SYNSENSO

Cell-free synthetic biology for combinatorial biosensor design

Project no. 101072980

Deliverable D1.2
Toolbox of well-characterized inteins

Version 1.1

WP 1 – Building cell-free computing circuits

Lead Participant	Imperial College London (ICL)
Contributors	Karen Polizzi (ICL), Rafael Montenegro (ICL)
Delivery date	29 February 2024
Dissemination level	Public
Type	Report

Funded by the European Union

Engineering and Physical Sciences Research Council

Funded by the European Union (GA 101072980) and supported by the UK Engineering and Physical Sciences Research Council. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Revision history

Date	Authors	Revision
14/02/2024	Rafael Montenegro	Draft version
15/02/2024	Karen Polizzi	Revised version
16/02/2024	Karen Polizzi & Rafael Montenegro	Final version
20/02/2024	Heinz Koeppl, Christiane Hübner	Internal review
29/02/2024	Heinz Koeppl	Final version

Contents

Abbreviations

Abbreviations	Details
CFS	Cell-free system
PTS	Protein trans-splicing
GFP	Green fluorescent protein
NGS	Next-generation sequencing

Partner Short Names

Abbreviations	Details
TUDa	Technische Universität Darmstadt
TUE	Eindhoven University of Technology
IIT	Istituto Italiano di Technologia
TVU	Tor Vergata Universita degli Studi di Roma
DBS	Dynamic Biosensors GmbH
UBI	Ulisse Biomed S.p.A.
ICL	Imperial College London
ABV	Abvance Biotech S.L.
ACC	accelopment Schweiz AG
LEB	LenioBio GmbH
BRA	BRAIN Biotech AG
NUC	Nuclera Nucleics Ltd.
BME	Biomerieux SA
UNA	University of Naples Federico II

Executive Summary

This deliverable aims precisely to show the progress of the project “Intein-based logic gates for multiplexed designs” towards having the toolbox of well-characterized inteins. The first part of the report corresponds to more broad aspects of the project: justification of the research and this deliverable, the objectives reached so far, the outcomes, and the following steps to establish the intein toolbox. Next, the second part reaches more technical aspects of the project, going from the introduction and state of the art, the progress of the research, and the proposed plan to finish the initial toolbox characterization.

Overall, synthetic biology is an expanding field at the intersection of biology and engineering, where new approaches for producing goods and biological-based technologies are being developed. One of those technologies is biosensing, where biological components are engineered to detect new or known targets in novel ways, bringing advantages over traditional detection systems. In addition to new biologically engineered approaches in synthetic biology, new biological platforms, or chassis, are also in development, cell-free systems being one of them. Cell-free systems are in vitro systems where the cells’ membrane is removed, and only the cellular machinery is kept to perform biochemical reactions, generally mediated by transcription and translation steps. The advantages of cell-free systems over cell-based ones are that they are fast reactions, are highly versatile by having a less crowded environment, have higher tunability of the genetic circuits as there is no burden effect in the cellular economics, have portability, and have lower production cost. All those cell-free advantages enhance the biosensing applications as well. Therefore, cell-free biosensing is a promising platform for developing biosensors. In addition, synthetic biology research is also concerned with novel protein circuit designs. Recently, new approaches have increased the available protein circuits, adding new protein functions for new bioengineering approaches like biocomputation. Split inteins, proteins capable of ligating 2 peptides by a protein splicing reaction, are one case of such novel functions having logic gate behavior ideal for biocomputation. Split inteins have been described in addition to having fast reaction rates, increasing their applicability into logic gate computation for biotechnological applications. Consequently, we aim to develop novel biosensors by using split-intein logic gates that allow fast and multiplexed reactions.

The project’s first relevant outcome was to set up the genetic system to screen for the intein activity and to test the biosensor circuits using intein proteins as one of the components of the reaction. We named this system the 2Split system (Technical details are explained in the second part of the deliverable). We also replicate a 2-plasmid system to characterize the inteins, working as split inteins, meaning 2 independent units. In addition, we now have an optimized protocol to get the *Escherichia coli* lysates required for the cell-free reaction and a recipe for the cell-free reaction itself. This protocol optimization is fundamental to characterize any reaction during the project as a cell-free one. Finally, we are finishing the Golden Gate optimization in the laboratory to have modular and systematic cloning, allowing us to clone the toolbox of inteins and further on the biosensor designs. The next steps for the project are to (i) finish the optimization of the Golden Gate reaction to have a defined protocol for modular cloning in the laboratory and (ii) clone the split intein toolbox to characterize it. During the next months, we will thus characterize the intein toolbox as our first milestone and share this information with the network, as stipulated in the SYNSENSO timeline; this milestone will be September 2018 for us, considering that our project started with a 6-month delay for external reasons.

Need for the Deliverable

Split inteins are powerful molecular tools as they can be expressed as 2 separate transcripts, and once translated, they can ligate 2 sequences at the peptide level while splicing themselves out of the reaction. In other words, split inteins bring 2 peptides together while they are removed from the equation in a biochemical reaction. Split inteins thus bring another key function that can be engineered by synthetic biology, increasing the protein circuit toolbox available. Synthetic biology generally uses genetic regulation as the main driver for biocomputation applications. Even though biocomputation approaches started with genetic circuits, more recently, novel protein circuits, including intein ones, have also been explored. Split

inteins have already shown biocomputation applicability. Therefore, it is justified to explore their capability to develop protein circuits reaching biocomputational approaches. By designing and characterizing split intein logic gates, we aim thus to (i) increase the protein circuit toolbox by adding the ligation feature brought by the spit intein's protein trans-splicing reaction and (ii) improve the biocomputation capabilities reached by synthetic biology, and in this case of the SYNSENSO network, to improve the cell-free biosensing applications precisely.

Objectives of the Deliverable

This deliverable aims to define the project's scope, status, and the next steps to accomplish its goals. Therefore, the precise objectives are:

- Define the cell-free biosensing state of the art and the project's scope and goals by incorporating split inteins circuits to cell-free systems.
- Update the project's progress by reporting the results and protocols developed for the cell-free reactions and the split intein toolbox characterization.
- Outline the following steps to optimize the Golden Gate cloning and build the split intein toolbox next to characterize it.

Outcomes

Overall, the first months included more design and optimization steps for the project. Initially, we defined the genetic and screening systems' designs as the first mandatory step to start working on the project. Next, we moved to optimize the cell-free reaction, including its preparation and, following it, the golden gate optimization. The list of the outcomes achieved up to now are:

1. The design of a 2-plasmid modular screening system based on Golden Gate Cloning.
2. The 2Split system design for cell-free biosensing based on split inteins.
3. List of fast-acting split intein pairs constituting the intein toolbox.
4. Optimized homemade *Escherichia coli* cell-free lysate protocol.
5. Protocol for setting up cell-free reactions on a microplate reader.
6. Optimization of Golden Gate cloning for modular and systematic cloning (Ongoing).

Next steps

First, the following mandatory step is to finish the Golden Gate optimization to clone the toolbox library. Once the cloning is optimized, the focus can be directed to characterizing the split inteins of the toolbox and next to characterize the modules of the 2Split system to show the applicability of the toolbox and the system's potential for cell-free biocomputation leading to biosensing application. Those next steps are thus in detail:

1. Finish the Golden Gate optimization.
2. Clone the split intein pairs and characterize them to define the intein toolbox.
3. Clone the 2Split system modules and characterize them to have a proof of concept of our design for biosensing applications.

