGATTGATATTGAAGTTAGCGGTAACCACCTG
TTCTATGCCAATGATATTCTGACCCATAATTCTAGCAGTACCATTACAGAGCTGGAGAAG
AAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCAAGTTTAACCCGGACCTGGCCCCTG
GCACAGAAAAAGTGACACGTGAGGGCCAGAAGGGCGAGAAGACCATCACAACCCCGA
CCCTGAAAAATCCTCTGACCGGCGTGATTATTAGCAAAGGTGAACCGAAAGAGGAGAT
CACAAAAGATCCGATCAACGAGTTAACCGAGTACGGTCCGGAAACAATCGCACCGGGT
CACCGCGATGAGTTTGACCCGAAGTTACCGACCGGCGAAAAAGAGGAAGTGCCGGGT
AAACCGGGTATCAAAAATCCTGAGACCGGCGATGTGGTTCGCCCTCCGGTTGATAGCG
TGACCAAGTATGGTCCGGTGAAGGGCGACAGTATCGTGGAAAAAGAAGAGATCCCTTT
CGAGAAAGAGCGCAAATTCAACCCTGACCTGGCACCGGGTACCGAGAAAGTTACACGC
GAAGGTCAGAAGGGTGAGAAGACAATTACAACACCTACCCTGAAGAACCCTTTAACCG
GCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAAATCACCAAAGACCCGATTAACGA
ACTGACCGAATACGGTCCTGAGACCATTACCCCGGGCCATCGTGATGAGTTCGACCCG
AAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGGGCAAGCCGGGTATCAAGAACCCT
GAAACCGGTGACGTGGTTCGTCCGCCGGTTGACAGCGTTACCAAGTACGGCCCGGTG
AAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCCGTTCGAAAAGGAACGTAAGTTCAA
CCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCCGCGAGGGTCAGAAAGGTGAGAAA
ACCATCACCACCCCTACACTGAAAAATCCGCTGACAGGCGTGATCATTAGCAAGGGTG
AGCCGAAAGAGGAGATTACCAAAGATCCGATTAACGAGCTCACCGAGTATGGTCCGAA
TCCGTGTTGTCTGGTTGGTAGCAGCGAAATCATTACCCGTAATTATGGTAAAACCACCA
TCAAAGAAGTGGTCGAGATCTTCGATAACGACAAAAACATTCAGGTGCTGGCCTTTAAT
ACCCATACCGATAATATTGAATGGGCACCGATTAAAGCAGCACAGCTGACCCGTCCGAA
TGCAGAACTGGTTGAACTGGAAATTGATACCCTGCATGGTGTTAAAACCATTCGTTGTA
CACCGGATCATCCTGTGTATACCAAAAATCGTGGTTATGTTCGTGCAGATGAACTGACC
GATGATGATGAACTGGTGGTTGCAATTTAAGGAGGTACCGGATCCGGCTGCTAACAAA
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCC
TTGGGGCCTCTAAA

Expression Product Sequence
MLKKILKIEELDERELIDIEVSGNHLFYANDILTHNSSSTITELEKKVEEIPFKKERKFNPDLAP
GTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHRDEFD
PKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDL
APGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDE
FDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNP
DLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPNPCCLVGS
SEIITRNYGKTTIKEVVEIFDNDKNIQVLAFNTHTDNIEWAPIKAAQLTRPNAELVELEIDTLHG
VKTIRCTPDHPVYTKNRGYVRADELTDDDELVVAI*

Freezer Box	ET Plasmids	#	30
Name	NRDC_S3_H	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	6793 bp
Parent Vector	pET11	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-baNe5l1hxwa7zD4SUDyA		

Description
SasG3 with Nrdj-1C at N-terminus for multi-way assembly of SasG proteins via inteins. His-tag for purification. Construct used in 2020 intein paper.

Shading Key
Nrdj-1C-SasG3-His

DNA
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC
AATTCCCCTCTAGAAATAATATTAAAGAGGAGAAATACTAGATGGAAGCCAAAACCTATA
TCGGCAAACTGAAAAGCCGTAAAATTGTGAGCAACGAGGATACCTATGATATTCAGACC
AGCACCCATAACTTTTTCGCCAATGATATTCTGGTGCATAATTCAGAGATCACCATTACA
GAGCTGGAGAAGAAGGTTGAGGAAATCCCGTTCAAGAAGGAGCGCAAGTTTAACCCGG
ACCTGGCCCCTGGCACAGAAAAAGTGACACGTGAGGGCCAGAAGGGCGAGAAGACCA
TCACAACCCCGACCCTGAAAAATCCTCTGACCGGCGTGATTATTAGCAAAGGTGAACCG
AAAGAGGAGATCACAAAAGATCCGATCAACGAGTTAACCGAGTACGGTCCGGAAACAA
TCGCACCGGGTCACCGCGATGAGTTTGACCCGAAGTTACCGACCGGCGAAAAAGAGG
AAGTGCCGGGTAAACCGGGTATCAAAAATCCTGAGACCGGCGATGTGGTTCGCCCTCC
GGTTGATAGCGTGACCAAGTATGGTCCGGTGAAGGGCGACAGTATCGTGGAAAAAGAA
GAGATCCCTTTCGAGAAAGAGCGCAAATTCAACCCTGACCTGGCACCGGGTACCGAGA
AAGTTACACGCGAAGGTCAGAAGGGTGAGAAGACAATTACAACACCTACCCTGAAGAA
CCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAAGAGGAAATCACCAAAGAC
CCGATTAACGAACTGACCGAATACGGTCCTGAGACCATTACCCCGGGCCATCGTGATG
AGTTCGACCCGAAACTGCCGACAGGTGAAAAAGAAGAGGTGCCGGGCAAGCCGGGTA
TCAAGAACCCTGAAACCGGTGACGTGGTTCGTCCGCCGGTTGACAGCGTTACCAAGTA
CGGCCCGGTGAAAGGCGACAGCATTGTTGAAAAGGAAGAAATTCCGTTCGAAAAGGAA
CGTAAGTTCAACCCTGATCTGGCCCCTGGTACAGAGAAAGTTACCCGCGAGGGTCAGA
AAGGTGAGAAAACCATCACCACCCCTACACTGAAAAATCCGCTGACAGGCGTGATCATT
AGCAAGGGTGAGCCGAAAGAGGAGATTACCAAAGATCCGATTAACGAGCTCACCGAGT
ATGGTCCGCATCATCACCATCACCACTAAGGAGGTACCGGATCCGGCTGCTAACAAAG
CCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT
TGGGGCCTCTAAA

Expression Product Sequence
MEAKTYIGKLKSRKIVSNEDTYDIQTSTHNFFANDILVHNSEITITELEKKVEEIPFKKERKFNP
DLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPETIAPGHR
DEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKF
NPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPG
HRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKER
KFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELTEYGPHHHH
HH*

Freezer Box	ET Plasmids	**#**	35
Name	pSB4434	**Source**	Stanley Brown
Resistance	Ampicillin	**Total plasmid size**	2404 bp
Parent Vector	pBluescript KS(-)	**Seq Primers**	P125, P126, P127
Benchling link	https://benchling.com/s/seq-cMx37oZfBKuiSgwluh4l		

Description
Truncated version of phage genome that contains ssDNA ori for production of ssDNA ‘backbone’ of Mini-M13 phage. From paper “An easy-to-prepare mini-scaffold for DNA origami”. Does not contain any specific POI, as purpose is to produce ssDNA. Full sequence of plasmid listed below.

Shading Key
AmpR-fd-ori-ColEI-ori

DNA
GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGAT
CTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACG
GGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC
GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT
CCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGT
CACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTT
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGT
CAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA
TACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCA
CCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGG
AAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA
CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCC
TTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTT
TAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGT
CCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGG
TCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCT
GATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTCCATTCGC
CATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACT
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGC
CGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGA
CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC
CTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA
GTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCG
GTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTT
GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC
CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC

GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCAC
CTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATAT

Freezer Box	ET Plasmids	**#**	36
Name	pSB4423	**Source**	Stanley Brown
Resistance	Chloramphenicol	**Total plasmid size**	9619 bp
Parent Vector	pACYC184	**Seq Primers**	P113 – P124
Benchling link	https://benchling.com/s/seq-rtErF5D2tPPYMK5kTxvo		

Description
All necessary phage proteins for packaging of M13 coded for on this plasmid. To be used together with pSB4434, from paper “An easy-to-prepare mini-scaffold for DNA origami”. Contains proteins G1P-G10P, but I will only highlight G3P as this is what I will edit.

Shading Key
G3P

DNA
TCGGTATCAAGCTGTTTAAGAAATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTA
AAGGCTCCTTTTGGAGCCTTTTTTTTTGGAGATTTTCAACATGAAAAAATTATTATTCGCA
ATTCCTTTAGTTGTTCCTTTCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTTTAGCAA
AACCCCATACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCGTT
ACGCTAACTATGAGGGTTGTCTGTGGAATGCTACAGGCGTTGTAGTTTGTACTGGTGAC
GAAACTCAGTGTTACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGG
TGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAA
ACCTCCTGAGTACGGTGATACACCTATTCCGGGCTATACTTATATCAACCCTCTCGACG
GCACTTATCCGCCTGGTACTGAGCAAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAG
TCTCAGCCTCTTAATACTTTCATGTTTCAGAATAATAGGTTCCGAAATAGGCAGGGGGC
ATTAACTGTTTATACGGGCACTGTTACTCAAGGCACTGACCCCGTTAAAACTTATTACCA
GTACACTCCTGTATCATCAAAAGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAG
ACTGCGCTTTCCATTCTGGCTTTAATGAGGATCCATTCGTTTGTGAATATCAAGGCCAAT
CGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTTCTGG
TGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGA
GGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAAAAGATGGCA
AACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACG
CTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGTTTCATTG
GTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTCTAAT
TCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTCAA
TATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTAGCGCTGGTAAA
CCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTC
TTTTATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAA
GGAGTCTTAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTTTCCTCGGTTTC
CTTCTGGTAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTT

Expression Product Sequence
MKKLLFAIPLVVPFYSHSAETVESCLAKPHTENSFTNVWKDDKTLDRYANYEGCLWNATGV
VVCTGDETQCYGTWVPIGLAIPENEGGGSEGGGSEGGGSEGGGTKPPEYGDTPIPGYTYI
NPLDGTYPPGTEQNPANPNPSLEESQPLNTFMFQNNRFRNRQGALTVYTGTVTQGTDPVK
TYYQYTPVSSKAMYDAYWNGKFRDCAFHSGFNEDPFVCEYQGQSSDLPQPPVNAGGGS
GGGSGGGSEGGGSEGGGSEGGGSEGGGSGGGSGSGDFDYEKMANANKGAMTENADE
NALQSDAKGKLDSVATDYGAAIDGFIGDVSGLANGNGATGDFAGSNSQMAQVGDGDNSPL
MNNFRQYLPSLPQSVECRPFVFSAGKPYEFSIDCDKINLFRGVFAFLLYVATFMYVFSTFANI
LRNKES*

Freezer Box	ET Plasmids	**#**	38
Name	4423_SpyCatcher	**Source**	Self
Resistance	Chloramphenicol	**Total plasmid size**	9982 bp
Parent Vector	pACYC184	**Seq Primers**	P113 – P124
Benchling link	https://benchling.com/s/seq-QusN4Iz0v23AqZT2vLVS		

Description
SpyCatcher fused to N-terminus of G3P. All necessary phage proteins for packaging of M13 coded for on this plasmid. To be used together with pSB4434, from paper "An easy-to-prepare mini-scaffold for DNA origami". Contains proteins G1P-G10P, but I will only highlight G3P as this is what I will edit.

I designed the fusion G3P proteins to insert the modification after the G3P leader sequence, which is a design principle commonly used in commercial phage-display kits.

