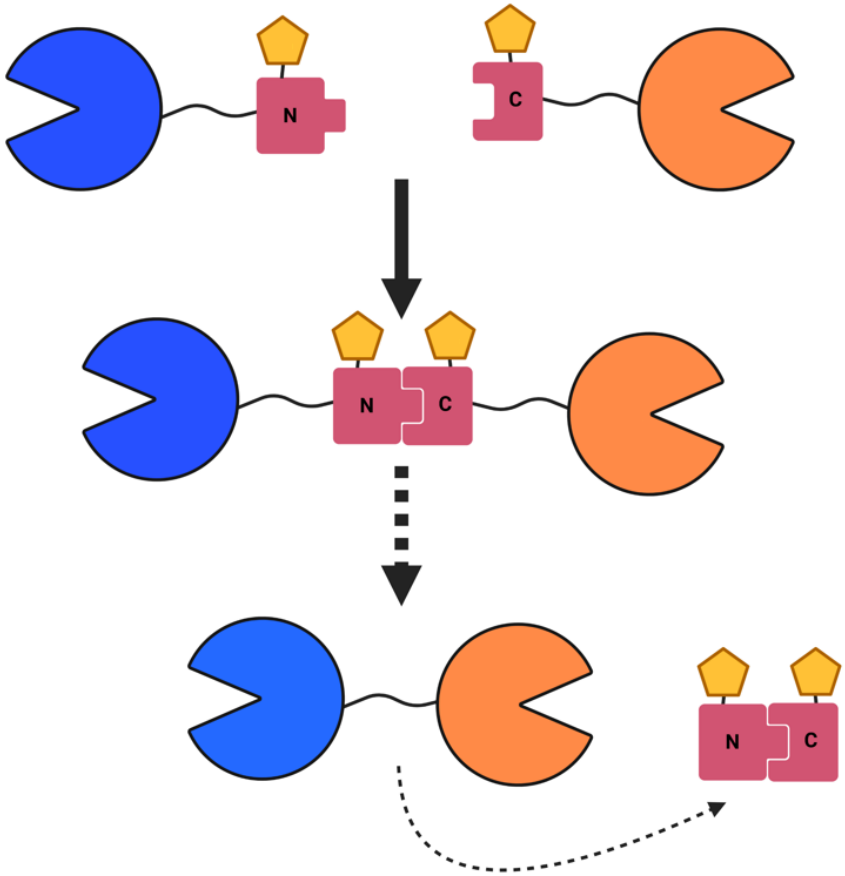
CHALMERS
UNIVERSITY OF TECHNOLOGY



A Modular Toolkit to Enable Intein-mediated Enzyme Fusions

Master's thesis in biotechnology

LINUS STORM

DEPARTMENT OF LIFE SCIENCES
DIVISION OF INDUSTRIAL BIOTECHNOLOGY

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

A Modular Toolkit to Enable Intein-mediated Enzyme Fusions

LINUS STORM

Department of Life Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2024

A Modular Toolkit to Enable Intein-mediated Enzyme Fusions
LINUS STORM

Department of Life Sciences
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Cover: Split intein fusion of two proteins (see p. 10).

A Modular Toolkit to Enable Intein-mediated Enzyme Fusions

LINUS STORM
Department of Life Sciences
Chalmers University of Technology

SUMMARY

In second generation biorefineries, lignocellulosic biomass is fermented by microbes to higher-value products. Hydrolysis of the substrate is necessary to break it into fragments that the microbes can utilize. Enzymatic hydrolysis is often preferred, but a bottleneck is the high costs of enzyme cocktails since lignocellulosic polymers require several types of enzymes to be degraded. Fusion enzymes have been produced to increase conversation rates, but the methods for creating them are either complex or limited. This project aimed to circumvent some of these limitations by creating a modular toolkit to produce any fusion enzymes, where the fusion is split intein-mediated *in vitro*. For evaluation of the toolkit, an α-xylosidase (*Bo*GH3B) and a β-glucosidase (*Bo*GH31A) involved in xyloglucan (XyG) degradation, were chosen.

The toolkit was developed to completion and consists of eight types of parts, excluding the protein of interest. Three assemblies were confirmed by sequencing, proving that the toolkit works. Aside from the toolkit-mediated assemblies, the enzymes on their own were expressed and tested in different combinations for XyG degradation. This was done to compare with the split intein-fused enzymes. However, when it came to expressing the protein assemblies the yield was low, and no fusion could be observed. Thus, the possibly synergistic effect of fusing *Bo*GH3B and *Bo*GH31A could not be investigated. A lot of research is still needed to determine the usefulness of split inteins for creating enzyme fusions. For future work, some experiments need redoing, and new experiments with different linkers and configurations of the assemblies could be performed.

Keywords: Modular cloning, DNA assembly, inteins, split inteins, fusion enzymes, xyloglucan

Table of Contents

1. Background

1.1 Introduction

A bioeconomy has been described as encompassing three main visions: the biotechnology, bio-resource, and bio-ecology visions [1]. The bio-resource vision concentrates on converting raw biological materials into higher-value products. A backbone of this vision is the biorefinery, where fermentation is used to process biological carbon sources into biomass, biofuels, food products, and pharmaceuticals using microorganisms [2]. In second generation biorefineries, the carbon source used as feedstock is lignocellulosic biomass, which is mainly comprised of polysaccharides (cellulose and hemicelluloses) and lignin [3]. Usually, the enzymes the microbes produce cannot directly access the sugars in their polymeric chains, therefore hydrolysis is necessary to convert the polymers into smaller fragments [4]. Enzymatic hydrolysis is often advantageous, as it can be more specific than non-enzymatic hydrolysis, reaches high yields, is more eco-friendly, and generates very few to no fermentation inhibitors [5]. A bottleneck with enzymatic hydrolysis, though, is the high costs of complex enzyme cocktails, contributing up to 30% of the total processing cost. To bring the bio-resource vision into fruition, this economic issue needs to be addressed. One approach could be to fuse enzymes involved in lignocellulose degradation to gain synergistic effects and make them more efficient, thus making the process cheaper by requiring less of the enzymes. Indeed, studies have shown that this approach can increase conversion rates [6-9].

Furthermore, multicatalytic or multifunctional enzymes have been shown to naturally evolve from individual enzymes in some biosynthetic pathways, suggesting that bringing together enzymes closely involved in bioprocesses may be advantageous [10]. Several strategies for mimicking this behaviour have been developed, such as encapsulation or covalent co-immobilization of enzymes, and gene fusion. Of the mentioned methods, one could argue that gene fusion is the simplest approach where the individual gene sequences are placed end-to-end while removing any intermediate stop codons, allowing for a single transcription and translation. Gene fusion is however very limited regarding the number of enzymes that can be feasibly fused, as the folding process becomes complicated very fast. Even a triple enzyme fusion might be too difficult for the cell to produce and problems with proteins precipitating is a possibility. This becomes a problem in the case of lignocellulose degradation, as oftentimes more than two enzymes are required [11]. Here, an alternative method is proposed to circumvent the gene fusion problem by using so-called split intervening proteins (inteins), which could enable *in vitro*-fusion of individually expressed proteins. To evaluate the functionality of split inteins for this application, two enzymes involved in the degradation of the hemicellulose xyloglucan will be fused. For any enzyme combination to be possible, a modular toolkit will also be developed.

1.2 Aims

This project aimed mainly to develop a modular toolkit for the construction of multi-catalytic enzymes by split intein fusion. It also aimed to evaluate the toolkit by fusing the enzymes *Bo*GH3B and *Bo*GH31A using two different split intein pairs, gp41-1 and VidaL, and setting up reactions with xyloglucan monomers. Additionally, the fusion enzymes were to be compared to the individual enzymes to see how the degradation efficiency was affected by the fusion. The final aims were to compare different configurations of the fused enzymes and to test the orthogonality of gp41-1 and VidaL.

1.3 Theoretical background

1.3.1 Composition and degradation of tamarind seed xyloglucan

Xyloglucan (XyG) is a major polysaccharide, more specifically hemicellulose, abundant in the cell walls of all land plants [12, 13]. Therefore, it may be regarded as a good target for degradation in lignocellulosic biowaste streams. The polymeric chain of XyG is held together by a β-1,4-D-glucan backbone, and attached to it are varying side chains depending on the type of XyG. Here, the focus is on tamarind seed XyG, which includes the side chains D-xylose and D-galactose. Unless otherwise specified, XyG will in this report imply the tamarind seed variant. The basal repeating structure of XyG is XXXG (see Figure 1A), where X signifies a xylosyl residue is linked to a glycosyl residue by an α-1,6 bond, and G indicates an unsubstituted glycosyl residue. Additionally, the second and third xylosyl residues may have galactosyl residues linked by a β-1,2 bond, denoted by an L (see Figure 1B-D). These structures, ranging from heptasaccharides to nonasaccharides will henceforth be referred to as XyG monomers.

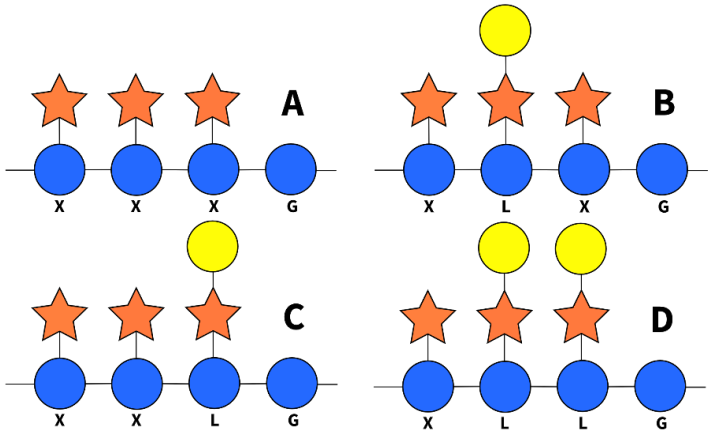


***Figure 1: Types of xyloglucan monomers**. The sugar residues in this figure are represented by standard symbols: glucosyl residues as blue circles, xylosyl residues as orange stars, and galactosyl residues as yellow circles [14]. (**A**) XXXG type XyG monomer: a heptasaccharide with four glucosyl residues in the backbone and three xylosyl residues. This structure may be referred to as the basal XyG monomer, and the sugar residues within it as the basal sugar residues. (**B**) XLXG type monomer: an octasaccharide with, in addition to the basal sugar residues, one galactosyl residue in the second position. (**C**) XXLG type monomer: an octasaccharide with, in addition to the basal sugar residues, one galactosyl residue in the third position. (**D**) XLLG type monomer: a nonasaccharide with, in addition to the basal sugar residues, two galactosyl residues in the second and third positions.*

To degrade XyG into its three monosaccharide building blocks, four types of enzymes are necessary [15]. Firstly, an *endo*-xyloglucanase cleaves the D-glucan backbone into XyG monomers by breaking the β-1,4 bonds between the glucosyl residues at the first and fourth positions (see Figure 2A). Secondly, a β-galactosidase cleaves the β-1,2 bond between the galactosyl and xylosyl residues. Lastly, an α-xylosidase and a β-glucosidase take turns in cleaving off the xylosyl and glucosyl residues, respectively, on the extreme non-reducing end of the XyG monomer (see Figure 2B). There are several different versions of these enzymes, originating from different species. The enzymes used in this project are *Bo*GH5A as the *endo*-xyloglucanase, *Cj*Bgl35A as the β-galactosidase, *Bo*GH31A as the α-xylosidase, and *Bo*GH3B as the β-glucosidase.