1 Introduction to the project and state of the art for: Cell-free synthetic biology for biosensing applications

Synthetic biology is a new field at the intersection between biology and engineering, where engineering principles are applied to biology[1,2]. This has included so far principles such as systematic design, computational modelling, and the engineering Design, Build, Test, Learn (DBTL) cycle to be applied to biological systems[3,4]. Such biological systems consist of genetic circuits by which genes are designed to express proteins working together to create a newly designed biological function. These systems contribute to biotechnological applications, where biological components are used with high technology to design biological systems, such as the biological production of fine chemicals and biosensor designs[1,2,5,6].

Gene expression is consequently a fundamental approach in biotechnology, where a genetic circuit, meaning an engineered DNA fragment acting as a gene, is created to express a desired protein with a final application, ranging from medical to industrial biotechnological products[1,7]. In traditional biotechnological applications, gene expression occurs typically inside a cell by adding an external genetic circuit[7–9]. Therefore, expressing the desired protein using the cellular machinery and recovering it from the cells is possible. However, gene expression in a cell causes a burden within this cell[10–12], reducing its growth capabilities, and depending on the expressed protein, it might be expensive to recover or challenging to track.

Cellular gene expression is widely used in biotechnological applications; however, cell-free expression has appeared as an alternative, especially for biosensing purposes. Cell-free systems (CFS) are protein expression platforms where the cells of a culture are disrupted to remove the cellular membrane and only keep the cellular machinery to perform a biochemical reaction[13,14]. These cell-free systems offer a range of advantages compared to cell-based ones, given the membrane-free conditions that allow fast reactions, high versatility, scalability, and portability[13,15–17], all desirable features for precise biosensing. Based on these advantages, CFS reactions have been used to express different synthetic genetic circuits, including biosensor applications[18]. CFS is thus an enhanced approach for biosensing as the target molecule triggering the detection signal is more accessible to track[18] and detect in the CFS's less-crowded environment[6].

A biosensor is a biological-based device or technology that detects a target biomolecule[6]. In addition, CFSs offer an open reaction condition, which is a crucial feature for the easier detection of biomolecules. Previous research has even proven the enhanced capabilities of CFSs for biosensing over cell-based applications, especially in shorter response times as the case for Zika virus[19]. However, precise genetic engineering is required to achieve this type of CFS biosensors. The aim of our research is thus to take the properties of CFS to develop a foundational technology to improve those biosensing capabilities. We plan to develop this foundational technology in detail by using a particular set of proteins, called inteins, to improve the genetic engineering tools available to detect biomolecules in CFS biosensors.

Inteins, on the other hand, are unique protein domains that can spontaneously excise themselves from a larger polypeptide[20,21]. In detail, inteins perform a splicing reaction at the peptide chain level in a process called protein splicing, meaning that they present themselves as molecular tools that can ligate two proteins together by a peptide bond while simultaneously separating themselves from the linked proteins[20]. An additional advantage of using and promoting inteins in synthetic biology is the presence of split inteins. Split inteins can be expressed separately by the N and C terminal parts, which increases their applicability[20,22–25]. Pinto et al.[24] demonstrated this ability by using split inteins in logic gates to express a fluorescent signal protein only when the two split intein circuits were in the reaction.

Thus, this project aims to develop a foundational technology that allows us to use split intein proteins as a component of biosensors following logic gates in cell-free systems. This technology will bring new genetic circuit designs, including split inteins as components, while foreseeing an improvement in logic

biocomputation to reach fast biosensing. My research's main objective is thus to develop biosensors implementing split intein-based logic gates to enable multiplexed designs and fast biosensing in cell-free systems. We expect to characterize a toolbox of split inteins first and later to build split intein-based logic gates allowing biosensing applications.

2 Project progress to date: towards a toolbox of split inteins for protein circuit designs

2.1 Genetic Design of the 2Split system

The project's first step was to set up a system and its design to use protein trans-splicing as a reaction trigger for logic gates thinking on a final biosensing application. The design was based on a top-down methodology, thinking first of the molecule to sense as a reporter (either amilCP or mCherry), and to add layers at each step to include the split inteins as triggering the fluorescent reporter first, and finally to include the proper logic gate system by setting a conditional protein trans-splicing triggered by a target recognition (Figure 1.A). Consequently, the multidomain protein design came up as the design for our system named "2Split" having split inteins in the middle of the protein complex and acting as key components to trigger the reaction (Figure 1.B).

The multidomain protein design has 2 independent units, each corresponding to the complex's N and C terminal parts. The N terminal part has the N part of the split reporter, followed by the N split intein, and finally, the N part of the recognition protein (Figure 1.B) - Left panel). The same logic and order applies to the C part of the complex. The target recognition will then bring the separate parts together, setting up a conditional protein splicing reaction triggered by the target recognition.

First, the idea was to set up the system with the rapamycin binding domains, FKBP and FRB[26], acting as the split recognition domains to validate the multidomain protein design of the 2Split system. Therefore, it will be possible to evaluate the behavior of the multidomain protein design considering basal "off-target" expression, sensitivity based on rapamycin concentration to signal response, fold change, and maximum reporter expression. By reaching the milestone of having a proof of concept of the system, it will be possible to design precise logic gates and start the bottom-up biosensing design specified in figure 1.A.

The multidomain protein design is illustrated with the ligation of reporter split halves, mCherry precisely (Figure 1). However, any peptide can be linked by the protein trans-splicing reaction (PTS). The rationale, and part of the justification, of using split inteins is the modularity it brings to the logic biocomputation. The molecule ligated by the PTS can also be intermediate proteins or regulators acting as a processing unit instead of the biosensor reporter. This principle is used to design the logic gates of the 2Split toolbox.

An ideal toolbox of logic gates has several logic operations in order to put in place more complex logic if needed. The YES, AND, and OR gates and their opposites (NOT, NAND, NOR) are typically necessary for complex logic[27]. The YES gate is the basic unit of the logic, and in the "2Split" system, the multidomain protein design unit corresponds to the YES gate, meaning that only when the system recognizes the target the output reaction is released, as illustrated on figures 1.B-C and 2.A.

Next, the AND gate was designed by adding an intermediate processing step. This modification was achieved by adding an additional step of split inteins and using two YES units, each of them recognizing a defined target. As illustrated in figure 2.B, the design has 2 YES units performing a PTS reaction, releasing the final split intein fused to the respective reporter half. The gate behaves as an AND logic gate because both targets must have the 2 intermediates to lead to the sequential and final PTS reconstituting the reporter.

The same principle of adding an intermediate step applies to the OR gate. However, the logic here means that any target will lead to the output signal. Thus, to put in place such a system, the two YES gates, having different targets each, will release the same intermediate. The other intermediate is expressed constitutively in the system, allowing the reporter reconstitution with at least one of the targets in the media, which is the OR gate behavior (Figure 2.C).

The novelty of the logic gate design is that it puts several orthogonal split inteins to work together in the same reaction as for the AND and OR gates. However, to put in place a NOT gate, meaning to separate two peptides instead of ligating them, it is difficult to put in place a system using only split inteins. In theory, it could be possible to do the reaction with split inteins by just linking a nonrelevant peptide, causing a disruption of the reporter protein. This idea of a NOT gate was considered as a first design draft (Data not shown). However, it was changed to a design including split proteases in the picture. An intein protein ligates two peptides, while a protease cleaves a peptide in two by identifying a precise recognition peptide. Thus, proteases are the logical opposite of inteins. The final NOT gate design thus reconstitutes a TEVP protease upon target binding by a PTS reaction. In the second step, the reconstituted TEVP will cleave the recognition site added previously to the reporter protein as a linker, thus disrupting the reporter protein as illustrated in figure 2.D.