Shading Key
SpyCatcher-G3P

DNA
TCGGTATCAAGCTGTTTAAGAAATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTA
AAGGCTCCTTTTGGAGCCTTTTTTTTTGGAGATTTTCAACATGAAAAAATTATTATTCGCA
ATTCCTTTAGTTGTTCCTTTCTATTCTCACTCCGCCATGGTTGATACCTTATCAGGTTTAT
CAAGTGAGCAAGGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATT
AAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGC
GTGATTCATCTGGTAAAACTATTAGTACATGGATTTCAGATGGACAAGTGAAAGATTTCT
ACCTGTATCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGGTA
GCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGC
AACTAAAGGTGACGCTCATATTGGCGGCAGCGGCGGCAGCGCTGAAACTGTTGAAAGT
TGTTTAGCAAAACCCCATACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACT
TTAGATCGTTACGCTAACTATGAGGGTTGTCTGTGGAATGCTACAGGCGTTGTAGTTTG
TACTGGTGACGAAACTCAGTGTTACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTG
AAAATGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTG
GCGGTACTAAACCTCCTGAGTACGGTGATACACCTATTCCGGGCTATACTTATATCAAC
CCTCTCGACGGCACTTATCCGCCTGGTACTGAGCAAAACCCCGCTAATCCTAATCCTTC
TCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGTTTCAGAATAATAGGTTCCGAAATAG
GCAGGGGGCATTAACTGTTTATACGGGCACTGTTACTCAAGGCACTGACCCCGTTAAAA
CTTATTACCAGTACACTCCTGTATCATCAAAAGCCATGTATGACGCTTACTGGAACGGTA
AATTCAGAGACTGCGCTTTCCATTCTGGCTTTAATGAGGATCCATTCGTTTGTGAATATC
AAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGG
TGGTTCTGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGG
CGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAA
AAGATGGCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTAC
AGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGAT
GGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGC
TGGCTCTAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAA
TTTCCGTCAATATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTAG
CGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGT
CTTTGCGTTTCTTTTATATGTTGCCACCTTTATGTATGTATTTTCTACGTTTGCTAACATA
CTGCGTAATAAGGAGTCTTAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTT
TCCTCGGTTTCCTTCTGGTAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCT
T

Expression Product Sequence

MKKLLFAIPLVVPFYSHSAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGA
TMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVN
GKATKGDAHIGGSGGSAETVESCLAKPHTENSFTNVWKDDKTLDRYANYEGCLWNATGVV
VCTGDETQCYGTWVPIGLAIPENEGGGSEGGGSEGGGSEGGGTKPPEYGDTPIPGYTYIN
PLDGTYPPGTEQNPANPNPSLEESQPLNTFMFQNNRFRNRQGALTVYTGTVTQGTDPVKT
YYQYTPVSSKAMYDAYWNGKFRDCAFHSGFNEDPFVCEYQGQSSDLPQPPVNAGGGSG
GGSGGGSEGGGSEGGGSEGGGSEGGGSGGGSGSGDFDYEKMANANKGAMTENADEN
ALQSDAKGKLDSVATDYGAAIDGFIGDVSGLANGNGATGDFAGSNSQMAQVGDGDNSPL
MNNFRQYLPSLPQSVECRPFVFSAGKPYEFSIDCDKINLFRGVFAFLLYVATFMYVFSTFANI
LRNKES*

Freezer Box	ET Plasmids	#	41
Name	pmVirD2	**Source**	Bjorn Hogbern (from AddGene)
Resistance	Ampicillin	**Total plasmid size**	6171 bp
Parent Vector	pSNAP-tag (T7)-2	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-15uoUbxdrcCrjeSblKqa		

Description
mVirD2 protein for covalent linkage of proteins to DNA. His-tag for purification purposes.

Shading Key
His-mVirD2

DNA
AAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATA
ACAATTTCACACAGATATCATTAAAGAGGAGAAATCTAGAATGCATCATCATCATCATCA
TCCCGATCGCGCTCAAGTAATCATTCGCATTGTGCCAGGAGGTGGAACCAAGACCCTT
CAGCAGATAATCAATCAGTTGGAGTACCTGTCCCGTAAGGGAAAGCTGGAACTGCAGC
GTTCAGCCCGGCATCTCGATATTCCCGTTCCGCCGGATCAAATCCGTGAGCTTGCCCA
AAGCTGGGTTACGGAGGCCGGGATTTATGACGAAAGTCAGTCAGACGATGATAGGCAA
CAAGACTTAACAACACACATTATTGTAAGCTTCCCCGCAGGTACCGACCAAACCGCAGC
TTATGAAGCCAGCCGGGAATGGGCAGCCGAGATGTTTGGGTCAGGATACGGGGGTGG
CCGCTATAACTATCTGACAGCCTACCACGTCGACCGCGATCATCCACATTTACATGTCG
TGGTCAATCGTCGGGAACTTCTGGGGCACGGGTGGCTGAAAATATCCAGGCGCCATCC
CCAGCTGAATTATGACGGCTTACGGAAAAAGATGGCAGAGATTTCACTTCGTCACGGCA
TAGTCCTGGATGCGACTTCGCGAGCAGAAAGGGGAATAGCAGAGCGACCAATCACATA
TGCTGAACATCGCCGCCTTGAGCGGATGCAGCCATGGTAACGCAAAAAACCCCGCTTC
GGCGGGGTTTTTTCGCGGATCCCCCGGGCTCGAGGTTAATTAAGCGGCCGCATTGATC
CGGCTGCTAACAAAGCCCGAAAGG

Expression Product Sequence
MHHHHHHPDRAQVIIRIVPGGGTKTLQQIINQLEYLSRKGKLELQRSARHLDIPVPPDQIREL
AQSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGG
RYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVL
DATSRAERGIAERPITYAEHRRLERMQPW*

Freezer Box	ET Plasmids	#	42
Name	VirD_SC	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	6534 bp
Parent Vector	pSNAP-tag (T7)-2	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-pJauhEJtMmk94RYtF7sK		

Description
mVirD2 protein for covalent linkage of proteins to DNA. C-terminus SpyCatcher fusion to attach mVirD2 to other proteins via SpyTag. His-tag for purification purposes.

Shading Key
His-mVirD2-SpyCatcher

DNA
AAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATA
ACAATTTCACACAGATATCATTAAAGAGGAGAAATCTAGAATGCATCATCATCATCATCA
TCCCGATCGCGCTCAAGTAATCATTCGCATTGTGCCAGGAGGTGGAACCAAGACCCTT
CAGCAGATAATCAATCAGTTGGAGTACCTGTCCCGTAAGGGAAAGCTGGAACTGCAGC
GTTCAGCCCGGCATCTCGATATTCCCGTTCCGCCGGATCAAATCCGTGAGCTTGCCCA
AAGCTGGGTTACGGAGGCCGGGATTTATGACGAAAGTCAGTCAGACGATGATAGGCAA
CAAGACTTAACAACACACATTATTGTAAGCTTCCCCGCAGGTACCGACCAAACCGCAGC
TTATGAAGCCAGCCGGGAATGGGCAGCCGAGATGTTTGGGTCAGGATACGGGGGTGG
CCGCTATAACTATCTGACAGCCTACCACGTCGACCGCGATCATCCACATTTACATGTCG
TGGTCAATCGTCGGGAACTTCTGGGGCACGGGTGGCTGAAAATATCCAGGCGCCATCC
CCAGCTGAATTATGACGGCTTACGGAAAAAGATGGCAGAGATTTCACTTCGTCACGGCA
TAGTCCTGGATGCGACTTCGCGAGCAGAAAGGGGAATAGCAGAGCGACCAATCACATA
TGCTGAACATCGCCGCCTTGAGCGGATGCAGCCATGGGGCGGCAGCGGCGGCAGCG
CCATGGTTGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTGATATGACA
ATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGA
GTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGA
TTTCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAA
CCGCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAA
GGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTTAACGCAAAAA
ACCCCGCTTCGGCGGGGTTTTTTCGCGGATCCCCCGGGCTCGAGGTTAATTAAGCGGC
CGCATTGATCCGGCTGCTAACAAAGCCCGAAAGG

Expression Product Sequence
MHHHHHHPDRAQVIIRIVPGGGTKTLQQIINQLEYLSRKGKLELQRSARHLDIPVPPDQIREL
AQSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGG
RYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVL
DATSRAERGIAERPITYAEHRRLERMQPWGGSGGSAMVDTLSGLSSEQGQSGDMTIEEDS
ATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYE
VATAITFTVNEQGQVTVNGKATKGDAHI*

Freezer Box	ET Plasmids	#	44
Name	Survivin	**Source**	Jeyaprakash Arulanandam
Resistance	Kanamycin	**Total plasmid size**	Not determined
Parent Vector	pEC-K	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-Sn3bcWqJ0e3P36L3GinE		

Description
To express Survivin protein with cleavable His tag for purification purposes.

Shading Key
His-3C-Survivin

DNA
GGGTTTGACCGAACATTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACC
ATGAAACATCACCATCACCATCACTCCGCGGGTCTGGAAGTTCTGTTCCAGGGGCCCG
ACTCGATGGGTGCCCCGACGTTGCCCCCTGCCTGGCAGCCCTTTCTCAAGGACCACCG
CATCTCTACATTCAAGAACTGGCCCTTCTTGGAGGGCTGCGCCTGCACCCCGGAGCGG
ATGGCCGAGGCTGGCTTCATCCACTGCCCCACTGAGAACGAGCCAGACTTGGCCCAGT
GTTTCTTCTGCTTCAAGGAGCTGGAAGGCTGGGAGCCAGATGACGACCCCATAGAGGA
ACATAAAAAGCATTCGTCCGGTTGCGCTTTCCTTTCTGTCAAGAAGCAGTTTGAAGAATT
AACCCTTGGTGAATTTTTGAAACTGGACAGAGAAAGAGCCAAGAACAAAATTGCAAAGG
AAACCAACAATAAGAAGAAAGAATTTGAGGAAACTGCGAAGAAAGTGCGCCGTGCCATC
GAGCAGCTGGCTGCCATGGATTAAGCAGTCGGTGGCGGTCTGAACGGTACCGGATCC
GAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCAC
TGAGATCCG

Expression Product Sequence
MKHHHHHHSAGLEVLFQGPDSMGAPTLPPAWQPFLKDHRISTFKNWPFLEGCACTPE
RMAEAGFIHCPTENEPDLAQCFFCFKELEGWEPDDDPIEEHKKHSSGCAFLSVKKQFE
ELTLGEFLKLDRERAKNKIAKETNNKKKEFEETAKKVRRAIEQLAAMD*

Freezer Box	ET Plasmids	**#**	45
Name	Survivin_SC	**Source**	Self
Resistance	Kanamycin	**Total plasmid size**	Not determined
Parent Vector	pEC-K	**Seq Primers**	P21, P22
Benchling link	https://benchling.com/s/seq-WRobRJQW0hJSqskSaVdo		

Description
To express Survivin protein with cleavable His tag for purification purposes. SpyCatcher at C terminal.

Shading Key
His-3C-Survivin-SpyCatcher

DNA
GGGTTTGACCGAACATTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACC
ATGAAACATCACCATCACCATCACTCCGCGGGTCTGGAAGTTCTGTTCCAGGGGCCCG
ACTCGATGGGTGCCCCGACGTTGCCCCCTGCCTGGCAGCCCTTTCTCAAGGACCACCG
CATCTCTACATTCAAGAACTGGCCCTTCTTGGAGGGCTGCGCCTGCACCCCGGAGCGG
ATGGCCGAGGCTGGCTTCATCCACTGCCCCACTGAGAACGAGCCAGACTTGGCCCAGT
GTTTCTTCTGCTTCAAGGAGCTGGAAGGCTGGGAGCCAGATGACGACCCCATAGAGGA
ACATAAAAAGCATTCGTCCGGTTGCGCTTTCCTTTCTGTCAAGAAGCAGTTTGAAGAATT
AACCCTTGGTGAATTTTTGAAACTGGACAGAGAAAGAGCCAAGAACAAAATTGCAAAGG
AAACCAACAATAAGAAGAAAGAATTTGAGGAAACTGCGAAGAAAGTGCGCCGTGCCATC
GAGCAGCTGGCTGCCATGGATGCCATGGTTGATACCTTATCAGGTTTATCAAGTGAGCA
AGGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAA
ACGTGATGAGGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCT
GGTAAAACTATTAGTACATGGATTTCAGATGGACAAGTGAAAGATTTCTACCTGTATCCA
GGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGGTAGCAACTGCTAT
TACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTG
ACGCTCATATTTAAGCAGTCGGTGGCGGTCTGAACGGTACCGGATCCGAATTCGAGCT
CCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCG

Expression Product Sequence
MKHHHHHHSAGLEVLFQGPDSMGAPTLPPAWQPFLKDHRISTFKNWPFLEGCACTPERM
AEAGFIHCPTENEPDLAQCFFCFKELEGWEPDDDPIEEHKKHSSGCAFLSVKKQFEELTLG
EFLKLDRERAKNKIAKETNNKKKEFEETAKKVRRAIEQLAAMDAMVDTLSGLSSEQGQSGD
MTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVET
AAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHI*

Freezer Box	ET Plasmids	#	53
Name	VirD_mCherry_SnC	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	5981 bp
Parent Vector	pProEx-Hta	**Seq Primers**	P1 + pBAD rev
Benchling link	https://benchling.com/s/seq-M9c864twplzyyNUFbK4o?m=slm-aQfLVfhT2dbx9pTTzRLX		

Description
mCherry-SnoopCatcher fusion protein for testing of binding to BslA-SnoopTag surfaces. Contains mVirD recognition site in non-coding region of the plasmid.