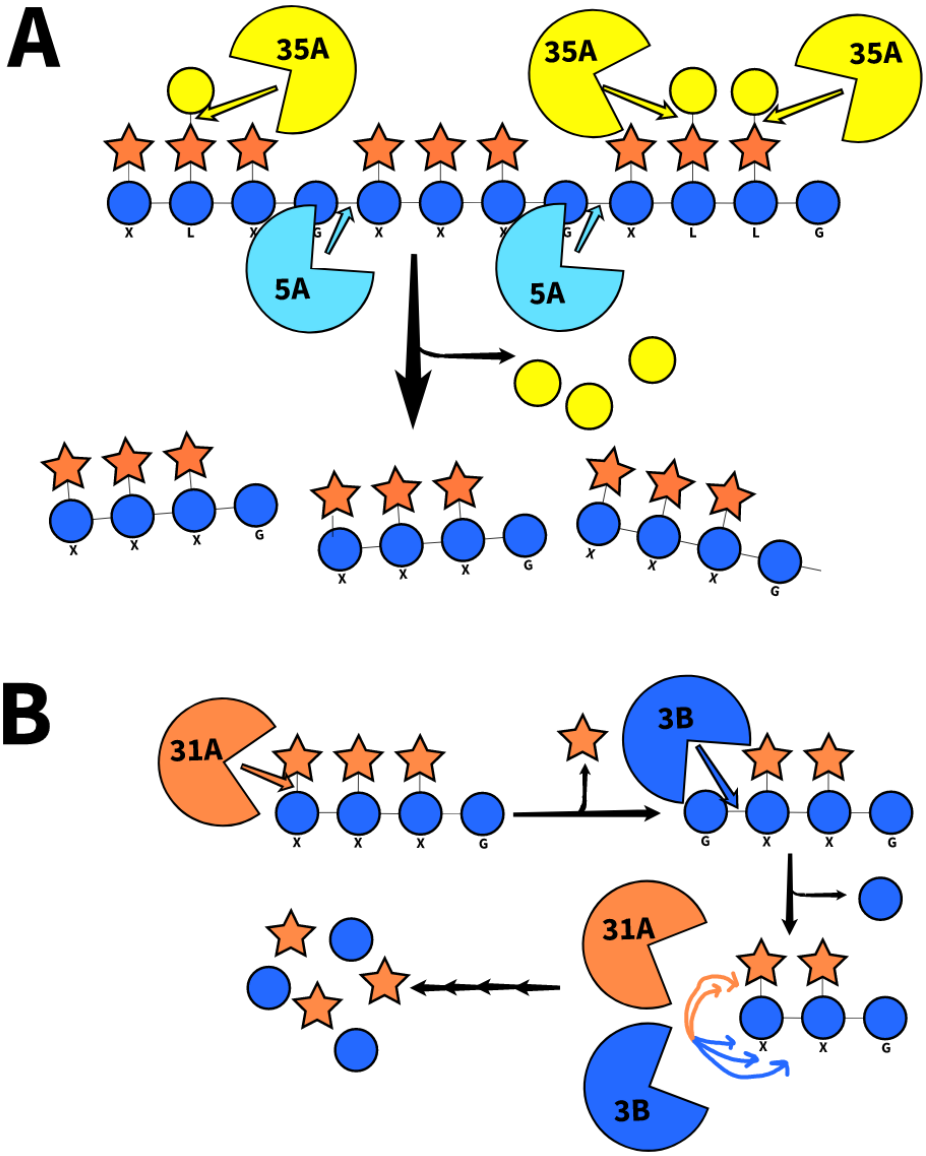


Figure 2: Tamarind xyloglucan degradation*. Four enzymes are involved in tamarind xyloglucan degradation: an endo-xyloglucanase (here BoGH5A or 5A), a β-galactosidase (here CjBgl35A or 35A), an α-xylosidase (here BoGH31A or 31A) and a β-glucosidase (here BoGH3B or 3B). (**A**) As a pre-treatment of the xyloglucan polysaccharide, it is subjected to BoGH5A and CjBgl35A to cleave the backbone and the galactose, respectively. The result is basal XyG monomers. (**B**) The xyloglucan monomers are finally subjected to BoGH31A and BoGH3B, which take turns in cleaving off xylose and glucose, respectively. In the end, only the monosaccharides xylose and glucose remain.*

The *Bo*GH enzymes all originate from *Bacteroides ovatus*, a human gut bacterium [15]. *Cj*Bgl35A originates from *Cellvibrio japonicus*, a bacterium first isolated from Japanese soil [16]. These specific enzymes were chosen due to already being characterized and having known functions. The nomenclature of these enzymes includes the species, enzyme family and number, and the variant. For example, *Bo*GH5A is variant A of an enzyme from glycoside hydrolase family 5, originating from *B. ovatus*. When discussing the genes or proteins in this project, they may be referred to as their full name or simply the enzyme family number and variant: 3B, 5A, 31A, and 35A.

1.3.2 Inteins and split inteins

Intervening proteins (inteins) are small segments of pre-proteins that can auto-excise themselves, while simultaneously ligating the flanking amino acid residues (exteins) in a process known as protein splicing [17]. The first intein was hinted to exist in 1988 when a gene in yeast was found to have regions homologous to genes coding for a vacuolar ATPase in both *Neurospora crassa* and carrot [18-20]. Although the expressed yeast protein had similar activity to that of the vacuolar ATPases, the open reading frame predicted a protein about 450 amino acids (AAs) larger. In 1990 this was investigated further where the yeast gene product and protein size were compared, and it was observed that the protein was about 450 AAs smaller than predicted [21]. Protein splicing was thus first theorized to exist, functioning like gene splicing where there are intervening segments that are spliced before the product is matured. In 1998, a split variant of inteins was discovered where the intein was split into an N-terminal intein (called C-intein) and a C-terminal intein (called N-intein), which were encoded by two separate genes [22]. The split inteins are inactive until they bind to each other after being translated; upon binding, the complete intein cuts itself out and repairs the break by forming a new peptide bond between the exteins. The N-terminal split intein is called a C-intein as it will be positioned on the C-terminal end of the whole intein once the two halves bind; the same goes for the C-terminal split intein. In biotechnological applications, split inteins may be used to create multi-catalytic enzymes by fusing different protein domains (see Figure 3).

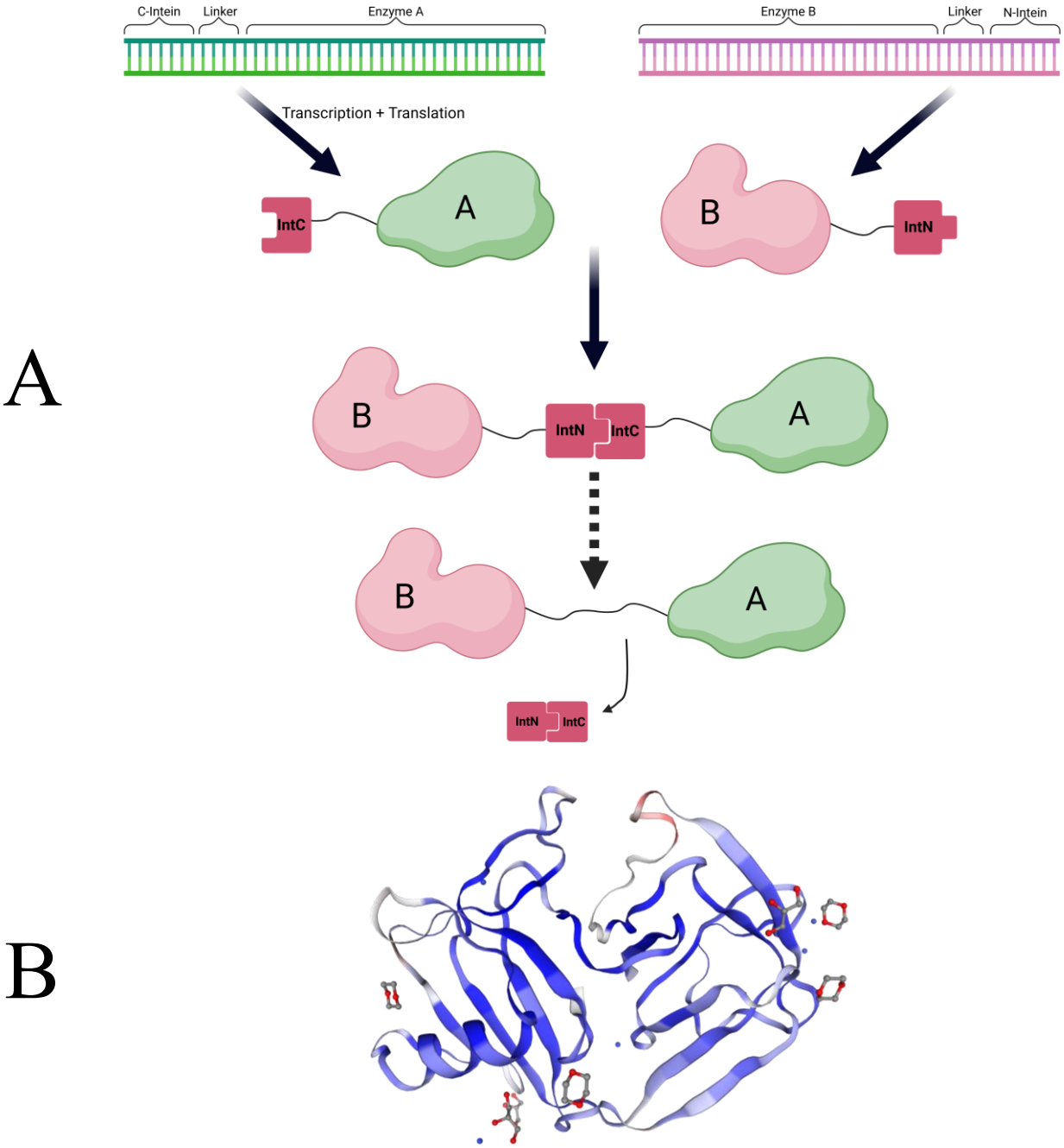


Figure 3: Bi-enzyme split intein fusion and crystal structure of intein. ***(A)*** *Two enzymes, A and B, are individually with a respective linker and split intein. The constructs are transcribed, translated, and then fused as the split inteins bind and auto-excise themselves. Since enzyme A is expressed with the C-intein (IntC), it will end up on the C-terminal end of the enzyme fusion. The opposite can be said for enzyme B, which is coupled with the N-intein (IntN). (Created with BioRender.com.)* ***(B)*** *The crystal structure of intein MjaTFIIB mini-intein (MjaTFIIBΔ155) [23].*

In this project, the bi-enzyme fusion was done with *Bo*GH3B and *Bo*GH31A for the degradation of basal XyG monomers. Although only a single split intein pair is needed for such a fusion, two will be evaluated: gp41-1 and VidaL. Gp41-1 was chosen for its high *trans*-splicing rate and yield, as well as tolerating a wide range of exteins [24]. VidaL was chosen for its high splicing rate and for its atypically short N-terminal fragment of 16 AAs which allows for an expanded range of N-extein cargoes [25]. Moreover, it has been observed that split inteins typically require up to three specific extein residues for an efficient post-translational splicing [25]. The residues for the N-intein are numbered -1, -2, -3,

and so on, while the C-intein residues are numbered in the positive. The native extein dependence for VidaL is ESGSGK (-3, -2, -1, +1, +2, +3), but the -3 and +3 extein residues may be changed to any AA. The native extein dependence for gp41-1 is SGYSSS [24, 26]. The orthogonality of the two inteins has not yet been tested but would be interesting to investigate, as it is required for simultaneous use in multi-enzyme fusions.

2. Methods

2.1 Methods background

2.1.1 Modular cloning

MoClo is short for modular cloning and several different MoClo toolkits exist for different organisms or purposes. One such toolkit was designed by Lee et al. for simultaneous multipart DNA assembly, and it is this toolkit that is considered when referring to MoClo (toolkit) in this work [27]. The MoClo method first involves cloning into *Escherichia coli* to mass-produce the plasmid containing the GOI(s), and then uses CRISPR-Cas9 technology to insert the GOI(s) into the yeast genome. In this project, only *E. coli* will be utilized which does not require CRISPR-Cas9. The MoClo toolkit contains 8 base types of parts and 96 parts in total, which are individually contained in plasmids named pYTK001 – pYTK096, respectively. pYTK001 is the so-called entry plasmid and contains a dropout green fluorescent protein (GFP) transcriptional unit. Whenever a new part is created, the entry plasmid is used as a backbone and the GFP is replaced with the respective GOI. In fact, all pYTK plasmids were entry plasmids at first, but have replaced GFP with whichever part that corresponds to that number. There are three levels of assemblies in MoClo. The first level is 0 but it is also called a part plasmid; it is created by ligating source DNA into pYTK001. The level 0 parts only contain one of the base types of parts, so it could be e.g., a promoter, terminator, or selection marker. One of each of the level 0 parts can be assembled to form a level 1 part. This is called a cassette plasmid, and it is a fully functional transcriptional unit. Multiple level 1 parts can be assembled into a level 2 part, called a multigene plasmid.

Except for pYTK001, none of the parts in the MoClo toolkit was utilized in this project. Instead, the synthetic biology principles used in it was applied to form new types of parts for the aims of this project. So, to understand how the parts of the intein-mediated protein fusion toolkit are designed, it is beneficial to know the MoClo part structure. When creating a new level 0 part, it is necessary to modify the template sequence to be compatible with MoClo. This is done using primers with special overhangs that contain one recognition site each for the type IIs restriction enzymes BsmBI and BsaI (see Figure 4) [27].