Overall, the design part of the project already has a multidomain protein design, the basic unit of our system named "2Split", and also the design for the split intein logic gates performing the YES, AND, OR, and NOT logic gates in theory. The NOT gate brings the advantage of releasing the TEV protease from the PTS reaction instead of the already mentioned reporter or intermediate circuit. In addition, if we couple the AND and OR gates to the release of this TEV protease, we will have the opposite gates, NAND and NOR, following the logic: AND+NOT=NAND. The next step will thus be to clone the genetic circuits to have the proof of concept of the overall system and the validation of the protein-based logic gates.

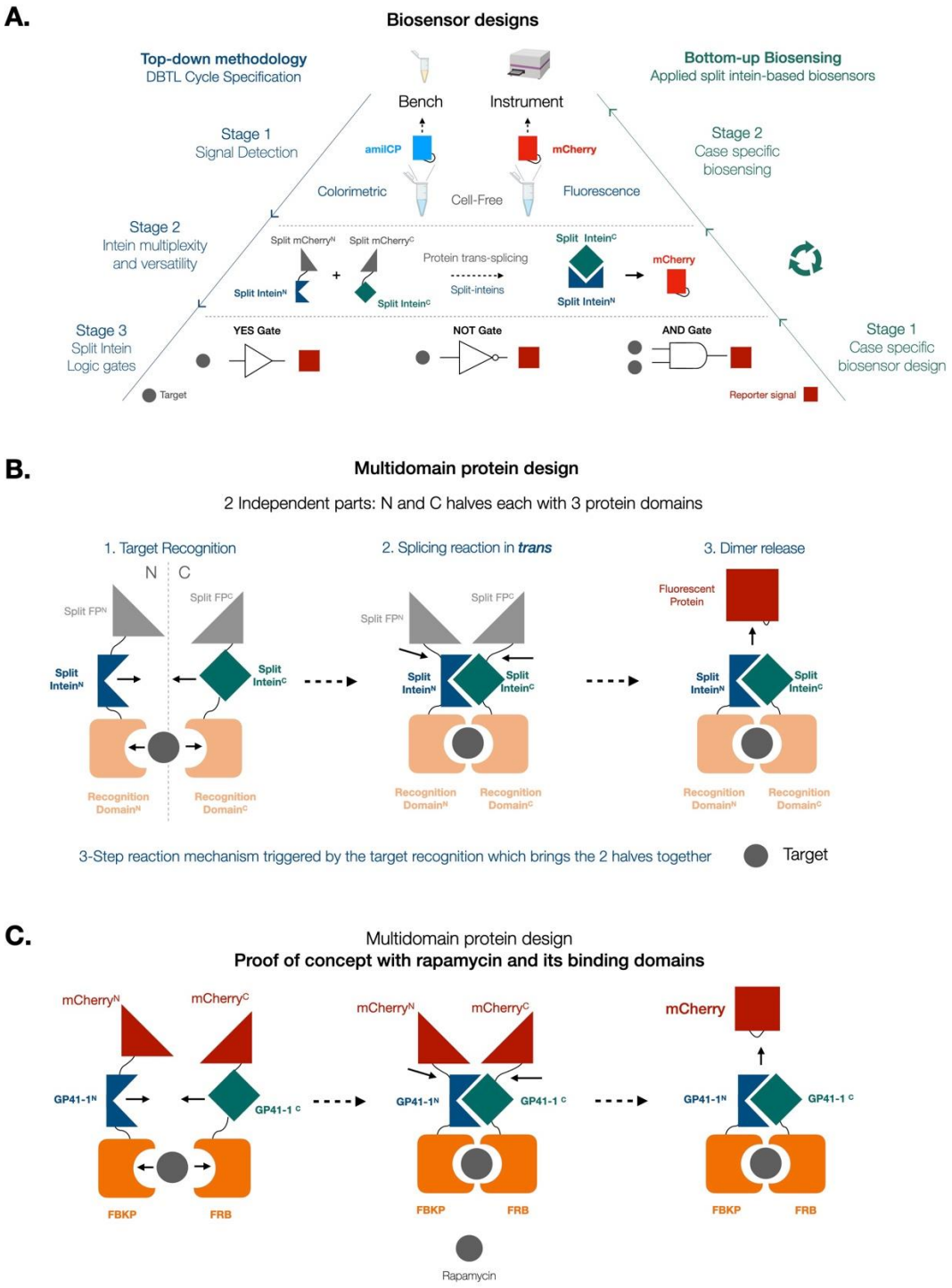

Figure 1: Project conception and design of the 2Split system for cell-free biosensing applications. (A) Project approach for the design of split intein logic gates starting with a top-down methodology by tracking the reporter protein and adding the split intein reactions in a stepwise manner at each time. Once the split inteins circuits are in place the idea is to pick an application and go for a bottom-up biosensing focus by coupling the technology to a particular target. (B) Design of the multidomain protein design of the 2Split system having two independent parts each of them with 3 domains: split reporter (Split FP: split fluorescent protein), split intein, and split recognition domain all of the same part, meaning N or C parts. The principle is that only when the target is recognize by the 2 half, the PTS reaction is triggered illustrated by the 3-step reaction having: target recognition, PTS reaction, and dimer release. (C) The proof-of-concept design for the 2Split system with the well characterized rapamycin domains FBKP and FRB, by adding rapamycin concentration it will be possible to have an idea on how the system performs with the trans conformation and optimize from there the system.