Shading Key
His-TEV-mCherry-SnoopCatcher

DNA
TCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCCGGTCCGTATAATCTGTG
GAATTGTGAGCGGATAACAATTTCACACAGGAAACAGACCATGCATCACCATCACCATC
ACGATTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCGCCATGGGATCCGG
AATTCAAAGGCCTACGTCGACGAGCTCAACTAGTGCGGCCGCTTTCGAATCTAGAATG
GCTAGCGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
TCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGG
GCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGT
GGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGG
CCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGG
CTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCA
GGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAA
CTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTC
CGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAA
GCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAA
GCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCAC
AACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACC
GGCGGCATGGACGAGCTGTACAAAGGCGGCTCCGGTGGTAGCATGAAGCCGCTGCGT
GGTGCCGTGTTTAGCCTGCAGAAACAGCATCCCGACTATCCCGATATCTATGGCGCGA
TTGATCAGAATGGGACCTATCAAAATGTGCGTACCGGCGAAGATGGTAAACTGACCTTT
AAGAATCTGAGCGATGGCAAATATCGCCTGTTTGAAAATAGCGAACCCGCTGGCTATAA
ACCGGTGCAGAATAAGCCGATTGTGGCGTTTCAGATTGTGAATGGCGAAGTGCGTGAT
GTGACCAGCATTGTGCCGCAGGATATTCCGGCTACATATGAATTTACCAACGGTAAACA
TTATATCACCAATGAACCGATACCGCCGAAATAATCGTACTACCATCACCATCACCATCA
CGATTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCGCCATGGATCCGGAAT
TCAAAGGCCTACGTC

Expression Product Sequence
MHHHHHHDYDIPTTENLYFQGAMGSGIQRPTSTSSTSAAAFESRMASVSKGEEDNMAIIKE
FMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSK
AYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPS
DGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLP
GAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKGGSGGSMKPLRGAVFSLQK
QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAF
QIVNGEVRDVTSIVPQDIPATYEFTNGKHYITNEPIPPK*

Freezer Box	ET Plasmids	#	54
Name	BslA_WT	**Source**	Danielle Williams
Resistance	Ampicillin	**Total plasmid size**	5416 bp
Parent Vector	pGEX-6P-1	**Seq Primers**	pGEX fwd + pGEX rev
Benchling link	https://benchling.com/s/seq-NZ6ufJB836Sq3JUe56E2?m=slm-h97r8cszfw7iz2QVSXsW		

Description
Wild-Type BslA for monolayer formation. Cleavable GST tag.

Shading Key
GST-3C-BslA

DNA
CTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGA
ATTGTGAGCGGATAACAATTTCACACAGGAAACAGTATTCATGTCCCCTATACTAGGTTA
TTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAA
ATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTG
AATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACA
GTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAA
AAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTT
TCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTA
CCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGAT
CATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGAC
CCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATC
CCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTG
GCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGAAGTTCTGTTC
CAGGGGCCCCTGGGATCCATGGCTGAATCTACATCAACTAAAGCTCATACTGAATCCAC
TATGAGAACACAGTCTACAGCTTCATTGTTCGCAACAATCACTGGCGCCAGCAAAACGG
AATGGTCTTTCTCAGATATCGAATTGACTTACCGTCCAAACACGCTTCTCAGCCTTGGC
GTTATGGAGTTTACATTGCCAAGCGGATTTACTGCAAACACGAAAGACACATTGAACGG
AAATGCCTTGCGTACAACACAGATCCTCAATAACGGGAAAACAGTAAGAGTTCCTTTGG
CACTTGATTTGTTAGGAGCTGGCGAATTCAAATTAAAACTGAATAACAAAACACTTCCTG
CCGCTGGTACATATACTTTCCGTGCGGAGAATAAATCATTAAGCATCGGAAATAAATTTT
ACGCAGAAGCCAGCATTGACGTGGCTAAGCGCAGCACTCCTCCGACTCAGTAACTCGA
GCGGCCGCATCGTGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAA
AACCTCTGACACATGCAGCTCCCGGAGACGGTCA

Expression Product Sequence
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD
VKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKL
PEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIP
QIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMAESTSTKAHTESTMR
TQSTASLFATITGASKTEWSFSDIELTYRPNTLLSLGVMEFTLPSGFTANTKDTLNGNALRTT
QILNNGKTVRVPLALDLLGAGEFKLKLNNKTLPAAGTYTFRAENKSLSIGNKFYAEASIDVAK
RSTPPTQ*

Freezer Box	ET Plasmids	#	55
Name	BslA_SpyTag	**Source**	Danielle Williams
Resistance	Ampicillin	**Total plasmid size**	5473 bp
Parent Vector	pGEX-6P-1	**Seq Primers**	pGEX fwd + pGEX rev
Benchling link	https://benchling.com/s/seq-p5vbF2jAyCHbT7sk33Rc?m=slm-ed9NvmArFXrd2xLCsRUQ		

Description
For monolayer formation and subsequent functionalization via SpyCatcher. Cleavable GST tag.

Shading Key
GST-3C-BslA-SpyTag
DNA
CTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGA
ATTGTGAGCGGATAACAATTTCACACAGGAAACAGTATTCATGTCCCCTATACTAGGTTA
TTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAA
ATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTG
AATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACA
GTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAA
AAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTT
TCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTA
CCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGAT
CATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGAC
CCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATC
CCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTG
GCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGAAGTTCTGTTC
CAGGGGCCCCTGGGATCCATGGCTGAATCTACATCAACTAAAGCTCATACTGAATCCAC
TATGAGAACACAGTCTACAGCTTCATTGTTCGCAACAATCACTGGCGCCAGCAAAACGG
AATGGTCTTTCTCAGATATCGAATTGACTTACCGTCCAAACACGCTTCTCAGCCTTGGC
GTTATGGAGTTTACATTGCCAAGCGGATTTACTGCAAACACGAAAGACACATTGAACGG
AAATGCCTTGCGTACAACACAGATCCTCAATAACGGGAAAACAGTAAGAGTTCCTTTGG
CACTTGATTTGTTAGGAGCTGGCGAATTCAAATTAAAACTGAATAACAAAACACTTCCTG
CCGCTGGTACATATACTTTCCGTGCGGAGAATAAATCATTAAGCATCGGAAATAAATTTT
ACGCAGAAGCCAGCATTGACGTGGCTAAGCGCAGCACTCCTCCGACTCAGGGCGGCT
CCGGTGGTAGCGCGCATATCGTGATGGTCGACGCATACAAACCGACCAAATAACTCGA
GCGGCCGCATCGTGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAA
AACCTCTGACACATGCAGCTCCCGGAGACGGTCA

Expression Product Sequence
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD
VKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKL
PEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIP
QIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMAESTSTKAHTESTMR
TQSTASLFATITGASKTEWSFSDIELTYRPNTLLSLGVMEFTLPSGFTANTKDTLNGNALRTT
QILNNGKTVRVPLALDLLGAGEFKLKLNNKTLPAAGTYTFRAENKSLSIGNKFYAEASIDVAK
RSTPPTQGGSGGSAHIVMVDAYKPTK*

Freezer Box	ET Plasmids	#	57
Name	BslA_SnoopTag	**Source**	Danielle Williams
Resistance	Ampicillin	**Total plasmid size**	5479 bp
Parent Vector	pGEX-6P-1	**Seq Primers**	pGEX fwd + pGEX rev
Benchling link	https://benchling.com/s/seq-Noc2VLUMMRFKovVldmrK?m=slm-OPBKlZPfu8pjSAoNxKj8		

Description
For monolayer formation and subsequent functionalization via SnoopCatcher. Cleavable GST tag.

Shading Key
GST-3C-BslA-SnoopTag

DNA
CTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGA
ATTGTGAGCGGATAACAATTTCACACAGGAAACAGTATTCATGTCCCCTATACTAGGTTA
TTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAA
ATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTG
AATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACA
GTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAA
AAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTT
TCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTA
CCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGAT
CATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGAC
CCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATC
CCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTG
GCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGAAGTTCTGTTC
CAGGGGCCCCTGGGATCCATGGCTGAATCTACATCAACTAAAGCTCATACTGAATCCAC
TATGAGAACACAGTCTACAGCTTCATTGTTCGCAACAATCACTGGCGCCAGCAAAACGG
AATGGTCTTTCTCAGATATCGAATTGACTTACCGTCCAAACACGCTTCTCAGCCTTGGC
GTTATGGAGTTTACATTGCCAAGCGGATTTACTGCAAACACGAAAGACACATTGAACGG
AAATGCCTTGCGTACAACACAGATCCTCAATAACGGGAAAACAGTAAGAGTTCCTTTGG
CACTTGATTTGTTAGGAGCTGGCGAATTCAAATTAAAACTGAATAACAAAACACTTCCTG
CCGCTGGTACATATACTTTCCGTGCGGAGAATAAATCATTAAGCATCGGAAATAAATTTT
ACGCAGAAGCCAGCATTGACGTGGCTAAGCGCAGCACTCCTCCGACTCAGGGCGGCT
CCGGTGGTAGCAAACTGGGCGATATTGAATTTATTAAAGTGAACAAATAACTCGAGCGG
CCGCATCGTGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACC
TCTGACACATGCAGCTCCCGGAGACGGTCA

Expression Product Sequence
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD
VKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKL
PEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIP
QIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSMAESTSTKAHTESTMR
TQSTASLFATITGASKTEWSFSDIELTYRPNTLLSLGVMEFTLPSGFTANTKDTLNGNALRTT
QILNNGKTVRVPLALDLLGAGEFKLKLNNKTLPAAGTYTFRAENKSLSIGNKFYAEASIDVAK
RSTPPTQGGSGGS KLGDIEFIKVNK*

Freezer Box	ET Plasmids	**#**	59
Name	GFP_SpyCatcher	**Source**	Danielle Williams
Resistance	Ampicillin	**Total plasmid size**	5760 bp
Parent Vector	pProEx-Hta	**Seq Primers**	P1 + pBAD rev
Benchling link	https://benchling.com/s/seq-ayZDRtzOXjYtXTwEvNWS?m=slm-2hjvCkwkuU50ScyzwZU6		

Description
GFP SpyCatcher fusion protein for binding to BslA_SpyTag monolayers. Cleavable GST tag.

Shading Key
His-TEV-GFP-SpyCatcher

DNA
TCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCCGGTCCGTATAATCTGTG
GAATTGTGAGCGGATAACAATTTCACACAGGAAACAGACCATGTCGTACTACCATCACC
ATCACCATCACGATTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCGCCATG
GGATCCAGTAAGGGTGAAGAACTGTTCACAGGAGTCGTCCCTATCTTGGTTGAGTTAGA
CGGCGATGTAAATGGTCACAAATTCTCTGTATCAGGGGAAGGCGAAGGCGACGCCACG
TATGGCAAACTGACCTTAAAGTTTATTTGCACTACGGGTAAACTGCCGGTGCCATGGCC
CACGCTGGTGACCACTCTGACCTATGGTGTTCAGTGTTTCTCCCGTTATCCGGATCACA
TGAAACAGCACGATTTCTTCAAATCTGCCATGCCGGAAGGCTACGTACAGGAACGGAC
AATCTTTTTCAAGGACGATGGAAACTACAAAACACGTGCCGAAGTCAAATTTGAAGGAG
ATACGTTGGTGAATCGCATTGAGCTGAAAGGGATTGATTTTAAAGAAGACGGTAATATC
TTAGGTCATAAACTGGAATATAACTACAATAGCCACAATGTGTATATTATGGCGGACAAA
CAGAAAAACGGAATCAAAGTAAATTTTAAGATTCGCCACAACATCGAAGATGGCAGCGT
CCAGTTGGCAGACCACTATCAGCAAAACACCCCTATTGGCGATGGGCCGGTACTCTCA
CCAGATAACCACTATCTCTCGACACAATCTAAACTGAGCAAAGACCCAAATGAGAAACG
CGATCACATGGTTCTGCTGGAGTTCGTTACTGCTGCCGGTATTACCCACGGTATGGAC
GAACTGTACAAGGGCGGCTCCGGTGGTAGCGCCATGGTTGATACCTTATCAGGTTTAT
CAAGTGAGCAAGGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATT
AAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGC
GTGATTCATCTGGTAAAACTATTAGTACATGGATTTCAGATGGACAAGTGAAAGATTTCT
ACCTGTATCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGGTA
GCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGC
AACTAAAGGTGACGCTCATATTAGATCTTAGTGAATTAACAAGCTTGGCTGTTTTGGCG
GATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATA
AAACAGAATTTGC

Expression Product Sequence
MSYYHHHHHHDYDIPTTENLYFQGAMGSSKGEELFTGVVPILVELDGDVNGHKFSVSGEG
EGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQ
ERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQ
KNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLSPDNHYLSTQSKLSKDPNEKRDHM
VLLEFVTAAGITHGMDELYKGGSGGSAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
EDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVN
EQGQVTVNGKATKGDAHIRS*

Freezer Box	ET Plasmids	#	64
Name	SpyCatcher_SasG7_gp411n	**Source**	Self
Resistance	Kanamycin	**Total plasmid size**	8472 bp
Parent Vector	pET28	**Seq Primers**	P21 + P22
Benchling link	https://benchling.com/s/seq-JtOSFtlbDbnMXn6c6Qti?m=slm-0IR8iRrDdePlmDJ7Uz6p		

Description
Single SasG7 unit with SpyCatcher for binding to SpyTag. Gp411 split intein available for trans-splicing with other SasG units if wanted. His-tag for purification purposes.