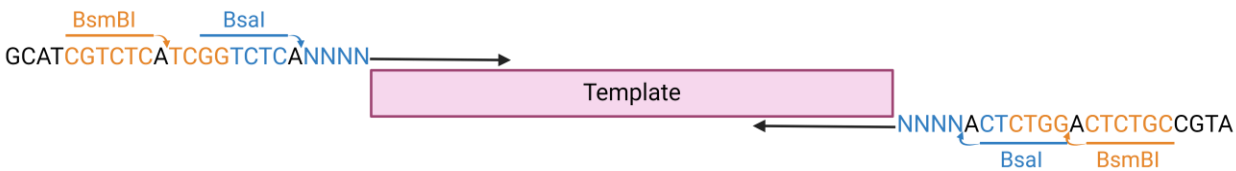


Figure 4: Creation of MoClo-compatible source DNA. *Forward and reverse primers used to amplify template DNA in preparation for assembling a new level 0 part. The primers consist of a complementary sequence to the template and an overhang which contains recognition sites for BsmBI and BsaI. The type-specific bases are denoted as 'NNNN'.*

Type IIs restriction enzymes cut outside their recognition site, which in combination with type-specific overhangs, enables a one-pot assembly [27, 28]. When making level 0 parts, BsmBI is used to enable ligation with the pYTK001 backbone; when making level 1 parts, BsaI is used to assemble several

level 0s in a specific order. For the intein-mediated protein fusion toolkit, these concepts were used to design custom level 0 parts and level 1 assemblies.

2.1.2 Immobilized metal ion affinity chromatography

Immobilized metal ion affinity chromatography (IMAC) is a protein purification technique based on interactions between an immobilized metal ion, in this case Ni(II), and electron donor atoms in the side chains of certain amino acid residues [29]. IMAC is usually performed in a fast protein liquid chromatography system (FPLC). As with any chromatographic method, IMAC involves the use of a stationary and mobile phase. Here, the stationary phase consists of Ni Sepharose™ excel resin packed inside HisTrap™ excel columns. The mobile phase here includes two different buffers, A and B, and *E. coli* lysate. Buffer A consists of 50 mM Tris, 250 mM NaCl and is set to pH 8; buffer B has the same reagents and pH, but also has 250 mM of imidazole. The separation of proteins occurs due to different affinities to the resin, so when running the cell lysate through the column, proteins that have more interactions with the resin bind more strongly and will be retained. To separate the protein of interest (POI) from other proteins, it is expressed with a hexa-histidine tag. The side chain of His is an imidazole, which has a high affinity to the resin [30]. By using a gradient elution program that slowly increases the percentage of buffer B, the number of unspecifically bound proteins (UBPs) in the elution fraction is reduced. The workflow for IMAC is described in the illustration below (see Figure 5).

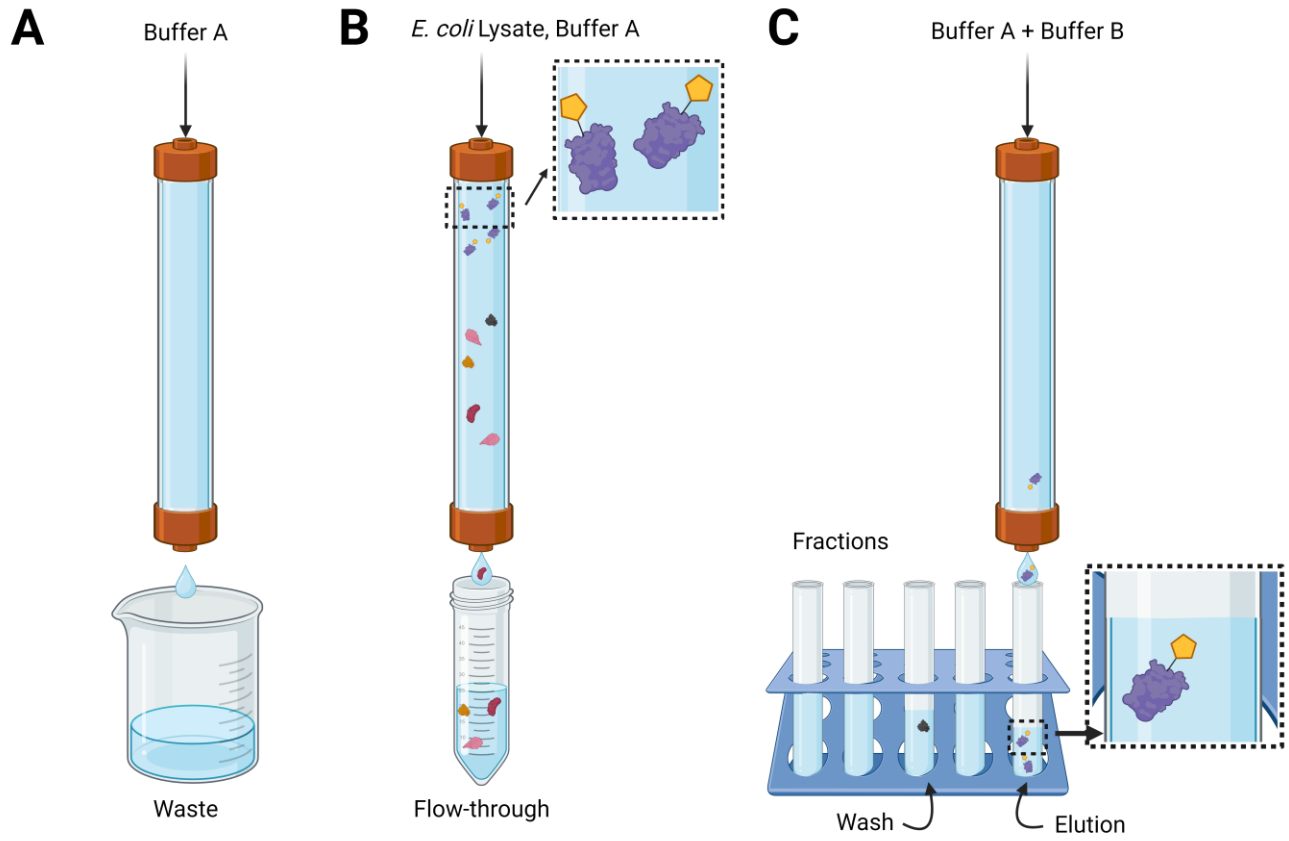


Figure 5: Schematic IMAC purification setup. ***(A)*** *Buffer A is run through the HisTrap™ excel column containing Ni Sepharose™ excel resin.* ***(B)*** *The E. coli lysate is loaded onto the column, and buffer A is used to push it through. The flowthrough is collected in a falcon tube. The POI is tagged with histidine (shown in the magnification) and will be retained in the column. Once most UBPs have passed the detector in the FPLC, the UV signal will drop, and the elution program is started.* ***(C)*** *The gradient elution program starts at 0% buffer B and increases it linearly to wash out any UBPs still bound to the resin. Presumably, these UBPs still have a lower affinity to the resin than the POI and will elute first. In the end of the protocol, the percentage amount of buffer B spikes rapidly to elute the POI. The proteins are detected by measuring the absorbance of the sample at 280 nm UV light. (Created with BioRender.com.)*

2.2 Project workflow

The methods described in the subsequent sections are in a summarized form. For more in depth descriptions of the protocols, see Appendices III – V.

2.2.1 *In silico* work – toolkit and primer design

Before beginning the lab-work, the toolkit was designed *in silico*, considering the desired order of the parts once they are assembled into level 1 parts. The type-specific overhangs were chosen mostly at random from the MoClo parts, but making sure that the 3' overhang of say part A matched the 5' overhang of part B. Also, the 5' overhang of the first part in the assembly was designed to include a start codon, while the last part of the assembly was designed to include a stop codon. The parts were designed in Benchling, an online platform for biotechnology research (www.benchling.com). The gene sequences for *Bo*GH3B, *Bo*GH5A, *Bo*GH31A, and *Cj*Bgl35A were imported into Benchling, and primers were designed to amplify them with specific overhangs to make them compatible with either In-Fusion cloning or MoClo. In-Fusion cloning is a quick and seamless cloning method that does not require digestion of the insert, thanks to the one-step annealing between the overhangs of the insert and the linearized vector [31].

2.2.2 Cloning and production of individual enzymes

In addition to assembling the fused enzymes, the individual ones need to be produced in part to act as pre-treatment of the xyloglucan substrate (*Bo*GH5A and *Cj*Bgl35A) and in part to compare with the split intein fusions (*Bo*GH3B and *Bo*GH31A). First, *B. ovatus* and *C. japonicus* were cultivated, the former anaerobically, and their genomic DNA (gDNA) were extracted. The *Bo*GH genes and *Cj*Bgl35A were PCR-amplified and purified from the respective gDNA using primers with In-Fusion overhangs. For verification of correct PCR products, their size was analysed by gel electrophoresis. Then, the genes were In-Fusion cloned separately into pET28a-TEVc-TIR fixed vectors and transformed by heat shock into stellar competent *E. coli* HST08 according to standard protocols [31, 32]. The transformed cells were plated on lysogeny broth (LB) agar plates (15 g/L agar) with 50 ng/mL neomycin for selection. Colonies were then inoculated in LB + 50 ng/mL neomycin and incubated overnight (ON) at 37 °C, whereafter the plasmids were extracted using the GeneJET plasmid miniprep kit. For verification of the plasmids, they were restriction digested with FastDigest enzymes and the fragments were analysed by gel electrophoresis according to standard protocols [33, 34]. The same verification and sequencing steps were used throughout the project for all PCR products and plasmids. A lot of troubleshooting was needed for *Cj*Bgl35A, for example using different DNA polymerases and gradient annealing temperatures when amplifying the gene from the gDNA. In the end, none of the colonies after transformation could be verified as having the correct insert. Later in the project, the 16S rRNA gene was amplified from the *C. japonicus* gDNA and sent for sequencing, and it turned out that the gDNA was not from *C. japonicus* but from an *E. coli* contaminant. Thus, *Cj*Bgl35A was disregarded for the remaining experiments.

After sequencing the vectors containing *Bo*GH3B, *Bo*GH5A and *Bo*GH31A, they were transformed into *E. coli* BL21 to be used as starting cultures for protein expression cultivations. The protein expression cultivation was based on an existing protocol [35]. After growing the BL21 cells at 37 °C in LB + 50 ng/mL neomycin to an OD_{600} of 0.5, 200 µL of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce gene expression, and the temperature was set to 16 °C. After one day, the cells were pelleted by centrifugation and the supernatant was discarded. The cells were then resuspended in buffer (20 mM HNa_2O_4P, 500 mM NaCl, 20 mM imidazole, pH 7.4), and stored at -20 °C. When it came time to purify the proteins, the cell suspensions were thawed, sonicated to break the cell membrane, and centrifuged. The supernatant was then used as the cell lysate for IMAC-

purification. The chromatograms from the IMACs were analysed to establish which fractions contained the wash and elution peaks. The wash and elution fractions, cell lysates, and flowthroughs for each protein, as well as the cell pellet of *Bo*GH31A were analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). The elution fractions of each protein were concentrated in centrifugal filters while adding buffer A to remove the imidazole.

2.2.3 Cloning and production of level 0 and level 1 parts

Most of the source DNA for the level 0 parts were ordered as forward and reverse oligos with flanking MoClo overhangs and then annealed. The C-intein of VidaL (VidaL-C) was ordered as DNA as it was too long. Gp41-1C and gp41-1N were available in plasmids in the lab, so they were PCR-amplified using primers with MoClo overhangs. Similarly, *Bo*GH3B and *Bo*GH31A were PCR-amplified from gDNA of *B. ovatus* using primers with MoClo overhangs. All the level 0 parts were then assembled by ligating each respective source DNA into pYTK001, after digesting both with BsmBI. The level 0 parts were then transformed into stellar competent cells and plated on LB + agar with 100 ng/mL chloramphenicol as pYTK001 has a gene that confers resistance to chloramphenicol. Through green/white screening, successful colonies were picked to be inoculated ON in LB + 100 ng/mL chloramphenicol, and miniprepped the following day.