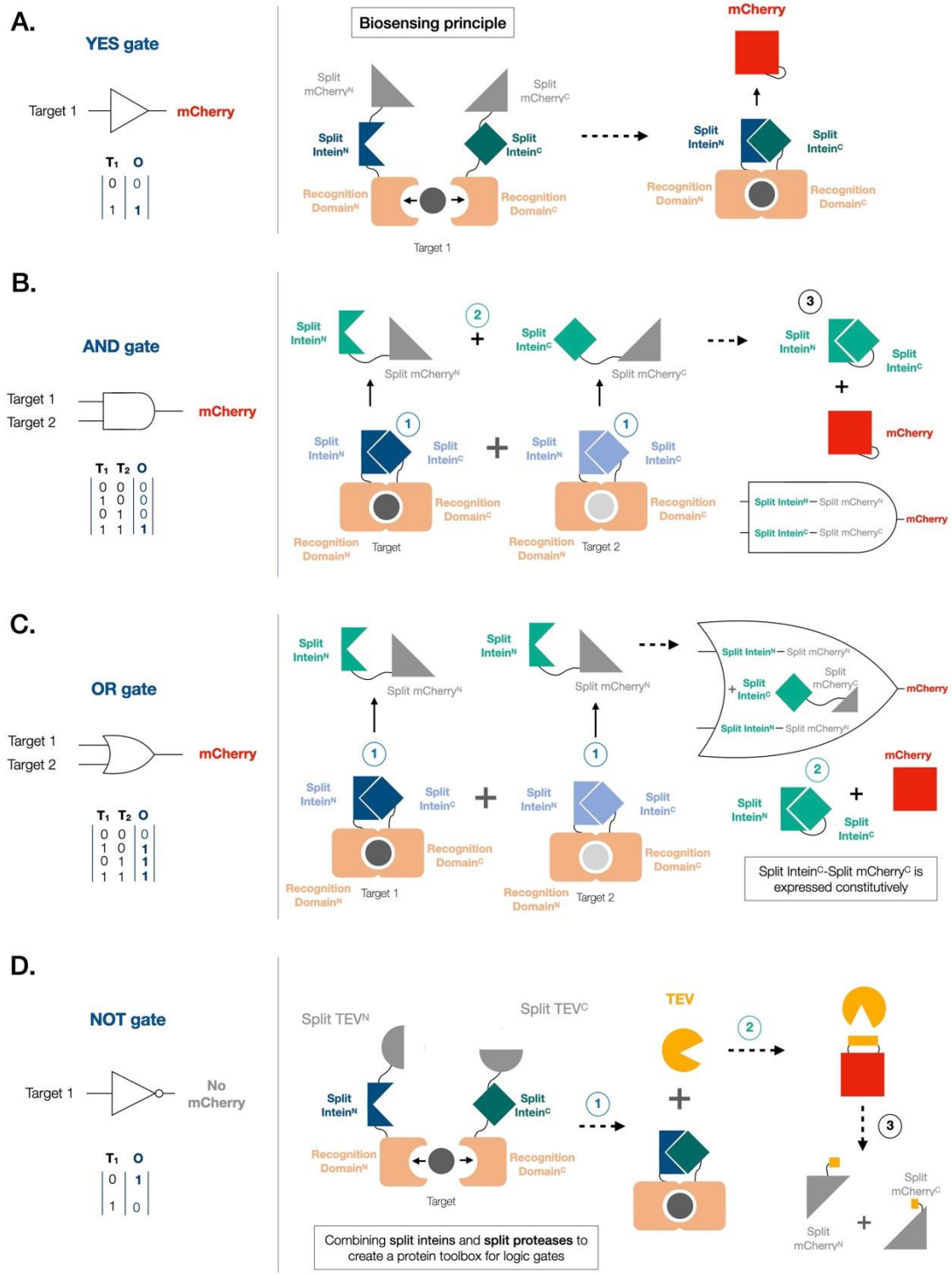

Figure 2: Split intein logic gates designs for the 2Split system. (A) The multidomain protein design correspond to the YES gate of the system. In order to have more complex gates like AND and OR, two multidomain protein units need to interact together. (B) The AND gate has 2 multidomain protein units recognizing 2 different targets, only when both of them are recognize the intermediate split intein-split reporter will encounter releasing the mCherry signal in this case. (C) The OR gate has 2 units recognizing different targets, however, the intermediate released by each unit is the same. Thus, the other intermediate is constitutively expressed in the system (Split intein C - Split mCherry C), allowing to have the OR gate behavior. (D) Finally, for the NOT gate a split intein and a split protease are combined to achieve the NOT behavior. Inteins ligate domain and proteases cut the link, performing thus opposites reactions, with the reasoning we designed the NOT gate by reconstituting a TEVP protease cutting the reporter protein by adding a recognition peptide at the fluorescent reporter.

2.2 *Escherichia coli*-based Cell-Free reaction optimization

The project's goal is to develop cell-free biosensors. Thus, a key step was to optimize the cell-free preparation and reaction. The well-known TX-TL method developed by Sun et a.[28] was used to optimize the reaction as a reference for the *E. coli* lysate and reaction preparation. The main modification was to use sonication to disrupt the cell membrane instead of the bead beating or cell disruption by a homogenizer.

The cell-free lysate preparation was then optimized by sonication, and the quality of the lysate was tested by tracking the expression of a Green Fluorescent Protein (GFP) reporter as illustrated in figure 3.A. The first step in the optimization was to test different sonication conditions by regulating the exposure time and the total energy applied. The heatmap of the first exploratory cell-free experiment showed clear optimal conditions of 500 J of total energy applied in exposures of 10 seconds and 1 minute of resting time until reaching the desired energy (Figure 3.B). However, based on the sonicator setting, it was still possible to test different values close to the optimal 500J. Consequently, we tested values of energy between 400 and 500J and found an ideal sonication condition at 445 J of total energy (Figure 3.C). Thus, because the GFP expression reached a maximal value with less energy applied, assuming a lower exposure, this should ensure less damage to the lysate biomolecules.

Any type of physical disruption of the cell will cause damage to the cellular machinery. However, the process can be optimized to reduce the impact on the cellular machinery to perform the cell-free reaction. This effect on the cellular machinery follows a Gaussian distribution with optimal conditions to set a fully functional lysate. Before and after reaching that level, the expression is lower than the maximal value. This effect can be explained by the fact that after reaching a maximum of cell membrane lysis, at some point, the lysis reaches a maximum for the number of cells in the sample, and the energy applied might affect the biomolecule function, especially via heat denaturation, leading to a reduction in the expression of the cell-free system. This effect can be seen after adding an energy higher than 500 J (Figure 3.B-C).

After defining the best conditions to get the lysate, the next step was to optimize the reaction and get an insight into the reaction over time. Any cell-free lysate requires optimizing key components to ensure a maximal expression of the genetic circuit. Normally, the elements to optimize for a particular lysate are potassium and magnesium as cofactors and the template DNA. Theoretically, any cell-free component can be optimized for a maximal expression. Potassium and magnesium concentrations were fixed to theoretical optimal values for *E. coli* lysate[29–31]. By fixing all the variables, it was possible to evaluate the effect of DNA template concentration on GFP output signal level (Figure 3.D). A maximum of GFP expression was reached by using 50nM to 250nM; adding the maximum 500nM of plasmid DNA template showed a 2-fold drop of the signal. The cell-free balance of TX-TL of the molecular machinery can explain this drop. The polymerases and ribosomes might be limited by the highest amount of DNA at the highest concentration, limiting their fast dynamics and justifying a lower output signal with the same amount of molecular machinery.

Regarding the GFP reporter signal behavior, the expression followed an expected pattern with a rapid increase of the signal reaching a maximum after 3 hours of the reaction (Figure 3.E). Having a fast reaction is ideal for biosensors, as one of the general goals is to have fast reaction to get fast biosensing, factor that suits well with the portability of cell-free systems as well. It is also possible to see a clear difference between the samples and the negative control even before the first hour of the reaction. Nevertheless, there is still room to optimize the reaction by reducing the lag phase as much as possible. Currently, the lag phase lasts 30 minutes. However, the lysate pre-mix and the DNA are added at time 0. Consequently, the lag phase time is expected. However, it could be reduced by simultaneously increasing the scope of the curve, which means having a faster reaction in general.

We also evaluate the effect of using different strains for the lysate to achieve a higher output signal, meaning a better lysate for heterologous expression of genetic circuits. The three factors evaluated were: induction OD_{600}, final OD_{600}, and run-off time. The run-off time corresponds to an incubation time before the final clarification of the lysate, where the idea is to activate and release the molecular machinery. By increasing this time to 60 min, the tendency was to see a decrease or no change in the GFP expression signal, indicating that 20 minutes of run-off is ideal and enough to have an effect on the lysate quality (Figure 3.F). In addition, a lower induction OD_{600} closer to ≤ 0.6 and a final OD_{600} closer to 2 seem to be ideal, combined with the 20 minutes run-off and particularly for BL21 Star strain, reaching higher GFP expression levels (Figure 3.F). Consequently, these 3 factors are fundamental ones to optimize to achieve an optimal lysate.

Finally, the last idea with the optimization was to switch the cell-free reaction conditions to the final application condition, always keeping the top-down methodology. The rationale was to set the final biosensor conditions to evaluate the cell-free performance by tracking the reporter protein. The cell-free reaction was run at 25°C instead of the original 30°C and without agitation. By having the final biosensing conditions, the GFP expression had the same expression pattern. As seen in figure 3.G., it was even possible to have a clear signal significantly different than the negative control at 30 minutes with the lysates prepared for the BL21 vs BL21 Star experiment. The cell-free lysate is validated for cell-free biosensing at room temperature conditions and can be further optimized to achieve improved performance.