Shading Key
SpyCatcher-SasG7-gp411N-His

DNA
ATGGCCATGGTTGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTGATAT
GACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCA
AAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACAT
GGATTTCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCG
AAACCGCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAG
CAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTGGTGGTAG
TGGTGGTAGTTCAACCATCACCGAGCTGGAGAAGAAAGTTGAAGAGATTCCTTTTAAAA
AAGAACGTAAGTTCAACCCGGATCTGGCACCTGGTACCGAGAAGGTGACACGTGAGGG
CCAGAAAGGCGAGAAGACAATCACAACACCTACCTTAAAAAATCCTTTAACAGGCGTGA
TTATCAGTAAAGGTGAACCGAAAGAGGAGATTACCAAAGACCCGATCAACGAGCTGACA
GAGTATGGCCCGGAAACCATCGCACCGGGCCATCGTGATGAGTTTGATCCGAAGTTAC
CTACCGGTGAGAAGGAAGAAGTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGG
CGATGTGGTTCGTCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGAC
AGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCCTGATTTA
GCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGTGAGAAGACCATTACAA
CCCCTACACTGAAAAACCCGCTGACCGGCGAGATCATTAGCAAGGGTGAGAGTAAGGA
AGAGATCACAAAGGACCCTATTAACGAACTGACCGAATACGGCCCGGAGACCATTACC
CCGGGTCACCGTGACGAGTTCGATCCTAAGCTGCCTACAGGCGAAAAGGAAGAAGTGC
CTGGTAAACCGGGCATCAAGAACCCTGAAACCGGCGACGTGGTTCGTCCGCCGGTTGA
CAGTGTTACCAAGTACGGCCCGGTGAAGGGCGATAGCATCGTTGAGAAGGAAGAAATC
CCGTTTGAGAAAGAGCGCAAATTCAATCCGGATCTGGCCCCTGGTACCGAGAAGGTTA
CCCGCGAAGGCCAAAAGGGTGAAAAAACAATTACAACACCTACACTGAAAAATCCGCTG
ACCGGTGTTATTATTAGTAAGGGTGAGCCGAAAGAGGAAATTACCAAAGATCCGATCAA
CGAATTAACCGAATACGGTCCGGAAACAATCACCCCGGGCCATCGCGATGAATTTGAT
CCTAAATTACCGACAGGCGAGAAAGAGGAAGTGCCGGGCAAGCCTGGTATTAAGAATC
CGGAGACAGGTGATGTTGTGCGCCCGCCGGTTGATAGCGTGACAAAGTATGGCCCTGT
GAAGGGCGACAGCATCGTGGAAAAAGAGGAGATCCCGTTCAAAAAGGAGCGCAAATTT
AATCCGGACTTAGCACCGGGTACAGAGAAAGTTACCCGTGAAGGTCAAAAAGGCGAGA
AGACCATTACCACCCCTACCTTAAAGAACCCTCTGACAGGTGAAATCATCAGTAAAGGC
GAAAGCAAAGAAGAGATCACCAAAGATCCGATCAATGAGTTAACAGAATATGGTCCGGA
GACAATCACACCGGGCCACCGCGATGAGTTTGACCCGAAGCTGCCGACAGGTGAAAAA
GAAGAGGTGCCTGGCAAACCGGGTATCAAGAACCCGGAGACCGGTGATGTTGTTCGC
CCTCCGGTGGATAGTGTGACAAAATACGGTCCTGTGAAGGGTGATAGTATTGTTGAAAA
GGAAGAAATTCCGTTTGAAAAAGAGCGTAAGTTCAATCCTGATTTAGCCCCTGGCACAG
AGAAGGTTACACGCGAAGGTCAGAAAGGTGAGAAAACAATCACCACCCCGACCCTGAA
GAATCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAGGAAGAAATTACAAAAG
ACCCTATTAATGAATTAACAGAGTATGGTCCTGAAACAATCACACCTGGCCACCGTGAC
GAATTCGATCCGAAACTGCCTACCGGTGAAAAAGAAGAGGTTCCTGGCAAGCCTGGCA
TTAAGAACCCGGAAACCGGCGATGTGGTGCGTCCGCCTGTGGACAGTGTTACAAAATA

TGGCCCGGTGAAAGGCGATAGCATTGTGGAAAAAGAGGAGATTCCTTTTAAGAAGGAG
CGTAAATTCAACCCTGACCTGGCCCCGGGTACAGAAAAGGTGACCCGCGAGGGCCAAA
AGGGCGAAAAAACCATCACCACACCGACATTAAAAAACCCTTTAACAGGCGAGATCATC
AGCAAAGGTGAGAGCAAAGAAGAAATCACCAAAGACCCGATTAATGAACTGACAGAGTA
CGGCCCTGAGACAATCACCCCTGGTCACCGCGACGAGTTCGACCCTAAGTTACCGACC
GGTGAAAAAGAAGAAGTTCCGGGTAAACCTGGCATCAAGAATCCTGAGACCGGCGACG
TTGTTCGCCCTCCGGTGGATAGCGTGACCAAATATGGTCCGGTTAAAGGTGACAGTATC
GTGGAGAAGGAAGAGATTCCTTTCGAGAAAGAGCGCAAGTTTAATCCGGACCTGGCCC
CTGGCACCGAGAAAGTTACACGCGAGGGCCAGAAGGGTGAAAAGACCATCACAACCC
CTACCCTGAAGAACCCGTTAACCGGTGAAATTATCAGCAAGGGTGAAAGTAAAGAGGA
GATCACCAAAGATCCTATCAATGAGCTCACCGAGTATGGTCCGAGCGGCTATTGTCTGG
ATCTGAAAACCCAGGTTCAGACACCGCAGGGTATGAAAGAAATTTCAAATATTCAGGTG
GGTGATCTGGTTCTGAGCAATACCGGTTATAATGAAGTGCTGAATGTGTTTCCGAAAAG
CAAAAAAAAAAGCTATAAAATCACCCTGGAAGATGGCAAAGAAATCATTTGTAGCGAAG
AACACCTGTTTCCGACCCAGACCGGTGAAATGAATATTAGCGGTGGTCTGAAAGAAGGT
ATGTGCCTGTATGTTAAAGAAGGCGGCCATCATCATCATCATCATTAA

Expression Product Sequence

MAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWI
SDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIGGSGGS
STITELEKKVEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKE
EITKDPINELTEYGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKY
GPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGES
KEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKG
EPKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDS
VTKYGPVKGDSIVEKEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISK
GESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPV
DSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEII
SKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLT
GEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVV
RPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNP
LTGEIISKGESKEEITKDPINELTEYGPSGYCLDLKTQVQTPQGMKEISNIQVGDLVLSNTGY
NEVLNVFPKSKKKSYKITLEDGKEIICSEEHLFPTQTGEMNISGGLKEGMCLYVKEGGHHHH
HH*

Freezer Box	ET Plasmids	#	79
Name	H_pHRed_SnoopCatcher	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	3852 bp
Parent Vector	pRSET	**Seq Primers**	P21 + P22
Benchling link	https://benchling.com/s/seq-e10wdiOgI1AyjXKUB307?m=slm-viGjDYEfvDYNGr2vhicM		

Description
Biosensor for pH measurement of solution. Fused to SnoopCatcher for attachment to BslA-SpyTag surface

Shading Key
His-pHRed-SnoopCatcher

DNA
GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAAACATCACCATCACCA
TCACCATGGAGCTAGCATGGTTTCCGTTATCGCAAAGCAAATGACCTATAAAGTTTATAT
GTCTGGCACCGTTAACGGCCACTACTTCGAGGTTGAAGGTGACGGCAAGGGCAAGCCT
TACGAGGGCGAGCAGACCGTTAAGCTGACCGTTACCAAAGGCGGCCCACTGCCTTTCG
CCTGGGATATCCTGTCCCCTCAGCTCCAATACGGCTCCATCCCTTTCACCAAATACCCC
GAGGACATCCCTGACTACTTTAAGCAGAGCTTCCCAGAGGGCTACACCTGGGAGCGCT
CTATGAACTTCGAGGACGGCGCAGTTTGTACCGTTTCCAACGACAGCTCCATCCAGGG
TAACTGCTTCATCTACAACGTTAAGATTTCCGGCGAGAACTTCCCACCTAACGGCCCAG
TTATGCAGAAGAAAACCCAAGGCTGGGAGCCCAGCACCGAACGCCTGTTTGCCCGCGA
CGGTATGCTGATTGGTAACGATTATATGGCCCTGAAACTGGAAGGCGGCGGTCACTAC
CTGTGCGAGTTCAAGAGCACTTACAAGGCTAAAAAGCCTGTTCGCATGCCTGGCCGCC
ACGAAATCGACCGCAAGCTGGATGTTACCTCTCACAACCGCGACTACACCAGCGTTGA
GCAGTGTGAAATCTCTATCGCCCGCCACTCTCTGCTGGGGACAGGCAACTCCGCCGAC
GGTGGTGGCTCTGGTGGTTCTAGAATGAAGCCGCTGCGTGGTGCCGTGTTTAGCCTGC
AGAAACAGCATCCCGACTATCCCGATATCTATGGCGCGATTGATCAGAATGGGACCTAT
CAAAATGTGCGTACCGGCGAAGATGGTAAACTGACCTTTAAGAATCTGAGCGATGGCAA
ATATCGCCTGTTTGAAAATAGCGAACCCGCTGGCTATAAACCGGTGCAGAATAAGCCGA
TTGTGGCGTTTCAGATTGTGAATGGCGAAGTGCGTGATGTGACCAGCATTGTGCCGCA
GGATATTCCGGCTACATATGAATTTACCAACGGTAAACATTATATCACCAATGAACCGAT
ACCGCCGAAATGACTCGAGGGGCCCAAGCTTGATCCGGCTGCTAACAAAGCCCGAAAG
GAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG

Expression Product Sequence
MKHHHHHHHGASMVSVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTK
GGPLPFAWDILSPQLQYGSIPFTKYPEDIPDYFKQSFPEGYTWERSMNFEDGAVCTVSNDS
SIQGNCFIYNVKISGENFPPNGPVMQKKTQGWEPSTERLFARDGMLIGNDYMALKLEGGG
HYLCEFKSTYKAKKPVRMPGRHEIDRKLDVTSHNRDYTSVEQCEISIARHSLLGTGNSADG
GGSGGSRMKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYR
LFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIVPQDIPATYEFTNGKHYITNEPIPPK*

Freezer Box	ET Plasmids	**#**	80
Name	H_SpyCatcher_NAD_cpVenus	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	4851 bp
Parent Vector	pRSET	**Seq Primers**	P21 + P22
Benchling link	https://benchling.com/s/seq-sgUjOffPkO8uBRHXiDRj?m=slm-lWdQOViHlF3b3yewzzJ2		

Description
Biosensor for NAD+ sensing. Fused to SpyCatcher for attachment to BslA-SpyTag surfaces. His tag for purification purposes.

Shading Key
His-SpyCatcher-NADcpVenus

DNA
GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAAACATCACCATCACCA
TCACCATGCCATGGTTGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTG
ATATGACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACG
GCAAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGT
ACATGGATTTCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTT
GTCGAAACCGCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAA
TGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTGGCG
GCAGCGGCGGCAGCATGGAAAAGGCCCCGCATGACATCCCAATGTACAGTTTAAACGA
CGGCTTTAGCAAGGAAGACATTTTCGCATTTGATGAGCGTGTGCGTAAGGCAATCGGG
AAGCCTGTTGCATACTGTTGCGAGCTCCTGATTGACGGCCTCGCAATCAGCCTGCGTTA
CGAGAATGGTGTGTTTGTGCGCGGAGCGACCCGCGGCGATGGTACCGTTGGTGAAAA
CATTACGGAGAATCTCCGTACAGTGCGTAGTGTTCCTATGCGCCTGACGGAGCCTATTT
CAGTTGAAGTTCGCGGCGAATGCTACATGCCTAAGCAAAGTTTTGTAGCACTGAACGAG
GAACGCGAGGAGAATGGTCAAGACATTTTCGCCAACCCACGCAACGCAGCAGCCGGTT
CATTGCGTCAATTAGACACCAAGATTGTCGCAAAGCGTAATCTGAATACGTTTCTTTATA
CAGTCGCGGATTTCGGTCCGATGAAGGCGAAGACCCAATTTGAGGCGCTTGAAGAGCT
TAGTGCTATTGGCTTCCGTACGAATCCGGAGCGTCAACTCTGCCAATCGATCGACGAA
GTATGGGCTTACATTGAGGAGTATCACGAGAAGCGTTCGACACTTCCGTACGAGATCAA
CGGAATCGTTATCAAGGTGAACGAGTTTGCATTACAAGATGAATTGGGCTTCACGGTGA
AAGCACCGCGTTGGGCCATCGCCTACAAGTTCCCCTACAATAGCGACAATGTTTATATC
ACCGCTGACAAGCAAAAGAATGGCATCAAGGCGAACTTTAAGATTCGCCATAACATTGA
GGACGGTGGCGTTCAATTAGCGGATCACTATCAACAAAACACTCCAATTGGTGACGGG
CCCGTGTTACTGCCCGATAATCATTATCTTTCGTTTCAATCTAAATTATCGAAAGATCCC
AATGAGAAACGTGATCACATGGTTTTATTAGAGTTCGTGACAGCCGCCGGCATTACGCT
GGGCATGGATGAGTTGTACAAGGGCGGCAGCGGCGGCATGGTAAGCAAAGGCGAAGA
GCTGTTTACCGGCGTTGTACCCATCCTCGTGGAATTAGATGGTGATGTTAATGGCCATA
AGTTTTCAGTGAGCGGGGAAGGCGAGGGCGACGCAACTTATGGCAAGTTGACCCTTAA
ACTTATCTGCACAACAGGCAAGCTCCCTGTCCCCTGGCCGACGCTGGTAACCACCTTG
GGCTATGGACTCCAATGCTTCGCTCGCTACCCCGACCACATGAAGCAGCATGACTTCTT
TAAATCAGCGATGCCTGAAGGCTACGTCCAAGAGCGCACCATCTTCTTCAAGGACGAT
GGGAACTATAAGACGCGTGCAGAAGTGAAGTTCGAAGGTGATACATTGGTAAATCGTAT
TGAATTGAAGGGAATCGACTTCAAAGAGGACGGTAATATCCTGGGCCACAAGCTCGAG
TATAATGGCACCATCGTTTTAGAGGGCACGCGTTCAGAGCAACAACCTCTGACATTAAC
GGCGGCCACAACCCGTGCCCAAGAACTGCGCAAGCAGCTGAACCAGTACTCTCATGAG
TACTACGTCAAGGACCAGCCGAGCGTCGAGGACTACGTGTATGACCGCCTGTACAAGG
AGCTTGTTGATATTGAAACCGAGTTCCCGGACCTTATTACTCCGGATTCACCCACACAG
CGCGTAGGTTGACTCGAGGGGCCCAAGCTTGATCCGGCTGCTAACAAAGCCCGAAAG
GAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG

Expression Product Sequence

MKHHHHHHHAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDS
SGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGD
AHIGGSGGSMEKAPHDIPMYSLNDGFSKEDIFAFDERVRKAIGKPVAYCCELLIDGLAISLRY
ENGVFVRGATRGDGTVGENITENLRTVRSVPMRLTEPISVEVRGECYMPKQSFVALNEERE
ENGQDIFANPRNAAAGSLRQLDTKIVAKRNLNTFLYTVADFGPMKAKTQFEALEELSAIGFR
TNPERQLCQSIDEVWAYIEEYHEKRSTLPYEINGIVIKVNEFALQDELGFTVKAPRWAIAYKF
PYNSDNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSFQS
KLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGSGGMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGLQCFARYPDHMKQHDF
FKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNG
TIVLEGTRSEQQPLTLTAATTRAQELRKQLNQYSHEYYVKDQPSVEDYVYDRLYKELVDIET
EFPDLITPDSPTQRVG*

Freezer Box	ET Plasmids	#	81
Name	H_FLIPPi_5uM_SpyCatcher	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	5652 bp
Parent Vector	pRSET	**Seq Primers**	P21 + P22
Benchling link	https://benchling.com/s/seq-7yapiHRt4aPRzlpgalOs?m=slm-GxRToRS26thhKyEqoQYH		

Description
Biosensor for inorganic phosphate sensing with kD of 5uM. Fused to SpyCatcher for attachment to BslA-SpyTag. His tag for purification purposes, cleavable with enterokinase.

Shading Key
His-EK-FLIPPi5uM-SpyCatcher

DNA
GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCGGGGTTCTCATCATC
ATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTA
CGACGATGACGATAAGGATCCGGGCCGCATGGTGAGCAAGGGCGAGGAGCTGTTCAC
CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCAT
CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTG
GGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACATCAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA
GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCGCCGGGATCGGTACCGTAGGATTTCTAACAGCGACC
TCGGCTCAAGCCCAAACCGTGCAAATCTCCGGGGCGGGCGCGACCTTTGCGGCTCCT
TTGCTGCAACGTTGGTTTGACGCCTACAACCGCACCGTAGACCCCACTGTGCAAGTCA
GCTATCAGTCTGTCGGTGCTGGTGCTGGCCTAGAGCAGGTGATCAATGGCACTGTGGA
CTTCGGCGCTTCCGAGGCGCCTTTCTCCGGTGCTCGCCTGGAGAGCTTCCGAGCTAAA
TACGGCTATGATCCCCTACAGTTGCCTCTGGCGGGAGGGGCCATCGAGTTTGCCTATA
ACCTGCCCGGCATTGAAGACGGAGAGCTCATCCTGAAGCGGAAAACCTACTGCGGCAT
CGTGACCGGCGAGATCACTCGCTGGGACGACATTCGCATCAAGGCCGAGAACCCAGG
TATAGCAAACAAGCTGCCACCCCTGGACATCACCTGGGTACACCGCTCTGATGGTTCT
GGGACTACCTTTGTGTTCACCAACCACATCAGAACTGTCTGCCCTAATTGGACAGCCGG
TGCTGGTACTTCTGTCGAGTGGCCTGTTGGTATTGGAGCCCAAGGGAATGAGGGCGTA
GCCGCCACCATCAAGCAGGAGCCAGGGGCGATTGGCTACGTGAACCAGTCCTATGCC
AAGCTGGAAAAGATGGCCACTGCTCGCTTGGAAAACAAAGCGGGCAACATTGTTGAGT
TCTCGACTGAGGCAGCTACCTCGGCGCTGGATGCTCCCATTCCTGATGACTTTGCGCT
GTTGGTGCCCGACCCTGAAGGGCCAAATGACTACCCAATCGTGGGCTTGTTCTGGGTG
ATGCTGTACCGCGAGTATCCCGATCAGCAGAAGCTGACCAAGCTGGTGGAGGCTCTGA
AGTGGACCCAGGGGCCAGAGGGTCAAGCCATCACCAAGGAGCTGGACTATATCCCTAT
GCCTGAGGCGGTTATCCAGCGGATCTTTGCAGAGCTGGATTCCATCACCGTTAACGGT
ACCGGTGGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTG
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCCTGCAGTGCTTCGCCC
GCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTA
CGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA

GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCCC
TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGTGGCTCTGGTGGTTCTGC
CATGGTTGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTGATATGACAA
TTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAG
TTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATT
TCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAACC
GCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGG
TCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTTAAAAGCTTGATCC
GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA
CTAGCATAACCCCTTGGGGCCTCTAAACGGG

Expression Product Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPGRMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHD
FFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN
YISHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL
SKDPNEKRDHMVLLEFVTAAGIGTVGFLTATSAQAQTVQISGAGATFAAPLLQRWFDAYNR
TVDPTVQVSYQSVGAGAGLEQVINGTVDFGASEAPFSGARLESFRAKYGYDPLQLPLAGG
AIEFAYNLPGIEDGELILKRKTYCGIVTGEITRWDDIRIKAENPGIANKLPPLDITWVHRSDGS
GTTFVFTNHIRTVCPNWTAGAGTSVEWPVGIGAQGNEGVAATIKQEPGAIGYVNQSYAKLE
KMATARLENKAGNIVEFSTEAATSALDAPIPDDFALLVPDPEGPNDYPIVGLFWVMLYREYP
DQQKLTKLVEALKWTQGPEGQAITKELDYIPMPEAVIQRIFAELDSITVNGTGGMVSKGEEL
FTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGLQ
CFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF
KEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGSGGSAMVDTLSGL
SSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYL
YPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHI*

Freezer Box	ET Plasmids	#	82
Name	H_FLIPPi_200uM_SpyCatcher	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	5652 bp
Parent Vector	pRSET	**Seq Primers**	P21 + P22
Benchling link	https://benchling.com/s/seq-AjKwTspQ0HacSV50b7S6?m=slm-aHeFUNpofZIHakHCeHSp		

Description
Biosensor for inorganic phosphate sensing with kD of 200uM. Fused to SpyCatcher for attachment to BslA-SpyTag. His tag for purification purposes, cleavable with enterokinase.

Shading Key
His-EK-FLIPPi200uM-SpyCatcher

DNA
GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCGGGGTTCTCATCATC
ATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTA
CGACGATGACGATAAGGATCCGGGCCGCATGGTGAGCAAGGGCGAGGAGCTGTTCAC
CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCAT
CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTG
GGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACATCAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA
GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCGCCGGGATCGGTACCGTAGGATTTCTAACAGCGACC
TCGGCTCAAGCCCAAACCGTGCAAATCTCCGGGGCGGGCGCGACCTTTGCGGCTCCT
TTGCTGCAACGTTGGTTTGACGCCTACAACCGCACCGTAGACCCCACTGTGCAAGTCA
GCTATCAGTCTGTCGGTAGTGGTGCTGGCCTAGAGCAGGTGATCAATGGCACTGTGGA
CTTCGGCGCTTCCGAGGCGCCTTTCTCCGGTGCTCGCCTGGAGAGCTTCCGAGCTAAA
TACGGCTATGATCCCCTACAGTTGCCTCTGGCGGGAGGGGCCATCGAGTTTGCCTATA
ACCTGCCCGGCATTGAAGACGGAGAGCTCATCCTGAAGCGGAAAACCTACTGCGGCAT
CGTGACCGGCGAGATCACTCGCTGGGACGACATTCGCATCAAGGCCGAGAACCCAGG
TATAGCAAACAAGCTGCCACCCCTGGACATCACCTGGGTACACCGCTCTGATGGTTCT
GCGACTACCTTTGTGTTCACCAACCACATCAGAACTGTCTGCCCTAATTGGACAGCCGG
TGCTGGTACTTCTGTCGAGTGGCCTGTTGGTATTGGAGCCCAAGGGAATGAGGGCGTA
GCCGCCACCATCAAGCAGGAGCCAGGGGCGATTGGCTACGTGAACCAGTCCTATGCC
AAGCTGGAAAAGATGGCCACTGCTCGCTTGGAAAACAAAGCGGGCAACATTGTTGAGT
TCTCGACTGAGGCAGCTACCTCGGCGCTGGATGCTCCCATTCCTGATGACTTTGCGCT
GTTGGTGCCCGACCCTGAAGGGCCAAATGACTACCCAATCGTGGGCTTGTTCTGGGTG
ATGCTGTACCGCGAGTATCCCGATCAGCAGAAGCTGACCAAGCTGGTGGAGGCTCTGA
AGTGGACCCAGGGGCCAGAGGGTCAAGCCATCACCAAGGAGCTGGACTATATCCCTAT
GCCTGAGGCGGTTATCCAGCGGATCTTTGCAGAGCTGGATTCCATCACCGTTAACGGT
ACCGGTGGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTG
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCCTGCAGTGCTTCGCCC
GCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTA
CGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA

GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCCC
TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGTGGCTCTGGTGGTTCTGC
CATGGTTGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTGATATGACAA
TTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAG
TTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATT
TCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAACC
GCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGG
TCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTTAAAAGCTTGATCC
GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA
CTAGCATAACCCCTTGGGGCCTCTAAACGGG

Expression Product Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPGRMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHD
FFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN
YISHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL
SKDPNEKRDHMVLLEFVTAAGIGTVGFLTATSAQAQTVQISGAGATFAAPLLQRWFDAYNR
TVDPTVQVSYQSVGSGAGLEQVINGTVDFGASEAPFSGARLESFRAKYGYDPLQLPLAGG
AIEFAYNLPGIEDGELILKRKTYCGIVTGEITRWDDIRIKAENPGIANKLPPLDITWVHRSDGSA
TTFVFTNHIRTVCPNWTAGAGTSVEWPVGIGAQGNEGVAATIKQEPGAIGYVNQSYAKLEK
MATARLENKAGNIVEFSTEAATSALDAPIPDDFALLVPDPEGPNDYPIVGLFWVMLYREYPD
QQKLTKLVEALKWTQGPEGQAITKELDYIPMPEAVIQRIFAELDSITVNGTGGMVSKGEELFT
GVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGLQCF
ARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVL
LPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGSGGSAMVDTLSGLSS
EQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYP
GKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHI*

Freezer Box	ET Plasmids	#	83
Name	H_FLIPPi_30mM_SpyCatcher	**Source**	Self
Resistance	Ampicillin	**Total plasmid size**	5652 bp
Parent Vector	pRSET	**Seq Primers**	P21 + P22
Benchling link	https://benchling.com/s/seq-L4kTd4JJqfUxNrcIwLdq?m=slm-3JIO4BVFPVJkabk3CRvL		

Description
Biosensor for inorganic phosphate sensing with kD of 200uM. Fused to SpyCatcher for attachment to BslA-SpyTag. His tag for purification purposes, cleavable with enterokinase.