For the level 1 assemblies, a vector suitable for protein expression was needed instead of the pYTK001 backbone. Therefore, the promoter, gene, and terminator for GFP was PCR-amplified in one continuous sequence from pYTK001 using primers with In-Fusion overhangs. Then, GFP was In-Fusion cloned into a pET28a-TEVc vector. Next, the plasmid was mutated by PCR to include a BsaI recognition site upstream of the GFP promoter, and another one downstream of the terminator. This vector was named pET28a_Mut2. When performing the level 1 assemblies, the same procedure as for the level 0 ones was used. Instead of pYTK001 as the backbone and BsmBI as the restriction enzyme, pET28a_Mut2 and BsaI were used, respectively. After transforming stellar competent cells with the level 1 assemblies, colony PCR (cPCR) was performed to verify the inserts. Correct colonies were cultured ON and miniprepped the next day. Like the individual enzymes, the level 1 assemblies were transformed into BL21, expressed through IPTG-induction, IMAC-purified, and analysed on SDS-PAGE. Due to time restraints, only a select few assemblies were chosen and a crude IMAC-purification using gravity columns was performed instead of using the FPLC. Fusion reactions were carried out by mixing two matching assemblies together at room temperature for 40 minutes, whereafter they were analysed by SDS-PAGE to determine if fusion occurred.

2.2.4 Evaluation of xyloglucan degradation

To evaluate *Bo*GH3B, *Bo*GH5A, and *Bo*GH31A's ability to degrade xyloglucan, reactions with different combinations of the enzymes were set up. *Bo*GH3B, *Bo*GH5A and *Bo*GH31A had concentrations of 8.75, 26.6, and 6.69 mg/mL, respectively. The reactions were done by mixing 1 µL of unfused enzymes, 0.1 g/L XyG and phosphate buffer (50 mM, pH 7) and incubating at 37 °C and 1000 rpm for 16 h. The enzyme combinations were 5+3+31, 5+3, 5+31, and then 5, 3, 31 on their own. The mono and oligosaccharides were analysed by ion chromatography, using a PA200 column and standards of xylose and glucose.

3. Results and discussion

3.1 Design of the modular toolkit

The toolkit that was developed includes eight types of parts: a His tag, split intein, linker, and a connector for both the C-terminal and N-terminal side of the protein of interest (see Figure 6). The part names were abbreviated to His, Int, Link, Con and then C or N, depending on the side of the POI. For example, LinkC is the linker to the C-terminal protein and IntN is the N-intein.

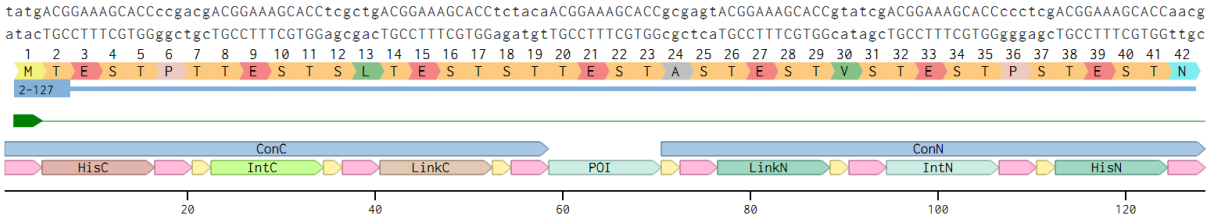


Figure 6: Overview of toolkit parts. *A model view of how the toolkit parts fit together in a level 1 assembly, where every part is represented by a 'TEST' amino acid sequence. All the parts and the POI are annotated with their abbreviated name, but there are also pink and yellow annotations between them. The pink annotations are the four base-long BsaI overhangs that are specific to each part type (See Appendix I for a full list). The yellow annotations are bases that were added to fix the reading frame.*

This design allows, in theory, for any number of enzymes or proteins to be fused if there are enough split inteins to facilitate the fusions. For a bi-enzyme fusion, as was the case with *Bo*GH3B and *Bo*GH31A in this project, the required assemblies would be:

(i) HisC – IntC – LinkC – *Bo*GH3B – ConN, and
(ii) ConC – *Bo*GH31A – LinkN – IntN – HisN.

Note that the enzymes could be swapped to control which enzyme will end up on which side once they fuse. Assemblies (i) and (ii) would produce an enzyme fusion of *Bo*GH3B and *Bo*GH31A with the former towards the C-terminus and the latter towards the N-terminus of the fusion protein (see Figure 7).

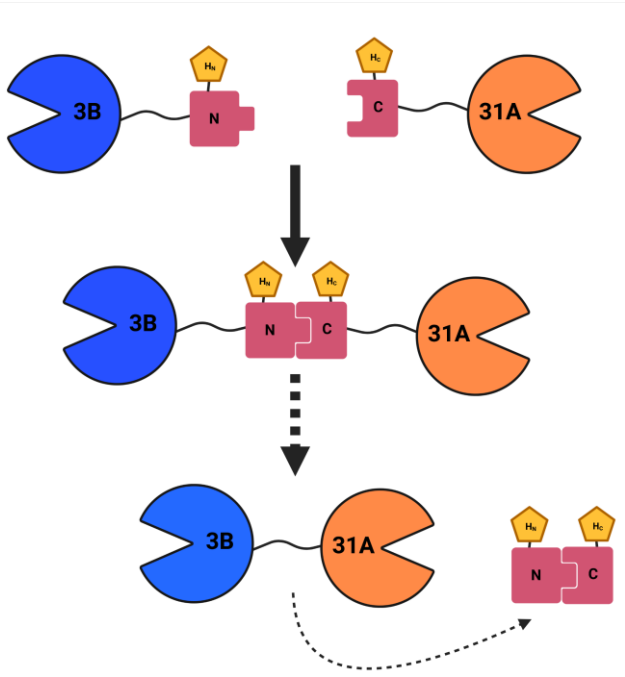


Figure 7: Split Intein fusion of BoGH3B and BoGH31A. *After assembling, cloning, expressing, and purifying assemblies (i) and (ii), they are mixed and incubated at room temperature. The split inteins fuse, auto-excise themselves and ligate the exteins, resulting in a fusion protein. (Created with BioRender.com.)*

For constructing fusion enzymes of three or more protein domains, the outer-most enzymes would be assembled in the same configuration as (i) and (ii). The inner enzymes, however, would need to be assembled on the form:

(iii) HisC – IntC – LinkC – Inner Enzyme – LinkN – IntN – HisN.

Since each split intein fuses with its other half, this needs to be taken into consideration when creating multi-enzyme fusions (see example in Appendix I). Now, the purpose of each part will be discussed. Each part was designed and ordered with MoClo overhangs to be cloned into pYTK001 and pET28a_Mut2 by BsmBI and BsaI assembly, respectively. The His tags are included to enable IMAC-purification, and to anti-purify the fused enzymes. Since the split inteins with the His tags disconnect when fused, the successfully fused enzymes would not bind to the column and would come out in the flowthrough. As described in Figure 6, each part has a type-specific 5' and 3' BsaI overhang, which is to enable level 1 assembly in a specific order. HisC and ConC both include the start codon in its 5' overhang to serve as a starting point for the transcription, while the sequences of HisN and ConN include a stop codon. The purpose of the connectors is to replace the His – Int – Link segment when the POI is not going to be connected to another enzyme on that end. ConC was created by copying a part of the pET28a-TEVc-T7-TIR fixed vector with small, non-reactive AAs and ConN starts with a stop codon so the bases after are trivial.

The split intein parts are of course included to enable split intein-mediated fusion of the enzymes, there are however some considerations to be made. For a bi-enzyme fusion, a single split intein pair is needed. Then, for every additional enzyme another pair is necessary, meaning that for a fusion of N enzymes, N – 1 split intein pairs are needed. Another criterion for the fusions to work is that all the split inteins need to be orthogonal, meaning they are not cross-reactive and can only fuse to their respective halves and not to any other split intein. In the assembly, some nucleotides were added to correct the reading frame (see Figure 6). When bases flanking the IntC and IntN were added, they were chosen to produce fitting AAs. Since both the -3 and +3 extein residues for gp41-1 were serine, and any AA was possible in the same positions for VidaL, these added bases were chosen to produce an S together with one base from the respective BsaI overhang. For the other bases that were added that were not in connection to the extein dependencies, the idea here was to produce small, non-polar and non-reactive AAs to not interfere with the function of the other parts. Finally, flexible $(Gly_4Ser_1)_3$ linkers were chosen as they were documented to possibly improve folding and stability [36]. Some of the annotated toolkit parts can be found in Appendix II.

3.2 Purification of unfused enzymes and evaluation of their XyG degradation

In this section, the results for the unfused enzymes will be presented and discussed. These experiments were performed to later compare with the split intein fused versions. *Bo*GH3B, *Bo*GH5A and *Bo*GH31A were successfully cloned into the pET28a-TEVc-T7-TIR fixed vector, while *Cj*Bgl35A was not. It turned out that the *C. japonicus* colony which was provided in the lab was an *E. coli* contamination, as confirmed by 16S sequencing. This was discovered late in the project, so due to time restrictions it was not remedied. Moreover, *Bo*GH3B, *Bo*GH5A and *Bo*GH31A were expressed in *E. coli* BL21, extracted, and IMAC-purified. The IMAC chromatograms for all three enzymes are shown below (see Figure 8).

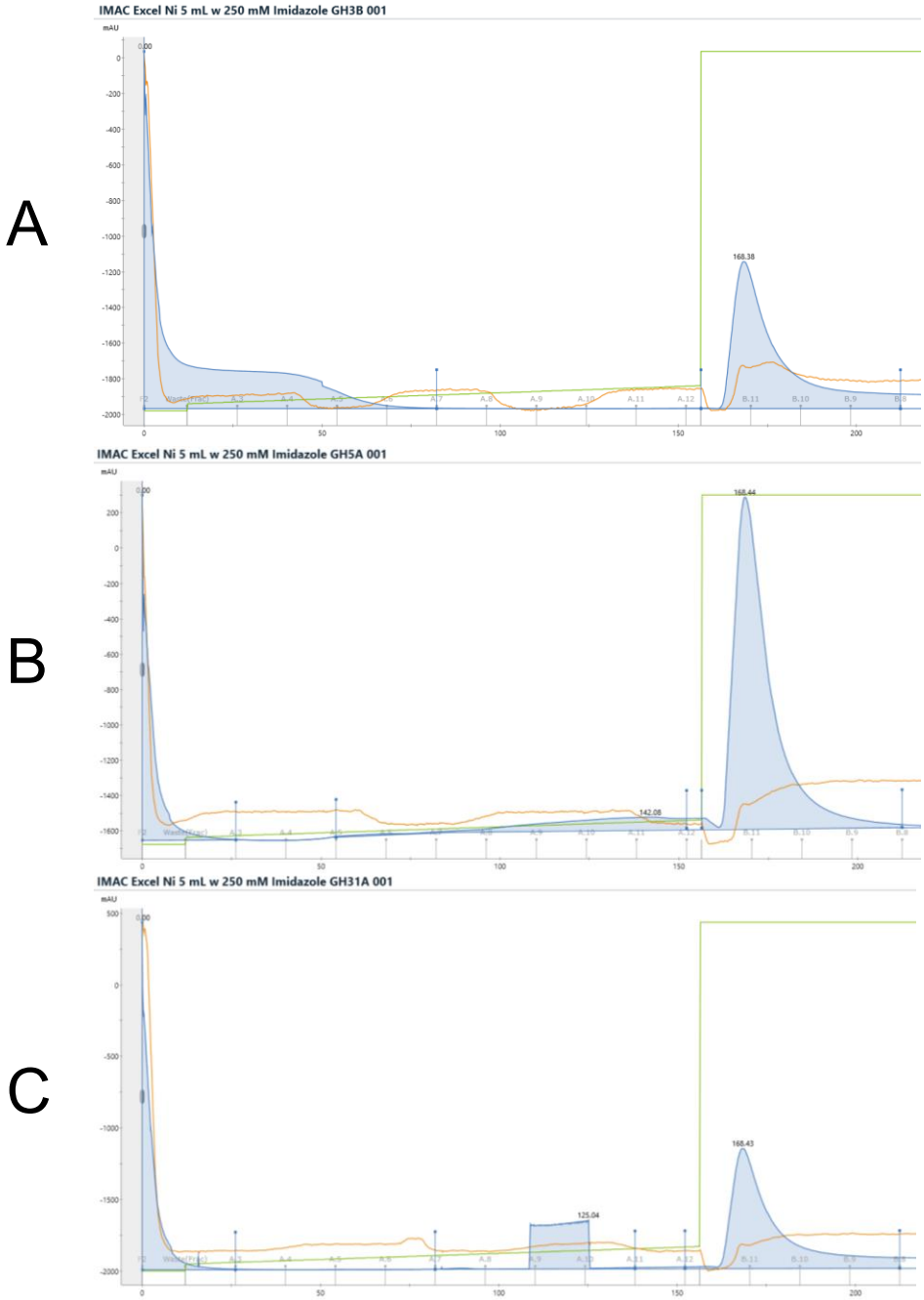


Figure 8: IMAC chromatograms for (A) BoGH3B, (B) BoGH5A and (C) BoGH31A. *The chromatograms depict the absorbance (mAU) at UV_{280} as a function of the volume (mL), shown by the blue curve. The green curve displays the concentration of buffer B, while the orange curve shows the conductivity. The same program and columns were used for all three runs: beginning at 0% buffer B and increasing it linearly until approximately 150 mL of mobile phase has passed through the column, then drastically increasing it to 100% to elute the POI. The fractions were collected in the order A1-A12, then B12-B1.*

For all three proteins, the bulk of the elution peak is contained within fractions B12 and B11. For *Bo*GH3B, there is a relatively large number of proteins exiting the column in the beginning when the % of buffer B and thus imidazole is very low, but there is still a prominent elution peak at B12 and B11. This might be due to variability in the proteins expressed by *E. coli*'s affinity to the column or an effect of the proteins binding to the POI. For *Bo*GH5A, there was a more expected pattern of a slowly increasing absorbance during the linear increase of buffer B, meaning more proteins are eluted as the amount of imidazole in the mobile phase increases. For *Bo*GH31A, there are seemingly no proteins being eluted during the linear gradient except towards the end when a lot of proteins are

suddenly detected in fraction A9. Since all three *Bo*GH enzymes were expressed using the same type of vector and *E. coli* strain, would they be removed the proteome should be identical. Therefore, the difference between the cell lysates is rather peculiar. There is a possibility that there were some protease activities in some of the samples, leading to partially denatured proteins or loss of His tags, which could affect how they are eluted and thus detected.