2.3 Golden gate cloning for the intein toolbox characterization

The plan was to set up a modular cloning system with Golden Gate to characterize the 2Split system and validate the logic gates. Golden Gate cloning is a method known for modular cloning[32–34]; thus, it was the ideal method to characterize different types of regulation and protein parts quickly. The proposal follows using pSEVA plasmids[35,36] (a standard collection of vectors for cloning and gene expression: Standard European Vector Architecture) and adapting them to Golden Gate cloning as illustrated in figure 4.A. The base plasmid has a mCherry oligo expressed constitutively and BsmBI inverted sites as illustrated in the upper middle section of figure 4.A. With this design, the mCherry will be lost when an insert is cloned, allowing for tracking a red-to-white colony screening and simplifying the workflow.

The advantage of using Golden Gate is the possibility of switching enzymes between cloning cycles. This advantage is the case for the design illustrated in figure 4.A. The insert replacing the mCherry corresponds to the split intein screening system where split inteins can be added to their respective half, meaning the split intein N half is cloned to the mCherry N half, and the same logic applies for the C halves (Figure 4.D).

The base plasmid cloning was straightforward and confirmed by DNA sequencing. However, this was not the case for the second step of the cloning, aiming to replace the mCherry oligo for several genetic circuits to put in place the split intein screening system and the circuits for the 2Split system proof of concept. As we were always getting red colonies instead of the red-to-white transition, we considered an alternative design for the Golden Gate adapter. Thus, we also consider having an alternative design where the adapter to be replaced is less than 50 base pairs. This modification means disrupting the amilCP reporter gene instead of replacing the whole gene as illustrated in figure 4.C.

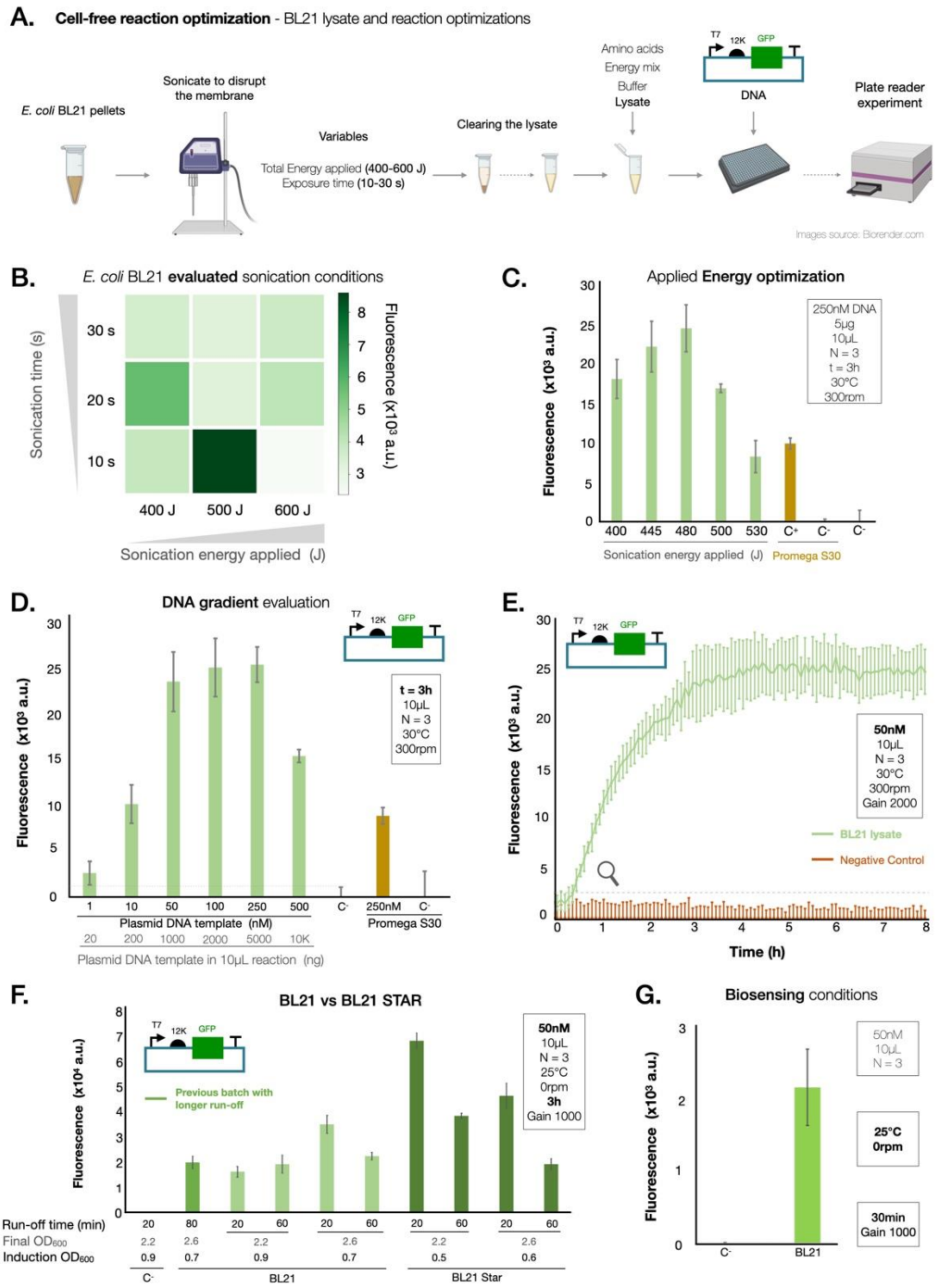

Figure 3: Cell-free reaction optimization for E. coli-based lysates. (A) Detailed methodology of the optimization process followed to find the best lysate conditions. The sonication conditions tested are specified and the characterization of such modification on the lysate were tested by tracking the GFP signal on a microplate reader. (B) Heatmap of the first optimization of the lysate testing different exposure times and total energy applied. The best performing lysate was the one having exposure times of 10 seconds with a total energy applied of 500J. (C) Zoom-in into the total energy applied by the sonicator with 10 seconds rounds of exposure, where 445 seemed like the best energy to apply. (D) DNA gradient added to the cell-free reaction from 1nM to 500nM in a 10 µL reaction. The optimal concentration for our home-made lysate was 50nM corresponding to the lowest DNA concentration added while keeping the highest expression value. (E) Detail of the GFP expression curve in the cell-free reaction overtime for the 8 hours of the run, where the curve reached a plateau after 4 hours of the run. (F) Comparison between the BL21 and BL21 STAR lysates. This result showed the relevance of having defined values for the induction and to stop the run, 0.6 and 2 OD_{600}. BL21 STAR seemed to reach higher expression values for the GFP suggesting a better performing lysate with a 20 minutes run-off reaction. (G) The cell-free reaction and lysate operates as expected switching to the desired biosensing conditions of no agitation and incubation at room temperature of 25°C.

Before moving to the alternative design, the idea was to find the issue with such an efficient method. The first step was to check the Golden Gate reaction by gel electrophoresis. This result showed that the expected linear ligation creating the insert to be added to the plasmid (Around 1.2Kb) was not happening. Consequently, we checked the digestion of the BsmBI-v2 and the Esp3I isoschizomer, meaning they have the same recognition sites on plasmid DNA. Both enzymes digested the plasmid, while BsmBI-v2 had a complete digestion. Finally, we checked the Golden Gate reaction by adding the Esp3I, and the expected band of 1.2 Kb appeared, confirming the reaction. This result suggests that Esp3I can digest the linear PCR template of the individual fragments, allowing the ligation, contrary to the reaction of BsmBI-v2. Finally, by figuring out the cloning issue, it is possible to resume the Golden Gate cloning with Esp3I instead of BmsBI to validate the 2Split system with the original adapter design having the mCherry.