Shading Key
His-EK-FLIPPi200uM-SpyCatcher

DNA
GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCGGGGTTCTCATCATC
ATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTA
CGACGATGACGATAAGGATCCGGGCCGCATGGTGAGCAAGGGCGAGGAGCTGTTCAC
CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCAT
CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTG
GGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACATCAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA
GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCGCCGGGATCGGTACCGTAGGATTTCTAACAGCGACC
TCGGCTCAAGCCCAAACCGTGCAAATCTCCGGGGCGGGCGCGGCCTTTGCGGCTCCT
TTGCTGCAACGTTGGTTTGACGCCTACAACCGCACCGTAGACCCCACTGTGCAAGTCA
GCTATCAGTCTGTCGGTAGTGGTGCTGGCCTAGAGCAGGTGATCAATGGCACTGTGGA
CTTCGGCGCTTCCGAGGCGCCTTTCTCCGGTGCTCGCCTGGAGAGCTTCCGAGCTAAA
TACGGCTATGATCCCCTACAGTTGCCTCTGGCGGGAGGGGCCATCGAGTTTGCCTATA
ACCTGCCCGGCATTGAAGACGGAGAGCTCATCCTGAAGCGGAAAACCTACTGCGGCAT
CGTGACCGGCGAGATCACTCGCTGGGACGACATTCGCATCAAGGCCGAGAACCCAGG
TATAGCAAACAAGCTGCCACCCCTGGACATCACCTGGGTACACCGCTCTGATGGTTCT
GGGACTACCTTTGTGTTCACCAACCACATCAGAACTGTCTGCCCTAATTGGACAGCCGG
TGCTGGTACTTCTGTCGAGTGGCCTGTTGGTATTGGAGCCCAAGGGAATGAGGGCGTA
GCCGCCACCATCAAGCAGGAGCCAGGGGCGATTGGCTACGTGAACCAGTCCTATGCC
AAGCTGGAAAAGATGGCCACTGCTCGCTTGGAAAACAAAGCGGGCAACATTGTTGAGT
TCTCGACTGAGGCAGCTACCTCGGCGCTGGATGCTCCCATTCCTGATGACTTTGCGCT
GTTGGTGCCCGACCCTGAAGGGCCAAATGACTACCCAATCGTGGGCTTGTTCTGGGTG
ATGCTGTACCGCGAGTATCCCGATCAGCAGAAGCTGACCAAGCTGGTGGAGGCTCTGA
AGTGGACCCAGGGGCCAGAGGGTCAAGCCATCACCAAGGAGCTGGACTATATCCCTAT
GCCTGAGGCGGTTATCCAGCGGATCTTTGCAGAGCTGGATTCCATCACCGTTAACGGT
ACCGGTGGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTG
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCCTGCAGTGCTTCGCCC
GCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTA
CGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA

GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCCC
TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGTGGCTCTGGTGGTTCTGC
CATGGTTGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTGATATGACAA
TTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAG
TTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATT
TCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAATATACATTTGTCGAAACC
GCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGG
TCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCTCATATTTAAAAGCTTGATCC
GGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA
CTAGCATAACCCCTTGGGGCCTCTAAACGGG

Expression Product Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPGRMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHD
FFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN
YISHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL
SKDPNEKRDHMVLLEFVTAAGIGTVGFLTATSAQAQTVQISGAGAAFAAPLLQRWFDAYNR
TVDPTVQVSYQSVGSGAGLEQVINGTVDFGASEAPFSGARLESFRAKYGYDPLQLPLAGG
AIEFAYNLPGIEDGELILKRKTYCGIVTGEITRWDDIRIKAENPGIANKLPPLDITWVHRSDGS
GTTFVFTNHIRTVCPNWTAGAGTSVEWPVGIGAQGNEGVAATIKQEPGAIGYVNQSYAKLE
KMATARLENKAGNIVEFSTEAATSALDAPIPDDFALLVPDPEGPNDYPIVGLFWVMLYREYP
DQQKLTKLVEALKWTQGPEGQAITKELDYIPMPEAVIQRIFAELDSITVNGTGGMVSKGEEL
FTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGLQ
CFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDF
KEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGSGGSAMVDTLSGL
SSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYL
YPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAH!*

8 References

Abascal, N. C., & Regan, L. (2018). The past, present and future of protein- based materials. *Open Biology*, *8*(10).

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., Hassan, N., Butt, H., Kababji, A., Hamdan, S. M., & Mahfouz, M. M. (2020). Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. *Communications Biology*, *3*(1).

Ambrose, A. J., Pham, N. T., Sivinski, J., Guimarães, L., Mollasalehi, N., Jimenez, P., Abad, M. A., Jeyaprakash, A. A., Shave, S., Costa-Lotufo, L. V., La Clair, J. J., Auer, M., & Chapman, E. (2021). A two-step resin based approach to reveal survivin-selective fluorescent probes. *RSC Chemical Biology*, *2*(1), 181–186.

Appleby-Tagoe, J. H., Thiel, I. V., Wang, Y., Wang, Y., Mootz, H. D., & Liu, X. Q. (2011). Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. *Journal of Biological Chemistry*, *286*(39), 34440–34447.

Aranko, A. S., Wlodawer, A., & Iwaï, H. (2014). Nature's recipe for splitting inteins. *Protein Engineering, Design and Selection*, *27*(8), 263–271.

Arpino, J. A. J., & Polizzi, K. M. (2020). A modular method for directing protein self-assembly. *ACS Synthetic Biology*, *9*(5), 993–1002.

Balleza, E., Kim, J. M., & Cluzel, P. (2018). Systematic characterization of maturation time of fluorescent proteins in living cells. *Nature Methods*, *15*(1), 47–51.

Batista, A. C., Soudier, P., Kushwaha, M., & Faulon, J. (2021). Optimising protein synthesis in cell-free systems, a review. *Engineering Biology*, *5*(1), 10–19.

Bednar, R. M., Golbek, T. W., Kean, K. M., Brown, W. J., Jana, S., Baio, J. E., Karplus, P. A., & Mehl, R. A. (2019). Immobilization of Proteins with Controlled Load and Orientation. *ACS Applied Materials and Interfaces*, *11*(40), 36391–36398.

Benner, S. A., & Sismour, A. M. (2005). Synthetic biology. *Nature Reviews Genetics*, *6*(7), 533–543.

Berckman, E. A., & Chen, W. (2021). Self-assembling protein nanocages for modular enzyme assembly by orthogonal bioconjugation. *Biotechnology Progress*, *37*(5), 3–9.

Bernardinelli, G., & Högberg, B. (2017). Entirely enzymatic nanofabrication of DNA-protein conjugates. *Nucleic Acids Research*, *45*(18), 1–9.

Berrade, L., Garcia, A. E., & Camarero, J. A. (2011). Protein microarrays: Novel developments and applications. *Pharmaceutical Research*, *28*(7), 1480–1499.

Bhagawati, M., Terhorst, T. M. E., Füsser, F., Hoffmann, S., Pasch, T., Pietrokovski, S., & Mootz, H. D. (2019). A mesophilic cysteine-less split intein for protein trans-splicing applications under oxidizing conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(44), 22164–22172.

Bonnet, J., Cartannaz, J., Tourcier, G., Contreras-Martel, C., Kleman, J. P., Morlot, C., Vernet, T., & Di Guilmi, A. M. (2017). Autocatalytic association of proteins by covalent bond formation: A Bio Molecular Welding toolbox derived from a bacterial adhesin. *Scientific Reports*, *7*(January), 1–14.

Bowen, C. H., Dai, B., Sargent, C. J., Bai, W., Ladiwala, P., Feng, H., Huang, W., Kaplan, D. L., Galazka, J. M., & Zhang, F. (2018). Recombinant Spidroins Fully Replicate Primary Mechanical Properties of Natural Spider Silk. *Biomacromolecules*, *19*(9), 3853–3860.

Bowen, C. H., Sargent, C. J., Wang, A., Zhu, Y., Chang, X., Li, J., Mu, X., Galazka, J. M., Jun, Y. S., Keten, S., & Zhang, F. (2021). Microbial production of megadalton titin yields fibers

with advantageous mechanical properties. *Nature Communications*, *12*(1), 1–12.

Bromley, K. M., Morris, R. J., Hobley, L., Brandani, G., Gillespie, R. M. C., McCluskey, M., Zachariae, U., Marenduzzo, D., Stanley-Wall, N. R., & MacPhee, C. E. (2015). Interfacial self-assembly of a bacterial hydrophobin. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5419–5424.

Brooks, S. M., & Alper, H. S. (2021). Applications, challenges, and needs for employing synthetic biology beyond the lab. *Nature Communications*, *12*(1), 1–16.

Brown, S., Majikes, J., Martínez, A., Girón, T. M., Fennell, H., Samano, E. C., & LaBean, T. H. (2015). An easy-to-prepare mini-scaffold for DNA origami. *Nanoscale*, *7*(40), 16621–16624.

Brune, K. D., Leneghan, D. B., Brian, I. J., Ishizuka, A. S., Bachmann, M. F., Draper, S. J., Biswas, S., & Howarth, M. (2016). Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Scientific Reports*, *6*, 1–13.

Cambronne, X. A., Stewart, M. L., Kim, D., Jones-Brunette, A. M., Morgan, R. K., Farrens, D. L., Cohen, M. S., & Goodman, R. H. (2016). NAD+ biosensor reveals multiple sources for mitochondrial NAD+. *Science*, *352*(6292), 1474–1477.

Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology*, *12*(5), 381–390.

Carvajal-Vallejos, P., Palliss, R., Mootz, H. D., & Schmidt, S. R. (2012). Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *Journal of Biological Chemistry*, *287*(34), 28686–28696.

Caschera, F., & Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie*, *99*(1), 162–168.

Chen, Z., Boyken, S. E., Jia, M., Busch, F., Flores-Solis, D., Bick, M. J., Lu, P., VanAernum, Z. L., Sahasrabuddhe, A., Langan, R. A., Bermeo, S., Brunette, T. J., Mulligan, V. K., Carter, L. P., DiMaio, F., Sgourakis, N. G., Wysocki, V. H., & Baker, D. (2019). Programmable design of orthogonal protein heterodimers. *Nature*, *565*(7737), 106–111.

Chen, Z., Kibler, R. D., Hunt, A., Busch, F., Pearl, J., Jia, M., VanAernum, Z. L., Wicky, B. I. M., Dods, G., Liao, H., Wilken, M. S., Ciarlo, C., Green, S., El-Samad, H., Stamatoyannopoulos, J., Wysocki, V. H., Jewett, M. C., Boyken, S. E., & Baker, D. (2020). De novo design of protein logic gates. *Science*, *368*(6486), 78–84.

Cheriyan, M., Pedamallu, C. S., Tori, K., & Perler, F. (2013). Faster protein splicing with the nostoc punctiforme DnaE intein using non-native extein residues. *Journal of Biological Chemistry*, *288*(9), 6202–6211.

Clancy, K. F. A., Dery, S., Laforte, V., Shetty, P., Juncker, D., & Nicolau, D. V. (2019). Protein microarray spots are modulated by patterning method, surface chemistry and processing conditions. *Biosensors and Bioelectronics*, *130*(September 2018), 397–407.

Cole, S. D., Beabout, K., Turner, K. B., Smith, Z. K., Funk, V. L., Harbaugh, S. V., Liem, A. T., Roth, P. A., Geier, B. A., Emanuel, P. A., Walper, S. A., Chávez, J. L., & Lux, M. W. (2019). Quantification of Interlaboratory Cell-Free Protein Synthesis Variability. *ACS Synthetic Biology*, *8*(9), 2080–2091.

Cole, S. D., Miklos, A. E., Chiao, A. C., Sun, Z. Z., & Lux, M. W. (2020). Methodologies for preparation of prokaryotic extracts for cell-free expression systems. *Synthetic and Systems Biotechnology*, *5*(4), 252–267.

Cras, J. J., Rowe-Taitt, C. A., Nivens, D. A., & Ligler, F. S. (1999). Comparison of chemical cleaning methods of glass in preparation for silanization. *Biosensors and Bioelectronics*, *14*(8–9), 683–688.

de Marco, A. (2018). Nanomaterial bio-activation and macromolecules functionalization: The search for reliable protocols. In *Protein Expression and Purification* (Vol. 147, pp. 49–54). Academic Press Inc.

Des Soye, B. J., Davidson, S. R., Weinstock, M. T., Gibson, D. G., & Jewett, M. C. (2018). Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. *ACS Synthetic Biology*, *7*(9), 2245–2255.

Doerner, A., Rhiel, L., Zielonka, S., Kolmar, H., Stemmer, W. P., Krutzik, P. O., Clutter, M. R., Trejo, A., Nolan, G. P., Data, R., List, R., Lim, K. H., Huang, H. H., Pralle, A., Park, S., Fitzgerald, M. C., West, G. M., Terpe, K., LLC, S.-A. C., … Dolder, B. (2014). Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Current Protocols in Molecular Biology*, *108*, 16.30.1-16.30.11.

Dopp, J. L., Jo, Y. R., & Reuel, N. F. (2019). Methods to reduce variability in E. Coli-based cell-free protein expression experiments. *Synthetic and Systems Biotechnology*, *4*(4), 204–211.

Edgell, C. L., Smith, A. J., Beesley, J. L., Savery, N. J., & Woolfson, D. N. (2020). De novo designed protein-interaction modules for in-cell applications. *ACS Synthetic Biology*, *9*(2), 427–436.

Fages-Lartaud, M., Tietze, L., Elie, F., Lale, R., & Hohmann-Marriott, M. F. (2021). mCherry contains a fluorescent protein isoform that interferes with its reporter function. *BioRxiv*. doi: https://doi.org/10.1101/2021.12.07.471677

Fairhead, M., & Howarth, M. (2015). Site-Specific biotinylation of purified proteins using BirA. *Site-Specific Protein Labeling: Methods and Protocols*, *1266*, 171–184.

Fletcher, J. M., Boyle, A. L., Bruning, M., Bartlett, Sg. J., Vincent, T. L., Zaccai, N. R., Armstrong, C. T., Bromley, E. H. C., Booth, P. J., Brady, R. L., Thomson, A. R., & Woolfson, D. N. (2012). A basis set of de novo coiled-Coil peptide oligomers for rational protein design and synthetic biology. *ACS Synthetic Biology*, *1*(6), 240–250.

Gallup, O., Ming, H., & Ellis, T. (2021). Ten future challenges for synthetic biology. *Engineering Biology*, *5*(3), 51–59.

Garenne, D., Beisel, C. L., & Noireaux, V. (2019). Characterization of the all-E. coli transcription-translation system myTXTL by mass spectrometry. *Rapid Communications in Mass Spectrometry*, *33*(11), 1036–1048.

Green, N. M. (1975). Avidin: The use of [14C]Biotin for kinetic studies and for assay. *Advances in Protein Chemistry*, *29*(C), 85–133.

Gregorio, N. E., Kao, W. Y., Williams, L. C., Hight, C. M., Patel, P., Watts, K. R., & Oza, J. P. (2020). Unlocking Applications of Cell-Free Biotechnology through Enhanced Shelf Life and Productivity of E. coli Extracts. *ACS Synthetic Biology*, *9*(4), 766–778.

Grünberg, R., & Serrano, L. (2010). Strategies for protein synthetic biology. *Nucleic Acids Research*, *38*(8), 2663–2675.