Some of the interesting fractions, in addition to the cell lysates, flowthroughs and one of the cell pellets, were analysed by SDS-PAGE to determine whether the correct proteins had been expressed and purified (see Figure 9).

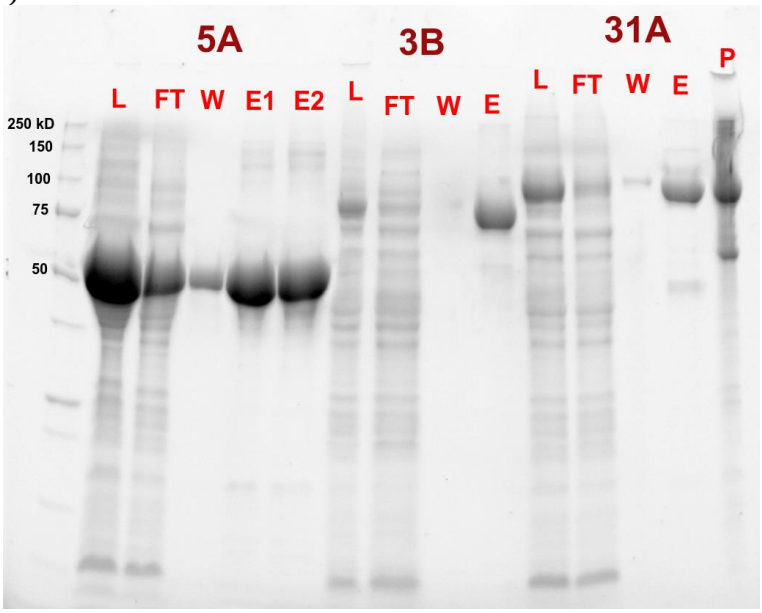


Figure 9: SDS-PAGE of individual BoGH enzymes. *The samples are lysates (L), flowthroughs (FT), wash fractions (W), elution fractions (E), and a cell pellet (P) from the individual proteins BoGH3B, BoGH5A, and BoGH31A. For BoGH5A, fraction A11 was used for the wash and B12 and B11 for the elution peak. For BoGH3B, A2 was used for the wash and B12 for the elution peak. For BoGH31A, A9 was used for the wash and B12 for the elution peak.*

The expected protein molecular weights were attained by analysing the sequences as translations in Benchling, where it was determined that *Bo*GH5A, *Bo*GH3B, and *Bo*GH31A ought to have molecular weights of approximately 54.21, 87.85, and 110.0 kDa, respectively. Comparing these to the SDS-PAGE, the elution fractions seem to have a large concentration of their respective POI. A general trend that can be seen is that the POI also can be found in the lysate in a large quantity, which is expected as the lysate should contain all proteins expressed by the cell. *Bo*GH5A is very prominent even in the flowthrough and wash, which might be a result of a compromised His tag due to protease activity or column overload. Since the POIs were very prominent in their respective elution fractions and relatively pure (no other strong bands), the pellets, lysates, and flowthroughs were discarded. The elution fractions were transferred to centrifugal filters, concentrated, and buffer-exchanged to remove imidazole. The measured concentrations afterwards were 8.75, 26.6, and 6.69 mg/mL for *Bo*GH3B, *Bo*GH5A, and *Bo*GH31A, respectively.

Reactions of different, unfused enzyme combinations with xyloglucan were set up and analysed by ion chromatography. The reactions were made by mixing 1 µL of enzymes(s) with 0.1 g/L or 0.01 g/L XyG and phosphate buffer to 500 µL total volume. The reactions were incubated at 37 °C and 1000 rpm for 16 h. It was decided that equal amounts of enzymes for the reactions was unnecessary, as the released monosaccharides and oligosaccharides were not to be quantified. The graphs used in this

report are from the reactions with 0.1 g/L XyG. First up are the samples with single enzymes (see Figure 10).

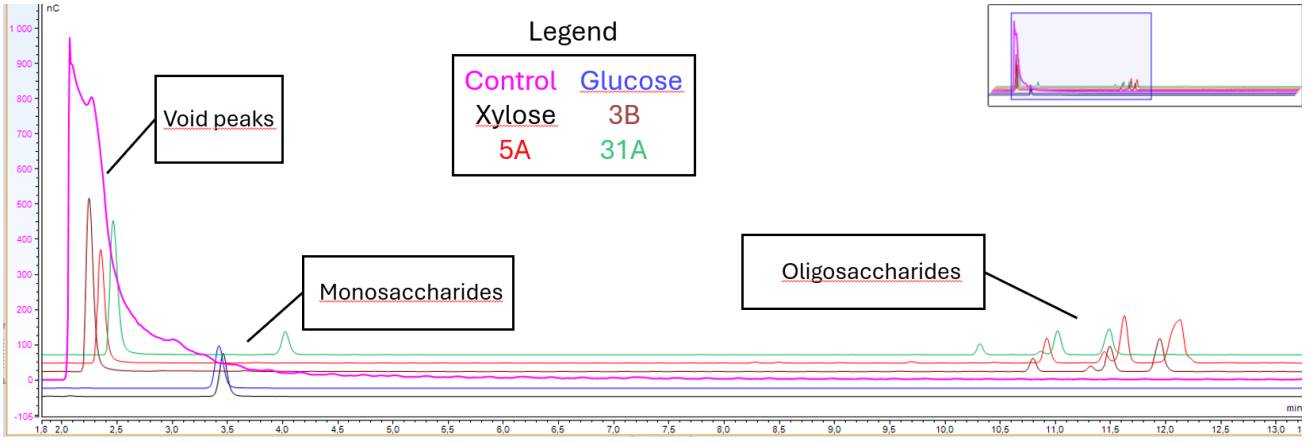


Figure 10: Chromatogram of single-enzyme reactions. *The chromatogram includes the three single-enzyme reactions of BoGH3B, BoGH5A, and BoGH31A, as well as a control (buffer with 0.1 g/L XyG), and glucose and xylose standards (100 µM). When comparing peaks of the same molecule, a larger peak means a higher concentration of the eluted compound.*

The reactions and the control all have a big peak around 2 min which is likely the void peaks that contain everything in the reactions that was not retained in the column. Here, it is likely that for example the buffer gets detected. Moreover, it was observed that xylose was released in the reaction with *Bo*GH31A (see Figure 11).

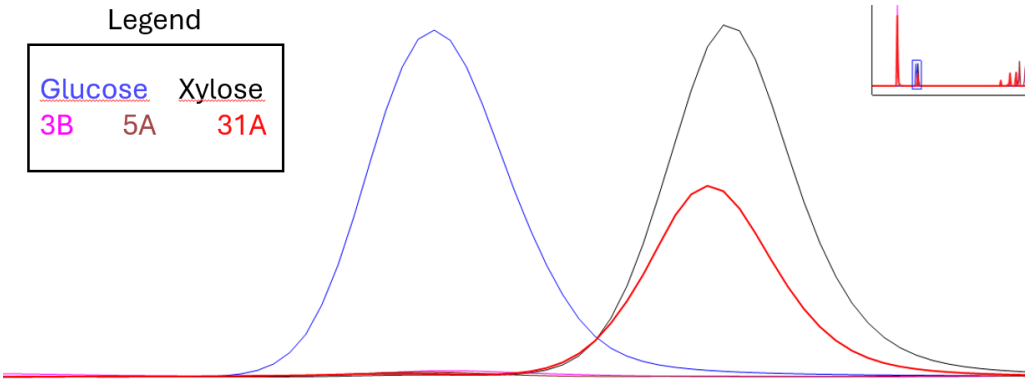


Figure 11: Released xylose and glucose for single-enzyme reactions. *BoGH31A was the only single-enzyme reaction that had any monosaccharide released.*

This result was not expected as the polymeric chain of XyG had not been fragmented by the *endo*-xyloglucanase yet. A full list of the possible sugars for each reaction is provided in Table 1.

Table 1: *Possible sugars in the XyG degradation reactions.*

Reaction	**Possible sugars**
*Bo*GH3B	none
*Bo*GH5A	XXXG, XLXG, XXLG, XLLG
*Bo*GH31A	none
*Bo*GH5A + *Bo*GH3B	XXXG, XLXG, XXLG, XLLG
*Bo*GH5A + *Bo*GH31A	Xyl, GXXG, GLXG, GXLG, GLLG
*Bo*GH5A + *Bo*GH3B + *Bo*GH31A	Glc, Xyl, LG, LXG, LLG

The result in Figure 11 suggests that the XyG was either already fragmented into oligosaccharides or *Bo*GH5A was mistakenly added to all the single-enzyme reactions. Since the control did not contain any oligosaccharides, the latter explanation is the most likely. When looking at the larger sugar residues in the reactions (see Figure 12), the results agree with this explanation as well.

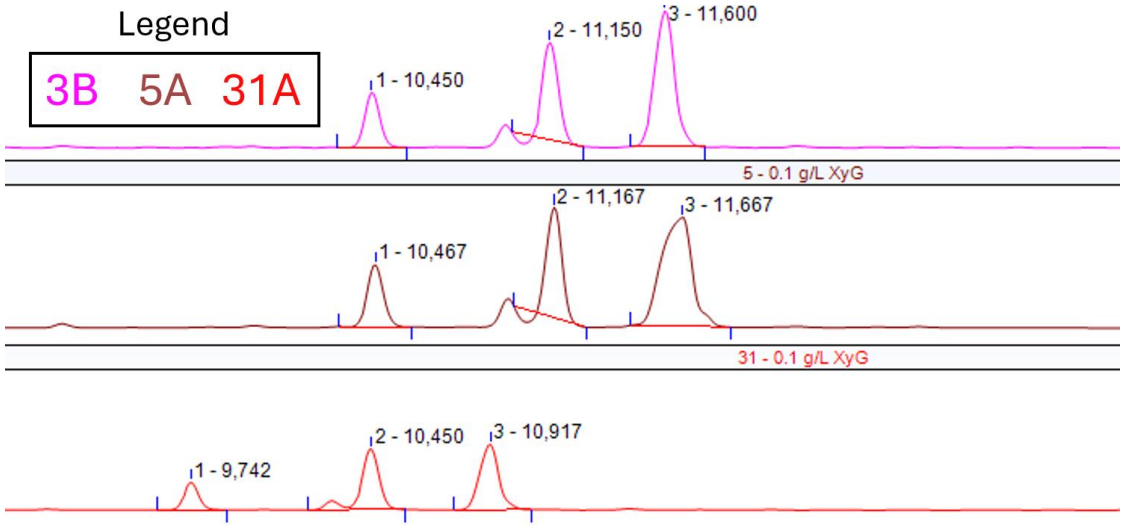


Figure 12: Sugar residues detected for single-enzyme reactions. *From top to bottom: BoGH3B, BoGH5A, BoGH31A. The BoGH3B and BoGH5A reactions seem to have the same residues, while BoGH31A has smaller ones.*

The chromatograms looked identical in the oligosaccharide region for *Bo*GH3B and *Bo*GH5A, while *Bo*GH31A had slightly smaller ones as inferred by the shorter retention times. From the possible sugars in Table 1, the only reaction here that should have any oligosaccharides is *Bo*GH5A. Four peaks were observed for this reaction, consistent with the sugar residues generated from degrading XyG into its four types of monomers. The same four peaks were detected in the *Bo*GH3B reaction. Reiterating the theory that *Bo*GH5A was mistakenly added to all single-enzyme reactions, if the *Bo*GH3B reaction in fact was a *Bo*GH5A + *Bo*GH31A reaction, the generated sugars would be the same as for the *Bo*GH5A reaction according to Table 1. This conclusion gains support when comparing the single-enzyme and bi-enzyme reactions (see Figure 13).