3 Research plan to have the toolbox of split intein for cell free biosensing

The project already has the designs of the split inteins logic gates aiming to achieve fast and multiplexed biosensing in CFS. Overall, the project outcome is positive so far by setting up and optimizing already 2 of the 3 protocols to be used during the PhD research. These optimized protocols are the in-house homemade cell-free lysate with its reaction and the golden gate cloning for a modular approach.

All the progress of the project, mentioned in the previous section of the report, corresponds exclusively to objective 1 of the Ph.D. project: Design and characterize a toolbox of inteins for protein circuit design. The goal for 2024 is to complete the first objective and start setting up the last protocol of the project to find split sites in the recognition proteins with biosensing interest. In addition, this year includes a SYNSENSO network event where a poster is needed to explain the project as part of the yearly update meeting of the SYNSENSO Marie Curie Network.

Next, two main tasks to accomplish and work in parallel to complete the split intein toolbox are to clone and characterize the split intein system, including the 2Split proof of concept, and to optimize the cell-free reactions for bench biosensing conditions. In addition, as the plate reader data tends to get more and more extensive, I will take time to learn Python and set up a pipeline to handle data more straightforwardly. This Python data analysis code will simplify the split intein toolbox cloning characterization and optimize the cell-free final optimization as the output is the same fluorescent protein.

The project's next key milestone is to complete this first split intein toolbox, estimated to be finished by the end of August 2024 (Figure 5). The first step is thus to finish the golden gate optimization to characterize all the cloned modules later to validate the system's proof of concept. This milestone deadline includes the Python pipeline learning plus implementation, as well as the design of slides and a poster to communicate the progress with the laboratory group, network meetings, and possible synthetic biology conferences.

The top priority is optimizing the Golden Gate reaction in general, particularly with the Esp3I enzyme. Consequently, during February 2024, I will focus on optimizing the Esp3I Golden Gate reaction. Once a single test reaction is sequence verified, I will clone the genetic circuits to test the 2Split proof of concept and its logic gates designs (Figure 5). I expect to have all the genetic circuits cloned by the end of May 2024 to test them. This proof of concept with rapamycin binding domains is critical to validate the multidomain protein design of the 2Split system and, thus, the project's outcome. There are a couple of examples using split inteins in trans-conformation where, in one, caffeine[25] binding is the trigger of the close proximity for the PTS but having high basal levels. The other case uses short peptides as molecular glues with a single protein target to trigger the PTS[37]. Thus, the idea is to characterize the proof of concept, knowing there are already a couple of validated cases, to find a wide dynamic range ideally.

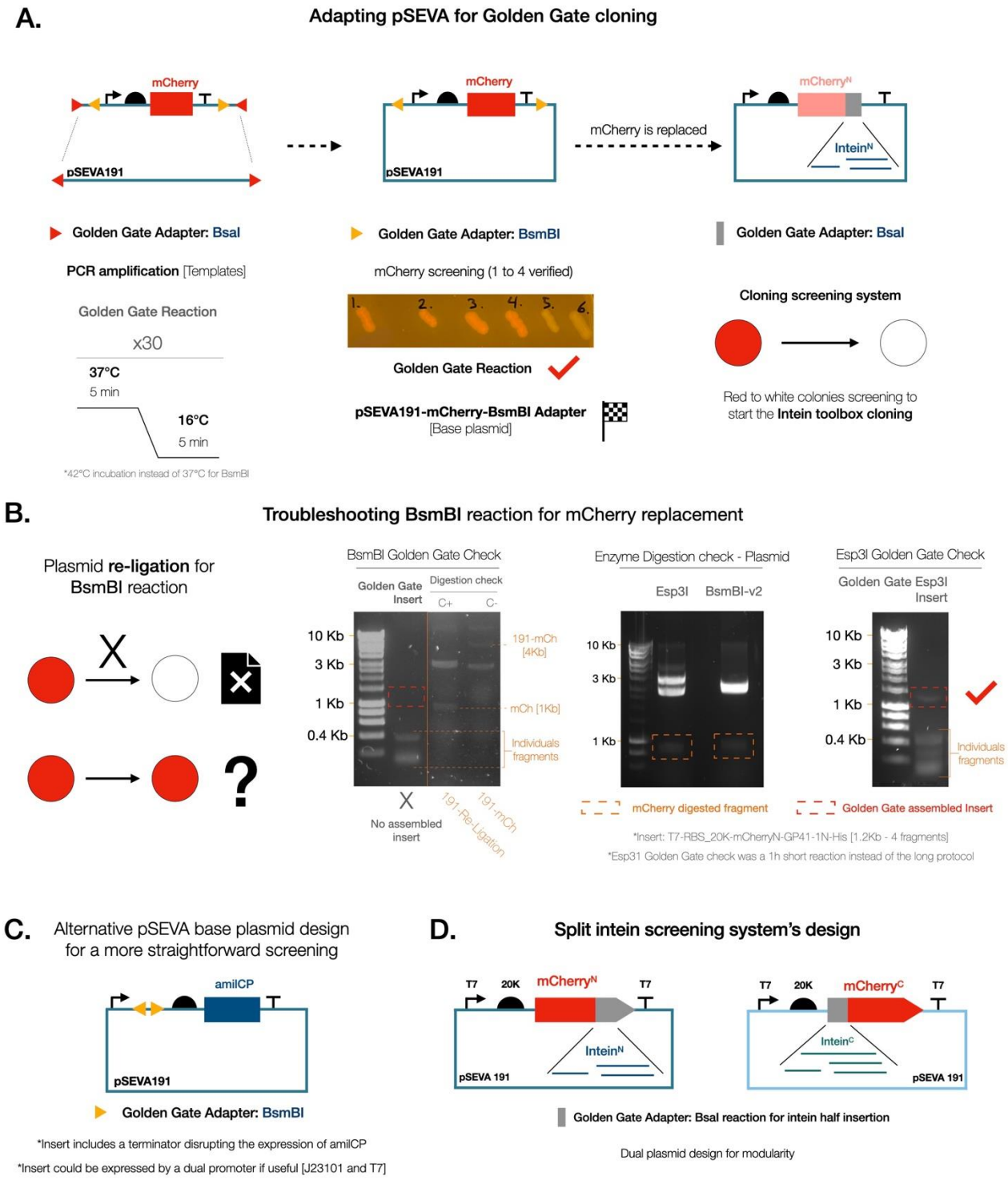

Figure 4: Golden gate setup and reaction for a modular cloning system. (A) Schematic for adapting pSEVA plasmids to a Golden Gate system by adding an adapter with a fluorescent reporter protein. Thus, when the desired insert is cloned the colonies will change from red to white color having then a screening system. The base plasmid pSEVA-mCherry-BsmBI Adapter was cloned and verified by sequencing. (B) The second step of the golden gate design didn't work as expected. There was not a switch from red to white colonies. After troubleshooting, we identified that BsmBI was not cutting the linear PCR template contrary to the Esp3I which allowed the ligation of the insert (Red dashed and square box). (C) Alternative design for the Golden Gate screening where the amilCP is disrupted instead of a full replacement as for the mCherry original design. This design might not be used, depending on the mCherry adapter performance. (D) The split intein screening system's design having a dual plasmid system. The idea is to have a modular system where to add the split intein halves directly to the plasmid with the corresponding split reporter halves and so characterize them on a modular approach.