Gruszka, D. T., Mendonça, C. A. T. F., Paci, E., Whelan, F., Hawkhead, J., Potts, J. R., & Clarke, J. (2016). Disorder drives cooperative folding in a multidomain protein. *Proceedings of the National Academy of Sciences*, *113*(42), 11841–11846.

Gruszka, D. T., Whelan, F., Farrance, O. E., Fung, H. K. H., Paci, E., Jeffries, C. M., Svergun, D. I., Baldock, C., Baumann, C. G., Brockwell, D. J., Potts, J. R., & Clarke, J. (2015). Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. *Nature Communications*, *6*, 7271.

Gruszka, D. T., Wojdyla, J. A., Bingham, R. J., Turkenburg, J. P., Manfield, I. W., Steward, A., Leech, A. P., Geoghegan, J. A., Foster, T. J., Clarke, J., & Potts, J. R. (2012). Staphylococcal biofilm-forming protein has a contiguous rod-like structure. *Proceedings of the National Academy of Sciences*, *109*(17), E1011–E1018.

Gu, H., S, L., S, O., LL, L., AM, S.-P., AR, G., J, K., I, J., & Frommer, W. B. (2006). A novel analytical method for in vivo phosphate tracking. *FEBS Letters*, *580*(25), 5885–5893.

Han, L., Chen, J., Ding, K., Zong, H., Xie, Y., Jiang, H., Zhang, B., Lu, H., Yin, W., Gilly, J., & Zhu, J. (2017). Efficient generation of bispecific IgG antibodies by split intein mediated

protein trans-splicing system. *Scientific Reports*, *7*(1), 8360.

Han, Z. W., Wang, Z., Feng, X. M., Li, B., Mu, Z. Z., Zhang, J. Q., Niu, S. C., & Ren, L. Q. (2016). Antireflective surface inspired from biology: A review. *Biosurface and Biotribology*, *2*(4), 137–150.

He, M. (2008). Cell-free protein synthesis: applications in proteomics and biotechnology. *New Biotechnology*, *25*(2–3), 126–132.

He, M., & Taussig, M. J. (2001). Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acids Research*, *29*(15).

Hinrichsen, M., Lenz, M., Edwards, J. M., Miller, O. K., Mochrie, S. G. J., Swain, P. S., Schwarz-Linek, U., & Regan, L. (2017). A new method for post-translationally labeling proteins in live cells for fluorescence imaging and tracking. *Protein Engineering, Design and Selection*, *30*(12), 771–780.

Ho, T. Y. H., Shao, A., Lu, Z., Savilahti, H., Menolascina, F., Wang, L., Dalchau, N., & Wang, B. (2021). A systematic approach to inserting split inteins for Boolean logic gate engineering and basal activity reduction. *Nature Communications*, *12*(1), 1–12.

Hobley, L., Ostrowski, A., Rao, F. V., Bromley, K. M., Porter, M., Prescott, A. R., MacPhee, C. E., van Aalten, D. M. F., & Stanley-Wall, N. R. (2013). BslA is a self-assembling bacterial hydrophobin that coats the Bacillus subtilis biofilm. *Proceedings of the National Academy of Sciences*, *110*(33), 13600–13605.

Hodgman, C. E., & Jewett, M. C. (2012). Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering*, *14*(3), 261–269.

Hsia, Y., Mout, R., Sheffler, W., Edman, N. I., Vulovic, I., Park, Y. J., Redler, R. L., Bick, M. J., Bera, A. K., Courbet, A., Kang, A., Brunette, T. J., Nattermann, U., Tsai, E., Saleem, A., Chow, C. M., Ekiert, D., Bhabha, G., Veesler, D., & Baker, D. (2021). Design of multi-scale protein complexes by hierarchical building block fusion. *Nature Communications*, *12*(1), 1–10.

Huang, P. S., Boyken, S. E., & Baker, D. (2016). The coming of age of de novo protein design. *Nature*, *537*(7620), 320–327.

Humenik, M., Smith, A. M., & Scheibel, T. (2011). Recombinant Spider Silks—Biopolymers with Potential for Future Applications. *Polymers*, *3*(1), 640–661.

Humenik, M., Winkler, A., & Scheibel, T. (2021). Patterning of protein-based materials. *Biopolymers*, *112*(2), 1–14.

Jacob, F. (1977). Evolution and tinkering. *Science*, *196*(4295), 1161–1166.

Jasaitis, L., Silver, C. D., Rawlings, A. E., Peters, D. T., Whelan, F., Regan, L., Pasquina-lemonche, L., Potts, J. R., Johnson, S. D., & Staniland, S. S. (2020). Rational design and self-assembly of coiled-coil linked SasG protein fibrils. *ACS Synthetic Biology*, *9*, 1599−1607.

Jeyaprakash, A. A., Basquin, C., Jayachandran, U., & Conti, E. (2011). Structural basis for the recognition of phosphorylated histone H3 by the Survivin subunit of the chromosomal passenger complex. *Structure*, *19*(11), 1625–1634.

Ji, S., Park, J., & Lim, H. (2012). Improved antireflection properties of moth eye mimicking nanopillars on transparent glass: flat antireflection and color tuning. *Nanoscale*, *4*(15), 4603.

Ji, S., Song, K., Nguyen, T. B., Kim, N., & Lim, H. (2013). Optimal moth eye nanostructure array on transparent glass towards broadband antireflection. *ACS Applied Materials and Interfaces*, *5*(21), 10731–10737.

Jones, J. A., Cristie-David, A. S., Andreas, M. P., & Giessen, T. W. (2021). Triggered Reversible Disassembly of an Engineered Protein Nanocage. *Angewandte Chemie* -

International Edition, *60*(47), 25034–25041.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Keeble, A. H., Banerjee, A., Ferla, M. P., Reddington, S. C., Anuar, I. N. A. K., & Howarth, M. (2017). Evolving Accelerated Amidation by SpyTag/SpyCatcher to Analyze Membrane Dynamics. *Angewandte Chemie - International Edition*, *56*(52), 16521–16525.

Keeble, A. H., & Howarth, M. (2020). Power to the protein: Enhancing and combining activities using the Spy toolbox. *Chemical Science*, *11*(28), 7281–7291.

Keeble, A. H., Turkki, P., Stokes, S., Anuar, I. N. A. K., Rahikainen, R., Hytönen, V. P., & Howarth, M. (2019). Approaching infinite affinity through engineering of peptide-protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(52), 26523–26533.

Khairil Anuar, I. N. A., Banerjee, A., Keeble, A. H., Carella, A., Nikov, G. I., & Howarth, M. (2019). Spy&Go purification of SpyTag-proteins using pseudo-SpyCatcher to access an oligomerization toolbox. *Nature Communications*, *10*(1), 1–13.

Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., & Singh, V. (2019). Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Frontiers in Bioengineering and Biotechnology*, *7*(October), 1–16.

Kilb, N., Burger, J., & Roth, G. (2014). Protein microarray generation by in situ protein expression from template DNA. *Engineering in Life Sciences*, *14*(4), 352–364.

Kim, D., & Herr, A. E. (2013). Protein immobilization techniques for microfluidic assays. *Biomicrofluidics*, *7*(4), 1–47.

Kim, E., Dai, B., Qiao, J. B., Li, W., Fortner, J. D., & Zhang, F. (2018). Microbially Synthesized Repeats of Mussel Foot Protein Display Enhanced Underwater Adhesion. *ACS Applied Materials and Interfaces*, *10*(49), 43003–43012.

Kim, J., Copeland, C. E., Padumane, S. R., & Kwon, Y. C. (2019). A crude extract preparation and optimization from a genomically engineered escherichia coli for the cell-free protein synthesis system: Practical laboratory guideline. *Methods and Protocols*, *2*(3), 1–15.

Kim, T. W., Keum, J. W., Oh, I. S., Choi, C. Y., Park, C. G., & Kim, D. M. (2006). Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of Biotechnology*, *126*(4), 554–561.

Kim, T. W., Oh, I. S., Keum, J. W., Kwon, Y. C., Byun, J. Y., Lee, K. H., Choi, C. Y., & Kim, D. M. (2007). Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. *Biotechnology and Bioengineering*, *97*(6), 1510–1515.

Ko, D. H., Tumbleston, J. R., Henderson, K. J., Euliss, L. E., Desimone, J. M., Lopez, R., & Samulski, E. T. (2011). Biomimetic microlens array with antireflective “moth-eye” surface. *Soft Matter*, *7*(14), 6404–6407.

Kryuchkov, M., Bilousov, O., Lehmann, J., Fiebig, M., & Katanaev, V. L. (2020). Reverse and forward engineering of Drosophila corneal nanocoatings. *Nature*, *585*(7825), 383–389.

Kwon, Y. C., & Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports*, *5*, 1–8.

Langan, R. A., Boyken, S. E., Ng, A. H., Samson, J. A., Dods, G., Westbrook, A. M., Nguyen, T. H., Lajoie, M. J., Chen, Z., Berger, S., Mulligan, V. K., Dueber, J. E., Novak, W. R. P., El-Samad, H., & Baker, D. (2019). De novo design of bioactive protein switches. *Nature*, *572*(7768), 205–210.

Lange, O. J., & Polizzi, K. M. (2021). Click it or stick it: Covalent and non-covalent methods for protein-self assembly. *Current Opinion in Systems Biology*, *28*, 100374.

Laohakunakorn, N. (2020). Cell-Free Systems: A Proving Ground for Rational Biodesign. *Frontiers in Bioengineering and Biotechnology*, *8*(July), 1–8.

Lebar, T., Lainšček, D., Merljak, E., Aupič, J., & Jerala, R. (2020). A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nature Chemical Biology*, *16*(5), 513–519.

Lee, J. H., Lee, J. H., Xiao, J., Desai, M. S., Zhang, X., & Lee, S. W. (2019). Vertical Self-Assembly of Polarized Phage Nanostructure for Energy Harvesting [Rapid-communication]. *Nano Letters*, *19*(4), 2661–2667.

Lee, J. W., Chan, C. T. Y., Slomovic, S., & Collins, J. J. (2018). Next-generation biocontainment systems for engineered organisms. *Nature Chemical Biology*, *14*(6), 530–537.

Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R., & Oza, J. P. (2019). Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. *Journal of Visualized Experiments : JoVE*, *144*, 1–11.

Li, J., Wang, H., Kwon, Y. C., & Jewett, M. C. (2017). Establishing a high yielding streptomyces-based cell-free protein synthesis system. *Biotechnology and Bioengineering*, *114*(6), 1343–1353.

Li, Y. (2015). Split-inteins and their bioapplications. *Biotechnology Letters*, *37*(11), 2121–2137.

Liu, D. V., Zawada, J. F., & Swartz, J. R. (2005). Streamlining Escherichia Coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnology Progress*, *21*(2), 460–465.

Livnah, O., Bayer, E. A., Wilchek, M., & Sussman, J. L. (1993). Three-dimensional structures of avidin and the avidin-biotin complex. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(11), 5076–5080.

López-Igual, R., Bernal-Bayard, J., Rodríguez-Patón, A., Ghigo, J. M., & Mazel, D. (2019). Engineered toxin–intein antimicrobials can selectively target and kill antibiotic-resistant bacteria in mixed populations. *Nature Biotechnology*, *37*(7), 755–760.

Lu, Y. (2017). Cell-free synthetic biology: Engineering in an open world. *Synthetic and Systems Biotechnology*, *2*(1), 23–27.

MacDonald, P. J., Chen, Y., & Mueller, J. D. (2012). Chromophore maturation and fluorescence fluctuation spectroscopy of fluorescent proteins in a cell-free expression system. *Analytical Biochemistry*, *421*(1), 291–298.

Manzano-Román, R., & Fuentes, M. (2019). A decade of Nucleic Acid Programmable Protein Arrays (NAPPA) availability: News, actors, progress, prospects and access. *Journal of Proteomics*, *198*(December 2018), 27–35.

Marttila, A. T., Laitinen, O. H., Airenne, K. J., Kulik, T., Bayer, E. A., Wilchek, M., & Kulomaa, M. S. (2000). Recombinant NeutraLite Avidin: A non-glycosylated, acidic mutant of chicken avidin that exhibits high affinity for biotin and low non-specific binding properties. *FEBS Letters*, *467*(1), 31–36.

Meldal, M., & Schoffelen, S. (2016). Recent advances in covalent, site-specific protein immobilization. *F1000Research*, *5*, 2303.

Meyer, C., Nakamura, Y., Rasor, B. J., Karim, A. S., Jewett, M. C., & Tan, C. (2021). Analysis of the innovation trend in cell-free synthetic biology. *Life*, *11*(6), 1–11.

Miguez, A. M., McNerney, M. P., & Styczynski, M. P. (2019). Metabolic Profiling of Escherichia coli-Based Cell-Free Expression Systems for Process Optimization. *Industrial and Engineering Chemistry Research*, *58*(50), 22472–22482.

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., Kylilis, N.,

Bell, D. J., McClymont, D. W., Jensen, K., Polizzi, K. M., Biedendieck, R., & Freemont, P. S. (2018). Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), 4340–4349.

Morris, R. J., Schor, M., Gillespie, R. M. C., Ferreira, A. S., Baldauf, L., Earl, C., Ostrowski, A., Hobley, L., Bromley, K. M., Sukhodub, T., Arnaouteli, S., Stanley-Wall, N. R., & MacPhee, C. E. (2017). Natural variations in the biofilm-associated protein BslA from the genus Bacillus. *Scientific Reports*, *7*(1), 1–13.