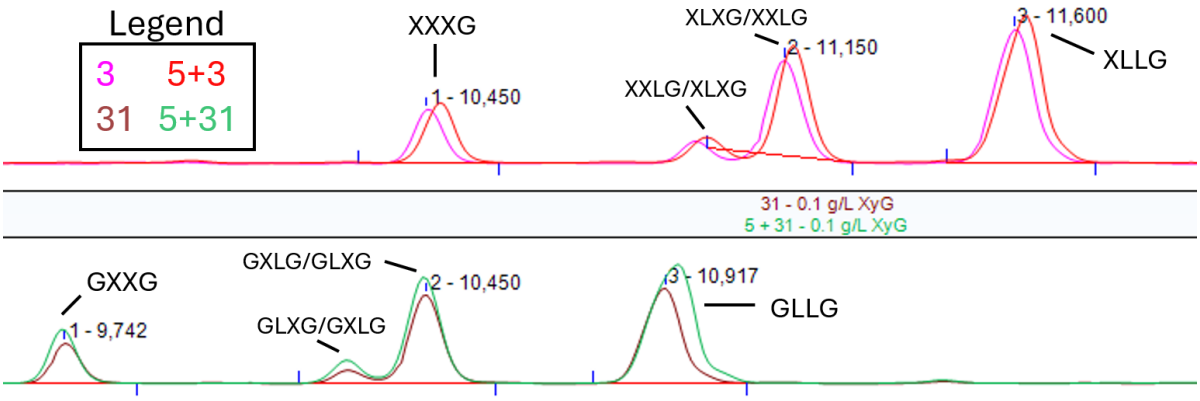


Figure 13: Sugar residues detected for single-enzyme and bi-enzyme reactions. *Four oligosaccharides were detected for the BoGH5A + BoGH3B reaction, and four slightly smaller oligosaccharides were detected for BoGH5A + BoGH31A. The identical sugars were observed for the reactions with only BoGH3B or BoGH31A, respectively.*

The sugar peaks in the bi-enzyme reactions look and elute identically to the respective single enzyme reactions. It is therefore considerably safe to say that the proposed explanation is true, but one would

have to redo the reactions entirely to be sure. The last IC result is from all three enzymes combined (see Figure 14 and Figure 15).

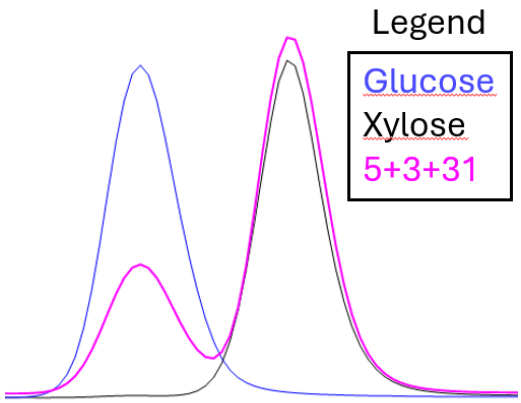


Figure 14: Released xylose and glucose for the tri-enzyme reaction. *For the combined reaction of BoGH5A, BoGH3B, and BoGH31A there is a release of both glucose and xylose, with the latter being in a higher concentration.*

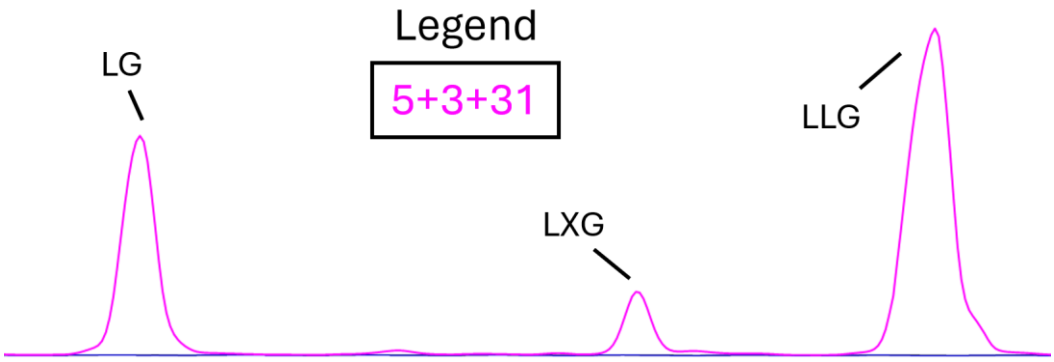


Figure 15: Oligosaccharides detected for the tri-enzyme reaction. *For the combined reaction of BoGH5A, BoGH3B, and BoGH31A there are three oligosaccharides detected, apart from xylose and glucose.*

According to Table 1, for the tri-enzyme reaction of *Bo*GH5A, *Bo*GH3B and *Bo*GH31A both xylose and glucose should be released, and three other sugars should be detected. This is what was observed in the chromatograms, so the three sugar residues were presumably LG, LXG, and LLG. This was supported by comparing their elution times to those measured for the sugar residues in the bi-enzyme reactions. These results seem to confirm the glycoside hydrolase activity of the *Bo*GH enzymes.

3.3 Split intein fusions

After assembling and sequencing the level 0 parts, the results were mostly good. All level 0 parts seemed correct, except for gp41-1C, gp41-1N, and *Bo*GH31A. For the enzyme, it seemed as if there was no insert at all. For the split inteins, only the sequence where the primer anneals to the template was correct, while the rest of the bases were either mismatched or undetermined. It was later discovered that there existed two different plasmids for each of the gp41-1 split inteins in the lab. The nucleotide sequences differ, while the amino acid sequence is the same. There is therefore a possibility that the wrong plasmid was chosen, resulting in poor primer annealing, however when running the PCR product on a gel the size seemed to be correct. As no cPCR was performed for the level 0 parts, a possibility is that the colonies picked still had the dropout GFP but was not noticeable enough under UV light. In hindsight, the sequencing results should not have been overlooked, but the decision was

made to continue with level 1 assemblies anyway. Eight different level 1 assemblies were created to evaluate different configurations of *Bo*GH3B and *Bo*GH31A and the function of the split inteins (see Table 2).

Table 2: *Level 1 assemblies for split intein fusions of BoGH3B and BoGH31A.*

Assembly #	Level 0 parts	Configuration, intein
1	HisC – VidC – LinkC – *Bo*GH31A – ConN	3B – 31A, VidaL
2	ConC – *Bo*GH3B – LinkN – VidN – HisN	
3	HisC – VidC – LinkC – *Bo*GH3B – ConN	31A – 3B, VidaL
4	ConC – *Bo*GH31A – LinkN – VidN – HisN	
5	HisC – Gp41-1C – LinkC – *Bo*GH31A – ConN	3B – 31A, Gp41-1
6	ConC – *Bo*GH3B – LinkN – Gp41-1N – HisN	
7	HisC – Gp41-1C – LinkC – *Bo*GH3B – ConN	31A – 3B, Gp41-1
8	ConC – *Bo*GH31A – LinkN – Gp41-1N – HisN	

Note that the orthogonality of VidaL and gp41-1 also may be tested, by for example mixing Assembly 3 with Assembly 6. During cPCR screening, it was evident that the level 1 assembly success rate was rather low for the ones containing *Bo*GH31A. This was not unexpected, as the sequencing of the level 0 *Bo*GH31A could not confirm its presence in the plasmid. Due to time-demanding cPCR screening and a lack of time overall, the decision was made to continue with fusing assemblies 2+3, and 6+7, which all had *Bo*GH3B as the POI. Sequencing showed that assembly 3 was successful, while assembly 2 was inconclusive, and assemblies 6 and 7 were correct except for the gp41-1 sections. Since these fusions would contain two *Bo*GH3Bs, a comparison to the unfused enzymes' degradation of xyloglucan was not possible. Instead, they were expressed in *E. coli* BL21, IMAC-purified, fused, and analysed on SDS-PAGE. The IMAC-purification of the level 1 assemblies was done crudely to save time, so no chromatograms were produced. After purification, the assemblies were checked by SDS-PAGE (see Figure 16).

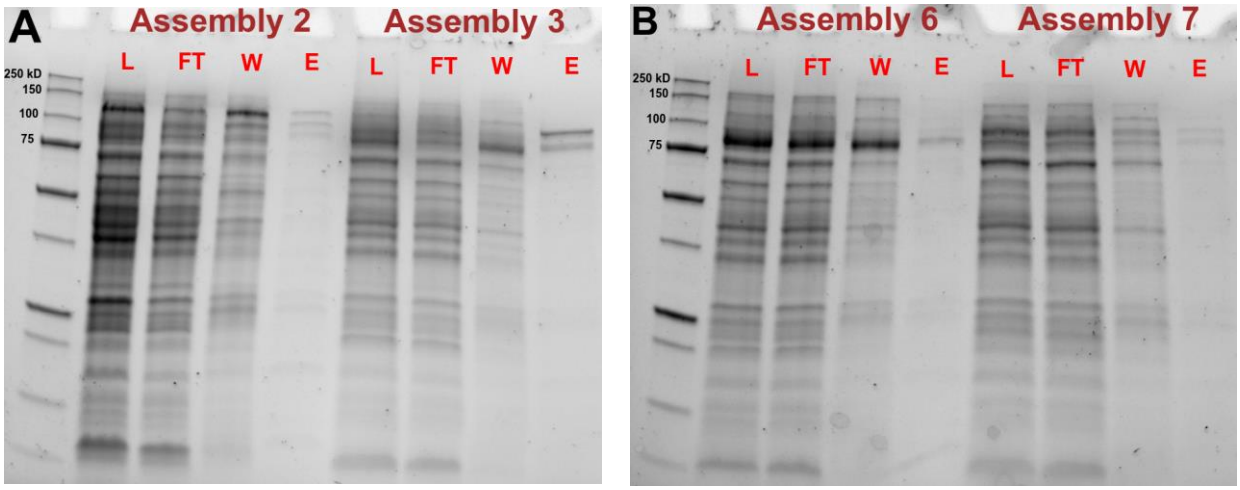


Figure 16: SDS-PAGE of level 1 assemblies. *Lysate, flowthrough, wash and elution fractions for assemblies* ***(A)*** *2 and 3* ***(B)*** *6 and 7.*

The expected protein molecular weights were 90.3, 102, 98.4, and 92.9 kDa for assembly 2, 3, 6 and 7, respectively. While the correct proteins seem to be present in the elutions, the bands are very weak compared to the SDS-PAGE of the individual enzymes. The presumed correct proteins have stronger bands in the wash, which might suggest a loss of the His-tags. Therefore, when preparing the split intein fusion reactions, the wash fraction was included. The fusions tested were assembly 2+3, 6+7, and 3+6 using only the elution fractions, and assembly 2+3 and 6+7 using a mix of equal amounts of the elution and wash fractions (see Figure 17). The 2+3 and 6+7 reactions were done to test the fusion

of VidaL and gp41-1, respectively. The 3+6 reaction was done to test the orthogonality of VidaL-C and gp41-1C.