After this validation, the next step will be implementing the split intein screening system with the 2-plasmid strategy. The plan is to test at least 5 split intein pairs and get their splicing kinetics, we plan to use the known fast split inteins: GP41-1, GP41-8, NrdJ-1, IMPDH-1[38], and CL[39] (Cystein-less intein) (Table 1). This characterization will follow the same plate reader experiment and data analysis as the previous one. The main difference here is that the reporter protein signal will be tracked by the time of appearance to get kinetics data, meaning, over time. Here, it is relevant to mention that this cloning can be done in parallel with the proof of concept. For the moment, the plan is to work stepwise, considering that the proof of concept of the 2Split system is more relevant for the outcome of the project and further design modifications if needed.

Table 1: List of the intein's names to be used and their source reference.

Number	Intein name	Source
1	GP41-1	Carvajal-Vallejos et al. (2012) – Pinto et al. (2020)
2	GP41-8	Carvajal-Vallejos et al. (2012) – Pinto et al. (2020)
3	NrdJ-1	Carvajal-Vallejos et al. (2012) – Pinto et al. (2020)
4	IMPDH-1	Carvajal-Vallejos et al. (2012) – Pinto et al. (2020)
5	CL (Cysteine-less)	Bhagawati et al. (2019)

This project stage is a crucial milestone because it will indicate the main direction to take to optimize the system or redirect it if the performance is not the best. Theoretically, the proof of concept with rapamycin and the logic gates should behave as expected based on previous work with split inteins. However, the characterization in our cell-free lysate and with our conditions is a mandatory first step. The actual critical part of the project is to link this type of biosensing with the desired target proteins. This critical point corresponds to the transition between the top-down methodology and the bottom-up biosensing, but this point comes further down the line.

In parallel, the idea is also to optimize the *cell-free E. coli* reaction further by adapting the reaction to bench conditions. These bench conditions refer to room temperature and no agitation mainly. The idea is to follow a Design of Experiment (DoE) approach to explore the design space of the cell-free by changing crowding agent types and concentrations, key cofactors' concentrations (Mg and K), DNA concentration, energy mix concentration, and lysate proportion (Standard percentage is 33%).

The rationale is to start with a DoE setup to explore the design space. However, the optimization to find the best conditions will follow a machine learning approach like Artificial Neural Networks[40,41] or Decision Three training and learning[42]. These two methods and codes are available online, and there is a postdoc in the laboratory who is an expert on machine learning and is willing to help with the learning part of machine learning.

Overall, this optimization aims to find the best conditions to create a cell-free milieu that facilitates fast biosensing. Thus, the hypothesis behind this DoE optimization setup is to change the cell-free milieu to achieve a molecular concentration and dynamics where the transcription and translation are maximized and accelerated as much as possible. This optimization has fixed conditions, far from standard 30 or 37°C, which could include agitation; the idea is thus to set up the reaction at 25°C without agitation. The goal here is also to maximize that reaction adjusted to a lower temperature so that we will evaluate several factors in the DoE in addition to temperature and agitation.

Next, this same year, the plan is to start the second objective and optimize the IBM method previously described[25,43]. Splitting the recognition protein with the IBM method is the third and last protocol to optimize to complete the project. The IBM method described by Ho et al.[25] requires screening from a library to

detect the best split sites by using an intein-transposon device. The IBM method starts by picking a target protein and letting introns pick random insertion sites. Next, the intron's cassette is replaced by another one expressing the splits of the protein fused to split inteins; this genetic circuit will allow screening for the functional insertion sites in the library. If the insertion site is functional, it is possible to find the positive clones to run them by Next Generation Sequencing (NGS) and finally identify the insertion sites. Thus, It is necessary to run and adapt the code available to analyze the NGS results. By running that code, it will be possible to identify the split sites on the target proteins and incorporate them into the split intein logic gates.

After finding the spit site in a target protein, the idea is thus to go back to the multidomain protein design and add the split recognition protein to the corresponding half. Once the two genetic circuits are cloned, the system will be evaluated as a YES gate, the same approach as the rapamycin proof of concept. Consequently, before starting this second objective, it is necessary to design and clone the positive control to optimize and validate the IBM method. The idea to validate this system is to add a known recognition domain like the caffeine one used by Ho et al.[25] or to fuse the rapamycin recognition domains (FBKP and FRK[26]) with a linker and find the normal split of the domains by the IBM method. In addition, and if needed, the mCherry sequence can be used as a control to replicate the IBM control as it is.

Simultaneously, while the IBM optimization will run, the idea is to define the candidate recognition proteins for the biosensing applications (Figure 5). At this point, it will be necessary to define the particular cases of biosensing; the idea so far is to work on two types of diseases: cancer and viral infection. We are aiming to make a breakthrough in biosensing by (i) reducing the biosensing time and (ii) implementing protein logic circuits to detect in a novel manner targets, and hopefully increasing the number of targets with the 2Split technology. Now, we have considered dengue virus detection and a cancer example case like prostate cancer. It will be necessary, too, for the cases to consider the type of sample required and the availability of the biosensing targets in the sample. Overall, the principle is to have that bottom-up biosensing perspective, mentioned in Figure 1.A, to ensure the biosensing possibility from the design itself.

The research plan for the 2024 year ends in theory with the selection of the candidate recognition proteins. However, the projects' upcoming main steps and key milestones are to optimize the Golden Gate reaction to have the libraries for the intein toolbox and the 2Split system's proof-of-concept. By achieving these milestones, it will be possible to foresee the performance of our 2Split system for cell-free biosensing and the subsequent design consideration for the accomplishment and success of the project. Consequently, the intein toolbox milestone will also be characterized and shared with the SYNSENSO network for the expected 18 months after the start of the project; that date will be September 2024 for us, considering that it started in April 2023.