Novikova, O., Jayachandran, P., Kelley, D. S., Morton, Z., Merwin, S., Topilina, N. I., & Belfort, M. (2016). Intein clustering suggests functional importance in different domains of life. *Molecular Biology and Evolution*, *33*(3), 783–799.

Oeemig, J. S., Zhou, D., Kajander, T., Wlodawer, A., & Iwaï, H. (2012). NMR and crystal structures of the Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *Journal of Molecular Biology*, *421*(1), 85–99.

Oi, C., Gidden, Z., Holyoake, L., Kantelberg, O., Mochrie, S., Horrocks, M. H., & Regan, L. (2020). LIVE-PAINT allows super-resolution microscopy inside living cells using reversible peptide-protein interactions. *Communications Biology*, *3*(1), 1–10.

Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T., & Kyogoku, Y. (1999). Improved segmental isotope labeling of proteins and application to a larger protein. *Journal of Biomolecular NMR*, *14*(2), 105–114.

Pansegrau, W., Schoumacher, F., Hohn, B., & Lanka, E. (1993). Site-specific cleavage and joining of single-stranded DNA by VirD2 protein of Agrobacterium tumefaciens Ti plasmids: Analogy to bacterial conjugation. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(24), 11538–11542.

Panthu, B., Ohlmann, T., Perrier, J., Schlattner, U., Jalinot, P., Elena-Herrmann, B., & Rautureau, G. J. P. (2017). Cell-Free Protein Synthesis Enhancement from Real-Time NMR Metabolite Kinetics: Redirecting Energy Fluxes in Hybrid RRL Systems. *ACS Synthetic Biology*, *7*(1), 218–226.

Papapostolou, D., & Howorka, S. (2009). Engineering and exploiting protein assemblies in synthetic biology. *Molecular BioSystems*, *5*(7), 723–732.

Pinto, F., Thornton, E. L., & Wang, B. (2020). An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nature Communications*, *11*(1), 1–15.

Piraner, D. I., Wu, Y., & Shapiro, M. G. (2019). Modular Thermal Control of Protein Dimerization. *ACS Synthetic Biology*, *8*(10), 2256–2262.

Purnick, P. E. M., & Weiss, R. (2009). The second wave of synthetic biology: From modules to systems. *Nature Reviews Molecular Cell Biology*, *10*(6), 410–422.

Qian, Z. G., Pan, F., & Xia, X. X. (2020). Synthetic biology for protein-based materials. *Current Opinion in Biotechnology*, *65*, 197–204.

Quijano-Rubio, A., Yeh, H. W., Park, J., Lee, H., Langan, R. A., Boyken, S. E., Lajoie, M. J., Cao, L., Chow, C. M., Miranda, M. C., Wi, J., Hong, H. J., Stewart, L., Oh, B. H., & Baker, D. (2021). De novo design of modular and tunable protein biosensors. *Nature*, *591*(7850), 482–487.

Raliski, B. K., Howard, C. A., & Young, D. D. (2014). Site-specific protein immobilization using unnatural amino acids. *Bioconjugate Chemistry*, *25*(11), 1916–1920.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., Walter, J. C., & LaBaer, J. (2004). Self-Assembling Protein Microarrays. *Science*, *305*(July), 86–91.

Raut, H. K., Ganesh, V. A., Nair, A. S., & Ramakrishna, S. (2011). Anti-reflective coatings: A critical, in-depth review. *Energy & Environmental Science*, *4*(10), 3779.

Saleh, L., & Perler, F. B. (2006). Protein splicing in cis and in trans. *Chemical Record*, *6*(4), 183–193.

Sano, T., Vajda, S., & Cantor, C. R. (1998). Genetic engineering of streptavidin, a versatile affinity tag. *Journal of Chromatography B: Biomedical Applications*, *715*(1), 85–91.

Sarkar, D., Harms, H., Galleano, I., Sheikh, Z. P., & Pless, S. A. (2021). Ion channel engineering using protein trans-splicing. In *Methods in Enzymology* (1st ed., Vol. 654). Elsevier Inc.

Schloss, A. C., Liu, W., Williams, D. M., Kaufman, G., Hendrickson, H. P., Rudshteyn, B., Fu, L., Wang, H., Batista, V. S., Osuji, C., Yan, E. C. Y., & Regan, L. (2016). Fabrication of Modularly Functionalizable Microcapsules Using Protein-Based Technologies. *ACS Biomaterials Science and Engineering*, *2*(11), 1856–1861.

Shah, N. H., & Muir, T. W. (2014). Inteins: Nature's gift to protein chemists. *Chemical Science*, *5*(2), 446–461.

Shen, H., Fallas, J. A., Lynch, E., Sheffler, W., Parry, B., Jannetty, N., Decarreau, J., Wagenbach, M., Vicente, J. J., Wang, L., Dowling, Q., Oberdorfer, G., Stewart, L., Yoreo, J. De, Jacobs-wagner, C., Kollman, J., & Baker, D. (2019). *De novo design of self-assembling helical protein filaments*. *362*(6415), 705–709.

Siddique, R. H., Gomard, G., & Hölscher, H. (2015). The role of random nanostructures for the omnidirectional anti-reflection properties of the glasswing butterfly. *Nature Communications*, *6*, 1–8.

Silverman, A. D., Kelley-Loughnane, N., Lucks, J. B., & Jewett, M. C. (2019). Deconstructing Cell-Free Extract Preparation for in Vitro Activation of Transcriptional Genetic Circuitry. *ACS Synthetic Biology*, *8*(2), 403–414.

Spicer, C. D., & Davis, B. G. (2014). Selective chemical protein modification. *Nature Communications*, *5*.

Stavenga, D. G., Foletti, S., Palasantzas, G., & Arikawa, K. (2006). Light on the moth-eye corneal nipple array of butterflies. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1587), 661–667.

Steen Redeker, E., Ta, D. T., Cortens, D., Billen, B., Guedens, W., & Adriaensens, P. (2013). Protein engineering for directed immobilization. *Bioconjugate Chemistry*, *24*(11), 1761–1777.

Stevens, A. J., Sekar, G., Shah, N. H., Mostafavi, A. Z., Cowburn, D., & Muir, T. W. (2017). A promiscuous split intein with expanded protein engineering applications. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(32), 8538–8543.

Sun, F., Zhang, W. Bin, Mahdavi, A., Arnold, F. H., & Tirrell, D. A. (2014). Synthesis of bioactive protein hydrogels by genetically encoded SpyTag-SpyCatcher chemistry. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(31), 11269–11274.

Sun, J., Wang, X., Wu, J., Jiang, C., Shen, J., Cooper, M. A., Zheng, X., Liu, Y., Yang, Z., & Wu, D. (2018). Biomimetic Moth-eye Nanofabrication: Enhanced Antireflection with Superior Self-cleaning Characteristic. *Scientific Reports*, *8*(1), 1–10.

Sun, X. B., Cao, J. W., Wang, J. K., Lin, H. Z., Gao, D. Y., Qian, G. Y., Park, Y. D., Chen, Z. F., & Wang, Q. (2019). SpyTag/SpyCatcher molecular cyclization confers protein stability and resilience to aggregation. *New Biotechnology*, *49*(August 2018), 28–36.

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., & Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments*, *79*, 1–14.

Swank, Z., Laohakunakorn, N., & Maerkl, S. J. (2019). Cell-free gene-regulatory network engineering with synthetic transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(13), 5892–5901.

Takahashi, K., Sato, G., Doi, N., & Fujiwara, K. (2021). A relationship between ntp and cell extract concentration for cell-free protein expression. *Life*, *11*(3).

Tamsir, A., Tabor, J. J., & Voigt, C. A. (2011). Robust multicellular computing using genetically encoded NOR gates and chemical "wires." *Nature*, *469*(7329), 212–215.

Tantama, M., Hung, Y. P., & Yellen, G. (2011). Imaging intracellular pH in live cells with a genetically encoded red fluorescent protein sensor. *Journal of the American Chemical Society*, *133*(26), 10034–10037.

Thompson, K. E., Bashor, C. J., Lim, W. A., & Keating, A. E. (2012). Synzip protein interaction toolbox: In vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synthetic Biology*, *1*(4), 118–129.

Thornton, J. M., Laskowski, R. A., & Borkakoti, N. (2021). AlphaFold heralds a data-driven revolution in biology and medicine. *Nature Medicine*, *27*(10), 1666–1669.

Tinafar, A., Jaenes, K., & Pardee, K. (2019). Synthetic Biology Goes Cell-Free. *BMC Biology*, *17*(1), 1–14.

Tommila, J., Polojrvi, V., Aho, A., Tukiainen, A., Viheril, J., Salmi, J., Schramm, A., Kontio, J. M., Turtiainen, A., Niemi, T., & Guina, M. (2010). Nanostructured broadband antireflection coatings on AlInP fabricated by nanoimprint lithography. *Solar Energy Materials and Solar Cells*, *94*(10), 1845–1848.

Truong, D. J. J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W., & Ortiz, O. (2015). Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Research*, *43*(13), 6450–6458.

Van Kregten, M., Lindhout, B. I., Hooykaas, P. J. J., & Van Der Zaal, B. J. (2009). Agrobacterium-mediated T-DNA transfer and integration by minimal vird2 consisting of the relaxase domain and a type IV secretion system translocation signal. *Molecular Plant-Microbe Interactions*, *22*(11), 1356–1365.

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.

Velculescu, V. E., Madden, S. L., Zhang, L., Lash, A. E., Yu, J., Rago, C., Lal, A., Wang, C. J., Beaudry, G. A., Ciriello, K. M., Cook, B. P., Dufault, M. R., Ferguson, A. T., Gao, Y., He, T.-C., Hermeking, H., Hiraldo, S. K., Hwang, P. M., Lopez, M. A., … Kinzler, K. W. (1999). Analysis of human transcriptomes. *Nature Genetics*, *23*(december), 387–388.

Verdecia, M. A., Huang, H. K., Dutil, E., Kaiser, D. A., Hunter, T., & Noel, J. P. (2000). Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement. *Nature Structural Biology*, *7*(7), 602–608.

Vezeau, G. E., & Salis, H. M. (2021). Tuning Cell-Free Composition Controls the Time Delay, Dynamics, and Productivity of TX-TL Expression. *ACS Synthetic Biology*, *10*(10), 2508–2519.

Villarreal, F., & Tan, C. (2017). Cell-free systems in the new age of synthetic biology. *Frontiers of Chemical Science and Engineering*, *11*(1), 58–65.

Wandera, K. G., Collins, S. P., Wimmer, F., Marshall, R., Noireaux, V., & Beisel, C. L. (2020). An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. *Methods*, *172*(May 2019), 42–50.

Wang, B., Kitney, R. I., Joly, N., & Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, *2*(1).

Wheatley, S. P., & Altieri, D. C. (2019). Survivin at a glance. *Journal of Cell Science*, *132*(7).

Williams, D. M., Kaufman, G., Izadi, H., Gahm, A. E., Prophet, S. M., Vanderlick, K. T., Osuji, C. O., & Regan, L. (2018). Facile Protein Immobilization Using Engineered Surface-

Active Biofilm Proteins [Rapid-communication]. *ACS Applied Nano Materials*, *1*(6), 2483–2488.

Winter, D. L., Iranmanesh, H., Clark, D. S., & Glover, D. J. (2020). Design of Tunable Protein Interfaces Controlled by Post-Translational Modifications. *ACS Synthetic Biology*, *9*(8), 2132–2143.

Wood, D. W., & Camarero, J. A. (2014). Intein applications: From protein purification and labeling to metabolic control methods. *Journal of Biological Chemistry*, *289*(21), 14512–14519.

Xia, X. X., Qian, Z. G., Ki, C. S., Park, Y. H., Kaplan, D. L., & Lee, S. Y. (2010). Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14059–14063.

Yang, F., Zuo, X., Fan, C., & Zhang, X. E. (2018). Biomacromolecular nanostructures-based interfacial engineering: From precise assembly to precision biosensing. *National Science Review*, *5*(5), 740–755.

Yazaki, J., Galli, M., Kim, A. Y., Nito, K., Aleman, F., Chang, K. N., Carvunis, A. R., Quan, R., Nguyen, H., Song, L., Alvarez, J. M., Huang, S. S. C., Chen, H., Ramachandran, N., Altmann, S., Gutiérrez, R. A., Hill, D. E., Schroeder, J. I., Chory, J., … Ecker, J. R. (2016). Mapping transcription factor interactome networks using HaloTag protein arrays. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), E4238–E4247.

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*, *109*(12), E690–E697.

Zakharova, M. Y., Kozyr, A. V., Ignatova, A. N., Vinnikov, I. A., Shemyakin, I. G., & Kolesnikov, A. V. (2005). Purification of filamentous bacteriophage for phage display using size-exclusion chromatography. *BioTechniques*, *38*(2), 194–198.

Zhang, C., Yi, P., Peng, L., & Ni, J. (2017). Optimization and continuous fabrication of moth-eye nanostructure array on flexible polyethylene terephthalate substrate towards broadband antireflection. *Applied Optics*, *56*(10), 2901.

Zubay, G. (1973). In Vitro Synthesis of Protein in Microbial Systems. *Annual Review of Genetics*, *5*, 267–287.