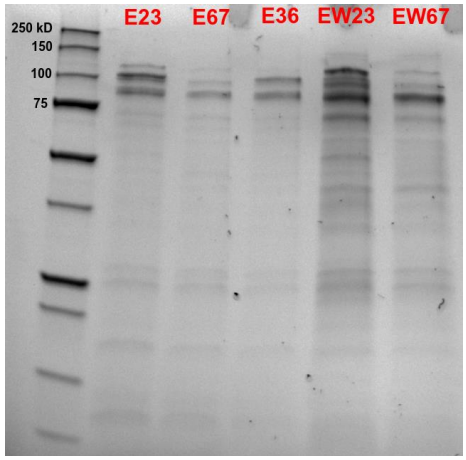


Figure 17: SDS-PAGE of split intein fusions. *E stands for elution fraction, EW stands for elution + wash fraction, and the numbers correspond to the assembly combinations.*

An approximative protein molecular weight for the fusions can be made by doubling the *Bo*GH3B molecular weight, which comes out to 176 kDa. The linkers also contribute to the molecular weight, but this number serves as a base test for seeing if the fusions have occurred. When looking at the SDS-PAGE results, it is evident that none of the fusions have succeeded because they do not go over 150 kDa. There may be many reasons as to why the fusions did not work. Firstly, the 6+7 fusion revolves around gp41-1 which may not be present in the level 0 plasmid. That does not explain why 2+3 did not fuse though, but here the protein expression might be a factor. The starting OD_{600} was not optimal for the cultivations because the volume of starter culture was mistakenly halved. Furthermore, the growth was not homogenous between the cultures so the OD_{600} was not optimal at the time of addition of IPTG either. This in addition to the crudely done IMAC-purification might have caused a great decrease in heterologously produced protein, causing the enzyme fusions to be non-detectable in the SDS-PAGE. Another variable to consider is the linkers, as they might have been too short to enable proper separation of the split intein and POI.

3.4 Concluding remarks and future outlook

This project aimed to develop and evaluate a toolkit for the construction of multi-catalytic enzymes using split intein-mediated fusion. A modular toolkit was developed, and from the level 1 assemblies' sequencing results it was shown to be functional. The toolkit could not be further evaluated though, as no fusions could be observed *in vitro*. The configuration of *Bo*GH3B and *Bo*GH31A and the orthogonality of gp41-1 and VidaL could not be tested either. To continue this project, the focus could first be on successfully cloning and assembling the level 0 parts for gp41-1C, gp41-1N, and *Bo*GH31A. For *Bo*GH31A it is a matter of re-doing the BsmBI assembly and screen for correct colonies, while for the split inteins, new primers that align to the in-lab plasmids would have to be designed. The protein expression for the level 1 assemblies was far from optimal, so it could benefit from some optimization by e.g., trying different IPTG concentrations, adding IPTG at different OD_{600}, or changing the cultivation time. It would also be beneficial to try out different positions of the hexa-histidine tag, as it is unknown whether it could affect the function of the split inteins. Furthermore, the linkers could be varied by changing the length and or the rigidity. Finally, it would also be interesting to see what enzyme combination and configuration would be optimal for XyG degradation, including all four enzymes and not only *Bo*GH3B and *Bo*GH31A.

References

[1] M. Bugge, T. Hansen, and A. Klitkou, "What Is the Bioeconomy? A Review of the Literature," *Sustainability,* vol. 8, no. 7, p. 691, 2016, doi: 10.3390/su8070691.

[2] H. Ohara, "Biorefinery," *Applied Microbiology and Biotechnology,* vol. 62, no. 5-6, pp. 474-477, 2003, doi: 10.1007/s00253-003-1383-7.

[3] A. Patel and A. R. Shah, "Integrated lignocellulosic biorefinery: Gateway for production of second generation ethanol and value added products," *Journal of Bioresources and Bioproducts,* vol. 6, no. 2, pp. 108-128, 2021, doi: 10.1016/j.jobab.2021.02.001.

[4] F. R. Amin *et al.*, "Pretreatment methods of lignocellulosic biomass for anaerobic digestion," *AMB Express,* vol. 7, no. 1, 2017, doi: 10.1186/s13568-017-0375-4.

[5] A. K. Chandel, V. K. Garlapati, A. K. Singh, F. A. F. Antunes, and S. S. Da Silva, "The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization," *Bioresource Technology,* vol. 264, pp. 370-381, 2018, doi: 10.1016/j.biortech.2018.06.004.

[6] F. S. Aalbers and M. W. Fraaije, "Enzyme Fusions in Biocatalysis: Coupling Reactions by Pairing Enzymes," *ChemBioChem,* vol. 20, no. 1, pp. 20-28, 2019, doi: 10.1002/cbic.201800394.

[7] S. Elleuche, "Bringing functions together with fusion enzymes—from nature's inventions to biotechnological applications," *Applied Microbiology and Biotechnology,* vol. 99, no. 4, pp. 1545-1556, 2015, doi: 10.1007/s00253-014-6315-1.

[8] R. Brunecky *et al.*, "Revealing Nature's Cellulase Diversity: The Digestion Mechanism of *Caldicellulosiruptor bescii* CelA," *Science,* vol. 342, no. 6165, pp. 1513-1516, 2013, doi: 10.1126/science.1244273.

[9] E. Glasgow, K. Vander Meulen, N. Kuch, and B. G. Fox, "Multifunctional cellulases are potent, versatile tools for a renewable bioeconomy," *Current Opinion in Biotechnology,* vol. 67, pp. 141-148, 2021, doi: 10.1016/j.copbio.2020.12.020.

[10] D. T. Monterrey, I. Ayuso-Fernández, I. Oroz-Guinea, and E. García-Junceda, "Design and biocatalytic applications of genetically fused multifunctional enzymes," *Biotechnology Advances,* vol. 60, p. 108016, 2022, doi: 10.1016/j.biotechadv.2022.108016.

[11] A. Kumar and R. Chandra, "Ligninolytic enzymes and its mechanisms for degradation of lignocellulosic waste in environment," *Heliyon,* vol. 6, no. 2, p. e03170, 2020, doi: 10.1016/j.heliyon.2020.e03170.

[12] K. Nishinari, M. Takemasa, Y. Suzuki, and K. Yamatoya, "Xyloglucan," Elsevier, 2021, pp. 317-365.

[13] A. Schultink, L. Liu, L. Zhu, and M. Pauly, "Structural Diversity and Function of Xyloglucan Sidechain Substituents," *Plants,* vol. 3, no. 4, pp. 526-542, 2014, doi: 10.3390/plants3040526.

[14] A. Varki *et al.*, "Symbol Nomenclature for Graphical Representations of Glycans," *Glycobiology,* vol. 25, no. 12, pp. 1323-1324, 2015, doi: 10.1093/glycob/cwv091.

[15] J. Larsbrink *et al.*, "A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes," *Nature,* vol. 506, no. 7489, pp. 498-502, 2014, doi: 10.1038/nature12907.

[16] J. Larsbrink, A. J. Thompson, M. Lundqvist, J. G. Gardner, G. J. Davies, and H. Brumer, "A complex gene locus enables xyloglucan utilization in the model saprophyte *Cellvibrio japonicus*," *Molecular Microbiology,* vol. 94, no. 2, pp. 418-433, 2014, doi: 10.1111/mmi.12776.

[17] F. Pinto, E. L. Thornton, and B. Wang, "An expanded library of orthogonal split inteins enables modular multi-peptide assemblies," *Nature Communications,* vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-15272-2.

[18] C. K. Shih, R. Wagner, S. Feinstein, C. Kanik-Ennulat, and N. Neff, "A dominant trifluoperazine resistance gene from *Saccharomyces cerevisiae* has homology with F0F1 ATP synthase and confers calcium-sensitive growth," *Molecular and Cellular Biology,* vol. 8, no. 8, pp. 3094-3103, 1988, doi: 10.1128/mcb.8.8.3094.

[19] E. J. Bowman, K. Tenney, and B. J. Bowman, "Isolation of Genes Encoding the *Neurospora* Vacuolar ATPase," *Journal of Biological Chemistry,* vol. 263, no. 28, pp. 13994-14001, 1988, doi: 10.1016%2FS0021-9258%2818%2968175-X.

[20] L. Zimniak, P. Dittrich, J. P. Gogarten, H. Kibak, and L. Taiz, "The cDNA Sequence of the 69-kDa Subunit of the Carrot Vacuolar H^+-ATPase," *Journal of Biological Chemistry,* vol. 263, no. 19, pp. 9102-9112, 1988, doi: 10.1016%2FS0021-9258%2819%2976514-4.

[21] R. Hirata, Y. Ohsumi, A. Nakano, H. Kawasaki, K. Suzuki, and Y. Anraku, "Molecular Structure of a Gene, *VMA1*, Encoding the Catalytic Subunit of H^+-Translocating Adenosine Triphosphatase from Vacuolar Membranes of *Saccharomyces cerevisiae*," *Journal of Biological Chemistry,* vol. 265, no. 12, pp. 6726-6733, 1990, doi: 10.1016%2FS0021-9258%2819%2939210-5.

[22] H. Wu, Z. Hu, and X.-Q. Liu, "Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803," *Proceedings of the National Academy of Sciences,* vol. 95, no. 16, pp. 9226-9231, 1998, doi: 10.1073/pnas.95.16.9226.

[23] H. Iwaï, K. M. Mikula, J. S. Oeemig, D. Zhou, M. Li, and A. Wlodawer, "Structural Basis for the Persistence of Homing Endonucleases in Transcription Factor IIB Inteins," *Journal of Molecular Biology,* vol. 429, no. 24, pp. 3942-3956, 2017, doi: 10.1016/j.jmb.2017.10.016.

[24] P. Carvajal-Vallejos, R. Pallissé, H. D. Mootz, and S. R. Schmidt, "Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources," *Journal of Biological Chemistry,* vol. 287, no. 34, pp. 28686-28696, 2012, doi: 10.1074/jbc.m112.372680.

[25] A. J. Burton, M. Haugbro, E. Parisi, and T. W. Muir, "Live-cell protein engineering with an ultra-short split intein," *Proceedings of the National Academy of Sciences,* vol. 117, no. 22, pp. 12041-12049, 2020, doi: 10.1073/pnas.2003613117.

[26] H. M. Beyer, K. M. Mikula, M. Li, A. Wlodawer, and H. Iwaï, "The crystal structure of the naturally split gp41-1 intein guides the engineering of orthogonal split inteins from *cis*-splicing inteins," *The FEBS Journal,* vol. 287, no. 9, pp. 1886-1898, 2020, doi: 10.1111/febs.15113.

[27] M. E. Lee, W. C. Deloache, B. Cervantes, and J. E. Dueber, "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly," *ACS Synthetic Biology,* vol. 4, no. 9, pp. 975-986, 2015, doi: 10.1021/sb500366v.

[28] C. Engler, R. Kandzia, and S. Marillonnet, "A One Pot, One Step, Precision Cloning Method with High Throughput Capability," *PLoS ONE,* vol. 3, no. 11, p. e3647, 2008, doi: 10.1371/journal.pone.0003647.

[29] R. C. F. Cheung, J. H. Wong, and T. B. Ng, "Immobilized metal ion affinity chromatography: a review on its applications," *Applied Microbiology and Biotechnology,* vol. 96, no. 6, pp. 1411-1420, 2012, doi: 10.1007/s00253-012-4507-0.

[30] J. J. Lee, D. F. Bruley, and K. A. Kang, "Manipulation of the Affinity Between Protein and Metal Ions by Imidazole and PHfor Metal Affinity Purification of Protein c from Cohn

Fraction IV-1," in *Oxygen Transport to Tissue XXIX*, Boston, MA, K. A. Kang, D. K. Harrison, and D. F. Bruley, Eds., 2008: Springer US, pp. 93-100. [Online]. Available: https://dx.doi.org/10.1007/978-0-387-74911-2_11

[31] TakaraBio. "In-Fusion Cloning overview." https://www.takarabio.com/learning-centers/cloning/in-fusion-cloning-general-information/in-fusion-cloning-overview (accessed March 19th, 2024).

[32] A. Froger and J. E. Hall, "Transformation of Plasmid DNA into E. coli Using the Heat Shock Method," *Journal of Visualized Experiments,* no. 6, 2007, doi: 10.3791/253.

[33] Thermo Scientific. "Fast Digestion of DNA." https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0012413_Fast_Digestion_DNA_UG.pdf (accessed April 24th, 2024).

[34] P. Y. Lee, J. Costumbrado, C.-Y. Hsu, and Y. H. Kim, "Agarose Gel Electrophoresis for the Separation of DNA Fragments," *Journal of Visualized Experiments,* no. 62, 2012, doi: 10.3791/3923.

[35] N. E. Biolabs. "Protocol for Protein Expression Using BL21 (C2530)." https://www.neb.com/en/protocols/0001/01/01/protocol-for-protein-expression-using-bl21-c2530 (accessed April 24th, 2024).

[36] X. Chen, J. L. Zaro, and W.-C. Shen, "Fusion protein linkers: Property, design and functionality," *Advanced Drug Delivery Reviews,* vol. 65, no. 10, pp. 1357-1369, 2013, doi: 10.1016/j.addr.2012.09.039.