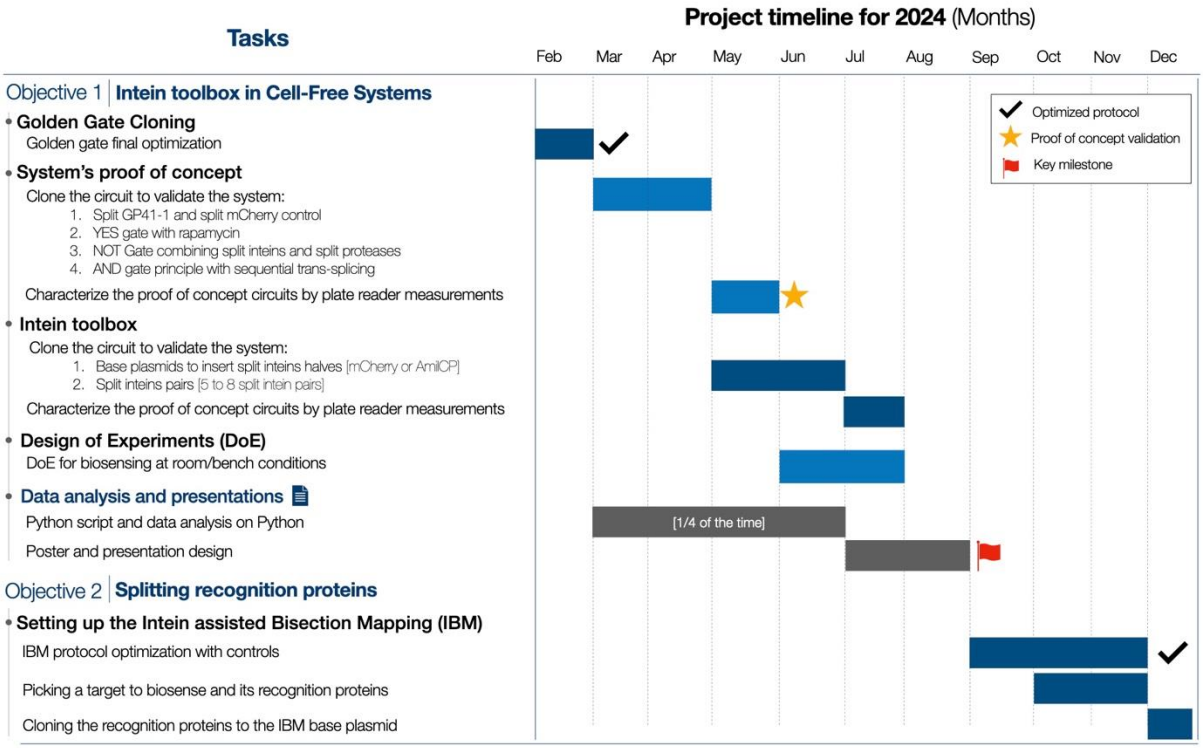

Figure 5: 2Split Project Gantt Chart for the next year. The plan for this upcoming year is to finish the objective 1 and to start the objective 2. To finish the objective 1 the first step is to finish the Golden Gate cloning optimization. Once this is done, it will be possible to clone the library to have the system's proof of concept and the split inteins to be tested (5 split inteins pairs corresponding to fast reacting ones. In parallel to this cloning and characterization, a design of experiment approach combined with machine learning will be put in place to optimize the cell-free reaction for room condition biosensing. Regarding the objective 2, the idea is to set up the IBM protocol and optimize by using control reactions first. Ideally this same year, the goal is to have the defined applications to biosense and thus identify the recognition proteins to split.

Bibliography

1. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367–379 (2010).
2. Voigt, C. A. Synthetic biology 2020–2030: six commercially-available products that are changing our world. *Nat Commun* **11**, 6379 (2020).
3. Baldwin, G. *et al. Synthetic Biology: A Primer*. (Imperial College Press, London, 2016).
4. Kitano, S., Lin, C., Foo, J. L. & Chang, M. W. Synthetic biology: Learning the way toward high-precision biological design. *PLOS Biology* **21**, e3002116 (2023).
5. Meng, F. & Ellis, T. The second decade of synthetic biology: 2010–2020. *Nat Commun* **11**, 5174 (2020).
6. Zhang, C., Liu, H., Li, X., Xu, F. & Li, Z. Modularized synthetic biology enabled intelligent biosensors. *Trends in Biotechnology* **41**, 1055–1065 (2023).
7. Bradley, R. W., Buck, M. & Wang, B. Tools and Principles for Microbial Gene Circuit Engineering. *Journal of Molecular Biology* **428**, 862–888 (2016).
8. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341–aac7341 (2016).
9. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11**, 508–520 (2014).
10. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat Methods* **12**, 415–418 (2015).
11. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat Methods* **15**, 387–393 (2018).
12. Borkowski, O. *et al.* Cell-free prediction of protein expression costs for growing cells. *Nat Commun* **9**, 1457 (2018).
13. Garenne, D. *et al.* Cell-free gene expression. *Nat Rev Methods Primers* **1**, 1–18 (2021).
14. Wagner, L., Jules, M. & Borkowski, O. What remains from living cells in bacterial lysate-based cell-free systems. *Computational and Structural Biotechnology Journal* **21**, 3173–3182 (2023).
15. Batista, A. C., Soudier, P., Kushwaha, M. & Faulon, J. Optimising protein synthesis in cell-free systems, a review. *Engineering Biology* **5**, 10–19 (2021).

16. Lee, S. J. & Kim, D.-M. Cell-free synthetic biology: Navigating the new frontiers of biomanufacturing and biological engineering. *Current Opinion in Systems Biology* **37**, 100488 (2024).
17. Amalfitano, E. & Pardee, K. Logic invades cell-free biosensing. *Nat Chem Biol* **18**, 355–359 (2022).
18. Tinafar, A. *et al.* Cell-Free Biosensors: Synthetic Biology Without Borders. in *Handbook of Cell Biosensors* (ed. Thouand, G.) 1–39 (Springer International Publishing, Cham, 2019). doi:10.1007/978-3-319-47405-2_130-1.
19. Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165**, 1255–1266 (2016).
20. Shah, N. H. & Muir, T. W. Inteins: nature's gift to protein chemists. *Chemical Science* 16 (2014).
21. Wang, H., Wang, L., Zhong, B. & Dai, Z. Protein Splicing of Inteins: A Powerful Tool in Synthetic Biology. *Front. Bioeng. Biotechnol.* **10**, 810180 (2022).
22. Di Ventura, B. & Mootz, H. D. Switchable inteins for conditional protein splicing. *Biological Chemistry* **400**, 467–475 (2019).
23. Lange, O. J. & Polizzi, K. M. Click it or stick it: Covalent and non-covalent methods for protein-self assembly. *Current Opinion in Systems Biology* **28**, 100374 (2021).
24. Pinto, F., Thornton, E. L. & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat Commun* **11**, 1529 (2020).
25. Ho, T. Y. H. *et al.* A systematic approach to inserting split inteins for Boolean logic gate engineering and basal activity reduction. *Nat Commun* **12**, 2200 (2021).
26. Sonntag, T. & Mootz, H. D. An intein-cassette integration approach used for the generation of a split TEV protease activated by conditional protein splicing. *Mol. BioSyst.* **7**, 2031–2039 (2011).
27. Miyamoto, T., Razavi, S., DeRose, R. & Inoue, T. Synthesizing Biomolecule-Based Boolean Logic Gates. *ACS Synth. Biol.* **2**, 72–82 (2013).
28. Sun, Z. Z. *et al.* Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *JoVE (Journal of Visualized Experiments)* e50762 (2013) doi:10.3791/50762.
29. Batista, A. C. *et al.* Differentially Optimized Cell-Free Buffer Enables Robust Expression from Unprotected Linear DNA in Exonuclease-Deficient Extracts. *ACS Synth. Biol.* **11**, 732–746 (2022).
30. Azad, M. S. *et al.* Cell-Free Protein Synthesis from Exonuclease-Deficient Cellular Extracts Utilizing Linear DNA Templates. *JoVE (Journal of Visualized Experiments)* e64236 (2022) doi:10.3791/64236.
31. Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R. & Oza, J. P. Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. *JoVE* 58882 (2019) doi:10.3791/58882.
32. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE* **6**, e16765 (2011).
33. Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S. M. & Cai, Y. EMMA: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. *ACS Synth. Biol.* **6**, 1380–1392 (2017).
34. Pryor, J. M. *et al.* Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLOS ONE* **15**, e0238592 (2020).
35. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research* **41**, D666–D675 (2013).
36. Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & de Lorenzo, V. SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Research* **43**, D1183–D1189 (2015).
37. King, A. M. *et al.* Selection for constrained peptides that bind to a single target protein. *Nat Commun* **12**, 6343 (2021).
38. Carvajal-Vallejos, P., Pallissé, R., Mootz, H. D. & Schmidt, S. R. Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources *. *Journal of Biological Chemistry* **287**, 28686–28696 (2012).
39. Bhagawati, M. *et al.* A mesophilic cysteine-less split intein for protein trans-splicing applications under oxidizing conditions. *Proceedings of the National Academy of Sciences* **116**, 22164–22172 (2019).
40. Pandi, A. *et al.* Metabolic perceptrons for neural computing in biological systems. *Nat Commun* **10**, 3880 (2019).
41. Pandi, A. *et al.* Cell-free biosynthesis combined with deep learning accelerates de novo-development of antimicrobial peptides. *Nat Commun* **14**, 7197 (2023).
42. Pandi, A. *et al.* A versatile active learning workflow for optimization of genetic and metabolic networks. *Nat Commun* **13**, 3876 (2022).
43. Ho, T. & Wang, B. Intein-assisted Bisection Mapping (IBM). (2020).