Appendix I – Part-specific overhangs and multi-enzyme fusion

***Table A1**: 5' and 3' BsaI overhangs for all toolkit parts.*

Part	5' overhang	3' overhang
ConC	TATG	TACA
ConN	GAGT	AACG
HisC	TATG	CCGA
HisN	CCCT	AACG
IntC	CCGA	GCTG
IntN	ATCG	CCCT
LinkC	GCTG	TACA
LinkN	GAGT	ATCG
POI	TACA	GAGT

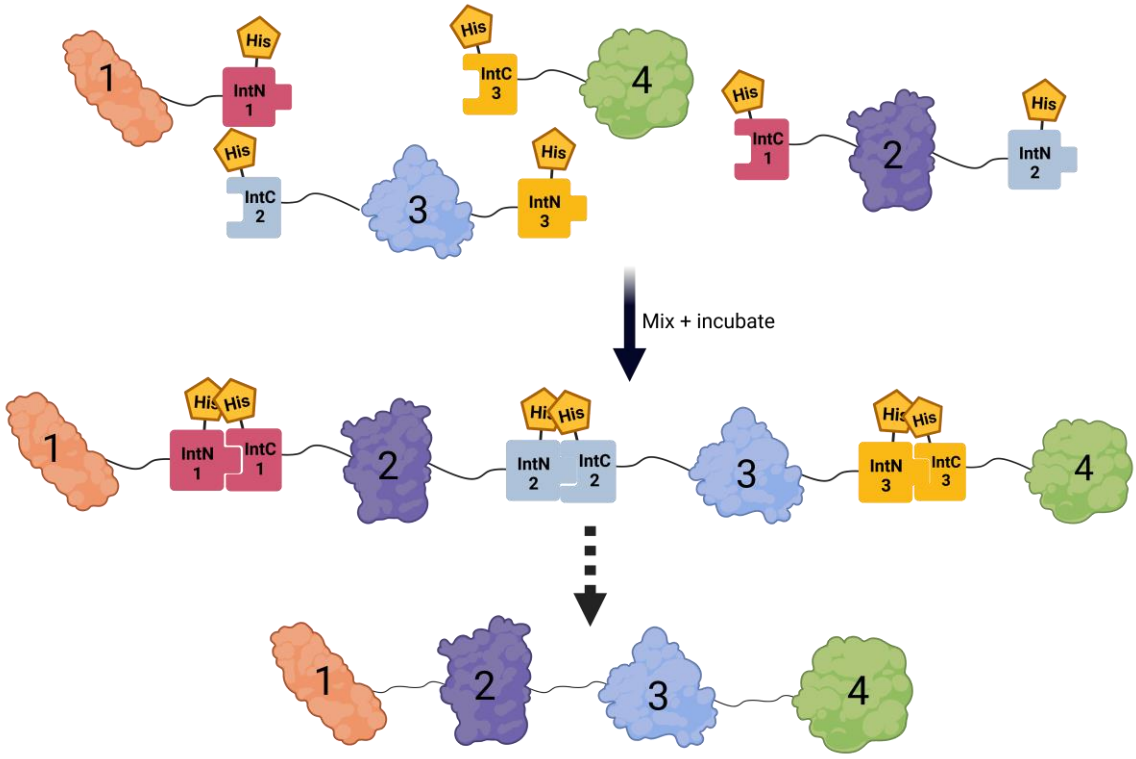


***Figure A1: Split intein fusion of four enzymes.** All protein assemblies are expressed individually, then mixed and incubated to mediate the fusion. (Created with BioRender.com.)*

Appendix II – Annotated sequences of toolkit parts

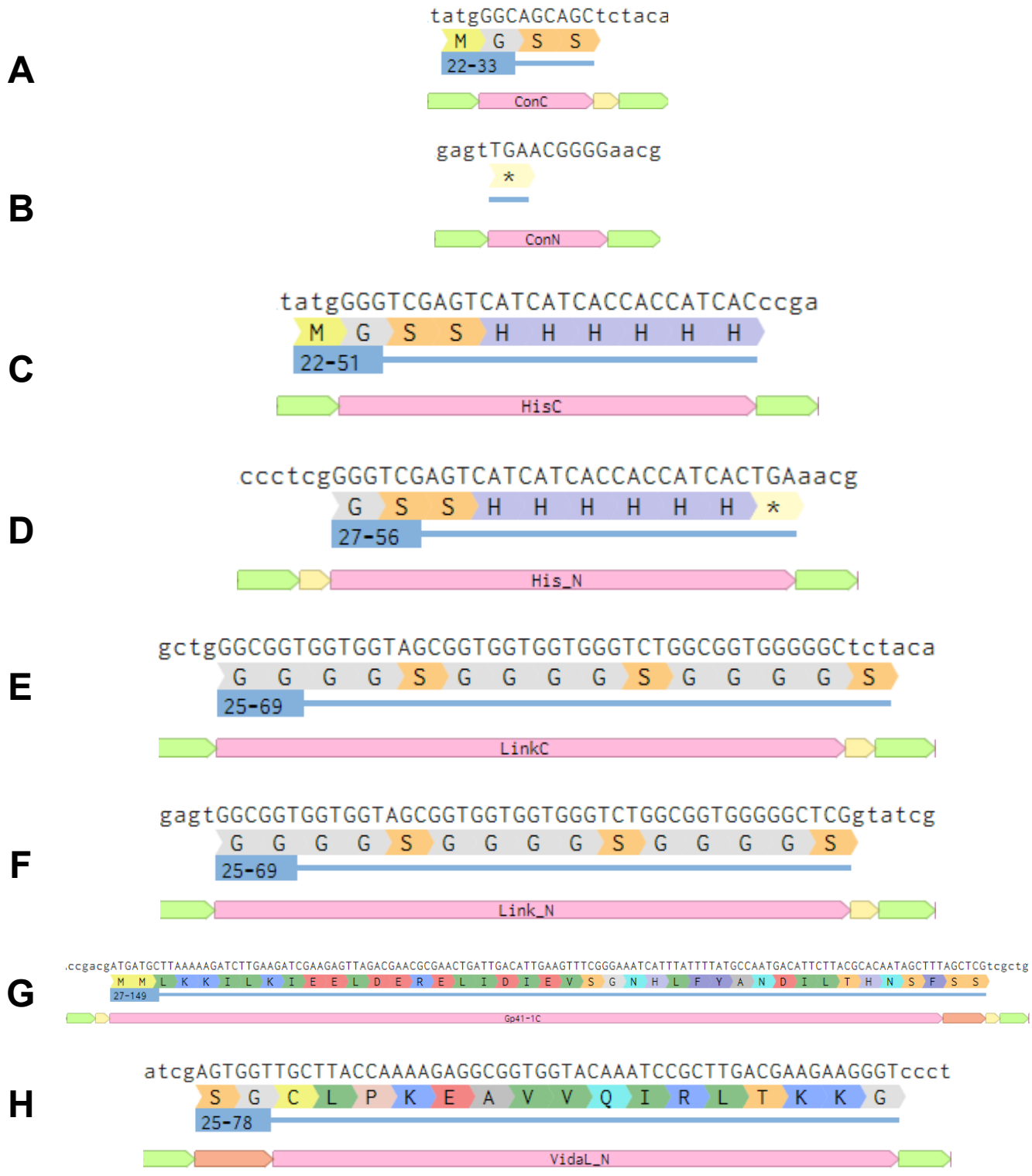


Figure A2: Annotated sequences of a few toolkit parts. *The sequences have been cropped to only include the four-base long type-specific BsaI overhangs (green), the coding sequence (pink), and added bases for reading frame correction (yellow) or extein dependence (orange).*

Appendix III – Protocols part I

Genome extraction

Centrifuge 1 mL of ON cultured cells and remove supernatant until there is < 80 µL left. Resuspend the pellet in 180 µL Buffer ATL, then add 200 µL Buffer AL and 200 µL 100% ethanol, vortex immediately. Transfer contents to a spin column and centrifuge for 1 min at 12000 g, discard the flowthrough. Add 500 µL Buffer AW1 to the spin column and centrifuge at the same conditions, discard the flowthrough. Repeat with 500 µL Buffer AW2 and discard the flowthrough. Centrifuge the column empty for 2 min, then move it to a clean Eppendorf tube, adding 50 µL of mQ water. Centrifuge and collect the flowthrough.

Polymerase chain reaction

In general, all PCR reactions were done using Phusion High-Fidelity DNA Polymerase, except when troubleshooting or doing cPCR. In this case, DreamTaq DNA Polymerase or Phire Hot Start II DNA Polymerase were utilized. For Phusion, the reaction setup is detailed in Tables A2 – A4.

Table A2: *Phusion PCR components.*

Compound	Amount
Phusion HF Buffer (5X)	10 µL
dNTPs (10 mM)	1 µL
Forward primer (10 µM)	2.5 µL
Reverse primer (10 µM)	2.5 µL
Template genomic DNA, or	50 – 250 ng, or
Template plasmid DNA	1 pg – 10 ng
Phusion HF DNA Polymerase	0.5 µL
MQ water	to 50 µL

Table A3: *General thermocycler conditions for Phusion PCR.*

Step		Temperature (°C)	Time
Initial denaturation		98	30 s
35 cycles	Denaturation	98	20 s
	Annealing	$T_{Annealing}$	20 s
	Extension	72	15 – 30 s/kb
Final Extension		72	10 min
Hold		4	∞

Table A4: *Thermocycler conditions for mutating the pET28a-TEVc vector.*

Step		Temperature (°C)	Time
Initial denaturation		95	5 min
5 cycles	Denaturation	95	30 s
	Annealing	48	30 s
	Extension	72	3.5 min
20 cycles	Denaturation	95	30 s
	Annealing	60	30 s
	Extension	72	3.5 min
Hold		10	∞

Appendix IV – Protocols part II

Polymerase chain reaction, continued
For cPCR, Phire Hot Start II DNA Polymerase was used. To prepare the template, a small part of the bacterial colony was picked and resuspended in 50 µL mQ water, then it was boiled at 95 °C for 5 min. The experimental setup is detailed in Tables A5 and A6.

Table A5: *Phire cPCR components.*

Compound	Volume (µL)
Phire Buffer (5X)	4
dNTPs (10 mM)	0.5
Forward primer (10 µM)	1
Reverse primer (10 µM)	1
Template DNA	2
Phire Hot Start II DNA Polymerase	0.5
MQ water	11

Table A6: *Thermocycler conditions for Phire cPCR.*

Step		Temperature (°C)	Time
Initial denaturation		98	3 min
35 cycles	Denaturation	98	5 s
	Annealing	$T_{Annealing}$	5 s
	Extension	72	10 – 15 s/kb
Final Extension		72	1 min
Hold		4	∞

In-Fusion cloning
Mix by pipetting the following reagents: 50 ng insert, 22.5 ng vector, 0.5 µL In-Fusion 5X mix, and mQ water to a total volume of 2.5 µL. Incubate mixture in a water bath at 50 °C for 15 min. Transform 1.25 µL of mixture into 25 µL of competent cells.

***E. coli* heat shock transformation**
Thaw cells (25 µL per reaction) on ice. Mix cells with 1.25 µL of plasmid/In-Fusion/ligation product and keep on ice for 30 – 45 min. Heat shock at 42 °C in a water bath for 45 s, then put back on ice for 1 min. Transfer cells to Eppendorf tubes with 1 mL of pre-warmed (37 °C) SOC media and incubate at 37 °C for 1 h. Plate 100 µL of culture on LB agar plates. For transformation of In-Fusion plasmids, concentrate the cells by centrifugation before plating.

Annealing of oligonucleotides
The oligonucleotides for MoClo assembly were annealed by mixing 22.5 µL of 100 µM forward and reverse primer with 5 µL of T4 DNA Ligase Buffer (5X), then running the mixture in a thermocycler at 98 °C for 5 min, then decreasing the temperature by 1 °C in 30 s increments down to 12 °C.

Appendix V – Protocols part III

BsaI/BsmBI assembly

The BsaI/BsmBI assemblies were done by mixing components together and running them in a thermocycler. The experimental setup is detailed in Tables A7 and A8.

Table A7: *BsaI/BsmBI assembly components.*

Compound	Volume (µL)
Insert 1,	0.5
Insert 2,	0.5
.....	
Insert n,	0.5
Receiver-vector	0.5
T4 DNA Ligase	0.5
BsaI or BsmBI	0.5
T4 DNA Ligase Buffer	1.0
MQ water	to 10 µL

Table A8: *BsaI/BsmBI assembly thermocycler conditions.*

Step		Time
Initial restriction (37 °C)		4 min
50 cycles	Restriction (37 °C)	1 min
	Ligation (16 °C)	2 min
Final restriction (37 °C)		4 min
Heat inactivation (80 °C)		10 min
Hold (12 °C)		∞

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY